INSTITUTO SUPERIOR DE AGRONOMIA

Ano Lectivo - 2014/2015

ESTATÍSTICA

Aulas práticas

Exercícios com algumas soluções

Professores das aulas práticas:
Fernanda Valente, Maria Emília Pinto,
Maria João Martins, Mariana Mota e
Maria Manuela Neves.

Nota Introdutória

Apresentamos aqui um conjunto de exercícios preparados para as aulas práticas da unidade curricular *Estatística*, leccionada no 3 º semestre de todas as licenciaturas de Bolonha do ISA, com excepção de Arquitectura Paisagista.

É por demais reconhecida a importância crescente da Estatística como ferramenta imprescindível na análise, interpretação e previsão de resultados em todas as áreas no domínio das Ciências e Engenharias, nomeadamente as que constituem a formação disponibilizada pelo Instituto Superior de Agronomia

Na reestruturação actual esta unidade curricular passa a ter 5h/semana sendo 2h teóricas e 3h práticas (até 2009/2010 dispunha de 2h teóricas e 4h práticas). Tem como principais objectivos:

- Apresentar os conceitos básicos de Probabilidade (com os principais modelos probabilísticos, imprescindíveis para a modelação e inferência de características de interesse e que estão sujeitas à intervenção do acaso - aleatórias).

- Ensinar métodos de interpretação, descrição, análise e inferência realizados sobre um conjunto de observações obtidas em trabalhos de campo nas diferentes unidades curriculares dos planos curriculares do ISA.

- Introduzir um *software* estatístico adequado, que possua técnicas actualizadas para os tratamentos estatísticos que os alunos necessitem de realizar. Muitos são os programas informáticos existentes. Entendemos, porém, que ao introduzir os nossos alunos no *s*oftware R estamos a colocá-los na vanguarda do que hoje é usado quer nas aplicações quer na investigação.

Estes exercícios integram três capítulos: Estatística Descritiva, Introdução à Teoria da Probabilidade e Introdução à Inferência Estatística. No final de cada capítulo de exercícios encontram-se tópicos de resolução ou algumas soluções.

Para preparar desde já os alunos na introdução e utilização do *software* estatístico que vamos usar, na 1ª parte destas folhas de exercícios, dedicadas à Estatística Descritiva, apresentamos **uma lista de exercícios de treino do R**. Alguns exercícios de Estatística Descritiva serão também resolvidos com apoio do R.

Como material de apoio sobre o R pode utilizar os textos referenciados na página *web* da *Estatística* (`https://fenix.isa.ulisboa.pt/qubEdu/disciplinas/estat-0/2014-2015/1-semestre`).

Antes de cada aula prática os alunos deverão tentar resolver os exercícios da matéria já leccionada, principalmente os indicados pelos professores.

Capítulo 1 Estatística Descritiva

Nota: Todos os ficheiros e/ou objectos referidos nos exercícios encontram-se na página *web* da *Estatística* (`https://fenix.isa.ulisboa.pt/qubEdu/disciplinas/estat-0/2014-2015/1-semestre`).

Exercícios de Introdução ao *software* R com algumas aplicações em Estatística

Os exercícios seguintes ilustram e fazem a aplicação da breve introdução ao *software* R feita na primeira semana de aulas. Este *software* será utilizado ao longo de todo o semestre, como apoio na resolução de exercícios.

1.1. É desencadeado um programa de controlo da poluição de um rio em que são efectuadas medições antes de lançar a campanha antipoluição (Ano0) e um ano após (Ano1). Os resultados destas medições são os seguintes:

Ponto de controlo	1	2	3	4	5	6	7	8	9	10
Ano0	68	88	101	82	96	74	65	74	52	99
Ano1	67	87	90	76	98	69	68	65	59	70

a) Usando a linguagem R crie dois vectores `Ano0` e `Ano1`, com os dados observados.

b) Verifique que os dois vectores têm o mesmo número de registos.

c) Crie o vector `dif` com as diferenças entre os valores de `Ano0` e `Ano1`.

d) Calcule medidas de localização e dispersão para os dados do Ano0.

1.2. a) Crie os vectores `nomes`, `idades` e `alturas`, com os nomes e as respectivas idades e alturas de dez dos seus colegas.

b) Construa o vector lógico `cartac` em que cada componente indica se o aluno da componente homóloga do vector `nomes` possui ou não carta de condução.

c) Determine o máximo do vector `idades` e o mínimo do vector `alturas`.

d) Determine a média das alturas e identifique os alunos que têm altura superior a este valor.

e) Identifique os colegas com carta de condução e determine o seu número.

f) Construa a *data frame* `colegas` que contenha os vectores `nomes`, `idades`, `alturas` e `cartac`. Resolva novamente as alíneas c), d) e e) utilizando este novo objecto.

1.3. O ficheiro "concelho.txt" contém o nome de todos os concelhos de Portugal Continental, o respectivo número de freguesias e o nome do distrito a que pertencem.

a) Leia os valores do ficheiro "concelho.txt" e guarde-os na *data frame* `conc`.

b) Qual ou quais os concelhos com maior número de freguesias?

c) Guarde no objecto `totc` o número total de concelhos existentes em Portugal Continental.

d) Crie a tabela `dist` com o nome de todos os distritos de Portugal Continental e o respectivo número de concelhos.

e) Construa um diagrama de barras com o número de concelhos por distrito de Portugal Continental.

f) Calcule o número médio de concelhos por distrito.

1.4. a) Crie o vector `v` com o resultado de 30 lançamentos de um dado equilibrado de seis faces (sugestão: utilize a instrução `sample(1:6,30,rep=TRUE)`)

b) Calcule a média $\overline{v} = \frac{\sum_{i=1}^{30} v_i}{30}$ e a variância $s^2 = \frac{\sum_{i=1}^{30}(v_i-\overline{v})^2}{30}$ destes 30 valores.

c) Utilize os comandos `mean(v)` e `var(v)` para resolver a questão anterior. Comente os resultados obtidos.

d) Faça um diagrama de barras de `v`.

e) Crie os vectores `w1` e `w2` com, respectivamente, os resultados de 300 e 3000 lançamentos de um dado de seis faces equilibrado. Construa um diagrama de barras para os valores de cada um destes vectores e compare-os com o diagrama obtido na alínea anterior.

1.5. Importe o ficheiro de objectos "SementesOzono.RData". Este ficheiro contém os vectores `sementes` e `ozono`. O vector `sementes` contém o número de sementes de um certo cereal que germinaram em cada um de 50 vasos iguais (inicialmente foram semeadas 5 sementes em cada vaso). O vector `ozono` contém 78 valores de concentração de ozono na atmosfera (ppm) obtidos numa dada cidade.

Calcule o número médio de sementes que germinaram por vaso e a concentração média de ozono observada.

1.6. Importe o ficheiro de objectos "DadosMeteo.RData". Este ficheiro contém os vectores `precip`, `temp` e `vento`, com dados de precipitação (mm), temperatura do ar (^{o}C) e velocidade do vento (ms^{-1}), respectivamente, medidos numa estação meteorológica em Évora.

a) Indique o número total de elementos e o número total de observações não disponíveis (NA) de cada vector.

b) Determine indicadores numéricos de localização e de dispersão para estes dados.

Exercícios de Estatística Descritiva a uma dimensão

Nota: Recorde-se que os ficheiros e/ou objectos referidos nos exercícios se encontram na página *web* da *Estatística* (`https://fenix.isa.ulisboa.pt/qubEdu/disciplinas/estat-0/2014-2015/1-semestre`).

1.7. Classifique, justificando, cada uma das seguintes variáveis quanto ao tipo: qualitativo/quantitativo, contínuas/discretas.

a) Cor do cabelo de uma pessoa;

b) Peso de um bebé à nascença;

c) Número de automóveis que passaram na portagem nos domingos de verão;

d) Qualidade da comida numa cantina (má, razoável, boa, muito boa);

e) Temperatura máxima diária em Agosto deste ano;

f) Número de sementes que germinaram num vaso onde foram semeadas 5 sementes.

1.8. Para cada um dos conjuntos de dados apresentados abaixo, construa uma tabela de frequências absolutas, relativas, relativas acumuladas e absolutas acumuladas.

a) Conjunto 1 - - número de nemátodos contados em cada uma de 60 placas observadas ao microscópio.

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

b) Conjunto 2 - - número de laranjas de cada uma das 40 árvores de um laranjal

131	136	150	152	155	156	162	169	170	177
188	196	201	201	205	210	210	211	214	216
217	220	225	226	231	238	240	244	244	247
251	262	268	275	288	297	300	302	303	305

c) Conjunto 3 - - peso (kg) de 56 ovelhas, após administração de um dado tratamento. Estes dados encontram-se no ficheiro "ovelhas.txt".

30.28	27.58	27.91	29.33	31.20	28.40	33.3	25.40
34.26	32.55	21.78	25.59	35.08	26.86	33.20	29.70
39.47	30.15	33.40	27.38	30.39	25.85	29.11	26.22
33.54	30.40	29.60	28.82	30.70	30.83	33.84	27.58
29.46	36.15	23.40	24.48	30.35	23.85	27.12	26.42
36.54	20.40	23.30	38.42	31.10	25.83	31.84	22.58
31.54	22.42	33.35	28.22	34.15	26.83	21.24	30.14

1.9. Considere os vectores $\boldsymbol{x} = (x_1, x_2, x_3, x_4, x_5) = (2, 5, 6, 10, 15)$ e $\boldsymbol{y} = (y_1, y_2, y_3, y_4, y_5) = (-1, 2, 0, 3, 4)$. Calcule:

a) $\sum_{i=1}^{5} x_i$	b) $\sum_{i=3}^{5} x_i$	c) $\sum_{k=1}^{5} x_3$
d) $\sum_{j=1}^{5} 2x_j$	e) $\sum_{k=2}^{5} x_k - 4$	f) $\sum_{i=1}^{5} x_i^2$
g) $(\sum_{i=1}^{5} x_i)^2$	h) $\sum_{i=1}^{5} x_i y_i$	i) $\sum_{i=1}^{5} x_i \sum_{i=1}^{5} y_i$.

1.10. Um viticultor registou o peso diário das uvas recolhidas durante os 15 dias de uma vindima, mas no fim só forneceu o peso médio diário - 515 kg.

a) Qual foi a produção total (peso em kg) daquele período?

b) Alguém comentou que naqueles 15 dias o peso mínimo diário colhido tinha sido 150 kg e o peso máximo diário 475 kg. O que pensa destas afirmações?

c) Constatou-se que num dos dias tinha havido erro no registo do peso de uvas colhidas. Por engano o registo desse dia foi de 20 kg. Qual o valor do peso médio diário, depois de retirado aquele registo?

1.11. Considere dois conjuntos de dados: o primeiro contendo o registo diário, efectuado durante 50 dias, do número de casos de intoxicação ocorridos numa fábrica e o segundo com o registo das preferências relativamente a 5 tipos de mistura de café (designadas por A, B, C, D e E) manifestadas num inquérito feito a 1000 consumidores.

Nº de casos	0	1	2	3	4	5	6
Nº de dias	13	15	8	6	5	2	1

Misturas de café	A	B	C	D	E
Nº de respostas	190	210	180	205	215

a) Indique a variável considerada em cada um dos casos e classifique-a.

b) Determine uma medida de localização adequada a cada um dos conjuntos de dados.

c) Indique o valor do mínimo e do máximo de cada conjunto de dados, caso existam.

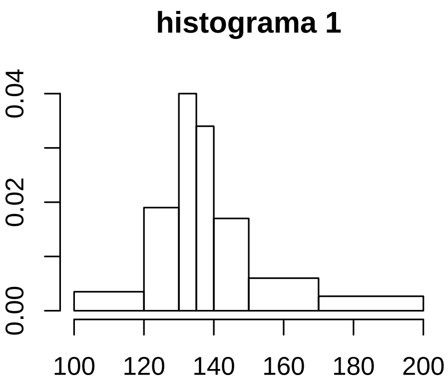
1.12. Na tabela que se segue apresenta-se o agrupamento dos dados relativos a uma amostra de alturas (em dm) de 100 árvores de uma mesma espécie.

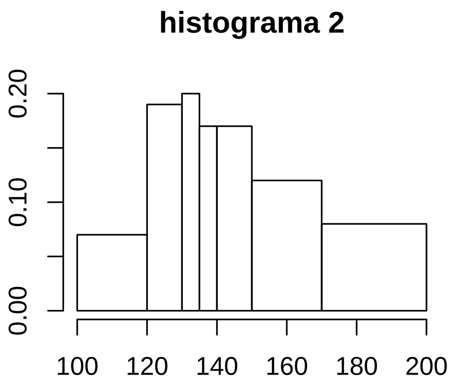
classe	[100;120[	[120;130[	[130;135[	[135;140[	[140;150[	[150;170[	[170;200[
nºde árvores	7	19	20	17	17	12	8

a) Elabore uma tabela de frequências relativas e frequências relativas acumuladas, dos valores observados.

b) Indique, justificando, qual dos histogramas apresentado a seguir se pode considerar o mais adequado para descrever o agrupamento de dados apresentado?

c) Determine a mediana, aproximada, da altura das árvores observadas.





1.13. De um total de N números há uma proporção p de 1's e uma proporção $q = 1 - p$ de 0's.

a) Calcule a média do conjunto dos N números.

b) Supondo N grande prove que o desvio padrão é aproximadamente $\sqrt{pq}$.

1.14. Considere o quadro seguinte com os dados da altitude das principais serras do Continente (Fonte: Instituto Geográfico e Cadastral e Centro de Estudos Geográficos; dados reproduzidos no *Anuário Estatístico*, I.N.E., Lisboa, 1980):

Designação	Altitude (m)	Designação	Altitude (m)
Peneda	1416	Gardunha	1227
Soajo	1415	Leomil	1008
Gerês	1507	Lapa	953
Barroso	1208	Marofa	973
Larouco	1525	Malcata	1075
Cabreira	1261	Grândola	325
Alvão	1283	Cercal	372
Marão	1415	Espinhaço de Cão	297
Padrela	1146	Monchique	902
Coroa	1273	Caldeirão	577
Montezinho	1438	Mendro	412
Nogueira	1318	Ossa	653
Bornes	1200	S.Mamede	1025
Mogadouro	993	Adiça	522
Montemuro	1382	Sicó	553
Arada	1116	Aire	679
Caramulo	1071	Candeeiros	613
Buçaco	549	Montejunto	664
Lousã	1204	Sintra	528
Açor	1340	Arrábida	501
Estrela	1991	Monte Figo	411
Alvelos	1084		

Os dados foram introduzidos no *software* R e estão disponíveis no objecto `serras`, no ficheiro "Serras.RData".

a) Agrupe os dados em classes. Faça a sua representação gráfica.

b) Comente a distribuição das altitudes das serras.

c) Averigúe se há candidatos a "outlier" no conjunto dos dados.

1.15. Os valores da precipitação (em mm) registada na Estação Meteorológica de Lisboa, nos 31 dias do mês de Janeiro de um dado ano, foram os seguintes (dados do Instituto de Meteorologia):

Dia	Precip.	Dia	Precip.	Dia	Precip.
1	0.0	11	3.8	21	0.9
2	0.0	12	0.3	22	0.3
3	0.0	13	0.0	23	18.2
4	0.0	14	0.0	24	4.0
5	4.7	15	0.5	25	4.6
6	0.6	16	7.0	26	22.0
7	17.2	17	0.0	27	15.6
8	1.4	18	0.0	28	0.0
9	11.2	19	3.3	29	3.4
10	1.0	20	7.6	30	0.0
				31	0.0

a) Construa o histograma para os dados da precipitação e comente-o.

b) Obtenha a caixa-de-bigodes dos dados e comente-a.

c) Calcule a precipitação média e mediana diária em Lisboa, naquele mês. Compare os valores obtidos da média e da mediana e comente, tendo em atenção que ambos são indicadores de localização.

d) Introduza os dados no *software* R. Responda às questões anteriores utilizando o R.

1.16. Num pomar de pêra-rocha registou-se o número de pêras que foram colhidas no ano passado em cada uma das 60 pereiras. Os dados recolhidos foram introduzidos no ficheiro "peras.dat".

a) Construa uma tabela de frequências e faça a representação gráfica do número de pêras em cada pereira daquele pomar.

b) Qual a produção média e a produção mediana de cada pereira? Calcule ainda o desvio padrão da produção de cada pereira. Comente.

c) Construa a caixa de bigodes dos dados apresentados.

1.17. Um biólogo está a estudar uma unidade de aquicultura de criação de douradas. Num dado dia recolheu 15 douradas no viveiro A e obteve como peso médio e variância $\overline{x_1} = 235\ g$ e $s_1^2 = 254\ g^2$; recolheu 20 douradas no viveiro B e obteve $\overline{x_2} = 245\ g$ e $s_2^2 = 267\ g^2$.

Indique a média e a variância do conjunto das 35 douradas observadas nos dois viveiros.

1.18. Diga **justificando** se são verdadeiras ou falsas as afirmações que se seguem:

a) A amplitude interquartil é metade da amplitude total.

b) A média está sempre entre o primeiro e o terceiro quartil.

c) A mediana está sempre entre o primeiro e o terceiro quartil.

d) O desvio padrão é sempre igual à amplitude interquartil.

e) O desvio padrão é menor do que a média dos desvios relativos à média.

1.19. Num estudo, efectuado nos dois últimos anos, sobre o número de golfinhos observados em cada passeio organizado pela empresa OlhóGolfinho no estuário do Sado, obtiveram-se os seguintes dados:

nº de golfinhos	0	1	2	3	4	5	6	8
nº de passeios	17	45	84	52	23	11	2	1

a) Diga qual a variável em estudo e classifique-a.

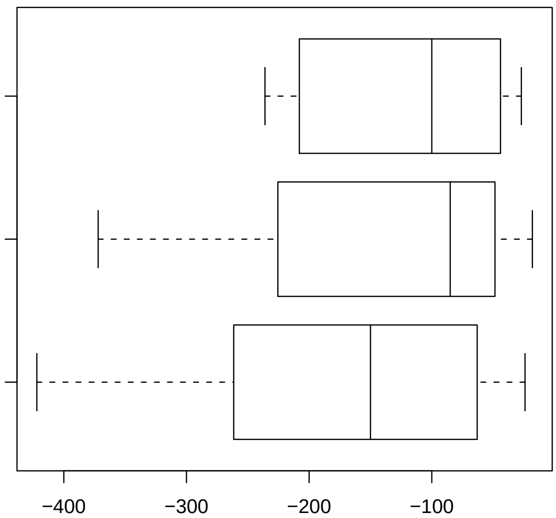
b) Apresente os dados numa tabela de frequências relativas e faça a representação gráfica adequada. Comente.

c) Descreva a amostra indicando medidas de localização central e de dispersão.

d) Qual é a percentagem de passeios em que no máximo se observaram 2 golfinhos?

1.20. Numa experiência medem-se fluxos de calor de meia em meia hora, das 7h às 18h (inclusivé), durante três dias consecutivos. Os resultados obtidos (em W m^{-2}) são indicados na tabela em baixo. Ao lado da tabela estão as caixas-de-bigodes dos três dias, sem qualquer ordem aparente. Os dados foram introduzidos no *software* e estão disponíveis no objecto `fluxoCalor`, no ficheiro “FluxoCalor.RData”.

DIA 1	DIA 2	DIA 3
-27	-24	-85
-32	-38	-74
-31	-61	-49
-53	-54	-31
-67	-59	-18
-48	-65	-32
-38	-67	-33
-47	-74	-57
-41	-120	-34
-41	-150	-59
-63	-171	-48
-114	-50	-92
-100	-98	-138
-100	-175	-74
-175	-184	-103
-208	-178	-196
-228	-228	-194
-208	-295	-259
-208	-320	-255
-196	-359	-284
-236	-401	-324
-210	-422	-294
-216	-405	-372



a) Associe cada diagrama ao respectivo dia. Justifique.

b) Sem fazer contas, diga se a média correspondente ao diagrama do topo será inferior ou superior a -100. Justifique.

c) Considere agora o conjunto das observações nos 3 dias.

   i) Calcule os indicadores de localização e dispersão destes dados.

   ii) Desenhe o *boxplot* dos dados e compare com os que lhe são fornecidos.

   iii) Construa uma tabela de frequências para dados agrupados em classes de amplitude 50.

   iv) Use a tabela da alínea anterior para calcular valores aproximados da média e da mediana das observações nos três dias. Comente os resultados.

1.21. Registaram-se os atrasos nas chegadas, em minutos, em voos europeus num dado dia do mês de Julho de 2009 no aeroporto da Portela. Os dados recolhidos em 100 voos foram organizados na seguinte tabela:

Atraso	]0;10]	]10;20]	]20;30]	]30;40]	]40;50]	]50;60]
Nº de voos	24	23	33	9	7	4

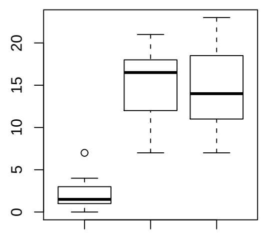
a) Identifique e classifique, justificando, a variável em estudo.

b) Determine valores aproximados da média, desvio padrão e coeficiente de variação da variável em estudo.

c) Determine valores aproximados dos quartis da variável em estudo. Poderão existir *outliers* na amostra observada? Justifique.

1.22. Num estudo realizado para avaliar o efeito de três *sprays*, A, B e C, em insectos, organizaram-se 3 grupos de 12 recipientes cada, nos quais se colocou o mesmo número de insectos a que se aplicaram aqueles insecticidas. Indicadores relativos ao nº de insectos mortos em cada um deles, encontram-se no quadro e diagrama seguintes.

Spray	$\sum_{i=1}^{12} x_i$	$\sum_{i=1}^{12} x_i^2$	min	Q1	Q2	Q3	max
A	174	2768	7	11.5	14.0	17.8	23
B	184	3022	7	12.5	16.5	17.5	21
C	25	95	0	1.0	1.5	3.0	7



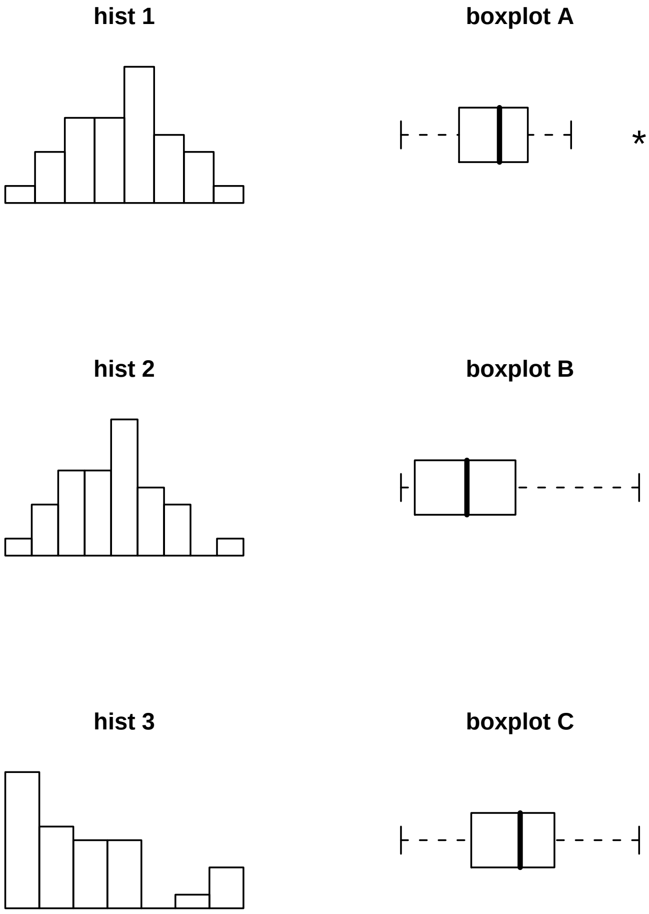
a) Associe cada *boxplot* a cada *spray*, indicando o valor das barreiras de *outliers* no primeiro diagrama. Justifique.

b) Compare os três conjuntos de dados quanto à localização, dispersão e simetria.

c) Para a totalidade das observações calcule a média, a variância e a amplitude total.

d) A totalidade dos dados foi agrupada em classes, com extremos 0, 5, 10, 15, 20 e 25, e frequencias absolutas observadas, respectivamente, 11, 5, 9, 8 e 3. Calcule a média, mediana e variância para os dados agrupados.

1.23. Na figura que se segue apresentam-se, para 3 conjuntos de dados, os histogramas e respectivas caixas de bigodes sem qualquer ordem. Associe cada histograma à caixa de bigodes relativa ao mesmo conjunto de dados.



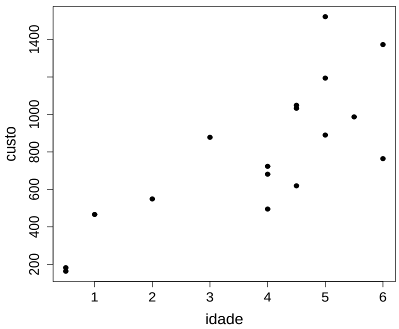
1.24. Pensa-se que o custo de manutenção (em euros) de tractores aumenta com a idade (em anos) do tractor. Para modelar a relação entre o custo e a idade registaram-se observações de idade de 17 tractores e respectivo custo de manutenção. Os dados foram introduzidos no programa R, nos objectos `idade` e `custo`. Apresentam-se alguns resultados obtidos.

```
> sort(custo)
[1]  163  182  466  495  549  619  681  723  764  878  890  987 1033 1049
```

```
[15] 1194  1373  1522
> summary(idade)
    Min. 1st Qu.  Median   Mean  3rd Qu.    Max.
  0.500   3.000   4.500   3.824   5.000   6.000
> summary(custo)
   Min. 1st Qu.  Median   Mean   3rd Qu.    Max.
  163.0    A        B    798.1   1033.0   1522.0

> var(idade)    > var(custo)     > cor(idade,custo)
[1] 3.248162   [1] 142478.4     [1] 0.7831099

> plot(idade,custo,pch=20)
```



a) Classifique a variável "custo de manutenção". Justifique. Face ao objectivo do estudo trata-se da variável resposta ou da variável independente?

b) No *output* do R, obtido com o comando `summary(custo)` faltam dois valores. Calcule-os.

c) Esboce a caixa-de-bigodes do custo de manutenção. Comente.

d) Determine a variância da idade dos tractores observados expressa em meses.

e) $^{(**)}$ Poder-se-á admitir a existência de uma relação linear entre as variáveis? Justifique. Independentemente da sua resposta determine a equação da recta de regressão que modela a relação entre as variáveis em estudo.

f) $^{(**)}$ Calcule a precisão da recta e interprete o seu significado.

g) $^{(**)}$ Qual a variação anual média do custo de manutenção estimada pela recta de regressão quando os tractores têm idade entre 0.5 e 6 anos?

h) $^{(**)}$ Determine os coeficientes da recta de regressão se a idade fosse dada em meses. Justifique.

$^{(**)}$ – Alínea a resolver na secção "Estatística Descritiva a duas dimensões".

Exercícios de Estatística Descritiva a duas dimensões

1.24. [A] As colunas da *data frame* `cor` (no ficheiro "Cor.RData") contêm, respectivamente, a cor dos olhos e do cabelo de 300 portugueses. Com estes dados, construa uma tabela de contingência. Determine as frequências marginais e indique o seu significado.

1.25. A tabela seguinte mostra os valores do índice de preços ao consumidor (IPC) em Portugal nos últimos anos.

i	1	2	3	4	5
Ano(x)	2002	2003	2004	2005	2006
IPC(y)	100	103.3	105.7	108.1	111.5

a) Calcule $cov(x, y)$ e a média e a variância de x e y.

b) Se aos anos de observação x se tivesse aplicado a transformação $x/2 - 1000$, i.e., se se considerasse os valores 1, 1.5, 2, 2.5, 3, qual seria o valor de $cov(x', y)$, com $x' = x/2 - 1000$?

c) Comente os resultados obtidos em a) e b) e diga se poderá considerar-se a covariância um bom indicador da existência de uma relação forte entre x e y. Justifique.

d) Independentemente das respostas anteriores, determine a equação da recta de regressão de y sobre x.

e) Calcule a precisão da recta e interprete o seu significado.

f) Qual a variação anual média dos preços, estimada pela regressão, no período 2002-06?

g) Mantendo-se a actual tendência, qual prevê que seja o IPC em 2007?

h) Quais seriam os valores dos coeficientes da recta de regressão se o índice de preços ao consumidor tivesse como base o ano de 2003, i.e. $y_i' = 100 y_i / y_2$?

1.26. Para $n = 20$ pares de observações (x_i, y_i), seja $y = -5.6 + 0.7\ x$, a equação da recta de regressão dos mínimos quadrados de y sobre x.

a) Comente as seguintes afirmações:

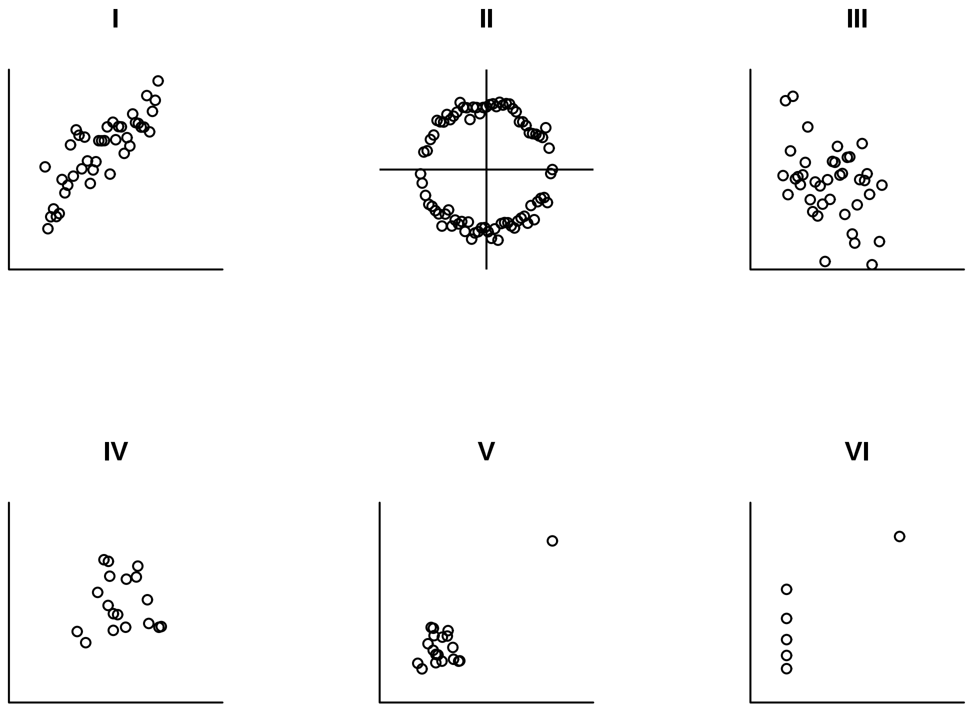
i) O coeficiente de correlação entre y e x é positivo porque o declive da recta é positivo.

ii) Em média, quando x aumenta y não aumenta, pois o declive da recta é menor do que 1.

b) Sendo $\sum_{i=1}^{20} x_i = 200$ determine $\sum_{i=1}^{20} y_i$.

1.27. Indique qual dos valores abaixo indicados se aproxima mais do coeficiente de correlação dos dados descritos nas seguintes nuvens de pontos:

a) 0 b) 0.8 c) -0.5 d)2.0



1.28. Num estudo sobre o consumo de gasolina de vários modelos de automóveis ligeiros de passageiros e a cilindrada do respectivo motor, foi estabelecida a seguinte equação da recta de regressão dos mínimos quadrados

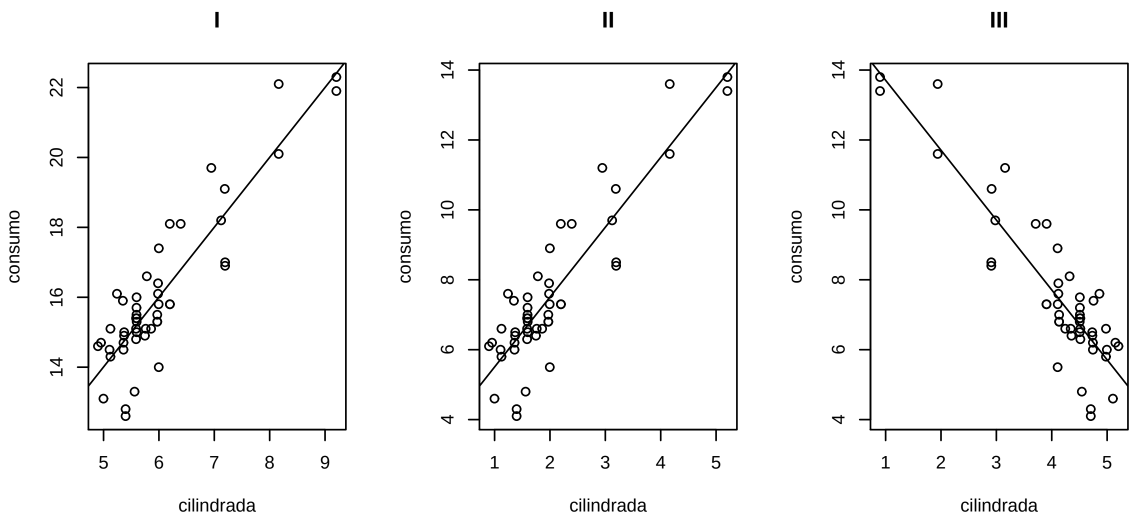
$$y = 3.5 + 2x$$

em que x é a cilindrada (em 10^3 cm^3) e y é o consumo (em litros por 100 km percorridos). Sabendo que a precisão desta recta é de 0.803 e que a média e o desvio padrão das cilindradas observadas foram de 2.027 e 0.994 (10^3 cm^3), respectivamente, responda às seguintes questões:

a) Determine a média e o desvio padrão dos consumos de gasolina dos automóveis observados.

b) Qual é a variação esperada para o consumo de gasolina quando se aumenta a cilindrada de 1000 cm^3?

c) Qual dos seguintes gráficos corresponde à nuvem de pontos e à respectiva recta de regressão do estudo descrito?



d) Parece-lhe adequada a utilização do modelo linear para descrever a relação entre o consumo de gasolina e a cilindrada do motor nos modelos de automóveis analisados? Justifique.

1.29. A evaporação dos solventes que se usam nas tintas depende da humidade ambiente. O conhecimento desta relação poderá ser útil para melhorar a qualidade da operação de pintura. Foi realizado um estudo para examinar a relação entre x - "humidade relativa ambiente (%)" e y - "quantidade de um determinado solvente evaporado durante a pintura (% do peso)". Desse estudo resultaram os seguintes dados:

$n = 20$; $\bar{x} = 52.5$; $\bar{y} = 9.5$; $s_x^2 = 256.5789$; $s_y^2 = 10.2632$; $cov(x,y) = -46.0526$

a) Classifique, justificando, a variável x - "humidade relativa ambiente".

b) Poder-se-á admitir a existência de uma relação linear entre as variáveis? Justifique.

c) Independentemente da resposta à alínea anterior, determine a recta de regressão dos mínimos quadrados de y sobre x. Indique uma medida da precisão dessa recta e interprete o seu valor.

d) Suponha que foi registado o resíduo $e = 0.34$, associado à observação $x = 55$, relativamente à recta de regressão definida em c). Qual o correspondente valor observado para y?

1.30. A medição directa do calor específico de ramos de macieira é difícil de efectuar. Um investigador propõe predizer o calor específico de ramos individuais a partir de medições (muito mais simples de efectuar) da percentagem de água no ramo, em vez de medir directamente o calor específico.

Para isso recolheu observações da percentagem de água (x) e do calor específico (y) de 21 ramos. Os valores obtidos (registados no ficheiro "CalorEspecifico.RData") são os seguintes :

x	y	x	y	x	y
49	46	53	57	62	119
58	90	50	44	63	131
59	104	57	100	52	53
51	65	53	89	51	70
56	85	60	96	65	131
61	113	52	69	52	66
56	96	58	111	54	69

a) Desenhe o diagrama de extremos e quartis para os valores do calor específico observados. Comente a distribuição dos dados.

b) Parece-lhe adequada a existência de uma relação linear entre x e y? Porquê? Independentemente da sua resposta ajuste aos dados a recta de regressão dos mínimos quadrados.

c) Qual o valor que se prevê para o calor específico quando a percentagem de água é de 60? Justifique.

d) Sabe-se que, para facilitar os cálculos, os valores originais obtidos para o calor específico dos ramos ($y^{'}$) foram transformados de acordo com a expressão $y = 1000\ y^{'} - 600$, sendo os valores de y os registados na tabela dada acima. Suponha que lhe era pedido para escrever a regressão linear entre x e $y^{'}$; deduza a relação existente entre os coeficientes da nova recta e os da recta que obteve em b). Haverá alteração na precisão da regressão?

1.31. Numa dada região, registou-se anualmente entre 1998 e 2006 a produção de trigo. Designando por x o ano e por y a produção de trigo, em milhares de toneladas, obtiveram-se os seguintes valores para os 9 pares de observações efectuadas:

$\overline{x} = 2002;$ $\quad \overline{y} = 270.5;$ $\quad \sum_{i=1}^{9}(x_i - \overline{x})^2 = 60$

$\sum_{i=1}^{9}(y_i - \overline{y})^2 = 1416.2;$ $\quad \sum_{i=1}^{9}(x_i - \overline{x})(y_i - \overline{y}) = -203$

a) Determine a recta de regressão dos mínimos quadrados da evolução da produção de trigo em função do tempo. Indique a sua precisão.

b) Se se decidisse identificar os anos por 1,...,9 respectivamente, qual seria a precisão da recta de regressão que se obteria considerando esta transformação? Justifique convenientemente.

1.32. A seguinte tabela apresenta o período de gestação (x), em dias, e o tempo médio de vida (y), em anos, registados em 10 mamíferos.

	urso	hipopótamo	canguru	leopardo	leão	macaco	rato	porco	cão	gato
x_i	219	238	42	98	100	164	21	112	61	63
y_i	18	25	7	12	15	15	3	10	12	12

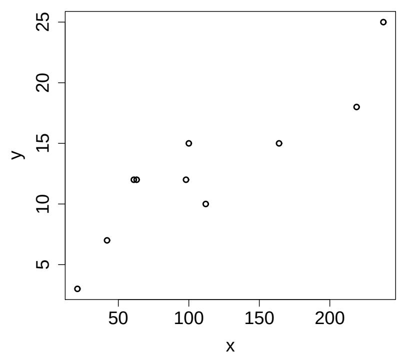
Os dados foram introduzidos no *software* R, apresentando-se abaixo os resultados obtidos:

```
x<-c(219,238,42,98,100,164,21,112,61,63)
y<-c(18,25,7,12,15,15,3,10,12,12)

>mean(x)              > mean(y)
[1] 111.8             [1] 12.9

> var(x)              > var(y)
[1] 5394.622          [1] 36.1

> cov(x,y)            > plot(x,y)
[1] 396.7556
```



a) Parece-lhe adequada a existência de uma relação linear entre x e y? Justifique.

b) Independentemente da resposta à alínea anterior determine a recta de regressão dos mínimos quadrados de y sobre x. Calcule a precisão da recta e interprete o seu significado.

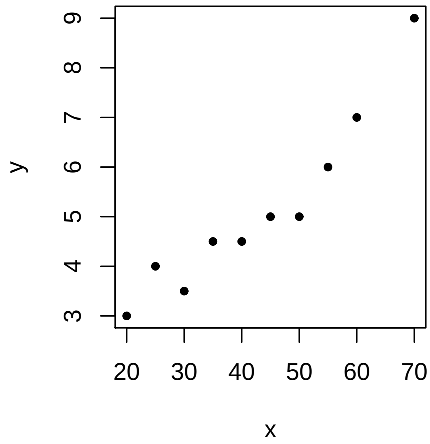
c) Interprete, no contexto do problema, o significado do coeficiente de regressão de y sobre x.

d) O período de gestação de uma girafa é de 425 dias. Se usasse a recta determinada em b) que previsão obteria para o seu tempo médio de vida? Critique o resultado obtido, sabendo que o tempo médio de vida de uma girafa é de 10 anos.

e) Determine a recta de regressão dos mínimos quadrados de "tempo médio de vida" sobre "tempo de gestação", sendo agora o tempo de gestação, x', dado em meses ($x' = x/30$). Qual a precisão desta recta?

1.33. Num estudo em que se pretende avaliar a influência da velocidade do vento (m/s) na quantidade de água (centenas de litro) evaporada por dia na albufeira de uma barragem, obtiveram-se os seguintes dados que foram introduzidos no *software* R.

```
> x<-c(20, 50,30,55,70,45,60,25,40,35) #vel. vento (m/s)
> y<-c(3,5,3.5,6,9,5,7,4,4.5,4.5)      #agua evaporada
> mean(x)
[1] 43
> mean(y)
[1] 5.15
> var(x)
[1] 256.6667
> var(y)
[1] 3.169444
> cov(x,y)
[1] 27
> plot(x,y,pch=20)
```



Com base nos resultados apresentados, responda às seguintes questões.

a) Parece-lhe adequada a existência de uma relação linear entre x e y? Justifique.

b) Independentemente da resposta à alínea anterior determine a recta de regressão dos mínimos quadrados de y sobre x. Calcule a precisão da recta e interprete o seu significado.

c) Determine a equação da recta de regressão de "quantidade de água evaporada" sobre "velocidade do vento" no caso de os valores da velocidade do vento serem dados em km/h. Qual será a precisão desta recta? Justifique. (**Note que** $1\ m/s = 3.6\ km/h$).

d) Determine $cov(x', y)$, sendo x' a variável em km/h.

1.34. Foram seleccionadas aleatoriamente 20 folhas de videira da casta Água Santa, tendo sido medidos, para cada folha, os comprimentos (em mm) da nervura principal (variável *NP*) e das nervuras laterais esquerda (variável *NLesq*) e direita (variável *NLdir*), bem como a área foliar (variável *Area*, em mm^2). Alguns indicadores associados aos valores observados são:

```
     NLesq              NP                NLdir              Area
 Min.   : 8.20    Min.   : 8.80    Min.   : 8.90    Min.   : 134.00
 Median :10.70    Median :12.05    Median :10.80    Median : 199.00
 Mean   :10.70    Mean   :11.97    Mean   :10.71    Mean   : 208.45
 Max.   :15.10    Max.   :15.70    Max.   :14.10    Max.   : 356.50
 Var    : 3.0011  Var    : 3.0314  Var    : 1.8047  Var    :3188.0763
```

```
> var(NLdir-NLesq)
[1] 0.8626053
```

a) Calcule o primeiro quartil da variável *NP*, sabendo que os valores observados foram:

```
15.7 15.4 14.0 12.7 12.6 12.6 12.6 12.6 12.5 12.4 11.7 11.6 11.5
11.1 10.8 10.5 10.5 10.2  9.7  8.8
```

b) Ajustou-se uma recta de regressão de área foliar (*Area*) sobre comprimento da nervura principal (*NP*), tendo-se obtido a equação $Area = -137.951 + 28.927NP$.

   i) Qual a variação esperada na área foliar associada a um aumento de 1 mm no comprimento da nervura principal, estimada pela regressão?

   ii) Sabe-se que uma das 20 folhas observadas no ajustamento tinha 11.3 mm de nervura principal e uma área foliar de 190.0 mm^2. Qual o resíduo associado a esta folha?

   iii) Determine a precisão da recta de regressão e interprete o valor obtido.

1.35. Realizou-se um estudo para averiguar a percentagem (P) de resíduos sólidos eliminados por um sistema de filtragem em função da taxa (T) de fluxo de efluente. No quadro abaixo encontram-se alguns resultados observados:

Taxa de fluxo de efluente (T)	1	4	5	6	8	10	12
Percentagem de resíduos sólidos (P)	24	19	18	•	14	•	10

a) Identifique a variável dependente e a variável independente.

b) Considere os resultados obtidos no R, apresentados abaixo, para responder às seguintes perguntas:

   i) Parece-lhe admissível a existência de uma relação linear entre as variáveis? Porquê?

   ii) Escreva a equação da recta de regressão dos mínimos quadrados ajustada aos dados. Interprete, no contexto do problema, o significado do coeficiente de regressão.

```
lm(formula = P ~ T)
Residuals:
       1        2        3        4        5        6        7
 0.62628 -0.59556 -0.33618  0.42321 -0.55802 -0.03925  0.47952

Coefficients:
             Estimate  Std. Error  t value  Pr(>|t|)
(Intercept)  24.63311    0.45651   53.96    4.13e-08 ***
T            -1.25939    0.06148  -20.49    5.13e-06 ***
---
Multiple R-squared: 0.9882
```

1.36. Considere os quatro conjuntos de dados seguintes (dados de Anscombe, 1973):

x1	y1	x2	y2	x3	y3	x4	y4
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.10	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.10	4	5.39	19	12.5
12	10.84	12	9.13	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89

a) Calcule as médias e as variâncias de cada uma das oito variáveis. Comente.

b) Calcule os coeficientes de correlação entre as variáveis x e as variáveis y de cada um dos quatro pares de variáveis. Comente.

c) Calcule as rectas de regressão de y sobre x para cada um dos quatro pares de variáveis (x_i,y_i), (i=1,...,4).

d) Construa as quatro nuvens de pontos correspondentes aos pares de variáveis utilizados nas duas alíneas anteriores. Comente, à luz dos resultados das alíneas anteriores.

e) Construa os gráficos dos resíduos para cada conjunto de pares de observações e comente-os.

Exercícios de Revisão de Estatística Descritiva

R1.1. Considere n pares de observações (x_i, y_i). Seja $z_i = \dfrac{x_i - \overline{x}}{s_x}$, $\; i = 1, \cdots, n$.

a) Mostre que as observações z_i têm média nula e variância unitária.

b) Determinou-se a recta de regressão dos mínimos quadrados de y sobre x tendo-se obtido $\quad y = 3 + 1.5x$. Sabendo que $\overline{y} = 10.5$ e $s_x^2 = 0.25$, determine a equação da recta de regressão de y sobre z.

R1.2. Considere n pares de observações (x_i, y_i).

a) Sendo $\overline{x}$ a média das observações x_i, mostre que a média de $z_i = x_i/k - m$, (k e m números reais, $k \neq 0$ e $i = 1, \ldots, n$) é $\overline{z} = \overline{x}/k - m$.

b) Sejam $y = b_0 + b_1 x$ a recta de regressão dos mínimos quadrados de y em x e $\hat{y}_i$ os valores estimados pela recta, correspondentes aos valores observados x_i. Mostre que o coeficiente de determinação, R^2, é igual ao quadrado do coeficiente de correlação, r, isto é,

$$R^2 = \frac{\sum_{i=1}^{n} (\hat{y}_i - \overline{y})^2}{\sum_{i=1}^{n} (y_i - \overline{y})^2} = r^2.$$

R1.3. A densidade óptica (d) de uma solução de um dado produto químico, medida para oito níveis diferentes de concentração (c), está registada na seguinte tabela (considerando unidades de medição adequadas):

c_i	1	2	4	5	8	10	12	15
d_i	4	9	18	20	35	41	42	60

$\sum_{i=1}^{8} c_i = 57 \quad \sum_{i=1}^{8} d_i = 229 \quad \sum_{i=1}^{8} c_i d_i = 2288 \quad \sum_{i=1}^{8} c_i^2 = 579 \quad \sum_{i=1}^{8} d_i^2 = 9091$

a) Pretende-se ajustar uma recta de regressão aos dados obtidos. Parece-lhe admissível tal ajustamento? Justifique convenientemente.

b) Independentemente da sua resposta à alínea anterior, escreva a equação da recta de regressão dos mínimos quadrados que relaciona as variáveis envolvidas na experiência. Qual a precisão da recta que obteve? Comente.

c) Sem efectuar novos cálculos, altere <u>uma única</u> observação de c, de modo que se verifique uma diminuição da média $\overline{c}$, um aumento da variância s_c^2 e um valor idêntico para a mediana $\tilde{c}$. Justifique.

R1.4. Pretende-se estudar a relação existente entre a superfície florestal (y) e a superfície territorial (x), expressas em milhares de hectares, nos 18 distritos do Continente. A equação da recta de regressão ($y = b_0 + b_1 x$), calculada a partir dos dados das Estatísticas Agrícolas do INE, num dado ano, tem os seguintes coeficientes: $b_0 = 13.1$; $b_1 = 0.32$.

a) Comente as seguintes afirmações:

i) Em média, quando a superfície territorial aumenta, não aumenta a superfície florestal, pois b_1 é menor do que 1;

ii) O coeficiente de correlação entre a superfície florestal e a superfície territorial tem de ser positivo, porque o coeficiente b_0 é positivo.

b) Sendo a superfície territorial total do Continente 8892.7 milhares de hectares, diga qual a superfície florestal total.

R1.5. Num projecto de construção de mesas para computadores, verificou-se ter interesse avaliar a distância entre o assento e os cotovelos, estando uma pessoa sentada. Designando essa quantidade por y, procura-se relacioná-la com a altura total da pessoa (x). Os valores de uma amostra de dimensão $n = 22$ são dados na tabela seguinte:

altura(x) (cm)	distância ao cotovelo (y) (cm)			
159	22	23		
160	25	25	27	
161	24	27	25	26
162	23	26	27	29
166	27	23	28	
168	27	29	31	31
172	34	35		

Nota: $\sum_{i=1}^{22} x_i^2 = 590754 \qquad \sum_{i=1}^{22} y_i^2 = 16288 \qquad \sum_{i=1}^{22} x_i y_i = 97547$

a) Calcule a distância média entre os assentos e os cotovelos e a altura média dos indivíduos observados.

b) Calcule a mediana da variável altura.

c) Estime um modelo de regressão linear simples da distância (y) em função da altura das pessoas. Indique a precisão da recta e comente-a.

d) Considere que o par $(166, 23)$ resulta de uma medição errada e foi decidido retirá-lo. Deduza à custa dos somatórios dados na Nota os parâmetros da recta construída com as observações restantes.

e) Qual o aumento esperado para a distância entre o assento e o cotovelo por cada aumento unitário na altura de uma pessoa?

R1.6. Dados n pares de observações (x, y), seja $y = b_0 + b_1 x$ a recta de mínimos quadrados ajustada.

a) Defina coeficiente de correlação, $r_{x,y}$ e indique uma sua propriedade.

b) Sendo $s_x^2 = 5.1$; $b_1 = -3$; $\overline{x} = 3$; $\overline{y} = 2.8$ e $r^2 = 0.92$ determine s_y e a equação da recta de regressão.

c) Prove que o declive da recta é invariante quando se efectua uma mesma transformação de escala a ambas as variáveis.

R1.7. Foi efectuado um estudo para analisar a relação entre o número de dias após a eclosão do ovo (variável x, em dias) e o comprimento das asas de crias de pardal doméstico (*Passer domesticus*) (variável y, em cm). Alguns indicadores associados aos dados observados são:

	Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo	Variância
x	3	6	10	10	14	17	21.83333
y	1.400	2.400	3.200	3.415	4.500	5.200	1.638077

A covariância entre as variáveis x e y é 5.9 cm×dia.

a) Averigúe se existem candidatos a *outliers* para os valores observados do comprimento das asas (y) e desenhe a respectiva caixa de bigodes, indicando os valores utilizados na sua construção.

b) Poder-se-á admitir a existência de uma relação linear entre as variáveis? Justifique.

c) Independentemente da resposta à alínea anterior, determine a recta de regressão dos mínimos quadrados de y sobre x. Qual é a variação diária média do comprimento das asas de crias de pardal doméstico prevista pela regressão?

d) Para a recta de regressão determinada na alínea anterior obteve-se o resíduo -0.21538 associado à observação $x = 10$. Calcule o correspondente valor observado para y.

e) Suponha que os dados do comprimento das asas foram registados em dm. Deduza a relação entre os coeficientes da recta de regressão neste caso e a obtida na alínea c).

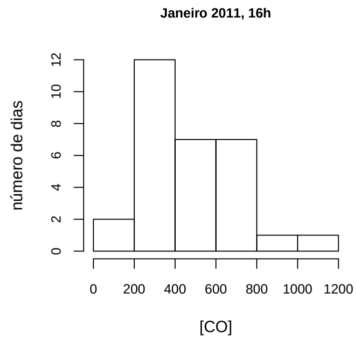
R1.8. Para oito pares de observações $\{(x_i, y_i)\}_{i=1}^{8}$ determinou-se a recta de regressão dos mínimos quadrados, $y = 2.45 - 1.2x$, cuja precisão é 0.9604. Responda, **justificando convenientemente**, se são verdadeiras ou falsas as afirmações nas seguintes alíneas.

a) Sabendo que $\sum_{i=1}^{8} x_i = 15$ então $\overline{y} = 0.2$.

b) O coeficiente de correlação é $r = 0.98$.

c) Se $s_x^2 = 0.5522$ então $s_y^2 = 0.828$.

R1.9. O gráfico mostra o histograma dos 30 registos diários disponíveis da concentração de CO, μg/m^3, medida às 16h, em Janeiro de 2011, numa avenida de Lisboa.



a) Identifique e classifique, justificando, a variável em estudo.

b) Elabore a tabela de frequências absolutas, relativas e relativas acumuladas associada ao histograma.

c) Determine valores aproximados para os seguintes indicadores:
i) mediana ii) média.

d) Em cada dia de registo da concentração de CO registou-se também o número de automóveis que passaram nessa avenida entre as 15h30 e as 16h. Com os dados disponíveis obteve-se a seguinte recta de regressão dos mínimos quadrados da concentração de CO (y) sobre o número de automóveis (x): $y = 468.25 + 0.021\, x$.

i) Calcule um valor aproximado do número total de automóveis registado.

ii) Interprete, no contexto do problema, o valor do coeficiente de regressão.

iii) Num dos dias registou-se $x = 282$, ao qual se verificou estar associado um resíduo igual a -12.37, para o modelo de regressão dado. Qual o valor observado para a concentração de CO naquele dia?

Soluções de alguns Exercícios

1.7. a) Qualitativa nominal
b) Quantitativa contínua
c) Quantitativa discreta
d) Qualitativa ordinal
e) Quantitativa contínua
f) Quantitativa discreta

1.9. a) 38 b) 31 c)30 d) 76 e) 32
f) 390 g) 1444 h) 98 i) 304

1.10. a) 7725 kg
c) 550.3571 kg

1.11. a) 1º conjunto - Conjunto da esquerda: variável - nº de casos de intoxicação em cada dia - quantitativa discreta

2º conjunto - Conjunto da direita: variável - tipo de mistura de café - variável qualitativa nominal

b) 1º conjunto: pode ser a média $\bar{x} = 1.7$ casos/dia ou a mediana $me = 1$ caso/dia ou a moda $mo = 1$ caso/dia
2º conjunto: moda $mo =$ mistura tipo E.

c) Só é possível indicar o mínimo e o máximo para o 1º conjunto:
mínimo - 0 casos/dia; máximos 6 casos/dia

1.12. b) As classes dadas têm amplitudes diferentes e como a área de cada rectângulo deverá representar a frequência relativa associada à respectiva classe, isto é, $A_i = f_i$, tem-se, por exemplo $A_1 = f_1 = 0.07$, $A_2 = f_2 = 0.19$, etc.

Como $f_3 = 0.2 \Rightarrow 0.2 = alt_3 \times 5 \Rightarrow alt_3 = 0.04$. Efectuando raciocínio análogo tem-se $alt_4 = 0.034$. Só poderá ser o histograma 1.

c) 1º calcular classe mediana que é $[135; 140[$.
$me \approx 136.1765$

1.17. Para as 35 douradas tem-se $\overline{x} = 240.71g$ e $s^2 = 290.0042g^2$ (quando se faz arredondamentos em cálculos intermédios pode obter-se resultados um pouco diferentes – por ex. $s^2 = 281.13g^2$)

1.19. a) variável - nº de golfinhos em cada passeio - variável discreta porque toma valores em $\mathbb{N}_0$.

c) média $\overline{x} = 2.28$, $mo = 2$ e $\tilde{x} = 2$ golfinhos por passeio.
Medidas de dispersão - variância e desvio padrão.

d) Pretende-se $F(2) = \frac{17+45+84}{235} = 0.6213$, frequência relativa acumulada, logo é 62%

1.20. a) Para Dia 1 tem-se $min(x_i) = -236$ e $max(x_i) = -27$; Dia 2 $min(x_i) = -422$ e $max(x_i) = -24$ e Dia 3 $min(x_i) = -372$ e $max(x_i) = -18$.

Então Dia 1 corresponde ao diagrama 1 (topo), Dia 2 corresponde ao diagrama 3 (o que está em baixo) e Dia 3 corresponde ao diagrama do meio.

1.23. O histograma 1 corresponde ao *boxplot* C; o histograma 2 corresponde ao *boxplot* A e o histograma 3 corresponde ao *boxplot* B.

1.25. a) $\sum_{i=1}^{5} x_i = 10020$; $\sum_{i=1}^{5} y_i = 528.6$; $\sum_{i=1}^{5} x_i^2 = 20080090$; $\sum_{i=1}^{5} y_i^2 = 55961.24$; $\sum_{i=1}^{5} x_i y_i = 1059342$
$cov(x,y) = 6.95$; $\overline{x} = 2004$; $\overline{y} = 105.72$; $s_x^2 = 2.5$; $s_y^2 = 19.412$

b) $cov(x^{'}, y) = 3.475$

d) $y = -5465.4 + 2.78x$

e) $r^2 = 0.9953$, i.e., 99% da variabilidade do índice é explicada pela recta de regressão, portanto como é próximo de 1, a recta é muito precisa.

f) A variação anual média estimada é dada por $b_1 = 2.78$

g) $\hat{y}_{2007} = 114.06$

h) $b_1^{'} = 2.69$; $b_0^{'} = -5290.8$

1.26. a) i) A afirmação é verdadeira. A relação $r = b_1\, s_x/s_y$, $(s_x > 0, s_y > 0)$ estabelece que r e b_1 têm o mesmo sinal. Logo $b_1 > 0 \Rightarrow r > 0$.
ii) A afirmação é falsa, pois b_1 é positivo. O coeficiente b_1 representa a variação esperada para y quando x aumenta de uma unidade.

b) Uma consequência da recta dos mínimos quadrados é: $\bar{y} = b_0 + b_1\bar{x}$, o que é equivalente a $\sum y_i = nb_0 + b_1 \sum x_i = 20 \times (-5.6) + 0.7 \times 200 = 28$.

1.27. Os valores que se aproximam mais dos coeficientes de correlação dos dados são: para a nuvem I $->$ b); para a nuvem II $->$ a); para a nuvem III $->$ c); para a nuvem IV $->$ a); para a nuvem V $->$ b); para a nuvem VI $->$ b).

1.28. a) $\bar{y} = 7.554$ litros por 100 Km; $s_y = 2.2185$ litros por 100 Km.

b) Espera-se que o consumo de gasolina aumente 2 litros por 100 Km.

c) Gráfico II.

1.31. a) $y = 7043.43 - 3.383\, x$; a precisão da recta é $R^2 = 0.485$.

b) A mesma precisão.

1.32. a) O diagrama de dispersão sugere a existência de uma relação linear entre as variáveis x e y. Como $r = 0.899$ se pode considerar não muito afastado de 1, é de admitir a existência de uma relação linear entre x e y.

b) $y = 4.677 + 0.0735x$.

A precisão da recta é dada por $r^2 = 0.899^2 = 0.808$, o que significa que 80.8% da variabilidade de y é explicada pela regressão de y sobre x.

c) O coeficiente de regressão de y sobre x, $b_1 = 0.0735$, significa que, para aqueles mamíferos, por cada dia de aumento no período de gestação se espera um aumento de 0.0735 anos no seu tempo médio de vida.

d) A previsão feita pela recta de regressão da alínea b) para o tempo médio de vida de uma girafa (sabendo que o seu período de gestação é de 425 dias) é

$$\hat{y} = 4.677 + 0.0735 \times 425 = 35.9 \text{ anos.}$$

Contudo, a utilização desta recta de regressão para prever o tempo médio de vida de uma girafa não é aconselhável, já que o valor da variável preditora ($x = 425$) não pertence à gama de valores observados de x ($[21, 238]$). Sendo assim, aquela recta não permite efectuar este previsão, pelo que, a grande diferença entre o valor ajustado e o valor real (10 anos) é justificável.

e) $y = 4.677 + 2.205x'$.

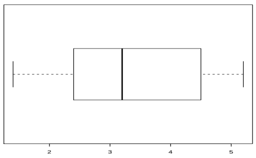
1.34. a) 10.65 mm.

b) i) Espera-se que a área foliar aumente 28.927 mm^2
ii) $e_{(NP=11.3)}$=1.076 mm.
iii) $R^2 = 0.798$.

R1.1. b) Seja $y = b_0' + b_1' z$ recta de y sobre z
$b_0' = \overline{y} - b_1'\overline{z}$ e como $\overline{z} = 0 \Rightarrow b_0' = \overline{y} = 10.5$
$b_1' = \frac{cov(z,y)}{s_z^2}$, como $z = \frac{x}{s_x} - \frac{\overline{x}}{s_x}$ $\quad cov(z,x) = \frac{1}{s_x}cov(x,y)$ $\quad$ e $\quad s_z^2 = 1$, então
$b_1' = \frac{1}{s_x}cov(x,y) = 0.75$

R1.7. a) Tem-se a barreira inferior $B_I = 2.4 - 1.5(4.5 - 2.4) = -0.75$ e a barreira superior $B_S = 4.5 + 1.5(4.5 - 2.4) = 7.65$, portanto não há, nos dados, valores inferiores à barreira inferior e também não há valores superiores à barreira superior, portanto não há candidatos a *outliers*.

Os valores necessários à construção da caixa de bigodes são:
$max(y) = 5.2$, $min(y) = 1.4$, $Q_1 = 2.4$, $Q_3 = 4.5$ e a mediana $\tilde{y} = 3.2$.



b) Como não dispomos dos dados não podemos construir a nuvem de pontos, mas o coeficiente de correlação, $r = \frac{cov(x,y)}{s_x \; s_y} = \frac{5.9}{\sqrt{21.83333 \times 1.638077}} = 0.9866$, apresenta um valor próximo de 1 ($-1 \leq r \leq 1$), pelo que podemos admitir a existência de uma relação linear entre as variáveis.

c) A recta de regressão é $y = 0.713 + 0.2702x$.

O comprimento das asas aumenta por dia, em média, 0.2702 cm.

d) Como $y_i = b_0 + b_1 \; x_i + e_i$, onde e_i é o resíduo, então para $x_i = 10$ tem-se $y_i = 0.713 + 0.2702 \times 10 - 0.21538 = 3.1996$ cm.

e) $y' = 0.1y$ designa o comprimento das asas expresso em dm.

Sendo assim, consideremos b'_1 e b'_0 os coeficientes da recta de regressão de y' em x.

$b'_1 = \frac{cov(x,y')}{s_x^2} = \frac{0.1 cov(x,y)}{s_x^2} = 0.1 \; b_1$

$b'_0 = \overline{y'} - b'_1\overline{x} = 0.1\overline{y} - 0.1 \; b_1 \; \overline{x} = 0.1(\overline{y} - b_1\overline{x}) = 0.1 \; b_0$.

Capítulo 2 Introdução à Teoria da Probabilidade

Exercícios de Introdução à Probabilidade

2.1. Considere os acontecimentos A e B tais que $P(A) = 0.4$, $P(B) = 0.6$ e $P(A \cup B) = 0.8$. Determine:

$$P(A \cap B); \quad P(A - B); \quad P(\overline{A} \cap B); \quad P(A \cup \overline{B}).$$

2.2. Considere o tempo de vida de uma lâmpada em centenas de horas.

Seja $\Omega = \{t : t > 0\}$ o espaço de resultados associado à duração de vida da lâmpada. Considere os acontecimentos:

$$A = \{t : t > 15\} \quad B = \{t : 2 < t < 10\} \quad C = \{t : t < 12\}$$

Caracterize os seguintes acontecimentos:

$A \cup B$ $\qquad A \cap C$ $\qquad A \cap \overline{B}$ $\qquad (A \cup B) \cap \overline{C}$ $\qquad \overline{A} \cup (B \cap C)$.

2.3. Sejam A e B acontecimentos aleatórios definidos num espaço de resultados Ω. Mostre que:

a) $P(\overline{A} \cap \overline{B}) = 1 - P(A) - P(B) + P(A \cap B)$;

b) A probabilidade de ocorrer um e apenas um dos acontecimentos A ou B é igual a $P(A) + P(B) - 2P(A \cap B)$.

2.4. Sejam A e B acontecimentos num espaço de resultados Ω tais que $P(A)+P(B) = x$ e $P(A \cap B) = y$. Determine em função de x e de y a probabilidade de:

a) Que se realize pelo menos um dos dois acontecimentos.

b) Que se realize um e um só dos dois acontecimentos.

c) Não se realizar qualquer dos dois acontecimentos.

d) Que se realize quanto muito um único acontecimento.

2.5. Sejam A e B acontecimentos tais que $P(A) = 0.5$ e $P(A \cup B) = 0.7$. Determine $P(B)$ sabendo que :

a) A e B são acontecimentos independentes;

b) A e B são acontecimentos mutuamente exclusivos.

2.6. Sejam A e B dois acontecimentos de um espaço de resultados Ω. Defina-se

$$A \vartriangle B = (A - B) \cup (B - A).$$

Escreva $P\,(A \vartriangle B)$ em termos de $P(A)$, $P(B)$ e $P(A - B)$.

2.7. Numa propriedade agrícola, sabe-se que 60%, 75% e 50% das árvores são de folha caduca, de fruto e de fruto com folha caduca, respectivamente. Calcule a probabilidade de uma árvore da propriedade, escolhida ao acaso:

a) não ser árvore de fruto;

b) ser árvore de fruto ou de folha caduca;

c) ser árvore de fruto, sabendo que tem folha caduca.

2.8. As probabilidades de três corredores de velocidade percorrerem 100 metros em menos de 10 segundos são respectivamente: 1/3, 1/5 e 1/10. Considerando que os tempos dos três atletas são independentes, calcule a probabilidade de, uma corrida em que participam apenas os três atletas, ser ganha em menos de 10 segundos.

2.9. Sejam A e B acontecimentos independentes definidos num espaço de resultados Ω.

a) Prove que A e $\overline{B}$ são acontecimentos independentes.

b) Se $P(A) > 0$ e $P(B) > 0$, prove que para qualquer acontecimento C se verifica:

$$P(C|A) = P(B)P(C|(A \cap B)) + P(\overline{B})P(C|(A \cap \overline{B})).$$

2.10. Sejam A, B e C acontecimentos independentes definidos em Ω.

a) O que significa dizer que A, B e C são acontecimentos independentes?

b) Mostre que:

i) A e $B \cup C$ são independentes;

ii) $P(A \cup B) = P(A)P(\bar{B}) + P(B)$;

iii) $P[(A \cap B) \cup C] = P(A)P(B)P(\overline{C}) + P(C)$.

2.11. Considere três acontecimentos A, B e C tais que $P(C) = 0.3$, $P(B|C) = 0.4$, $P(\overline{B}|\overline{C}) = 0.8$, $P(A|(B \cap C)) = P(A|\overline{(B \cap C)}) = 0.2$.

a) Calcule $P(C|B)$.

b) Calcule $P\,[(B \cap C)|A]$.

c) Diga, justificando, se os três acontecimentos são ou não independentes.

2.12. Sejam A, B e C três acontecimentos de um espaço de resultados Ω.

a) Sabendo que $A \cup B \cup C = \Omega$, $P(A) = 0.3$, $P(\overline{B}) = 0.7$, $P(C) = 0.5$ e $A \cap B = C \cap B = \emptyset$, determine $P(A \cap C)$.

b) Se $P(A) = 0.3$ e $P(\overline{B}) = 0.2$, mostre que A e B são acontecimentos não mutuamente exclusivos.

c) Mostre que se A e B constituem <u>uma partição de Ω</u> então $P(A - C) = P(\overline{C}) - P(\overline{C} \cap B)$, para qualquer acontecimento C.

2.13. Sejam A e B acontecimentos tais que $P(B|A) = 1/4$, $P(B|\overline{A}) = 1/3$ e $P(A) = 2/5$.

a) Determine $P(B)$ e $P(\overline{A}|B)$.

b) Serão A e B acontecimentos independentes? E mutuamente exclusivos? Justifique.

2.14. Um teste detecta a presença de um certo tipo de bactérias na água com uma probabilidade 0.9 no caso de o referido tipo existir. Se não existir, detecta a sua ausência com probabilidade 0.8. Sabendo que a probabilidade de que uma amostra de água contenha bactérias daquele tipo é 0.20, calcule a probabilidade de que, ao retirar uma amostra ao acaso:

a) o teste dê resultado positivo,i.e., detecte a presença de bactérias;

b) haja de facto bactérias quando o teste dá positivo;

c) o teste dê um resultado errado.

2.15. Um determinado tipo de peças é produzido pelas fábricas F_1, F_2 e F_3. Durante um certo período de tempo, F_1 produziu o dobro das peças de F_2 enquanto F_2 e F_3 produziram o mesmo número de peças. Sabe-se ainda que 2%, 2% e 4% das peças produzidas por F_1, F_2 e F_3, respectivamente, são defeituosas. Todas as peças produzidas nesse período de tempo foram colocadas num depósito.

a) Qual a percentagem de peças defeituosas e que são provenientes da fábrica F_2?

b) Qual a percentagem de peças defeituosas armazenadas?

c) Foi encontrada uma peça defeituosa no depósito. Qual a origem(fábrica) menos provável dessa peça?

2.16. Num dado país 10% da população sofre de uma determinada doença: 6% de forma grave e 4% de forma moderada. Para o seu diagnóstico é efectuado um teste que dá resultado positivo:

– com probabilidade 1 para um indivíduo com doença na forma grave;

– com probabilidade 0.75 para um indivíduo com doença na forma moderada;

– com probabilidade 0.05 para um indivíduo não doente.

a) Efectuando um teste num indivíduo ao acaso, qual a probabilidade de o resultado ser positivo?

b) Se, para um dado indivíduo, o resultado do teste foi positivo, qual a probabilidade de ele ter a doença?

c) Será que existe independência entre ter a doença na forma moderada e na forma grave? Justifique.

2.17. Os trabalhadores de uma empresa da área alimentar foram classificados em três níveis de acordo com o grau de instrução: formação mínima, formação média e formação superior.

Sabe-se que 55% dos trabalhadores tem salário superior a 1000 euros. Em particular, têm salário superior a 1000 euros 70% dos trabalhadores com formação superior, 40% dos trabalhadores com formação média e nenhum trabalhador com formação mínima. Sabe-se ainda que 60% dos trabalhadores têm formação superior.

Escolhendo um trabalhador ao acaso, calcule a probabilidade de:

a) Ter formação média.

b) Ter formação superior, sabendo que ganha mais de 1000 euros.

2.18. O estudo das soluções aquosas existentes num laboratório relativamente ao seu pH e comprimento de onda de absorção conduziu aos seguintes resultados:

- 40% das soluções são ácidas, 20% são alcalinas e as restantes são neutras;
- 25% das soluções ácidas absorvem na zona de comprimentos de onda do vermelho;
- 5% das soluções são neutras e absorvem na zona de comprimentos de onda do vermelho;
- a probabilidade de uma solução absorver na zona de comprimentos de onda do vermelho é 0.2.

Seleccionando ao acaso uma solução no laboratório, determine a probabilidade de a solução:

a) ser neutra e não absorver na zona de comprimentos de onda do vermelho;

b) absorver na zona de comprimentos de onda do vermelho sabendo que é alcalina;

c) não ser neutra e não absorver na zona de comprimentos de onda do vermelho.

2.19. Estudos recentes indicam que a probabilidade de existir petróleo numa determinada região do Algarve é de 0.2. Sabe-se que, caso haja petróleo, a probabilidade de este sair na primeira perfuração é de 0.4.

a) Qual a probabilidade de sair petróleo na primeira perfuração?

b) Se na primeira perfuração não apareceu petróleo, qual a probabilidade de existir petróleo na região?

2.20. Sejam A e B acontecimentos de um espaço de resultados Ω.

a) Defina os seguintes conceitos "os acontecimentos A e B são independentes" e "os acontecimentos A e B são mutuamente exclusivos".

b) Se A e B são acontecimentos tais que $P(\overline{A}) = 0.8$, $P(B) = \alpha$ e $P(\overline{A} \cap \overline{B}) = 0.6$, determine α de modo que:

i) A e B sejam independentes;

ii) A e B sejam mutuamente exclusivos.

c) Prove que:

$$P(\overline{A} \cup (B \cap \overline{A})) + P(A \cup (B \cap \overline{A})) = 1 + P(B \cap \overline{A}).$$

2.21. Uma embarcação de recreio encontra-se desaparecida no mar. Considera-se que o seu desaparecimento seja devido a uma de três causas equiprováveis: afundou-se, foi sequestrada ou foi destruída por um temporal. Enviou-se uma equipa de busca e salvamento. Sabe-se que a probabilidade da equipa de salvamento resgatar com vida todos os tripulantes é $4/5$, $3/5$ e $7/10$, respectivamente, nos casos de afundamento, sequestro e temporal. Determine a probabilidade de:

a) Todos os tripulantes serem salvos.

b) Todos os tripulantes serem salvos nos casos em que não houve temporal.

c) Ter ocorrido um sequestro sabendo que todos os tripulantes foram salvos.

2.22. Seja Ω o espaço de resultados associado a uma dada experiência aleatória e A e B acontecimentos em Ω com $P(A) > 0$.

a) Defina o seguinte conceito: "os acontecimentos A e B são mutuamente exclusivos".

b) Se A e B são acontecimentos tais que $P(A) = 0.4$, $P(B) = 0.3$ e $P(A|\overline{B}) = 0.5$

i) Determine a probabilidade de que pelo menos um dos acontecimentos, A ou B, se realize.

ii) Mostre que $\overline{A}$ e $\overline{B}$ não são mutuamente exclusivos.

c) Prove que $P(A) \times [P(B|A) - 1] + P(\overline{A} \cup \overline{B}) = P(\overline{A})$.

Exercícios de Variáveis Aleatórias e Pares Aleatórios

2.23. Uma caixa contém 10 iogurtes, estando 4 estragados. Retiram-se 3 sem reposição. Seja X a variável aleatória que representa o número de iogurtes estragados retirados.

a) Construa a distribuição de probabilidades de X.

b) Represente graficamente a distribuição obtida na alínea a).

c) Determine a função distribuição cumulativa de X e represente-a graficamente.

d) Calcule $P(1 \leq X \leq 3)$.

2.24. Cinco bolas são retiradas, com reposição, de uma urna que contém 4 bolas vermelhas e 6 bolas brancas. Seja X a variável aleatória que representa o total de bolas vermelhas retiradas.

a) Determine a função massa de probabilidades de X.

b) Determine a função distribuição cumulativa de X e represente-a graficamente.

c) Calcule $P(1 \leq X \leq 3)$.

2.25. Num laboratório existem 40 ratos dos quais 30% são machos. Caracterize a variável aleatória X que conta o número de fêmeas que aparecem numa amostra de 8 ratos quando a selecção da amostra é feita:

a) Com reposição.

b) Sem reposição.

Determine o valor médio e a variância de X em cada uma das situações anteriores.

2.26. Seja X uma variável aleatória discreta com a seguinte distribuição de probabilidades:

x_i	-2	-1	0	1	2
$P(X = x_i)$	0.1	0.3	0.1	0.2	0.3

a) Calcule $E(X)$ e $Var(X)$.

b) Determine a função distribuição cumulativa de X.

c) Calcule $P(X \geq 0 | X < 2)$.

d) Determine a distribuição de probabilidades da variável aleatória $Y = X^2$.

2.27. O número de televisores encomendados mensalmente em determinada loja é bem descrito por uma variável aleatória X com a seguinte função distribuição cumulativa:

$$F(x) = \begin{cases} 0 & \text{se } x < 0 \\ 0.1 & \text{se } 0 \leq x < 1 \\ 0.3 & \text{se } 1 \leq x < 2 \\ 0.6 & \text{se } 2 \leq x < 3 \\ 1 & \text{se } x \geq 3 \end{cases}$$

a) Determine a função massa de probabilidade da variável aleatória X.

b) Quantos televisores deve ter a loja em *stock*, por mês, para que a probabilidade de satisfazer todas as encomendas seja superior a 0.95?

c) Se num dado mês a loja só tiver 2 televisores em *stock*, determine a distribuição de probabilidades da variável aleatória que representa a diferença, em valor absoluto, entre as encomendas e o *stock*.

2.28. Seja X uma variável aleatória discreta com a seguinte distribuição de probabilidades:

x_i	$1-2k$	$k-1$	k	$2k$
p_i	p	$3p$	p	p

a) Sabendo que $E(X) = 1/3$ calcule o valor de p e k.

b) Calcule $Var[X]$.

c) Deduza a distribuição de probabilidades da variável aleatória $Y = X^2$.

2.29. Seja X uma variável aleatória com valor médio μ e variância σ^2. Seja $Y = \frac{X-\mu}{\sigma}$. Determine $E[Y]$, $Var[Y]$ e $E[Y^2]$.

2.30. Considere a função densidade de probabilidade de uma variável aleatória contínua X assim definida:

$$f(x) = \begin{cases} a\, x^2 & 0 \leq x \leq 1 \\ 0 & \text{outros valores de } x. \end{cases}$$

a) Determine a.

b) Determine a função de distribuição cumulativa de X.

c) Calcule o valor b tal que a probabilidade de X exceder b seja igual a 0.05.

2.31. Considere que a concentração de poluente (em unidades adequadas) nos gases emitidos por uma fábrica é uma variável aleatória X com função densidade

$$f(x) = \begin{cases} 0 & \text{se } x \leq 0 \\ \frac{1}{2}(1-e^{-2})x & \text{se } 0 < x < 2 \\ e^{-x} & \text{se } x \geq 2 \end{cases}$$

a) Determine a função distribuição cumulativa de X.

b) Indique duas propriedades que uma qualquer função distribuição cumulativa deve verificar. Averigue se a função que obteve na alínea anterior verifica essas propriedades.

c) Qual é a probabilidade de a concentração ser superior a 2?

d) A concentração mediana será superior ou inferior a 2? Justifique.

2.32. Considere a variável aleatória contínua X com a seguinte função distribuição cumulativa:

$$F(x) = \begin{cases} 0 & \text{se } x < 0 \\ ax + b & \text{se } 0 \leq x < \pi \\ 1 & \text{se } x \geq \pi. \end{cases}$$

a) Determine a e b.

b) Determine a função densidade de probabilidade de X.

c) Calcule o primeiro quartil da distribuição de X.

d) Calcule $P(X < \frac{\pi}{2} | X \geq \frac{\pi}{4})$.

e) Calcule:

i) $E(X)$ e $Var(X)$;

ii) $E\left[\dfrac{1}{X+2}\right]$ e $E\left[3X+2\right]$

2.33. Um fornecedor de um produto de laboratório tem uma capacidade de armazenamento de 150 kg. No início de cada mês é reposto o stock até à capacidade máxima de armazenamento. As vendas mensais deste produto em centenas de kg são dadas por uma variável aleatória X, cujo comportamento é bem descrito pela seguinte função densidade de probabilidade

$$f(x) = \begin{cases} x & \text{se} \quad 0 \leq x \leq 1 \\ 1 & \text{se} \quad 1 < x \leq 1.5 \\ 0 & \text{outros valores de } x. \end{cases}$$

a) Determine a função distribuição cumulativa da v.a. X.

b) Se a dada altura de um mês já foram vendidos 50 kg do produto, qual a probabilidade de se venderem mais de 100 kg no fim desse mês?

c) Indique o valor médio mensal e a mediana mensal das vendas daquele produto.

d) O lucro, Y, da venda do referido produto é função de vários factores. Considere a seguinte expressão (simplificada) do lucro em função das vendas $Y = 50X - 25$.

i) Qual o valor esperado do lucro mensal?

ii) Qual a probabilidade de, num dado mês, não haver prejuízo ?

2.34. Seja X uma variável aleatória contínua cuja função distribuição cumulativa é:

$$F(x) = \begin{cases} 0 & \text{se } x < 0 \\ 3x/4 & \text{se } 0 \leq x < 1 \\ kx - x^2/4 & \text{se } 1 \leq x < 2 \\ k & \text{se } x \geq 2 \end{cases}$$

a) Determine a constante k.

b) Obtenha a função densidade de probabilidade de X e faça a sua representação gráfica.

c) Calcule o valor esperado e a mediana de X.

d) Determine:

i) $P[X \geq 1]$ ii) $P[X < 3/2 | X > 1/2]$.

2.35. Num processo de inventário concluíu-se que a raridade de determinada espécie animal era directamente proporcional à área observada até que se avistasse um exemplar da espécie, associada ao percurso de amostragem. Considere então a v.a. X designando a distância percorrida até se avistar algum exemplar da espécie, com função densidade dada por

$$f(x) = \begin{cases} \frac{b}{k\ x^2} & 1 \leq x \leq b \\ 0 & \text{restantes valores de } x. \end{cases}$$

a) Indique quais as condições que k e b devem verificar de modo que $f(x)$ seja uma função densidade.

b) Determine a função de distribuição cumulativa de X.

c) Calcule a mediana de X.

d) Considere a v.a. $Y = C_0 + C_1 X$, que caracteriza o custo de amostragem, onde C_0. designa os custos fixos e C_1 o custo por unidade de percurso. Determine o custo esperado para o inventário.

2.36. A concentração (em ppm) de monóxido de carbono (CO) na atmosfera, numa determinada zona urbana, no período das 8h às 10h, pode ser considerada uma variável aleatória X com a seguinte função densidade

$$f(x) = \begin{cases} 0 & \text{se } x \leq 0 \\ 3.4e^{-3.4x} & \text{se } x > 0 \end{cases}.$$

a) Determine a função distribuição cumulativa de X.

b) Qual é a probabilidade de, numa amostra recolhida ao acaso naquele período, a concentração de CO exceder 2 ppm?

c) Mostre que a função geradora de momentos de X é

$$M_X(t) = \frac{3.4}{3.4 - t} \quad \text{para} \quad t < 3.4.$$

d) Determine as concentrações média e mediana de CO naquele período.

2.37. A proporção (em volume) de determinada substância química num produto pode ser considerada uma variável aleatória X com função densidade de probabilidade da forma:

$$f(x) = \begin{cases} ax^3 + x & 0 \leq x \leq 1 \\ 0 & \text{o. v. de } x \end{cases}$$

a) Determine a.

b) Determine a função distribuição cumulativa de X.

c) Sempre que a proporção da referida substância é inferior a 0.5 o produto é de qualidade inferior. Qual a probabilidade de num produto que foi considerado de qualidade inferior a proporção daquela substância ser superior a 0.25?

d) Determine o valor esperado e a variância da proporção de substância no produto.

e) Seja Y o valor comercial (em euros) de cada unidade de volume de produto, função da proporção da substância, assim definido $Y = 2X$.

i) Determine a função densidade de probabilidade de Y.

ii) Qual a probabilidade de uma unidade de volume daquele produto ser vendida por mais de 1 Euro?

iii) Exprima $Cov[X, Y]$ em termos de $Var[X]$ e determine o seu valor.

2.38. Considere a função real de variável real assim definida:

$$f(x) = ke^{-|x|}, \quad k \in \mathbb{R} \quad \text{e} \quad x \in \mathbb{R}.$$

a) Determine o valor de k de modo que seja função densidade de probabilidade de uma variável aleatória X.

b) Determine a função de distribuição cumulativa de X.

c) Calcule o valor médio, variância e mediana de X.

d) Calcule $P(|X - E(X)| < 1)$.

e) Mostre que a função geradora de momentos de X é $M_X(t) = \frac{1}{1-t^2}$, $|t| < 1$.

f) Determine a função densidade da v.a. $Y = X^2$.

2.39. Sejam X e Y o número de vitelos e leitões, respectivamente, que nascem anualmente numa pequena quinta de criação de gado.

O quadro seguinte dá a distribuição de probabilidades conjunta de X e Y.

X \ Y	6	7	8	9	10
2	0.05	0.04	0.04	0.10	0.02
3	0.13	0.10	0.10	0.10	0.07
4	0.07	0.06	0.06	0.05	0.01

a) Escreva a distribuição de probabilidades do número de leitões nascidos por ano nesta quinta.

b) Serão X e Y v.a. independentes? Justifique.

c) Considere o número total de leitões e vitelos que nascem anualmente. Qual o valor esperado para aquele total?

d) Determine a distribuição de probabilidades do número de vitelos que nascem num dado ano, sabendo que nasceram 7 leitões.

2.40. Num grupo de casais em que ambos os elementos estão empregados, a distribuição de probabilidade conjunta do salário da mulher (X) e do homem (Y) é dada por

X \ Y	1000	1500	2000
500	0.05	0.1	0.15
1000	0.1	0.2	0.1
1500	0.15	0.1	0.05

a) Num casal escolhido ao acaso, qual é a probabilidade de os salários do homem e da mulher serem iguais?

b) Determine as distribuições de probabilidade marginais.

c) Haverá independência entre os salários do homem e da mulher? Justifique.

d) Qual é o valor médio e a variância do salário de um casal?

e) Num casal em que o homem ganha 1000 euros, qual é a probabilidade de a mulher ganhar tanto ou mais do que o homem?

f) Determine a distribuição de probabilidade do salário de um casal.

2.41. Um cliente de uma livraria pode fazer encomendas de livros estrangeiros em inglês e francês que não existam em stock.

O número de livros em inglês e francês encomendados semanalmente é o par aleatório (X, Y) com a seguinte distribuição de probabilidades:

X \ Y	1	2	3	4
0	0.01	0.02	0.04	0.03
1	0.05	0.10	0.20	0.15
2	0.04	0.08	0.16	0.12

a) Qual a probabilidade de numa semana serem encomendados no máximo dois livros?

b) Qual a percentagem de semanas em que existe igualdade de livros ingleses e franceses encomendados?

c) Determine as funções distribuição cumulativa marginais das variáveis X e Y.

d) Calcule o coeficiente de correlação entre as variáveis X e Y e interprete o seu resultado.

e) Qual a distribuição de probabilidades da variável aleatória "Número total de livros em inglês e francês encomendados semanalmente"?

f) Qual a probabilidade de numa semana se encomendar pelo menos um livro em inglês sabendo que foram encomendados dois livros em francês?

2.42. Seja (X, Y) o par aleatório com a seguinte função densidade de probabilidade conjunta:

$$f(x,y) = \begin{cases} a\ (x+y) & 1 \leq x \leq 2 \quad 1 \leq y \leq 2 \\ 0 & \text{restantes valores de } (x,y). \end{cases}$$

a) Determine o valor da constante a.

b) X e Y são variáveis aleatórias independentes? Justifique.

c) Calcule $P[Y < X]$ e $P[Y > 3 - X]$. Comente.

d) Determine o valor médio de $1/X$.

e) Determine a função densidade condicional de $Y|X = 3/2$.

f) Calcule $P[Y < 3/2|X = 3/2]$.

2.43. Seja (X, Y) um par aleatório contínuo com função densidade de probabilidade conjunta assim definida:

$$f(x,y) = \begin{cases} 3y & \text{se } 0 \leq x \leq y \leq 1 \\ 0 & \text{outros valores de } (x,y). \end{cases}$$

a) Determine as densidades marginais de X e de Y.

b) X e Y são variáveis aleatórias independentes? Justifique a sua resposta.

c) Calcule a função distribuição cumulativa de Y.

d) Calcule $P[X < 1/4 | Y = 1/2]$.

2.44. Seja (X, Y) a variável aleatória bidimensional contínua com a seguinte função densidade conjunta:

$$f(x, y) = \begin{cases} 4xye^{-x^2-y^2} & \text{se } x > 0 \text{ e } y > 0 \\ 0 & \text{para outros valores.} \end{cases}$$

a) Determine as funções densidade marginais de X e Y.

b) X e Y são variáveis aleatórias independentes? Justifique.

c) Calcule $Cov(X, Y)$.

2.45. Um depósito de água para rega tem uma capacidade de 20000 litros. Seja X a variável aleatória que designa a quantidade de água, <u>em dezenas de milhares de litros</u>, que existe no depósito, no início de cada semana. Seja Y a quantidade, também <u>em dezenas de milhares de litros</u>, gasta na rega durante a semana. Suponha que X e Y têm função densidade de probabilidade conjunta assim definida:

$$f(x, y) = \begin{cases} xy/2 & \text{se} \quad 0 \leq y \leq x \leq 2 \\ 0 & \text{outros valores de } (x, y). \end{cases}$$

a) Determine as densidades marginais de X e de Y. Serão X e Y variáveis aleatórias independentes? Justifique.

b) Qual a probabilidade de a quantidade de água gasta na rega durante uma dada semana ser inferior a 5000 litros, sabendo que no início dessa semana:

  i) O depósito contém mais de 10000 litros de água?

  ii) O depósito contém exactamente 10000 litros de água?

c) Qual a probabilidade de numa dada semana se gastar exactamente a quantidade existente no depósito no início da semana?

d) Qual a quantidade média de água existente no depósito no início de cada semana?

2.46. Seja (X, Y) um par aleatório cuja função densidade de probabilidade conjunta é assim definida:

$$f(x, y) = \begin{cases} 2y\, e^{-x} & \text{se } x > 0 \text{ e } 0 < y \leq 1 \\ 0 & \text{outros valores de } (x, y). \end{cases}$$

a) Determine as densidades marginais de X e de Y e identifique a distribuição de X.

b) Serão X e Y variáveis aleatórias independentes? Justifique.

c) Determine $E[X]$, $E[Y]$ e $Cov[X,Y]$.

d) Calcule $P[Y > \sqrt{X}]$.

2.47. Sejam X e Y variáveis aleatórias independentes, com funções densidade de probabilidade definidas, respectivamente, por:

$$f_X(x) = \begin{cases} 2(1-x) & \text{se } 0 < x < 1 \\ 0 & \text{outros valores de } x \end{cases} \quad \text{e} \quad f_Y(y) = \begin{cases} 3y^2 & \text{se } 0 < y < 1 \\ 0 & \text{outros valores de } y. \end{cases}$$

a) Determine, justificando, a função densidade conjunta de (X,Y).

b) Determine $P[X > \frac{1}{2} | \frac{1}{3} < Y < \frac{2}{3}]$.

c) Calcule $E[XY]$.

2.48. Num projecto duas tarefas A e B decorrem de forma independente. Sejam X e Y as v.a.s contínuas que representam o tempo (em anos) de duração das tarefas A e B, respectivamente. As funções densidade de X e de Y são:

$$f_X(x) = \begin{cases} 6x(1-x) & \text{se } 0 \leq x \leq 1 \\ 0 & \text{outros valores de } x \end{cases} \; ; \qquad f_Y(y) = \begin{cases} 1/2 & \text{se } 0 \leq y \leq 2 \\ 0 & \text{outros valores de } y. \end{cases}$$

a) Determine a função distribuição cumulativa de X.

b) Qual é a probabilidade de a tarefa A durar mais de meio ano?

c) Qual é o tempo médio de duração da tarefa A?

d) Determine a função densidade conjunta do par (X,Y). Justifique.

e) Calcule a covariância entre as variáveis X e Y. Justifique.

f) Determine a probabilidade de a tarefa A durar mais tempo do que a tarefa B.

g) Determine a probabilidade de a tarefa A durar mais de meio ano se:

   i) a tarefa B durar menos de 1 ano.

   ii) a tarefa B durar 1 ano.

h) Determine:

   i) O valor médio do quociente $X/(Y+1)$.

   ii) $Var[X-Y]$

   iii) $Cov[X, Y-X]$

Principais Distribuições Discretas e Contínuas

2.49. Uma experiência aleatória pode dar dois resultados: êxito ou fracasso, sendo a probabilidade de resultar em êxito de 0.9. O custo de uma experiência que resulte em êxito é de 5 euros e em fracasso de 10 euros. A experiência é repetida 20 vezes, de forma independente. Seja X a variável aleatória que conta o número de êxitos.

a) Identifique a distribuição de X.

b) Calcule $P[X > 15]$.

c) Calcule a probabilidade de haver mais êxitos do que fracassos.

d) Mostre que o custo total C das 20 experiências pode ser expresso como $C = 200 - 5X$.

e) Calcule $E(C)$.

f) Calcule $P[C < 125]$.

2.50. Uma dada experiência biológica analisa cobaias. Cada vez que se repete a referida experiência, uma cobaia diferente é analisada e cada repetição só usa uma cobaia. Sabendo que a experiência é bem sucedida em 40% dos casos, calcule:

a) A probabilidade de ter pelo menos duas experiências bem sucedidas, se tiver 10 cobaias.

b) O número de cobaias necessário para que o número esperado de sucessos seja 24.

c) O número de cobaias necessário para que a probabilidade de obter pelo menos uma experiência com sucesso não seja inferior a 0.95.

2.51. Numa experiência laboratorial injectaram-se n cobaias com uma determinada droga que inibe a síntese de proteínas. A probabilidade de uma cobaia morrer durante a experiência, devido à droga, é 0.2. Seja X a v. a. que designa "o número de cobaias que morreram durante a experiência".

a) Identifique o modelo de probabilidade para X.

b) Supondo que se utilizaram 15 cobaias, qual a probabilidade de:

   i) Duas delas morrerem?

   ii) Pelo menos 8 delas sobreviverem?

c) Determine o número mínimo de cobaias que têm que ser utilizadas na experiência de modo que a probabilidade de morrer pelo menos uma cobaia seja superior a 0.95.

2.52. Seja X uma variável aleatória com distribuição $B(n;p)$ e Y a variável aleatória definida por $Y = \dfrac{X}{n}$. Calcule:

a) $E[Y]$, $Var[Y]$ e $E[Y^2]$;

b) A função geradora de momentos de Y.

2.53. Um instrumento para medição de humidade é constituído por 3 componentes instaladas de forma independente. Todas as componentes têm a mesma probabilidade de avariar, igual a 0.1.

a) Identifique e caracterize a variável aleatória que designa o número de componentes avariadas no referido instrumento.

b) Determine a probabilidade de duas ou mais componentes avariarem.

c) Para os instrumentos nas condições referidas, se nenhuma das 3 componentes está avariada, o instrumento funciona sempre, se uma única componente está avariada o instrumento tem probabilidade 0.6 de funcionar e se 2 ou mais componentes estão avariadas, não funciona. Determine a probabilidade de:

   i) O instrumento funcionar.

   ii) Existir uma componente avariada se o instrumento está a funcionar.

d) Numa encomenda de 10 dos instrumentos das alíneas anteriores para instalar em 10 locais diferentes, quantos instrumentos se espera virem a falhar?

2.54. A probabilidade de um atirador acertar num alvo é $p = 1/4$.

a) Seja X a variável aleatória que conta o número de tiros necessários até acertar, pela primeira vez, no alvo. Determine n tal que $P[X \leq n] > 0.8$.

b) Quantos tiros espera o atirador dar até acertar pela primeira vez no alvo?

2.55. O Duarte vai posicionar-se na linha de lançamento livre num campo de basquetebol e atirar até fazer um cesto. Se admitirmos que os lançamentos são independentes e de probabilidade de acertar constante e igual a 0.8, determine:

a) a probabilidade de necessitar de menos de 5 lançamentos para acertar;

b) o número esperado de lançamentos que tem que efectuar para acertar.

2.56. Sejam X_1 e X_2 duas variáveis aleatórias independentes. Determine a distribuição de $X_1 + X_2$ quando:

a) X_1 e X_2 têm distribuição binomial de parâmetros (n_1, p) e (n_2, p), respectivamente.

b) X_1 e X_2 têm distribuição geométrica, com o mesmo parâmetro p.

2.57. Uma pessoa planta 6 bolbos, escolhidos ao acaso de uma caixa que contém 5 bolbos de túlipa e 4 bolbos de junquilho. Qual a probabilidade de essa pessoa plantar 2 bolbos de junquilho e 4 de túlipa?

2.58. Um método frequentemente utilizado para estimar o número de animais de uma dada espécie num certo habitat é o da captura-recaptura. O método pode ser exemplificado pela seguinte situação:

> Num lago são capturados, marcados e devolvidos à água 5 peixes de uma certa espécie. Passado algum tempo (a fim de permitir que os peixes marcados se distribuam aleatoriamente pelo lago, embora não convenha deixar passar demasiado tempo, para se poder admitir que a dimensão da população permaneceu constante) são pescados 4 peixes dessa mesma espécie e conta-se quantos de entre eles estão marcados, o que será representado pela variável aleatória X.

a) Qual a probabilidade de nenhum dos 5 peixes marcados ser recapturado, se existirem 10 peixes da referida espécie no lago? E se existirem 100?

b) A ideia do método de captura-recaptura consiste em considerar o tamanho da população como sendo aquele que torna mais provável o valor de X que resultou de uma experiência deste tipo. Assim, por exemplo, qual dos 4 valores $N = 10$, $N = 20$, $N = 100$ ou $N = 1000$, considera mais plausível para o tamanho da população se:

   i) da experiência resultou $X = 1$;

   ii) da experiência resultou $X = 2$.

2.59. Num armazém encontra-se um lote de 10000 latas de um certo produto alimentar. Deste lote, 500 latas já ultrapassaram o prazo de validade. O organismo responsável pelo controle de qualidade vai analisar uma amostra. Considere que o processo de amostragem é feito <u>com reposição</u>.

a) Suponha que as latas são inspeccionadas sucessivamente <u>até encontrar uma fora do prazo de validade</u>.

   i) Identifique e caracterize a variável em estudo.

   ii) Qual a probabilidade de ser necessário inspeccionar 3 ou mais latas?

   iii) Qual o número esperado de latas inspeccionadas?

b) Suponha agora que retira uma amostra de 15 latas. O controle de qualidade rejeita o lote se nessa amostra forem encontradas mais do que duas latas fora do prazo de validade.

   i) Calcule a probabilidade de rejeição do lote.

   ii) O que pode dizer do valor calculado na alínea anterior no caso de o processo de amostragem ser feito <u>sem reposição</u>? Justifique.

2.60. O número de petroleiros que chega a uma certa refinaria, em cada dia, é uma v.a. X com distribuição de Poisson de parâmetro $\mu = 2$. As actuais instalações portuárias da refinaria podem atender até 3 petroleiros por dia. Se mais de 3 petroleiros chegam num dia, os petroleiros em excesso são enviados para outro porto.

a) Qual a probabilidade de, num dado dia, a refinaria ter de recusar petroleiros?

b) Qual deverá ser a capacidade de atendimento da refinaria para permitir o acolhimento de todos os petroleiros que chegam em cerca de 95% dos dias?

c) Qual o número esperado de petroleiros chegados por dia?

d) Qual o número mais provável de petroleiros chegados num dia?

e) Qual a probabilidade de em dois dias chegarem 5 petroleiros?

f) Qual o número esperado de petroleiros atendidos num dia?

g) Qual o número esperado de petroleiros recusados num dia?

2.61. Uma empresa de aluguer de autocarros para excursões de longo curso dispõe de 5 veículos. Sabe-se, pela análise do seu comportamento, que a procura semanal de veículos segue uma distribuição de Poisson de média 4.

a) Determine a probabilidade de, em certa semana, um dos autocarros não ser alugado.

b) Qual a probabilidade de, em duas semanas, serem procurados 6 veículos?

c) Determine o valor esperado do número de clientes que em certa semana não podem ser atendidos por já estarem alugados todos os autocarros.

2.62. Duas máquinas A e B produzem 10% e 90% da produção total de um dado artigo, respectivamente. Suponha que 5% dos artigos fabricados por cada uma das máquinas são defeituosos.

a) Qual a probabilidade de um artigo defeituoso ter sido fabricado pela máquina A?

b) De um lote bastante grande do referido artigo, é retirada uma amostra aleatória de 50 artigos. Qual a probabilidade de encontrar no máximo 10 artigos defeituosos? E 5?

c) Qual o número máximo de artigos que deverá tirar ao acaso da produção total para que a probabilidade de não encontrar defeituosos seja superior a 0.80?

2.63. Suponha que o tempo, em minutos, necessário para que um técnico afine cada componente de uma máquina numa linha de produção é bem descrito por uma variável aleatória X com distribuição $U(5;10)$.

a) Qual a probabilidade de que uma componente escolhida ao acaso:

i) Tenha necessitado de mais de 7 minutos para ser afinada.

ii) Tenha exigido ao técnico um tempo de afinação inferior a 9 minutos sabendo que aquele tempo foi superior a 7 minutos.

b) Seja $Y = 10 + 5X$ o custo em euros de afinação de cada componente. Qual o custo esperado de afinação de uma componente escolhida ao acaso?

2.64. Uma empresa agro-química fabrica mensalmente 90 toneladas de um dado produto. Sabendo que a procura mensal deste produto é uma variável aleatória aproximadamente normal de parâmetros μ =80 ton e σ =10 ton, calcule:

a) A probabilidade de a procura mensal do produto se situar entre 68 e 90 toneladas;

b) A probabilidade de haver num mês procura insatisfeita;

c) A produção necessária para que a probabilidade de haver procura mensal insatisfeita seja 0.025.

2.65. Um grossista de distribuição de fruta recebe do produtor pêssegos de quatro categorias: extra, A, B e C. Da experiência anterior, sabe-se que o diâmetro de um pêssego é uma variável aleatória com distribuição aproximadamente normal de média 64 mm e desvio padrão 3 mm. A classificação do referido fruto em função do seu diâmetro é a seguinte:

Categoria	Diâmetro (x) em mm
C	$x \leq 60$
B	$60 < x \leq 65$
A	$65 < x \leq 70$
Extra	$x > 70$

Atendendo aos custos de armazenamento e de distribuição, admite-se que o lucro líquido por tonelada é de 80 euros para a categoria extra, 50 euros para a categoria A, 10 euros para a categoria B e -5 euros para a categoria C.

Qual o lucro líquido esperado de um fornecimento constituído por uma tonelada de pêssegos?

2.66. O comprimento da raiz principal de um determinado arbusto é bem modelado por uma distribuição normal de média 50cm e desvio padrão 5cm.

a) Qual a probabilidade de um arbusto seleccionado ao acaso ter uma raíz principal com comprimento superior a 54cm?

b) Em 200 arbustos desta espécie, qual o número deles em que se espera que as raízes principais tenham comprimento superior a 54cm?

c) Qual a probabilidade de a raíz principal de um arbusto seleccionado ao acaso ter comprimento entre 50 e 56cm?

d) Qual a probabilidade de a média dos comprimentos da raiz principal de 9 arbustos seleccionados ao acaso ser superior a 52cm?

2.67. Admita que o conteúdo (em ml) de frascos de certo xarope segue uma distribuição normal de parâmetros $\mu = 200\ ml$ e $\sigma = 2\ ml$.

a) Qual a probabilidade de um frasco adquirido ao acaso ter menos de 195 ml de xarope?

b) Se comprar 5 frascos, qual a probabilidade de ficar em casa com menos de 0.99 l de xarope?

c) Num lote de 30 frascos qual a probabilidade, aproximada, de haver no máximo 5 com menos de 195 ml?

2.68. Uma fábrica produz motores cujo tempo de vida é uma variável aleatória com distribuição normal de parâmetros μ =10 anos e σ =2 anos. A fábrica quer criar um período de garantia para os motores de forma a que não mais de 3% tenham de ser substituídos. Qual deverá ser o período de garantia máximo oferecido pela fábrica?

2.69. Um produto pesa em média 10g com desvio padrão de 2g. Este produto é embalado em caixas com 50 unidades cada. Sabe-se que as caixas vazias pesam em média 500g com desvio padrão de 25g. Admita que as variáveis peso do produto e da caixa vazia são independentes com distribuição normal.

a) Qual é a probabilidade de numa caixa encontrar no máximo 40 unidades do referido produto com peso inferior a 8g cada?

b) Qual é a probabilidade de uma caixa cheia pesar mais do que 1050g?

2.70. Seja Y uma variável aleatória contínua com a seguinte função geradora de momentos:

$$M_Y(t) = e^{3t+8t^2}, \qquad t \in \mathbb{R}$$

a) Calcule a função geradora de momentos da variável aleatória $X = \frac{Y-3}{4}$.

b) Determine o valor médio e a variância de X.

c) Se $W \frown N(\mu, \sigma)$ então a função geradora de momentos de W é definida por

$$M_W(t) = e^{\mu t + \frac{1}{2}t^2\sigma^2}, \qquad t \in \mathbb{R}.$$

Identifique as distribuições de X e Y.

2.71. Um determinado modelo de avião pode transportar uma carga máxima (passageiros e bagagens) de 9000kg. Admita que o peso da bagagem de um passageiro é uma variável aleatória com distribuição $N(18, 5)$, que o peso de um passageiro-homem é uma variável aleatória com distribuição $N(70, 10)$ e que o peso de um passageiro-mulher é uma variável aleatória com distribuição $N(60, 10)$.

a) Qual é o peso da bagagem de um passageiro que é excedido por 20% dos passageiros, no máximo?

b) Considere um casal (homem e mulher) que entra no avião. Qual a probabilidade de o peso da mulher ser superior ao do homem? Que hipóteses tem de admitir para responder a esta questão?

c) Num determinado vôo a lotação do avião está completa com 80 homens e 20 mulheres, que levam a respectiva bagagem. Qual a probabilidade de o avião não poder partir por excesso de carga?

d) A companhia pratica a cobrança de uma taxa para bagagens com peso superior a 20kg. Havendo 60 passageiros num vôo, qual é a probabilidade de que mais de 10 passageiros paguem a referida taxa.

2.72. Num restaurante, a procura semanal (em litros) de uma determinada marca de cerveja segue uma distribuição normal com desvio padrão 4 litros. Sabe-se que em 50% das semanas a procura é inferior a 50 litros.

a) Determine o valor médio e a variância da variável aleatória que designa a procura semanal de cerveja daquela marca.

b) Calcule a probabilidade de, numa semana, a procura ser superior a 60 litros.

c) Qual deverá ser a quantidade de cerveja que deve existir em cada semana de modo que a probabilidade de haver ruptura de stock seja igual a 0.015?

d) Sendo cada litro de cerveja vendido a 5 euros, caracterize a lei da venda mensal (4 semanas) de cerveja.

2.73. O número de ovos postos por segundo em certo aviário pode ser considerado uma v.a. X com distribuição de Poisson de valor médio 1.

a) Mostre que $P[X = 0] = P[X = 1]$ e que para todo o $k > 1$ se tem $P[X = k] < P[X = k - 1]$. Comente.

b) Qual a probabilidade de num período de 5 segundos serem postos mais do que três ovos? Justifique.

c) Sabendo que numa dada região existem 40 aviários idênticos a este, qual a probabilidade, aproximada, de o número total de ovos postos por segundo nessa região ser inferior a 30? Justifique.

2.74. A produção semanal de uma plantação de bananas é bem modelada por uma variável aleatória com distribuição normal de média 5 t e desvio padrão 2 t. Admita que a produção é independente de semana para semana.

a) Numa semana escolhida ao acaso, qual a probabilidade de a produção de bananas ser inferior a 3 t? E superior a 7 t?

b) A plantação tem despesas fixas semanais no valor de 2000€. O produtor vende as bananas a 500€/t. Qual a probabilidade de, numa dada semana em que toda a produção é escoada, o valor da venda ser inferior às despesas fixas? Justifique.

c) Qual a probabilidade de num mês (4 semanas) se conseguir satisfazer uma encomenda de 23 t de bananas? Justifique.

d) Determine a probabilidade de, em 8 semanas escolhidas ao acaso, haver no máximo uma em que a produção foi inferior a 3 t.

2.75. Suponha que os elos de uma corrente de bicicleta têm comprimentos aleatórios com distribuição normal de média 0.5 cm e desvio padrão 0.04cm. As normas de um fabricante de bicicletas exigem que o comprimento de uma corrente esteja compreendido entre 49 e 50 cm .

a) Qual a percentagem de elos cujo comprimento excede 0.6cm?

b) Se uma corrente tiver 100 elos, qual a proporção de correntes a satisfazer as normas exigidas?

c) Utilizando apenas 99 elos, que valor deverá assumir o desvio padrão para que 90% das correntes satisfaça as normas do fabricante?

2.76. Suponha que o peso de um pacote de farinha é uma variável aleatória que se admite ter valor médio 1 kg e desvio padrão 0.05 kg. Pretende-se arrumar 99 pacotes numa prateleira que suporta 100 kg. Qual o risco de a prateleira desabar? Justifique.

2.77. A produção diária de azeite num lagar segue uma distribuição normal com desvio padrão 50 litros. Em 50% dos dias de laboração, a produção é inferior a 950 litros. Admita que a produção é independente de dia para dia.

a) Indique o valor da produção que não é excedido em 95% dos dias.

b) Qual a percentagem de dias em que a produção se afasta da média mais do que 100 litros?

c) Qual a probabilidade de a produção mensal (30 dias) ser inferior a 28000 litros?

2.78. A temperatura diária em graus Celsius registada durante o verão de um ano, numa certa localidade às 12h, é uma v.a. X que segue uma distribuição normal com $\mu = 27°\text{C}$ e $\sigma = 2°\text{C}$.

a) Qual a probabilidade de se registar uma temperatura inferior a 28.7°C?

b) Considere as temperaturas referentes a 10 anos, no verão, às 12h no mesmo local. Qual a probabilidade de se registar uma temperatura inferior a 28.7°C em mais de 3 anos?

c) Um turista americano que planeia visitar aquele local transforma a temperatura em graus Fahrenheit através da relação $Y = \frac{9}{5}X + 32$.

i) Qual é a distribuição da temperatura expressa em graus Fahrenheit? Justifique.

ii) Determine a probabilidade de o turista ter que enfrentar temperaturas superiores a 90°F.

2.79. Num edifício funcionam 7 elevadores. A carga máxima de cada elevador é de 320 kg. A dada altura entra um grupo de 4 pessoas em cada um dos elevadores. Calcule a probabilidade de no máximo 3 elevadores não funcionarem se o peso de uma pessoa for considerado uma variável aleatória com distribuição normal de média 71,75 kg e desvio padrão 10 kg.

2.80. O número de avarias por mês nos comboios da linha de Sintra que provocam a interrupção da circulação é uma variável aleatória X com distribuição de Poisson de parâmetro $\lambda = 3.5$. O número de avarias num dado mês é independente do número de avarias nos outros meses.
Por outro lado, o tempo necessário para restabelecer a circulação ferroviária após uma avaria é uma variável aleatória Y com distribuição $\mathcal{N}(2.5\ ,0.75)$ (em horas), e também aqui, o tempo de reparação após uma avaria é independente dos tempos de reparação após outras avarias.

a) Qual o número esperado de avarias num dado mês? E num ano?

b) Qual a probabilidade de a interrupção da circulação após uma avaria exceder 4.5 horas?

c) Qual a probabilidade de o tempo total de interrupção da circulação na linha de Sintra exceder 8 horas, se houver 2 avarias num mês?

d) Qual a distribuição de probabilidades da variável aleatória que conta o tempo total de interrupção da circulação num mês, se houver r avarias ? Qual o seu valor esperado?

2.81. As maçãs colhidas num pomar são classificadas, de acordo com o seu peso, em três categorias: pequenas, médias e grandes. As maçãs pequenas são aquelas cujo peso é inferior a 45 gramas e as grandes aquelas cujo peso excede 75 gramas. Suponha que o peso de uma maçã (em gramas) é uma variável aleatória X, que se admite seguir uma lei normal com valor médio 60 gramas e desvio padrão 15 gramas.

a) Qual a probabilidade de uma maçã escolhida ao acaso ser classificada como grande?

b) O lucro (em cêntimos) da venda de cada maçã está relacionado com o seu peso de acordo com a expressão $L = 12 - X^2/350$. Determine o lucro esperado da venda de cada maçã.

c) Enchem-se sacos com 10 maçãs, seleccionadas ao acaso. Qual a probabilidade de um saco, escolhido ao acaso, pesar mais de 0.5 kg?

d) Num dia em que se colheram 3000 maçãs qual a probabilidade, aproximada, de mais de 450 serem grandes?

e) As maçãs defeituosas vão para sumo. Sabe-se que vão para sumo 2% das maçãs grandes, 5% das médias e 15% das pequenas. Escolhida uma maçã ao acaso qual a probabilidade de ir para sumo?

2.82. Uma pessoa vai diariamente a uma pastelaria tomar o pequeno almoço. Suponha que o tempo de espera para ser atendida é uma variável aleatória com distribuição exponencial de média 4 minutos. Qual a probabilidade de:

a) A pessoa ser atendida em menos de três minutos?

b) Em pelo menos 4 dias de uma semana (considere a semana com 7 dias e o tempo de atendimento independente de dia para dia) a pessoa ser atendida em menos de três minutos?

2.83. Numa indústria alimentar o consumo diário de uma certa matéria prima (em toneladas) é uma variável aleatória com distribuição exponencial de valor médio 2.

a) Qual o valor do consumo diário mediano daquela matéria prima?

b) Qual deve ser o *stock* da matéria prima no início de cada dia de modo que a probabilidade de ruptura de *stock* seja 0.02?

c) Considere de novo a situação da alínea anterior e suponha que a ruptura de *stock* é independente de dia para dia.

   i) Identifique e caracterize a distribuição do número de dias de um mês (30 dias) em que há ruptura de *stock*.

   ii) Qual a probabilidade, aproximada, de num mês (30 dias) haver mais de 2 dias com ruptura de *stock*?

2.84. O concentrado de tomate é vendido em bisnagas que permitem uma menor redução da exposição ao ar. O diâmetro de abertura destas bisnagas é uma variável aleatória, X, com distribuição normal de parâmetros $\mu = 5$ mm e $\sigma = 0.6$ mm.

a) Calcule a probabilidade de, numa bisnaga escolhida ao acaso, o diâmetro de abertura medir
i) menos de 3.5 mm   ii) 5 mm.

b) Considere uma amostra de 100 bisnagas escolhidas ao acaso. Determine a probabilidade de:

   i) a média dos diâmetros de abertura ser superior a 5.1 mm.

   ii) 3 bisnagas terem diâmetro de abertura inferior a 3.5 mm.

c) O custo (em u.m. adequadas) de fabrico de uma bisnaga depende do diâmetro da abertura e pode considerar-se dado por $C = 20 + 2X$. Caracterize a lei do custo de fabrico de cada bisnaga.

2.85. Uma fábrica dispõe de duas máquinas, que trabalham independentemente, uma a produzir fichas e a outra a produzir tomadas. Sabe-se que o diâmetro de um pino de uma ficha segue uma distribuição normal com valor médio 2.0 mm e desvio padrão 0.03 mm, e que o diâmetro de um orifício de uma tomada segue uma distribuição normal com valor médio 2.04 mm e desvio padrão 0.04 mm. Calcule a probabilidade de:

a) o diâmetro de um pino (de uma ficha) ser superior a 2.01;

b) o diâmetro de um pino (de uma ficha) ser inferior a 2.04, sabendo que é superior a 2.01;

c) o diâmetro de um pino (de uma ficha) ser igual ao diâmetro de um orifício (de uma tomada);

d) um orifício (de uma tomada) ter diâmetro superior ao de um pino (de uma ficha).

Exercícios de Revisão de Introdução à Teoria da Probabilidade

R2.1. Sejam E, F e G acontecimentos de um espaço de resultados Ω.

a) Mostre que se E e F são acontecimentos mutuamente exclusivos se tem $P(\overline{E \cup F}|G) = 1 - P(E|G) - P(F|G)$, para $P(G) > 0$.

b) Se $P(E) = 0.3$, $P(F) = 0.8$ e $P(G) = 0.4$ podem os três acontecimentos constituir uma partição de Ω? Justifique.

R2.2. Sejam A e B acontecimentos de um certo espaço de resultados Ω, com $0 < P(A) < 1$. Mostre que:

a) Se $A \subset B$ então $P(B - A) = P(B) - P(A) \geq 0$.

b) Se $P(B|A) = P(B|\overline{A})$ então os acontecimentos A e B são independentes.

c) Sejam $B_1, B_2, ..., B_n$ acontecimentos constituindo uma partição de B.

i) Caracterize os acontecimentos B_j, $j = 1, ..., n$.

ii) Admita que $P(B_j) > 0$ e $P(A|B_j) = p$, $j = 1, 2, ..., n$.
Prove que $P(A|B) = p$.

R2.3. Um vendedor de bolbos prepara encomendas a partir de 3 lotes de bolbos que, por terem idades diferentes, não apresentam a mesma probabilidade de germinação. A probabilidade de germinação de um bolbo é de 0.80 se pertence ao lote A, de 0.85 se pertence ao lote B e de 0.90 se pertence ao lote C.

a) i) Qual a probabilidade de germinação de um bolbo retirado ao acaso de um lote escolhido ao acaso?

ii) Retirou-se um bolbo ao acaso de um lote escolhido ao acaso e verificou-se que não germinava. Qual a probabilidade de o bolbo ter sido retirado do lote C?

b) Se uma encomenda for constituída por um bolbo (retirado ao acaso) de cada lote, qual a probabilidade de pelo menos dois bolbos germinarem, admitindo a independência de germinação entre os bolbos retirados de lotes diferentes?

R2.4. Considere A, B e C acontecimentos de um espaço de resultados Ω.

a) Seja $1/4$ a probabilidade de pelo menos um dos acontecimentos B ou C ocorrer.

i) Qual a probabilidade de nenhum dos dois acontecimentos ocorrer?

ii) Sabendo que $P(A|(B \cup C)) = 0.3$ e $P(A|\overline{(B \cup C)}) = 0.7$, calcule $P(A)$.

iii) Qual a probabilidade de pelo menos um dos dois acontecimentos B ou C ocorrer, sabendo que A ocorreu?

b) Mostre que se A e B são independentes, A e C são independentes e B e C são mutuamente exclusivos, então A e $B \cup C$ são independentes.

R2.5. Um organismo estatal recorre a imagens de satélite para monitorizar a ocorrência de incêndios. Para tal, para cada parcela de terreno (que pode estar ardida ou não) é calculado um determinado índice a partir da informação adquirida pelo sensor do satélite. O valor deste índice pode ser considerado uma v.a. que no caso de parcelas não ardidas se admite ter distribuição normal de valor médio 2 e desvio padrão 5 e para parcelas ardidas se admite ter distribuição normal de valor médio 8 e desvio padrão 3.

Quando o sensor calcula um valor para o índice superior a 5, a parcela é registada pelo sensor como ardida.

Considere que a proporção total de parcelas ardidas é 6%.

a) Qual é a probabilidade de, numa parcela ardida, o sensor calcular um valor do índice superior a 5?

b) Qual a probabilidade de uma parcela escolhida ao acaso ser registada pelo sensor como ardida?

c) Para uma parcela escolhida ao acaso, qual a probabilidade de ser tomada uma decisão errada (o sensor registar como ardida uma parcela e ela não estar ardida ou o sensor registar como não ardida a parcela e ela estar ardida)?

R2.6. Responda, **justificando convenientemente**, se são verdadeiras ou falsas as afirmações nas seguintes alíneas.

a) Sejam A e B acontecimentos de um espaço de resultados Ω tais que $P(A) > 0$ e $P(B) > 0$.

   i) Se $A \cup B$ se realiza, então A também se realizou.

   ii) $P(A - B) \leq P(A)$.

   iii) Se A e B são mutuamente exclusivos, então $P(A|(A \cup B)) = \frac{P(A)}{P(A)+P(B)}$.

b) Seja X uma variável aleatória com função distribuição cumulativa F.

   i) Se $a > 0$, $F(x) \leq F(x + a)$, $\forall x \in \mathbb{R}$.

   ii) Se $X \frown Geo(p)$, $0 < p < 1$, então $P[X \leq 1|X \leq 2] = 1/q$, sendo $q = 1-p$.

R2.7. Uma empresa seguradora tem ao balcão dois vendedores de seguros de vida. A experiência tem revelado que 50% das pessoas que contactam o vendedor A e apenas 25% das pessoas que contactam o vendedor B fazem um seguro de vida. Considere o par aleatório (X, Y) que representa o número de apólices vendidas diariamente por A e B num dia em que cada vendedor atende 2 pessoas.

a) Admitindo que que cada pessoa contactou um só vendedor, determine a distribuição de probabilidades conjunta do par aleatório (X, Y).

b) Qual a probabilidade de se vender pelo menos um seguro de vida?

c) Qual a probabilidade de A vender pelo menos um seguro de vida sabendo que B vendeu dois seguros?

d) Calcule $E(X+Y)$ e $Var(X)$.

R2.8. Sejam X e Y variáveis aleatórias.

a) Defina $Var[X]$ e prove que $Var[b+X] = Var[X]$ com $b \in \mathbb{R}$.

b) Seja X uma v.a. discreta com distribuição de probabilidade (x_i, p_i), $i = 1, \ldots, n$. Defina $E[X]$ e prove que $E[b+X] = b + E[X]$ com $b \in \mathbb{R}$.

c) Se $X \frown Poisson(2)$ e $Y \frown Poisson(3)$, independentes, mostre que $X + Y \frown Poisson(5)$.

R2.9. Suponha que a dureza X e a perda abrasiva Y de um dado bloco de material, expressas em unidades adequadas, são variáveis aleatórias com função densidade de probabilidade conjunta assim definida

$$f(x,y) = \begin{cases} kxy + y & \text{se} \quad 0 \leq x \leq 1, \quad 0 \leq y \leq 1 \\ 0 & \text{outros valores de } (x,y). \end{cases} , k \in \mathbb{R}$$

a) Determine k.

b) Mostre, justificando, que as variáveis X e de Y são independentes.

c) Determine a perda abrasiva esperada de um bloco de material, se a sua dureza é de 0.4. Justifique.

R2.10. Seja X o tempo total desde a chegada de um cliente a uma estação de serviço até ao momento em que faz o pagamento, e seja Y o tempo que está em fila até efectuar o pagamento (medidos em unidades de 5 minutos). Suponha que as variáveis (X, Y) têm função densidade de probabilidade conjunta assim definida:

$$f(x,y) = \begin{cases} (x/2)e^{-x} & \text{se } 0 \leq y \leq x \\ 0 & \text{para outros valores de } (x,y). \end{cases}$$

Sugestão: sempre que necessário pode utilizar o seguinte resultado
$\int_0^{+\infty} x^n e^{-x} dx = n!, n \in \mathbb{N}_0.$

a) Calcule as funções densidade marginais de X e Y.

b) Calcule o tempo médio de serviço. Qual a variância do tempo de serviço.?

c) As variáveis aleatórias X e Y são independentes? Justifique.

R2.11. Considere a variável aleatória X com a seguinte função densidade

$$f(x) = \begin{cases} \alpha x^{\alpha-1} & \text{se } x \in [0,1] \qquad\qquad \alpha \in \mathbb{R}^+ \\ 0 & \text{restantes valores de } x. \end{cases}$$

a) Determime o valor de α para o qual $E[X] = 2/3$.

b) Determine a função distribuição cumulativa de X.

c) Determine o valor de m tal que $P[X < m] = P[X > m]$.

d) Seja Z uma variável aleatória, independente de X, com distribuição normal reduzida.

i) Determine a função densidade conjunta do par aleatório (X, Z).

ii) Determine $P[Z > 1|X < \frac{1}{2}]$.

iii) Determine o valor esperado e a variância de $X - Z^2$.

R2.12. Seja X uma variável aleatória que designa o tempo de vida (em anos) de organismos de uma dada população e Y um índice, comprendido entre 0 e 1, que exprime a velocidade de envelhecimento. Suponha que X e Y têm função densidade de probabilidade conjunta assim definida:

$$f(x,y) = \begin{cases} 2ky\, e^{-x} & \text{se } \ x > 0 \quad \text{e} \quad 0 < y \leq 1 \\ 0 & \text{outros valores de } (x,y). \end{cases}$$

a) Determine o valor da constante k.

b) Determine as densidades marginais de X e de Y. Serão X e Y variáveis aleatórias independentes? Justifique.

c) Qual a probabilidade de um organismo daquela população viver mais de 2 anos se o seu índice de envelhecimento é igual a 0.8?

d) Qual a probabilidade de um organismo daquela população ter um índice de envelhecimento igual a 0.5?

R2.13. Considere a v.a. discreta, X, com a seguinte distribuição de probabilidade

x_i	-1	0	1
p_i	0.2	0.3	0.5

a) Chama-se momento de ordem k de uma v.a. X a $E[X^k]$, caso exista. Calcule os momentos de ordem 1, 2 e 3 da v.a. X. Determine $Var[X]$.

b) Considere Y outra v. a. discreta, independente de X, tal que
$P[Y = 0] = 0.3$ e $P[Y = 1] = 0.7$.

i) Determine a distribuição de probabilidade conjunta do par aleatório (X, Y).

ii) Deduza a distribuição de probabilidade da v.a. $D = |X - Y|$.

R2.14. Dois estudantes A e B combinaram encontrar-se entre as 9h e as 10h para estudarem juntos. Sejam X e Y as horas de chegada do estudante A e B, respectivamente, ao local de encontro. Considerando que X e Y são variáveis aleatórias independentes com funções densidade, respectivamente

$$f_X(x) = \begin{cases} 2(x-9) & 9 \leq x \leq 10 \\ 0 & \text{restantes valores de } x \end{cases}$$

e

$$f_Y(y) = \begin{cases} 1 & 9 \leq y \leq 10 \\ 0 & \text{restantes valores de } y \end{cases}$$

a) Determine a função distribuição cumulativa de Y.

b) Determine a função densidade conjunta do par aleatório (X, Y).

c) Qual a probabilidade de ambos os estudantes chegarem entre as 9h e as 9h30?

d) Determine a probabilidade de o estudante A chegar antes do estudante B.

e) Qual a probabilidade de o estudante A chegar exactamente às 9h30?

f) Sabendo que o estudante A chegou às 9h15, qual a probabilidade do estudante B ter chegado antes dessa hora?

R2.15. Para avaliar uma determinada doença em perus efectuam-se dois testes. Designe-se por X e Y as v.a.'s associadas aos resultados de cada um dos testes. Sabe-se que aqueles testes dão resultado positivo (valor 1) ou negativo (valor 0) com as probabilidades indicadas na tabela seguinte. O valor da constante a varia consoante a gravidade da doença.

X \ Y	0	1
0	1/4-a	1/4+a
1	1/4	1/4

a) Quais os valores possíveis para a?

b) Determine a função distribuição cumulativa de X.

c) Para que valor(es) de a são as variáveis aleatórias independentes?

d) Considere $a = 0$. Calcule $E[X+Y]$, $Var[X+Y]$ e $\rho_{X,Y}$.

e) Determine a função massa de probabilidade de $Y|X=1$.

R2.16. Um inquérito realizado a um grande número de casais com dois ou menos filhos conduziu à seguinte distribuição de probabilidade conjunta para (X, Y), onde X representa a faixa etária média do casal ($X = 1 \rightarrow$ menos de 25 anos; $X = 2 \rightarrow$ 25 anos ou mais) e Y representa o número de filhos:

X \ Y	0	1	2
1	0.15	a	0.05
2	b	0.4	c

a) Complete a tabela sabendo que $P[X = 1] = 0.3$ e $E[Y] = 0.8$.

b) Determine a distribuição de probabilidade da faixa etária média do casal, para os casais com 1 filho.

c) Serão X e Y variáveis aleatórias independentes? Justifique.

R2.17. Considere-se um sistema eléctrico constituído por 100 componentes que funcionam de forma independente. Cada componente tem uma duração de vida (em anos) que é descrita por uma variável aleatória X com função densidade:

$$f(x) = \begin{cases} \frac{1}{2}e^{-x/2} & \text{se } x \geq 0 \\ 0 & \text{restantes valores.} \end{cases}$$

a) Determine a função distribuição cumulativa de X.

b) Determine o tempo mediano de vida de um componente.

c) Calcule a probabilidade de um componente ter uma duração de vida superior a 3 anos.

d) Qual a probabilidade de um componente durar pelo menos 4 anos sabendo que está a funcionar há pelo menos 1 ano? Comente o resultado obtido, caracterizando a lei da v.a. X.

e) Calcule, aproximadamente, a probabilidade de a duração média de vida das 100 componentes ser superior a 3 anos.

R2.18. Na época natalícia, certa pastelaria fabrica 3 tamanhos de bolo-rei: de 500g, de 750g e de 1000g. Nem todos os bolos fabricados contêm brinde. Este é colocado de tal forma que 20% dos bolos de 500g ficam sem brinde, o mesmo sucedendo com 10% dos bolos de 1000g e com 30% dos bolos de 750g. 25% dos bolos fabricados são de 500g e outros 25% de 1000g.

a) Qual a probabilidade de um bolo sem brinde ser de 750g?

b) A filha de um casal seu amigo apareceu-lhe com um brinde que lhe saíu no bolo-rei comprado na referida pastelaria. Qual dos bolos (tamanho) tem maior probabilidade de ter sido comprado pelo casal?

c) A referida pastelaria tem uma produção diária de 1000 bolos. Qual a probabilidade de uma pessoa que compra 10 desses bolos ter pelo menos 2 com brinde?

R2.19. Considere uma empresa agrícola que produz uvas e melões nas quantidades (em toneladas) X e Y, respectivamente. Devido às instáveis condições atmosféricas o valor das produções é aleatório com f.d.p. conjunta dada por:

$$f(x,y) = \begin{cases} k(1-x)(2-y) & 0 \leq x \leq 1, 0 \leq y \leq 2 \\ 0 & \text{outros valores.} \end{cases}$$

a) Calcule k.

b) Se num dado momento a produção de melões for de 1 ton, qual será a f.d.p. da produção de uvas?

c) Será que as quantidades produzidas de cada fruta são independentes? Justifique.

d) Escolhendo ao acaso 20 empresas nas condições anteriores, qual será a probabilidade de, em pelo menos 5 delas, a produção de uvas ser superior a 800 kg?

R2.20. Um laboratório exporta um certo produto químico para o mercado europeu. Este mercado exige que o produto fornecido tenha entre outras características, uma determinada coloração. Da produção do laboratório, 60% tem a coloração adequada, mas apenas metade desta quantidade satisfaz também as outras condições exigidas pelo referido mercado.

a) Qual a percentagem da produção do laboratório que satisfaz as condições exigidas pelo referido mercado ?

b) De um lote de 100 produtos em que 30 não estão em condições de exportação, retirou-se uma amostra de 10, sem reposição. Calcule a probabilidade de aparecer pelo menos um produto que não seja exportável.

R2.21. Numa linha de fabrico de uma determinada componente electrónica pode ocorrer um defeito muito raro mas causador de grandes prejuízos. Seja 0.01 a probabilidade de ocorrência desse defeito. Um teste muito simples é realizado para detecção do defeito. Apresenta, no entanto, probabilidades significativas de conduzir a conclusões erradas. Assim, cerca de 5% das vezes o teste indica a existência de defeito se não houver defeito e cerca de 3% das vezes indica ausência de defeito se houver defeito.

a) Qual a probabilidade de se ter uma conclusão incorrecta?

b) Determine a probabilidade de o teste indicar a existência de defeito.

c) São comercializadas embalagens contendo 80 daquelas componentes. Qual a probabilidade de, numa determinada embalagem, duas componentes apresentarem defeito?

d) A venda de cada embalagem referida na alínea anterior para o mercado é feita com um lucro Y , que é função de vários factores entre os quais o número de componentes defeituosas. Com o objectivo de simplificar os cálculos considere constante o efeito de todos os outros factores, sendo o lucro dado pela relação

$$Y = 0.02 - 0.1X$$

onde X é o número de componentes defeituosas em cada embalagem. Qual é nessa situação, a probabilidade de uma embalagem não dar prejuízo?

R2.22. Em certo bairro recentemente construído e constituído por prédios de duas, três ou quatro assoalhadas, verificou-se que em 37% dos apartamentos os moradores não têm filhos. A distribuição dos apartamentos por número de assoalhadas é a seguinte:

Nºde assoalhadas	2	3	4
Percentagem	30%	40%	30%

a) Determine a média e a variância do número de assoalhadas de um apartamento.

b) Admitindo que o número de filhos por apartamento tem uma distribuição de Poisson, determine a probabilidade de num certo apartamento haver pelo menos cinco filhos.

c) Sabendo que dos moradores em apartamentos de duas assoalhadas apenas 20% têm pelo menos um filho e que nos de três assoalhadas 30% não têm filhos, qual a probabilidade de num apartamento de 4 assoalhadas escolhido ao acaso haver pelo menos um filho.

R2.23. Sejam X_1 e X_2 variáveis aleatórias com distribuição de Poisson de parâmetros λ_1 e λ_2, respectivamente. Prove que se X_1 e X_2 forem independentes, a distribuição de $X_1|(X_1 + X_2 = k), \quad k \in \mathbb{N}$, é binomial.

R2.24. Um aviário embala os ovos que produz em três tipos de caixas, com capacidade para 6 ovos. Ao chegarem ao local de venda algumas caixas contêm ovos partidos. Sejam X, Y e W variáveis aleatórias que designam, respectivamente, o número de ovos partidos numa caixa de esferovite, plástico e cartão. Admite-se que X, Y e W têm as seguintes distribuições:

i	0	1	2	3	4	5	6
$P[X = i]$	0.73	0.17	0.07	0.03	0	0	0
$P[Y = i] = P[W = i]$	0.58	0.23	0.10	0.05	0.02	0.01	0.01

a) Determine a probabilidade de uma caixa de esferovite ter pelo menos um ovo partido.

b) Qual o valor esperado e a variância do número de ovos partidos numa caixa de esferovite?

c) Determine a probabilidade, aproximada, de o número total de ovos partidos em 100 caixas de esferovite ser superior a 50.

d) Pretende-se adquirir 30 embalagens de ovos em caixas de cartão. Qual a probabilidade de haver mais do que uma caixa com todos os ovos partidos?

e) Diariamente, 20% dos ovos produzidos são colocados em caixas de esferovite, 30% em caixas de plástico e os restantes 50% em caixas de cartão. Adquirida uma caixa ao acaso da produção diária deste aviário, determine a probabilidade dos acontecimentos:

   i) a caixa não ter nenhum ovo partido.

   ii) a caixa ser de cartão, tendo-se verificado que não tem ovos partidos.

R2.25. Para efeitos de comercialização, um dado fruto é classificado de acordo com o seu tamanho. Considera-se que o diâmetro de uma peça deste fruto é uma variável aleatória com distribuição normal de desvio padrão igual a 5 cm e média μ cm. A classificação, em categorias, do referido fruto é a seguinte:

Categoria	Diâmetro (x) em cm
C1	$x \leq 6$
C2	$6 < x < 12$
C3	$x \geq 12$

a) Sabendo que 30% dos frutos são da categoria C3, calcule o diâmetro médio dos frutos e a percentagem dos frutos das outras categorias.

b) Se os frutos forem vendidos em embalagens de 6 unidades, qual a probabilidade de uma embalagem ter pelo menos 2 frutos da categoria C3?

c) Sabendo que 10%, 8% e 2% dos frutos pertencentes respectivamente às categorias C1,C2 e C3 se apresentam em más condições, qual a probabilidade de um fruto retirado ao acaso não estar em condições de ser consumido?

R2.26. O número de peixes pescados por dia por um pescador é uma v.a. com distribuição de Poisson com média 10. Admita que o número de peixes pescados é independente de dia para dia.

a) Determine a probabilidade de, num dado dia, o pescador pescar 2 peixes.

b) Qual a probabilidade aproximada de numa semana (7 dias) o número total de peixes pescados ser superior a 80?

c) Num ano de pesca (365 dias), qual a probabilidade aproximada de haver pelo menos 3 dias em que são pescados 2 peixes em cada dia?

R2.27. O tempo de vida em horas de uma peça é uma variável aleatória com distribuição exponencial de média 1000. Determine a probabilidade de:

a) Uma peça estar ainda em funcionamento ao fim de 800 horas.

b) Uma peça durar pelo menos 900 horas, sabendo que já está a funcionar há pelo menos 100 horas. Interprete o resultado.

c) Num lote de 100 peças escolhidas ao acaso, o tempo médio de vida dessas peças ser inferior a 980 horas (Admita que o tempo de vida é independente de peça para peça).

R2.28. Um artesão executa peças de barro, sendo o tempo de execução uma variável aleatória X com distribuição exponencial de valor esperado 2 horas.

a) Determine a função distribuição cumulativa de X.

b) Qual a probabilidade de uma peça levar pelo menos 1h 30m a ser executada?

c) Verificando-se que em certo momento uma peça já está a ser produzida há 30m, qual a probabilidade de ser necessário esperar pelo menos mais 1h 30m para concluir a peça? Comente relacionando com o resultado da alínea anterior.

d) O artesão mantém os registos do tempo de execução de cada peça. Seis peças foram escolhidas ao acaso. Qual a probabilidade de nenhuma ter levado mais de 1h e 30m a ser executada?

e) Num dia em que não havia qualquer peça em stock foi aceite uma encomenda de 100 peças, tendo o artesão assumido o compromisso de fornecer as peças no prazo máximo de 30 dias (o que corresponde a 240 horas de trabalho). Qual a probabilidade de ele cumprir o seu compromisso? Justifique.

R2.29. Numa fábrica de rações produz-se diariamente 100 unidades de um dado tipo de ração. A quantidade de uma certa matéria prima incorporada em cada unidade é uma variável aleatória de valor médio 75g e variância $25g^2$. Admita a independência entre as quantidades de matéria prima necessárias para cada unidade.

a) Qual a probabilidade, aproximada, de a quantidade média de matéria prima por unidade, usada num dado dia, ser superior a 74g?

b) Determine a percentagem, aproximada, de dias em que a quantidade de matéria prima usada não excede 7.6kg.

c) Fazem-se lotes de 100 unidades de ração. Sabe-se que o custo da matéria prima é de 5 cêntimos/g. Qual deve ser o valor a pedir na venda de cada lote de modo a cobrir o custo da matéria prima em 95% das situações?

R2.30. O diâmetro de um certo tipo de peças é uma variável aleatória com distribuição normal. As peças são consideradas defeituosas se o seu diâmetro diferir do valor médio μ mais do que 1.25 mm. Sabe-se que 2.28% das peças possuem um diâmetro superior a 7 mm, sendo também esta percentagem a das peças com um diâmetro inferior a 5 mm.

Tendo-se extraído uma amostra de 100 peças de um grande lote, qual a probabilidade de aparecerem pelo menos 5 peças defeituosas.

R2.31. Seja T uma variável aleatória que representa a duração até falhar (tempo de vida) de uma componente. Chama-se fiabilidade da componente no tempo t ao valor $R(t) = P[T > t]$, i.e., à probabilidade de a componente não falhar durante o tempo t.

a) Exprima a derivada da fiabilidade em função de f, função densidade de probabilidade da variável aleatória T.

b) A distribuição do tempo de vida de uma lâmpada de um certo tipo é exponencial e a sua fiabilidade para o período de 72 horas é 95%.

i) Qual é o tempo médio de vida de uma lâmpada daquele tipo?

ii) Sabendo que uma determinada lâmpada não falhou durante as primeiras 72 horas de funcionamento, calcule a probabilidade de a mesma lâmpada não falhar nas 72 horas seguintes.

iii) Considerando que se dispõe de uma amostra aleatória de 50 lâmpadas, qual a probabilidade, aproximada, de a sua duração média ser superior a 1600 horas?

R2.32. Duas empresas, A e B, fazem a abertura de furos para captação de água numa dada região. A empresa A cobra 3500€ por furo, independentemente da sua profundidade. A empresa B cobra 1000€ mais 25€ por cada metro de profundidade do furo. Na região em questão, a profundidade dos furos segue uma distribuição normal com média 75 m e desvio padrão 12 m e a profundidade é independente de furo para furo.

a) Ao realizar um furo na referida região, qual é a probabilidade de este ter uma profundidade:

i) superior a 90 m?

ii) igual a 90 m?

b) Em média quantos furos é necessário realizar até ter dois furos com uma profundidade superior a 90 m?

c) Caracterize, justificando, a distribuição dos preços, P_A e P_B da empresa A e B respectivamente. Qual é o preço médio de um furo realizado pela empresa B?

d) Qual é a probabilidade da empresa A cobrar mais do que a empresa B pela abertura de um furo?

e) Sabendo que a empresa B abriu 10 furos num dado mês, qual a probabilidade de o preço de abertura desses furos ter sido superior ao valor que a empresa A cobraria se fosse ela a abrir os furos?

R2.33. Uma escola tem rapazes e raparigas de 18 anos na razão 3:1, respectivamente. Aos 18 anos a altura dos rapazes segue uma distribuição normal de valor médio 175 cm e desvio padrão 10 cm enquanto a altura das raparigas segue uma distribuição normal de valor médio 165 cm e desvio padrão 10 cm .

a) Qual a probabilidade de uma rapariga, escolhida ao acaso, ter altura superior a 170 cm?

b) Qual a probabilidade de um aluno qualquer (rapaz ou rapariga), escolhido ao acaso, ter altura superior a 170 cm?

c) Verificou-se que um aluno escolhido ao acaso tinha altura superior a 170 cm. Qual a probabilidade de ser rapariga?

d) Pretende-se seleccionar 10 raparigas para um campeonato de voleibol. Qual a probabilidade de a média das alturas ser superior a 170 cm?

R2.34. Seja X uma variável aleatória com distribuição uniforme contínua no intervalo $]-\theta, \theta[$, sendo θ um valor real positivo.

a) Calcule a função de distribuição cumulativa de X.

b) Calcule o valor médio, a mediana e a variância de X.

c) Considere a variável aleatória $Y = 2X + 1$.

i) Determine $P[1 < Y < \theta + 1]$.

ii) Calcule $Cov(X, Y)$.

iii) Serão Y e X variáveis aleatórias independentes? Justifique.

R2.35. O peso médio dos indivíduos duma certa espécie de bivalves é 31g e o respectivo desvio padrão é 2.4g. Recolhe-se uma amostra aleatória de 100 indivíduos desta espécie.

a) Qual a probabilidade, aproximada, de a média da amostra ser inferior a 30g?

b) Qual a probabilidade, aproximada, de a média da amostra estar compreendida entre 30 e 31g?

c) Qual a probabilidade, aproximada, de o peso total da amostra ser superior a 3150g?

R2.36. O tempo de duração T, em minutos, de uma chamada telefónica é uma variável aleatória com distribuição exponencial padrão. O custo, em euros, de cada chamada $C(T)$, função da duração, é dado por

$$C(T) = \begin{cases} 0.2 & 0 < T \leq 3 \\ 0.2 + 0.6(T-3) & T > 3 \end{cases}.$$

Determine o custo médio de cada chamada.

R2.37. Responda, **justificando convenientemente**, se são verdadeiras ou falsas as afirmações nas seguintes alíneas.

a) Seja X uma variável aleatória.

i) Se $X \frown Poisson(3)$ então $P[2 < X < 3] = 0$.

ii) Se $X \frown \mathcal{N}(0,1)$ então $P[X = 0] = 1/2$.

iii) Seja $X \frown \mathcal{N}(0,1)$ e $Y = 2X$. Então $Y \frown \mathcal{N}(0,4)$.

b) Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória de dimensão n, proveniente de uma população X com valor médio μ e variância σ^2. Seja $\overline{X}$ a média da amostra aleatória.

i) Se $X \frown Bernoulli(p)$ então $n\overline{X} \frown B(n,p)$.

ii) $E[\mu] = \overline{X}$.

iii) $\overline{X} \frown \mathcal{N}(\mu, \sigma/\sqrt{n})$.

R2.38. Seja X uma variável aleatória contínua com função distribuição cumulativa assim definida:

$$F(x) = \begin{cases} 0 & \text{se } x < b \\ 1 - \dfrac{b}{x^2} & \text{se } x \geq b, \end{cases}$$

com $b \in \mathbb{R} \setminus \{0\}$.

a) Determine, justificando, o valor de b. Esboce o gráfico da função distribuição cumulativa.

b) Determine a função densidade de probabilidade da v.a. X.

c) Define-se um par de variáveis aleatórias discretas (U,V) do seguinte modo:

- $U = 0$ se $X > 2$ e $U = 1$, caso contrário;
- V v.a. uniforme discreta tomando os valores 0 e 1;
- sabe-se que $P[U = 0|V = 1] = 1/2$.

i) Construa a distribuição de probabilidade conjunta do par (U,V).

ii) As variáveis aleatórias U e V são independentes? Justifique.

iii) Determine:
$P[V = 1|U = 0]$; $P[U + V \leq 1]$; $E[UV]$.

Soluções de alguns Exercícios

2.1. $P(A \cap B) = 0.2$; $P(A - B) = 0.2$; $P(\bar{A} - B) = 0.4$; $P(A \cup \bar{B}) = 0.6$

2.2. $A \cup B = \{t \in \mathbb{R}^+ : 2 < t < 10 \ \vee \ t > 15\}$; $A \cap C = \emptyset$; $A \cap \overline{B} = A - B = A$; $(A \cup B) \cap \overline{C} = A$; $\overline{A} \cup (B \cap C) = \overline{A}$

2.5.
a) $P(B) = 0.4$
b) $P(B) = 0.2$

2.7.
a) 0.25.
b) 0.85.
c) 5/6.

2.8. A- Acontecimento "atleta 1 percorre 100 m em menos de 10 s"

B- Acontecimento "atleta 2 percorre 100 m em menos de 10 s"

C- Acontecimento "atleta 3 percorre 100 m em menos de 10 s"

Pede-se $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$

ou $P(A \cup B \cup C) = 1 - P\left(\overline{A} \cap \overline{B} \cap \overline{C}\right) = \frac{78}{150} = 0.52$

2.11.
a) $P(C|B) = 0.4615$.
b) $P[(B \cap C)|A] = 0.12$.
c) São não independentes.

2.12.
a) 0.1

2.14.
a) 0.34
b) 0.53
c) 0.18

2.16.
a) 0.135
b) 0.667
c) Designando por G - "doença na forma grave" e RP - "resultado positivo", tem-se $P(G) = 0.06$, $P(RP) = 0.135$ e $P(G \cap RP) = 0.06$, logo não há independência.

2.17.
a) 0.325
b) $\simeq 0.764$

2.18. a) 0.35

b) 0.25

c) 0.45

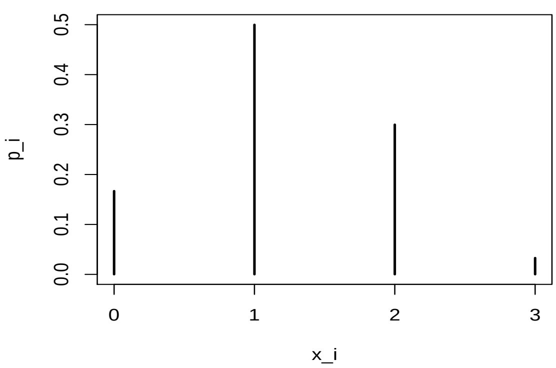
2.19. a) 0.08

b) 0.1304

2.23. a)

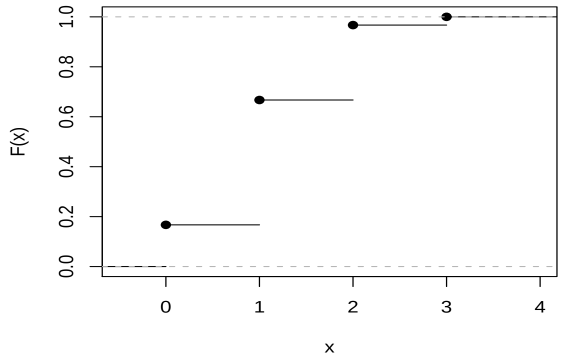
x_i	0	1	2	3
$P(X = x_i)$	0.167	0.500	0.300	0.033

b)



c)

$$F(x) = \begin{cases} 0 & x < 0 \\ 0.167 & 0 \leq x < 1 \\ 0.667 & 1 \leq x < 2 \\ 0.967 & 2 \leq x < 3 \\ 1 & x \geq 3. \end{cases}$$



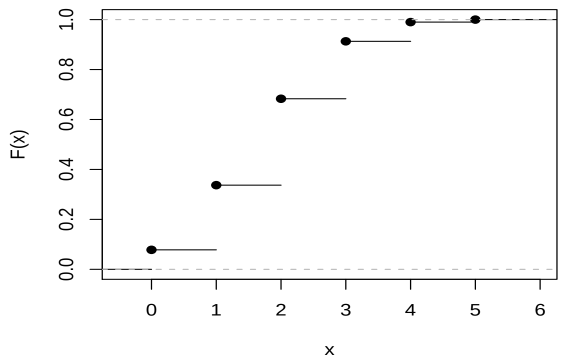
d) 0.833.

2.24. a)

x_i	0	1	2	3	4	5
$P(X = x_i)$	0.0778	0.2592	0.3456	0.2304	0.0768	0.0102

b)

$$F(x) = \begin{cases} 0 & x < 0 \\ 0.0778 & 0 \leq x < 1 \\ 0.3370 & 1 \leq x < 2 \\ 0.6826 & 2 \leq x < 3 \\ 0.9130 & 3 \leq x < 4 \\ 0.9898 & 4 \leq x < 5 \\ 1 & x \geq 5. \end{cases}$$



c) 0.8352.

2.25. a) Com reposição:

x_i	0	1	2	3	4	5	6	7	8
$P(X = x_i)$	0.00007	0.00123	0.01000	0.04668	0.13614	0.25412	0.29647	0.19765	0.05765

$E[X] = 5.6; \quad Var(X) = 1.68.$

b) Sem reposição:

x_i	0	1	2	3	4	5	6	7	8
$P(X = x_i)$	0.00001	0.00029	0.00454	0.03374	0.13179	0.28115	0.32332	0.18475	0.04042

$E[X] = 5.6; \quad Var(X) = 1.38.$

2.26. a) $E[X] = 0.3 \qquad Var(X) = 2.01.$

b)

$$F(x) = \begin{cases} 0 & x < -2 \\ 0.1 & -2 \leq x < -1 \\ 0.4 & -1 \leq x < 0 \\ 0.5 & 0 \leq x < 1 \\ 0.7 & 1 \leq x < 2 \\ 1 & x \geq 2. \end{cases}$$

c) 3/7.

d)

$y_i = {x_i}^2$	0	1	4
$P(Y = y_i)$	0.1	0.5	0.4

2.27. a)

x_i	0	1	2	3
$P(X = x_i)$	0.1	0.2	0.3	0.4

b) ≥ 3

c) $D = |X - 2|$

d_i	0	1	2
$P(D = d_i)$	0.3	0.6	0.1

2.28. a) $p = 1/6$ e $k = 1$.

b) 8/9.

c) $Y = X^2$

y_i	0	1	4
$P(Y = y_i)$	3/6	2/6	1/6

.

2.30. a) $a = 3$.

b)

$$F(x) = \begin{cases} 0 & x < 0 \\ x^3 & 0 \leq x < 1 \\ 1 & x \geq 1. \end{cases}$$

c) $\sqrt[3]{0.95}$.

2.32. a) Se X é v.a. contínua então terá que ter-se $F(x)$ contínua $\lim_{x\to 0^+} F(x) = 0$ e $\lim_{x\to\pi^-} F(x) = 1$, isto é,
$a.0 + b = 0$ e $a\pi + b = 1 \Rightarrow b = 0$ e $a = \dfrac{1}{\pi}$

b) $f(x) = \begin{cases} \dfrac{1}{\pi} & 0 < x < \pi \\ 0 & x \leq 0 \text{ ou } x \geq \pi \end{cases}$

c) $F(x) = \dfrac{1}{4} \Leftrightarrow x = \frac{\pi}{4} \Rightarrow \chi_{0.25} = \dfrac{\pi}{4}$.

d) $P(X < \dfrac{\pi}{2}|X \geq \dfrac{\pi}{4})\dfrac{P(\frac{\pi}{4} \leq X < \frac{\pi}{2})}{P(X \geq \frac{\pi}{4})} = \dfrac{1}{3}$

e) i) $E(X) = \displaystyle\int_0^\pi \frac{x}{\pi}dx = \frac{\pi}{2}$; $Var[X] = \frac{\pi^2}{12}$

ii) $[log(\pi + 2) - log2] /\pi$

2.33. a) $F(x) = \begin{cases} 0 & x < 0 \\ x^2/2 & 0 \leq x < 1 \\ x - 1/2 & 1 \leq x < 1.5 \\ 1 & x \geq 1.5. \end{cases}$

b) $P[X > 1|X > 1/2] = 4/7$

c) $E[X] = 23/24$ centenas de kg, i.e., o valor médio é aprox. 95.93 kg. A mediana M é tal que $F(M) = 1/2 \Leftrightarrow M = 1$. Portanto a mediana é 100kg.

d) i) $E[Y] = 275/12$

ii) $P[Y \geq 0] = 7/8$.

2.35. a) $k = b - 1$.

b) $F(x) = \begin{cases} 0 & x < 1 \\ \dfrac{b}{b-1}\left(1 - \dfrac{1}{x}\right) & 1 \leq x < b \\ 1 & x \geq b \end{cases}$.

c) $\chi_{0.5} = \dfrac{2b}{b+1}$.

d) $E[Y] = C_0 + C_1\left(\dfrac{b}{b-1}\ln b\right)$.

2.36. a) $F(x) = \begin{cases} 0 & \text{se } x < 0 \\ 1 - e^{-3.4x} & \text{se } x \geq 0 \end{cases}$.

b) $P[X > 2] = 1 - F(2) = 0.00111$.

d) $E[X] = M'_X(0) = \frac{1}{3.4} = 0.294$; $\chi_{0.5} = 0.204$.

2.37. a) $a = 2$

b) $F(x) = \begin{cases} 0 & \text{se } x < 0 \\ \frac{x^4+x^2}{2} & \text{se } 0 \leq x < 1 \\ 1 & \text{se } x \geq 1 \end{cases}$.

c) $P[X > 1/4 | X < 1/2] = \frac{F(1/2)-F(1/4)}{F(1/2)}$

e) i) $f(y) = \begin{cases} \frac{y^3+2y}{8} & \text{se } 0 < y < 2 \\ 0 & \text{o. v. de } y \end{cases}$.

ii) $P[Y > 1] = 27/32$

iii) $Cov[X, Y] = 2Var[X]$

2.38. a) $k = 1/2$

b) $F(x) = \begin{cases} \frac{1}{2}\, e^x & x < 0 \\ 1 - \frac{1}{2}\, e^{-x} & x \geq 0. \end{cases}$

c) $E[X] = 0$, $Var(X) = 2$ e $Me = 0$

d) $1 - 1/e$.

f) $f_Y(y) = \begin{cases} 0 & y \leq 0 \\ \dfrac{\exp(-\sqrt{y})}{2\sqrt{y}} & y > 0. \end{cases}$

2.39. a)

y_j	6	7	8	9	10
$p_{.j}$	0.25	0.2	0.2	0.25	0.1

b) As v.a.'s discretas X e Y são independentes sse $\forall (i,j)$ $p_{ij} = p_{i.}p_{.j}$. Observe-se, por exemplo, que $p_{11} \neq p_{1.}\ p_{.1}$, pois $p_{11} = P[X = 2, Y = 6] = 0.05$, $p_{1.} = P[X = 2] = 0.25$ e $p_{.1} = P[Y = 6] = 0.25$, portanto X e Y não são independentes.

c) $E[X + Y] = E[X] + E[Y] = 3 + 7.75 = 10.75$.

d) Designemos por W a variável $(X|Y = 7)$. A distribuição de probabilidades de W é

w_i	2	3	4
$P[W = w_i]$	0.2	0.5	0.3

2.40. a) $P[X = Y] = P[X = 1000, Y = 1000] + P[X = 1500, Y = 1500] = 0.2$

b) $X = \begin{cases} 500 & 1000 & 1500 \\ 0.3 & 0.4 & 0.3 \end{cases}$ e $Y = \begin{cases} 1000 & 1500 & 2000 \\ 0.3 & 0.4 & 0.3 \end{cases}$

c) Como $0.05 = P[X = 500, Y = 1000] \neq P[X = 500] \times P[Y = 1000] = 0.3 \times 0.3$, conclui-se que X e Y não são independentes.

d) $E[X + Y] = E[X] + E[Y] = 1000 + 1500 = 2500$ euros.

e) $P[X \geq 1000|Y = 1000] = 0.833$

2.41. a) $P(X + Y \leq 2) = 0.08$

b) $P(X = Y) = 0.13$

c) $F_X(x) = \begin{cases} 0 & \text{se} \;\; x < 0 \\ 0.1 & \text{se} \;\; 0 \leq x < 1 \\ 0.6 & \text{se} \;\; 1 \leq x < 2 \\ 1 & \text{se} \;\; x \geq 2 \end{cases}$ e $F_Y(y) = \begin{cases} 0 & \text{se} \;\; y < 1 \\ 0.1 & \text{se} \;\; 1 \leq y < 2 \\ 0.3 & \text{se} \;\; 2 \leq y < 3 \\ 0.7 & \text{se} \;\; 3 \leq y < 4 \\ 1 & \text{se} \;\; y \geq 4 \end{cases}$

d) $\rho = 0$

e) $Z = \begin{cases} 1 & 2 & 3 & 4 & 5 & 6 \\ 0.01 & 0.07 & 0.18 & 0.31 & 0.31 & 0.12 \end{cases}$

f) $P(X \geq 1|Y = 2) = 0.9$

2.42. a) $a = \frac{1}{3}$

b) X e Y não são v.a. independentes

c) $P[Y < X] = \frac{1}{2}$ e $P[Y > 3 - X] = \frac{5}{9}$

d) $E[1/X] = \frac{\ln(2)}{2} + \frac{1}{3}$

e) $f_{Y|X=\frac{3}{2}}(y) = \begin{cases} \dfrac{1}{2} + \dfrac{y}{3} & \text{se} \;\; y \in [1, 2] \\ 0 & \text{se} \;\; y \notin [1, 2] \end{cases}$

f) 11/24.

2.43. a)

$$f_X(x) = \begin{cases} \frac{3}{2}(1 - x^2), & 0 \leq x \leq 1 \\ 0, & \text{outros valores de } \; x \end{cases}$$

$$f_Y(y) = \begin{cases} 3y^2, & 0 \leq y \leq 1 \\ 0, & \text{outros valores de } \; y. \end{cases}$$

b) X e Y não são v.a. independentes

c) $F_Y(y) = \begin{cases} 0 & \text{se } y < 0 \\ y^3 & \text{se } 0 \leq y < 1 \\ 1 & \text{se } y \geq 1. \end{cases}$

d) $\frac{1}{2}$

2.44. a) $f_X(x) = \begin{cases} 2xe^{-x^2} & \text{se} \;\; x > 0 \\ 0 & \text{se} \;\; x \leq 0 \end{cases}$ e $f_Y(y) = \begin{cases} 2ye^{-y^2} & \text{se} \;\; y > 0 \\ 0 & \text{se} \;\; y \leq 0 \end{cases}$

b) X e Y são v.a. independentes

c) $COV(X, Y) = 0$

2.45. a) $f_X(x) = \begin{cases} \frac{x^3}{4} & \text{se } 0 \leq x \leq 2 \\ 0 & \text{o.v.de } x \end{cases}$ e $f_Y(y) = \begin{cases} \frac{y}{2}(2 - \frac{y^2}{2}) & \text{se } 0 \leq y \leq 2 \\ 0 & \text{o.v.de } y \end{cases}$

X e Y não são variáveis independentes.

b) i)1/10 ii)1/4

c) 0

d) 8/5 ou seja 16000 litros.

2.46. a) $f_X(x) = \begin{cases} e^{-x} & \text{se } x > 0 \\ 0 & \text{se } x \leq 0 \end{cases}$ e $f_Y(y) = \begin{cases} 2y & \text{se } 0 < y \leq 1 \\ 0 & \text{o.v.de } y \end{cases}$

b) X e Y são v.a. independentes

c) $E[X] = 1$, $E[Y] = 2/3$ e $Cov[X, Y] = 0$ pois X e Y são independentes.

d) 1/e

2.49. a) $X \frown B(20; 0.9)$.

b) 0.9568.

c) $\simeq 1$.

e) 110

f) 0.9568.

2.50. a) 0.9536.

b) 60 cobaias.

c) Pelo menos 6 cobaias.

2.51. a) $X \frown B(n, 0.2)$.

b) i) 0.2309
ii) 0.9958

c) O número mínimo de cobaias a utilizar na experiência deverá ser 14.

2.53. a) X é a variável aleatória que designa o nº de componentes avariadas no instrumento (instaladas de forma independente). $X \frown B\,(3, 0.1)$.

b) $P[X \geq 2] = 1 - P[X \leq 1] = 0.028$

c) Seja F o acontecimento "o instrumento funcionar".
$P\,(F|X = 0) = 1$ e $P\,[X = 0] = 0.7290$, $P\,(F|X = 1) = 0.6$ e $P\,[X = 1] = 0.243$, $P\,(F|X \geq 2) = 0$ e $P[X \geq 2] = 0.028$
i) $P\,(F) = 0.8748$.
ii) Pretende-se obter o valor de $P\,(X = 1|F)$. Pelo teorema de Bayes, tem-se $P\,(X = 1|F) = \frac{P(F|X=1)P[X=1]}{P(F)} = 0.1667$.

2.54. a) ≥ 6.

b) 4.

2.55. a) 0.9984.

b) 1.25.

2.60. a) 0.1429.

b) A capacidade de atendimento deverá ser de 4 petroleiros por dia.

c) 2.

d) 1 ou 2.

e) 0.156

f) 1.782.

g) 0.218.

2.61. a) 0.1954.

b) 0.1221.

c) 0.41.

2.63. a) 3/5

b) 2/3

c) 47.5 euros

2.64. a) 0.7262.

b) 0.1587.

c) 99.6 ton.

2.65. 24.135 euros

2.67. a) 0.0062.

b) ≈ 0.0125.

c) ≈ 1.

2.68. 6.24 anos

2.73. a) $P[X=0] = \frac{e^{-1}1^0}{0!} = e^{-1}$; $P[X=1] = \frac{e^{-1}1^1}{1!} = e^{-1}$.
Para $k > 1$, $P[X=k] - P[X=k-1] = \frac{e^{-1}}{k!} - \frac{e^{-1}}{(k-1)!} < 0$. Então pode-se dizer que $P[X=k]$ é decrescente, sendo estritamente decrescente para $k > 1$.

b) Considerando o nº de ovos postos por segundo independente do nº de ovos postos noutro segundo (disjunto do anterior) e sendo $Y = \sum_{i=1}^{5} X_i$, representando o nº de ovos postos em 5 segundos. $P(Y > 3) = 0.735$.

c) Seja W a v.a. que designa o número de ovos postos, por segundo, em 40 aviários idênticos. $W = \sum_{i=1}^{40} X_i$ (X_i o número de ovos postos por segundo no i-ésimo aviário). $P(W < 30) \simeq 0.0484$

2.74. a) Seja X a v.a. que representa a produção semanal (em t) de bananas; $X \frown \mathcal{N}(5, 2)$. $P[X < 3] = 0.1587$. Como X é simétrica relativamente a $\mu = 5$; $P[X > 7] = P[X < 3] = 0.1587$.

b) O valor da venda é $500X$. $P[500X < 2000] = 0.3085$.

c) Seja S_4 a v.a. que representa a produção de bananas (em t) em 4 semanas, $S_4 = \sum_{i=1}^{4} X_i$ em que $X_i \frown \mathcal{N}(5, 2)$ e X_i independentes $(i = 1, \ldots, 4)$.
$S_4 \frown \mathcal{N}(4 \times 5, 2 \times \sqrt{4}) \Leftrightarrow S_4 \frown \mathcal{N}(20, 4)$. $P[S_4 \geq 23] = 0.2266$.

d) Seja Y a v.a. que representa o número de semanas, em 8 escolhidas ao acaso, em que a produção foi inferior a 8 t; $Y \frown B(8, p)$ em que $p = P[X < 3]$ calculada em a). $P[Y \leq 1] = 0.5803$

2.76. O risco da prateleira desabar quando se empilham os 99 sacos é baixo, cerca de 2.2%.

2.79. 0.9998

2.80. a) 3.5 e 42

b) 0.0039

c) 0.0023

d) 1 ou 2.

d) A v.a. tem distribuição $N(2.5r; 0.75\sqrt{r})$. O valor esperado é $2.5r$

R2.3. a) i) $P(G) = P(G \cap A) + P(G \cap B) + P(G \cap C) = 0.85$.
ii) $P(C|\overline{G}) = 2/9$.

b) $P(G|A).P(G|B).P(\bar{G}|C) + P(G|A).P(\bar{G}|B).P(G|C)+$
$+P(\bar{G}|A).P(G|B).P(G|C) + P(G|A).P(G|B).P(G|C)$ =0.941.

R2.4. a) Tem-se $P(B \cup C) = 1/4$.

i) $P(\overline{B \cup C}) = 3/4$

ii) $P(A) = 0.6$

iii) $P(B \cup C|A) = 0.125$

R2.5. a) 0.8413

b) 0.3083

c) 0.2674

R2.7. a)

X \ Y	0	1	2
0	0.141	0.094	0.015
1	0.281	0.188	0.031
2	0.141	0.094	0.015

b) $P(X+Y \geq 1) = 0.859$

c) $P(X \geq 1 | Y = 2) = 0.75$

d) $E(X+Y) = 1.5$ e $Var(X) = 0.5$

R2.10. a) $f_X(x) = \begin{cases} \frac{x^2}{2}e^{-x}, & x > 0 \\ 0, & \text{outros valores de } x \end{cases}$

$f_Y(y) = \begin{cases} \frac{1}{2}e^{-y}(y+1), & y > 0 \\ 0, & \text{outros valores de } y. \end{cases}$

b) Nota: 1 unidade = 5 min; 1 unidade^2 = 25 min^2

$E[X-Y] = E[X] - E[Y] = 3 - 3/2 = 1.5$ unid. $\Leftrightarrow E[X] = 7.5$ min.

$Var[X-Y] = Var[X] + Var[Y] - 2Cov[X,Y]$

$Var[X-Y] = 3 + 7/4 - 2 \times 3/2 = 7/4$ unid.^{2} $\Leftrightarrow Var[X-Y] = 43.75$ min^2.

c) X e Y não são v.a. independentes.

R2.12. a) $k = 1$.

b) $f_X(x) = \begin{cases} e^{-x}, & x > 0 \\ 0, & \text{outros valores de } x \end{cases}$

$f_Y(y) = \begin{cases} 2y, & 0 < y \leq 1 \\ 0, & \text{outros valores de } y. \end{cases}$

Como, $f_{X,Y}(x,y) = 2ye^{-x} = f_X(x) \times f_Y(y)$, para $x > 0$ e $0 < y \leq 1$, e $f_{X,Y}(x,y) = 0 = f_X(x) \times f_Y(y)$, para os restantes valores de x e y, conclui-se que as variáveis aleatórias X e Y são independentes.

c) $P(X > 2 | Y = 0.8) = P(X > 2) = e^{-2}$.

d) Como Y é uma v.a. contínua então $P[Y = 0.5] = 0$.

R2.13. a) $E(X) = 0.3$, $E(X^2) = 0.7$, $E(X^3) = 0.3$ e $Var(X) = 0.61$.

b) i)

X \ Y	0	1
-1	0.06	0.14
0	0.09	0.21
1	0.15	0.35

ii) $D = |X - Y| = \begin{cases} 0 & 1 & 2 \\ 0.44 & 0.42 & 0.14 \end{cases}$

R2.14. a) $F_Y(y) = \int_{-\infty}^{y} f_Y(t)\,dt$.

$$F_Y(y) = \begin{cases} 0 & \text{se } y < 9 \\ y - 9 & \text{se } 9 \leq y \leq 10 \\ 1 & \text{se } x > 10. \end{cases}$$

b) Como as variáveis aleatórias X e Y são independentes,
$f(x,y) = f_X(x) \times f_Y(y)$, $\forall (x,y) \in \mathbb{R}^2$.
Portanto,

$$f(x,y) = \begin{cases} 2(x-9), & 9 \leq x \leq 10 \text{ e } 9 \leq y \leq 10 \\ 0, & \text{outros valores de } (x,y). \end{cases}$$

c) $P[9 \leq X \leq 9.5, 9 \leq Y \leq 9.5] = 0.125$.

d) $P[X < Y] = 1/3$.

R2.16. a) $a = 0.1$; $c = 0.1$; $b = 0.2$.

b) $X|(Y=1) = \begin{cases} 1 & 2 \\ 1/5 & 4/5. \end{cases}$

R2.18. a) 0.667.

b) 750 g.

c) $\simeq 1$.

R2.19. a) $k = 1$

b) $f(x/y=1) = \begin{cases} 2(1-x) & 0 \leq x \leq 1 \\ 0 & \text{o.v.} \end{cases}$

c) Sim

d) p =prob. de a produção ser superior a 800 kg
$p = \int_{0.8}^{1} 2(1-x)dx = 0.04$
$X \rightarrow$ nº de empresas naquelas condições $\quad X \frown B(20; 0.04)$
$P[X \geq 5] = 1 - P[X \leq 4]$

R2.20. a) 30%.

b) $\simeq 0.977$.

R2.21. a) 0.0498.

b) 0.0592.

c) 0.144.

d) 0.449.

R2.22. a) média = 3; variância = 0.6 .

b) 0.004.

c) 0.9667.

R2.24. a) 0.27.

b) $E[X] = 0.4 \quad Var(X) = 0.0592$.

c) $\simeq 0.0918$.

d) 0.037.

e) i) 0.61; ii) 0.475.

R2.25. a) Diâmetro médio dos frutos: 9.4 cm; 24.8% e 45.2% de frutos C1 e C2, respectivamente.

b) 0.58.

c)

Categorias	$P(C_i)$	$P(\text{Mau}/C_i)$
C_1	0.25	0.10
C_2	0.45	0.08
C_3	0.30	0.02.

Então $P(\text{Mau}) = 0.067$.

R2.27. a) $P[X > 800] \approx 0.449329$

b) 0.449329

c) ≈ 0.4207

R2.28. a) $F(x) = \begin{cases} (1 - e^{-x/2}) & x > 0 \\ 0 & x < 0 \end{cases}$

b) 0.472

c) $P[X > 2.0|X > 0.5] = 0.472$

d) Y - nºde peças para as quais o tempo de execução é ≥ 1.5 horas;
$Y \frown B(6; 0.472)$; $P[Y = 0] = 0.0216$.

e) ≈ 0.9772

R2.29. a) 0.9773

b) 0.9773

c) Deve pedir-se 379.11 euros

R2.31. b)

i) 1403.69 horas

ii) 0.95

iii) ≈ 0.1611

R2.34. Se $X \frown U(-\theta, \theta)$ $(\theta > 0)$

$$f(x) = \begin{cases} \frac{1}{2\theta}, & -\theta < x < \theta \\ 0, & \text{outros valores de } x. \end{cases}$$

a)

$$F(x)=\begin{cases} 0, & x<-\theta \\ \frac{x+\theta}{2\theta}, & -\theta\leq x<\theta \\ 1, & x\geq\theta. \end{cases}$$

b) $E[X]=\frac{-\theta+\theta}{2}=0$ e $Var[X]=\frac{[\theta-(-\theta)]^2}{12}=\frac{\theta^2}{3}$.

Como a distribuição uniforme é simétrica, a mediana, Me, é igual ao valor médio, logo $Me=0$.

Mas, podia calcular-se Me como o valor para o qual $F(Me)=0.5$.

$$F(Me)=\frac{Me+\theta}{2\theta}=0.5 \Leftrightarrow Me=0.$$

c) i) $P[1<Y<\theta+1]=F(\theta/2)-F(0)=1/4$.

ii) $Cov(X,Y)=Cov(X,2X+1)\overset{(*)}{=}2Cov(X,X)=2Var[X]=\frac{2\theta^2}{3}$.

$(*)$ propriedade da covariância.

iii) Dado que $Cov(X,Y)\neq 0$, X e Y são variáveis aleatórias não independentes.

Capítulo 3 Introdução à Inferência Estatística

Exercícios de Distribuições de Amostragem e Estimadores

3.1. De uma população normal de média 8 e desvio padrão 4, extraíu-se uma amostra aleatória de dimensão 100.

a) Qual a probabilidade de a diferença, em valor absoluto, entre a média da amostra e a média da população ser superior a 0.5?

b) Se a população não fosse normal, qual seria essa probabilidade? Justifique.

3.2. O peso médio dos indivíduos duma certa espécie de bivalves é 31g e o respectivo desvio padrão é 2.4g. Recolhe-se uma amostra aleatória de 100 indivíduos desta espécie.

a) Qual a probabilidade, aproximada, de a média da amostra ser inferior a 30g?

b) Qual a probabilidade, aproximada, de o peso total da amostra ser superior a 3150g?

3.3. Mediu-se o comprimento (em cm) de 25 coelhos adultos de uma dada raça escolhidos aleatoriamente. A média da amostra foi 56.2 cm e o desvio padrão 4 cm. Admita que a população dos comprimentos dos coelhos adultos daquela raça é bem modelada por uma distribuição normal.

a) Indique estimativas para a média e o desvio padrão da população.

b) Usando as estimativas da alínea anterior como os parâmetros da população, determine a probabilidade de o comprimento de um coelho adulto daquela raça, escolhido ao acaso, ser superior a 57 cm.

c) Chama-se erro padrão da média amostral a uma estimativa do desvio padrão da v.a. média da amostra. Com base nos resultados obtidos, determine o erro padrão da média amostral?

3.4. Seja X uma variável aleatória com distribuição de Bernoulli de parâmetro p desconhecido e (X_1, X_2, X_3) uma amostra aleatória associada a esta variável.

a) Quais as propriedades da amostra aleatória (X_1, X_2, X_3)?

b) Sugerem-se os dois estimadores para p, $\hat{P}_1 = \overline{X}$ e $\hat{P}_2 = (X_1 + 2X_2 + X_3)/4$. Determine o valor esperado de cada um dos estimadores? Interprete.

c) Analise os estimadores quanto à sua variância. Qual dos dois estimadores considera que deve ser escolhido para estimar p? Justifique.

d) Procedeu-se à recolha de uma amostra, nas condições do enunciado, tendo-se obtido os valores $(0, 1, 0)$. Obtenha duas estimativas para p.

3.5. Depois de fabricado e embalado, a actividade de um certo adubo segue aproximadamente uma distribuição normal com $\mu = 120$ dias e $\sigma = 40$ dias.

Pretende-se enviar um lote de embalagens do referido adubo de modo que a actividade média amostral $(\overline{X})$ seja superior a 118 dias com probabilidade de 0.95. Qual o tamanho do lote a enviar?

3.6. Seja $(X_1, ..., X_n)$ uma amostra aleatória de tamanho n proveniente de uma população X com distribuição $U(0, 1)$. Exprima, em função de n, a probabilidade, aproximada, de $\overline{X}$ ser superior 0.9.

3.7. Uma amostra aleatória de dimensão 50, $(X_1, ..., X_{50})$, é extraída de uma população de Poisson com $\lambda = 10$. Recorra à distribuição normal para calcular um valor aproximado de $P(\overline{X} > 11)$.

3.8. Seja $X_1, X_2, \ldots, X_n$ uma amostra de variáveis aleatórias independentes e com a mesma distribuição, de valor médio μ e variância σ^2. Considere os seguintes estimadores de μ e σ^2, respectivamente,

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \text{e} \qquad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

a) Prove que $E[\overline{X}] = \mu$ e $Var[\overline{X}] = \sigma^2/n$.

b) Prove que

$$\sum_{i=1}^{n}(X_i - \mu)^2 = \sum_{i=1}^{n}(X_i - \overline{X})^2 + n(\overline{X} - \mu)^2$$

c) Usando os resultados das alíneas anteriores, mostre que $E[S^2] = \sigma^2$.

3.9. Considere a amostra aleatória $X_1, X_2, \ldots, X_{100}$, retirada de uma população X para a qual se tem $E[X] = 0.5$ e $Var[X] = 0.0625$. Considere a variável aleatória $S = \sum_{i=1}^{100} X_i$.

a) Defina "amostra aleatória".

b) Determine o valor esperado e a variância de S. Justifique convenientemente.

c) Determine, aproximadamente, $P[S < 45]$.

3.10. Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória de dimensão n, proveniente de uma população X com valor médio μ e variância σ^2. Seja $\overline{X}$ a média da amostra aleatória.

Responda, **justificando convenientemente**, se são verdadeiras ou falsas as afirmações nas seguintes alíneas.

a) $\overline{X} \frown \mathcal{N}(\mu, \sigma/\sqrt{n})$.

b) $E[\mu] = \overline{X}$.

c) Se $X \frown Bernoulli(p)$ então $n\overline{X} \frown B(n,p)$.

3.11. Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória de dimensão n, proveniente de uma população com distribuição de Bernoulli de parâmetro p, desconhecido. Considere a v.a. $Y = \sum_{i=1}^{n} X_i$.

a) Caracterize, justificando, a distribuição da v.a. Y.

b) Seja $\widehat{P} = Y/n$. Mostre que $\widehat{P}$ é um estimador centrado de p.

Exercícios de Intervalos de Confiança e Testes de Hipóteses

3.12. Seja X uma população com distribuição normal, de média μ e desvio padrão σ=2. Uma amostra aleatória de dimensão 25 foi extraída desta população, tendo-se obtido $\overline{x} = 78.3$.

a) Calcule o intervalo de confiança a 99% para μ.

b) Qual o erro máximo cometido (a 99% de confiança) ao estimar μ por $\overline{x}$ =78.3?

c) Qual deverá ser a dimensão da amostra para que o erro máximo cometido, a 99% de confiança, ao estimar μ por $\overline{x}$, não exceda 0.1?

3.13. Para avaliar a tensão máxima suportada por uma barra de aço testaram-se n barras tendo-se obtido, nas unidades adequadas, $\overline{x}_n = 20$ e para extremo superior do intervalo de confiança a 95% para o verdadeiro médio da tensão obteve-se 21.7. Sabendo que se admite que a tensão suportada por uma barra de aço é uma v.a. normal com desvio padrão $\sigma = 3$, determine o extremo inferior do intervalo de confiança e a dimensão n da amostra.

3.14. Uma empresa de peixe congelado está a ser investigada com o objectivo de se verificar se cada embalagem pesa de facto 1 kg em média. Numa amostra aleatória de 100 embalagens de peixe registou-se o peso (kg) de cada embalagem, x_i $(i = 1, \ldots, 100)$, tendo-se obtido os seguintes valores:

$$\sum_{i=1}^{100} x_i = 95.9 \text{ kg} \qquad \sum_{i=1}^{100} x_i^2 = 93.12 \text{ kg}^2$$

a) Indique uma estimativa para a média e para a variância do peso de uma embalagem de peixe.

b) O que pode dizer sobre o resultado da investigação? Justifique convenientemente, especificando as hipóteses que necessitou de considerar.

c) Determine, explicitando as hipóteses necessárias, um intervalo a 95% de confiança para a variância do peso de uma embalagem de peixe.

3.15. O grau de acidez do azeite produzido em certa região supõe-se ter distribuição normal. Uma amostra da produção de dimensão 25 conduziu a uma acidez média de 1 grau e a um desvio padrão de 0.33 graus.

a) Face a estes valores, alguém sugeriu o intervalo]0.815, 1.185[para o valor esperado da acidez. Qual o nível de confiança associado a este intervalo?

b) Para o nível de confiança determinado na alínea anterior, qual o erro máximo cometido ao estimar o verdadeiro valor médio da acidez por 1 grau?

3.16. Numa amostra de 16 elementos, que se supõe ter sido retirada de uma população com distribuição normal, o desvio padrão obtido foi de 5.2

a) Foi calculado um intervalo de confiança para a média populacional tendo-se obtido $]24.2297; 29.7703[$. Indique justificando:

   i) Qual a média da amostra;

   ii) Qual o grau de confiança do intervalo calculado.

b) Qual o intervalo de confiança a 95% para a variância populacional?

3.17. Um silvicultor sabe que a altura das árvores para madeira é importante para os compradores. Afirmou a um comprador que a altura média das suas árvores era 28 metros. O comprador fechou negócio, mas aquando do abate escolheu ao acaso 65 árvores e mediu as alturas. Obteve $\sum_{i=1}^{65} x_i^2 = 43959.32\ m^2$, onde x_i designa a altura de cada árvore seleccionada.
O comprador obteve o seguinte intervalo de confiança para a altura média, em metros:

$$]25.1538; 26.5462[$$

a) Calcule uma estimativa para a altura média das árvores daquela floresta.

b) Determine a confiança do intervalo dado acima.

c) Considera que o comprador tem razão para reclamar? Justifique.

d) Pretende-se construir um intervalo de confiança para a variância da altura das árvores.

   i) Indique o estimador da variância usado na construção deste intervalo, apresentando a sua expressão.

   ii) Com base nos resultados fornecidos construa o intervalo de confiança a 95% para a variância das alturas, indicando o(s) pressuposto(s) necessário(s).

3.18. Suponha que o rendimento de um pé de tomateiro expresso em kg é uma variável aleatória com distribuição normal de valor médio 1 kg. Numa parte da produção foi utilizado um novo fertilizante. Observada uma amostra de 10 pés de tomateiro da parte da produção em que foi utilizado o novo fertilizante obtiveram-se os seguintes resultados:

1.375	1.223	1.773	1.752	0.779	1.407	1.068	1.633	1.201	1.042

expressos em kg. Que decisão se deverá tomar perante estes resultados, face ao novo fertilizante?

3.19. Um processo de determinação do conteúdo de enxofre de um determinado produto forneceu os seguintes resultados:

1.12 1.10 1.08 1.06 1.08 1.14 1.10 1.11 1.14

a) Encontre um intervalo de confiança a 95% para o verdadeiro conteúdo médio de enxofre, apresentando as hipóteses a fazer sobre a população que achar convenientes.

b) De acordo com certas normas, o conteúdo de enxofre não deve ter um desvio padrão superior a 0.02. Acha que a amostra recolhida permite afirmar que o produto está dentro das normas? Justifique convenientemente a resposta.

c) De uma outra fábrica que produz idêntico produto recolheu-se uma mostra de 20 elementos que proporcionou um desvio padrão de 0.085. Pode concluir que a variabilidade do conteúdo de enxofre é, nesta fábrica, três vezes superior à da outra?

3.20. Num laboratório foi recentemente desenvolvido um método de análise mais rápido e barato do que o método convencional. Realizaram-se testes com o objectivo de comparar a precisão do novo método com a do convencional, repetindo-se sucessivamente a mesma análise. No seguinte quadro encontram-se cálculos efectuados com os resultados dos testes, em unidades adequadas (x designa o resultado do teste pela aplicação de cada um dos métodos):

	nº de observações	$\sum_i (x_i - \overline{x})^2$
Método convencional	24	30.5
Método novo	30	55.7

Sendo a precisão de um método medida pela variância dos resultados da aplicação desse método, poder-se-á concluir que os dois métodos têm precisões distintas? Que hipóteses é necessário admitir para responder a esta questão?

3.21. Pretende-se comparar dois antibióticos A e B usados para o tratamento de um certo tipo de infecção em bovinos. Escolhem-se ao acaso dois grupos de bovinos doentes: a um grupo é administrado o antibiótico A e ao outro o antibiótico B, registando-se o tempo (h) até os sintomas desaparecerem.

a) Considere a v.a. X, de valor médio μ e variância σ^2 desconhecidos, que designa o tempo até os sintomas desaparecerem no grupo de bovinos a que se aplica o antibiótico A. Seja $(X_1, \ldots, X_{10})$ uma amostra aleatória retirada de X. Sugerem-se dois estimadores para μ,

$$T_1 = \overline{X} \qquad \text{e} \qquad T_2 = \frac{X_1 + 3X_{10}}{4},$$

com $\overline{X}$ média da amostra aleatória.

i) Defina "amostra aleatória de dimensão 10".

ii) Mostre que T_1 e T_2 são estimadores centrados de μ e obtenha as respectivas variâncias.

iii) Qual dos dois estimadores deve ser escolhido para estimar μ? Justifique.

b) Os resultados obtidos pela aplicação dos dois antibióticos foram introduzidos no *software* R. Utilize, sempre que possível, os resultados apresentados no Anexo para responder às seguintes questões.

i) Com base nos dados recolhidos indique estimativas para o tempo médio no caso de aplicação do antibiótico A, associadas aos estimadores dados em a).

ii) Face aos resultados obtidos poder-se-á concluir que o antibiótico A é mais lento do que o antibiótico B no desaparecimento dos sintomas da infecção? Explicite e valide os pressupostos necessários à resolução do problema.

Anexo

```
> A<-c(49.5,48,52.5,46.5,52,48.5,40,50.5,46,46.5)
> B<-c(47.5,48.5,40.5,43.5,49,43,42,45,48.5,40)

> var(A)          > var(B)          > var(A-B)
[1] 13.05556      [1] 11.84722      [1] 19.29167


> shapiro.test(A)
        Shapiro-Wilk normality test
data:  A
W = 0.9189, p-value = 0.3481

> shapiro.test(B)
        Shapiro-Wilk normality test
data:  B
W = 0.9009, p-value = 0.2244

> shapiro.test(A-B)
        Shapiro-Wilk normality test
data:  A - B
W = 0.9482, p-value = 0.6474
```

```
> var.test(A,B)
        F test to compare two variances
data:  A and B
F = 1.102, num df = 9, denom df = 9, p-value = 0.8873
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.2737195 4.4366172
sample estimates:
ratio of variances
          1.101993

> t.test(A,B,paired=T,alternative="two.sided")
        Paired t-test
data:  A and B
t = 2.3399, df = 9, p-value = 0.04403
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.1079894 6.3920106
sample estimates:
mean of the differences
                   3.25

> t.test(A,B,paired=T,alternative="less")
        Paired t-test
data:  A and B
t = 2.3399, df = 9, p-value = 0.978
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
     -Inf 5.796092
sample estimates:
mean of the differences
                   3.25

> t.test(A,B,paired=T,alternative="greater")
        Paired t-test
data:  A and B
t = 2.3399, df = 9, p-value = 0.02201
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 0.703908      Inf
sample estimates:
mean of the differences
                   3.25
```

```
> t.test(A,B,paired=F,var.equal=T,alternative="two.sided")
        Two Sample t-test
data:  A and B
t = 2.0595, df = 18, p-value = 0.05421
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.06538397  6.56538397
sample estimates:
mean of x mean of y
    48.00     44.75

> t.test(A,B,paired=F,var.equal=T,alternative="less")
        Two Sample t-test
data:  A and B
t = 2.0595, df = 18, p-value = 0.9729
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
     -Inf 5.986459
sample estimates:
mean of x mean of y
    48.00     44.75

> t.test(A,B,paired=F,var.equal=T,alternative="greater")
        Two Sample t-test
data:  A and B
t = 2.0595, df = 18, p-value = 0.02711
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 0.5135412       Inf
sample estimates:
mean of x mean of y
    48.00     44.75
```

3.22. Numa experiência agronómica pretende-se avaliar o crescimento total de uma certa espécie de plantas (expresso em peso seco) relativamente a dois regimes de fertilização A e B.

Ao fim de determinado tempo procedeu-se a medições, tendo-se obtido os seguintes resultados:

A	5.44	5.36	5.60	6.46	6.75	6.03	4.15	4.44
B	5.12	3.80	4.96	6.43	5.03	5.08	3.22	4.42

a) Numa experiência anterior (com um elevado número de plantas da mesma cultivar) relativa ao tratamento A, obteve-se uma variância de 0.42. Verifique se os dados actuais são consistentes com esse valor. Diga, justificando, se haveria alguma(s) hipótese(s) necessária(s) à resolução do problema.

b) Verifique se os dois regimes de fertilização A e B evidenciam diferenças significativas no que respeita ao crescimento das plantas. Explicite as hipóteses necessárias à resolução do problema.

3.23. Num estudo sobre o número de folhas por planta de tabaco, obtiveram-se os seguintes dados:

n º de folhas	17	18	19	20	21	22	23	24
n º de plantas	3	22	44	42	22	10	6	1

a) Indique a unidade estatística e a variável em estudo.

b) Determine a média e a mediana da amostra. Compare-as e comente.

c) Com base nesta amostra poder-se-á afirmar que 90% das plantas de tabaco têm 21 ou menos folhas? Justifique.

3.24. Duas marcas de chocolate (A e B) produzem chocolate rotulado “70% de cacau”. Seleccionaram-se aleatoriamente 15 chocolates de cada uma das marcas, que foram analisados quanto ao teor de cacau. Os resultados foram introduzidos no *software* R. Utilize, sempre que possível, os resultados apresentados abaixo para responder às seguintes questões (alguns cálculos apresentados são desnecessários ou inadequados).

a) Conjectura-se que as variabilidades do teor de cacau dos chocolates das duas marcas são diferentes. Será esta conjectura compatível com os dados obtidos? Explicite e valide os pressupostos necessários à resolução do problema.

b) Indique estimativas para o valor esperado do teor de cacau dos chocolates de cada uma das marcas.

c) Os dados recolhidos evidenciam que o teor médio de cacau do chocolate da marca A é significativamente superior ao da marca B? Explicite e valide os pressupostos necessários à resolução do problema.

d) O produtor de chocolate da marca A afirma que o teor médio de cacau do seu chocolate é superior a 70%. O que pode dizer sobre a afirmação do produtor? Justifique convenientemente.

Especifique as hipóteses nula e alternativa de um teste de hipóteses que permita averiguar a validade desta afirmação.

```
> A
 [1] 72.8 69.6 70.0 73.3 69.8 71.1 71.0 71.0 71.0 76.3 71.3 75.4 72.1 70.4 68.6
> B
```

```
 [1] 69.8 71.3 67.3 69.4 67.7 67.9 70.7 70.5 70.2 71.4 71.6 73.1 70.6 68.5 70.2

> var(A)               > var(B)                   > var(A-B)
[1] 4.483143           [1] 2.616952               [1] 3.643810


> shapiro.test(A)                          > shapiro.test(B)
        Shapiro-Wilk normality test                Shapiro-Wilk normality test
data:  A                                   data:  B
W = 0.9066, p-value = 0.1199               W = 0.9613, p-value = 0.7144

> shapiro.test(A-B)
        Shapiro-Wilk normality test
data:  A - B
W = 0.9754, p-value = 0.9283


> var.test(A,B)
        F test to compare two variances
data:  A and B
F = 1.7131, num df = 14, denom df = 14, p-value = 0.3254
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.5751437 5.1026658
sample estimates:
ratio of variances
          1.713116


> t.test(A,B,paired=T,alternative="less")
        Paired t-test
data:  A and B
t = 3.1787, df = 14, p-value = 0.9967
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
     -Inf 2.434763
sample estimates:
mean of the differences
               1.566667


> t.test(A,B,paired=T,alternative="greater")
        Paired t-test
data:  A and B
t = 3.1787, df = 14, p-value = 0.003349
```

```
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 0.6985701       Inf
sample estimates:
mean of the differences
               1.566667


> t.test(A,B,paired=F,var.equal=T,alternative="less")
        Two Sample t-test
data:  A and B
t = 2.2771, df = 28, p-value = 0.9847
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
     -Inf 2.737039
sample estimates:
mean of x mean of y
 71.58000  70.01333


> t.test(A,B,paired=F,var.equal=T,alternative="greater")
        Two Sample t-test
data:  A and B
t = 2.2771, df = 28, p-value = 0.01531
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 0.3962939       Inf
sample estimates:
mean of x mean of y
 71.58000  70.01333
```

3.25. Para decidir se deveria ou não lançar um novo produto no mercado, uma empresa de bens alimentares fez um inquérito em 10 supermercados do Sul e 20 do Norte do país, acerca do número de unidades X do referido produto que estes esperam poder vender semanalmente. Obtiveram-se os seguintes resultados:

	$\sum x_i$	$\sum x_i^2$
Sul	1000	102550
Norte	1200	75950

a) Determine estimativas para o número médio e para a variância das unidades que os supermercados do Sul e do Norte esperam vender.

b) "Embora a média de vendas difira significativamente no Norte e Sul, a variabilidade das vendas é praticamente igual". Critique fundamentadamente a afirmação anterior, justificando.

3.26. A fim de investigar os efeitos de ambientes nitrosos e de ambientes fosfatados no desenvolvimento de colónias de bactérias, contaminam-se 10 plaquetas envolvidas em cada um daqueles ambientes com as bactérias em estudo, e deixa-se incubar durante 24 horas. Após esse tempo, procede-se à contagem do número de colónias de bactérias em cada plaqueta, tendo-se obtido os seguintes resultados:

Ambiente nitroso	60	47	12	29	51	46	49	74	63	101
Ambiente fosfatado	8	46	21	13	58	33	20	46	31	38

a) Investigue a hipótese de o tipo de ambiente não influir no desenvolvimento das colónias de bactérias.

b) Que hipótese(s) foi necessário considerar para poder resolver a alínea a)?

3.27. Numa Estação Florestal estudam-se problemas de intercepção de precipitação. Nesse sentido em 12 dias de chuva (suponha que se trata de uma amostra aleatória de dias com precipitação) são colocados dois udómetros para medir a quantidade de precipitação: um numa zona desarborizada e outro sob as copas das árvores. As leituras da quantidade de água em cada udómetro (medidas em cm de altura) deram os seguintes valores (cada coluna corresponde a um dia):

zona descoberta	5.87	1.30	2.34	2.82	5.89	9.09	1.93	9.27	4.65	4.35	5.00	8.43
sob coberto	4.96	1.14	2.24	2.26	4.75	7.83	1.86	8.85	4.17	3.65	4.08	7.99

a) Estime a precipitação média na zona desarborizada e na zona sob coberto.

b) Será admíssivel supor que a 95% de confiança as precipitações médias nos dois casos são iguais? Explicite as hipóteses necessárias à resolução desta questão.

c) Tendo em conta que, pela própria natureza do problema, o nível de precipitação sob coberto não deverá ser superior ao nível da correspondente precipitação a descoberto, indique um procedimento estatístico mais adequado para avaliar se existem diferenças significativas nos dois casos.

3.28. Pretende-se testar se a proporção de ulmeiros afectados pela grafiose é idêntica em duas zonas A e B. Na zona A foi recolhida uma amostra aleatória de 30 ulmeiros e verificou-se que 20 estavam afectados pela grafiose. Na zona B recolheu-se uma amostra de 35 ulmeiros e verificou-se que 27 estavam afectados pela grafiose. Que conclusão se pode tirar ao nível de significância de 0.05?

3.29. Diga, justificando, se cada uma das afirmações seguintes é verdadeira ou falsa:

a) Num teste de hipóteses $H_0 : \mu = 0$ vs. $H_1 : \mu \neq 0$ em que se obteve um *p-value* $= 0.034$, não se rejeita a hipótese de a média populacional ser nula com um nível de significância de 0.01.

b) Se $]2.15, 3.24[$ é um intervalo de confiança a 95% para σ^2 então a probabilidade de σ^2 pertencer a este intervalo é igual a 0.95.

3.30. É desencadeado um programa de controlo da poluição de um rio em que são efectuadas medições, antes de lançar a campanha antipoluição e um ano após. As medições são combinações de vários índices; quanto maior for o valor resultante maior é a poluição. Obtiveram-se os seguintes resultados:

Ponto de controlo	1	2	3	4	5	6	7	8	9	10
Antes da campanha	68	88	101	82	96	74	65	74	52	99
Um ano após	67	87	90	76	98	69	68	65	59	70

Será que a campanha antipoluição reduziu de facto a poluição? Explicite e verifique todas as hipóteses necessárias à resolução do problema, justificando.

3.31. Para comparar dois tipos de máquina de ceifar (segadeiras) quanto à sua eficiência, foram seleccionadas ao acaso 9 searas tendo sido cada uma dividida em dois lotes. Em cada seara uma das segadeiras foi atribuída ao acaso a um dos lotes, ficando a outra para o outro lote.

A eficiência é avaliada num intervalo de valores de 0 (eficiência mínima) a 10 (eficência máxima). Os valores registados relativos à eficiência de cada segadeira foram introduzidos no *software* R. Responda às seguintes questões utilizando os resultados de alguns dos comandos apresentados abaixo.

a) Indique uma estimativa da eficiência média de cada uma das segadeiras.

b) Calculou-se um intervalo a 95% de confiança para a variância da eficiência da segadeira 1, tendo-se obtido $]0.5144; 4.1381[$. Qual o valor observado para a variância da segadeira 1?

c) Será que a segadeira 1 é mais eficiente do que a segadeira 2? Justifique convenientemente indicando e validando as condições necessárias à resolução.

```
> # seg1 designa os valores observados de eficiência da segadeira 1
> # seg2 designa os valores observados de eficiência da segadeira 2

> mean(seg1)   > mean(seg2)     > var(seg2)       > var(seg1-seg2)
[1] 7.4       [1] 6.488889     [1] 0.5861111      [1] 0.6761111


> shapiro.test(seg1)                          > shapiro.test(seg2)
        Shapiro-Wilk normality test                   Shapiro-Wilk normality test
data:  seg1                                   data:  seg2
W = 0.8848, p-value = 0.1762                  W = 0.9105, p-value = 0.3193
```

```
> shapiro.test(seg1-seg2)
        Shapiro-Wilk normality test
data:  seg1 - seg2
W = 0.9648, p-value = 0.8472


> t.test(seg1,seg2,paired=T,alternative="two.sided")
        Paired t-test
data:  seg1 and seg2
t = 3.3242, df = 8, p-value = 0.01047
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.2790663 1.5431559
sample estimates:
mean of the differences
              0.9111111


> t.test(seg1,seg2,paired=T,alternative="greater")
        Paired t-test
data:  seg1 and seg2
t = 3.3242, df = 8, p-value = 0.005237
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 0.4014339       Inf
sample estimates:
mean of the differences
              0.9111111


> t.test(seg1,seg2,paired=T,alternative="less")
        Paired t-test
data:  seg1 and seg2
t = 3.3242, df = 8, p-value = 0.9948
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
     -Inf 1.420788
sample estimates:
mean of the differences
              0.9111111

> t.test(seg1,seg2,paired=F,alternative="two.sided")
        Welch Two Sample t-test
data:  seg1 and seg2
t = 2.088, df = 14.548, p-value = 0.05482
```

```
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.02147036  1.84369258
sample estimates:
mean of x mean of y
 7.400000  6.488889


> t.test(seg1,seg2,paired=F,alternative="greater")
        Welch Two Sample t-test
data:  seg1 and seg2
t = 2.088, df = 14.548, p-value = 0.02741
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 0.1446037       Inf
sample estimates:
mean of x mean of y
 7.400000  6.488889


> t.test(seg1,seg2,paired=F,alternative="less")
        Welch Two Sample t-test
data:  seg1 and seg2
t = 2.088, df = 14.548, p-value = 0.9726
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
     -Inf 1.677619
sample estimates:
mean of x mean of y
 7.400000  6.488889
```

3.32. Uma máquina de ensacar está regulada para encher sacos com 16 kg de açúcar. Para controlar o seu funcionamento escolheram-se, aleatoriamente, 15 sacos, que foram pesados. Os valores obtidos foram introduzidos no *software* R. Responda às seguintes questões utilizando, sempre que possível, os resultados dos comandos apresentados.

a) Indique uma estimativa para o peso médio e para a variância do peso de cada saco.

b) Foi calculado um intervalo de confiança para a variância do peso de cada saco, σ^2, tendo-se obtido $]0.01368\ ;\ 0.04931[$.

   i) Qual o estimador que foi utilizado na construção do intervalo?

   ii) Determine o nível de confiança do intervalo dado.

c) Que conclusão pode tirar quanto à regulação da máquina? Justifique convenientemente, indicando e validando os pressupostos necessários à sua resposta.

```
> peso
 [1] 16.1 15.8 15.9 16.1 15.8 16.2 16.0 15.9 16.0 15.7 15.8 15.7 16.0 16.0 15.8

> mean(peso)                      > shapiro.test(peso)
[1] 15.92                                   Shapiro-Wilk normality test
                                     data:  peso
> var(peso)                         W = 0.9377, p-value = 0.3544
[1] 0.02314286
```

3.33. Um enólogo pretende avaliar a acidez total de um vinho. Para isso selecciona aleatoriamente 20 garrafas de vinho na adega e analisa o seu conteúdo através do método clássico e de um dispositivo de titulação automática. Alguns resultados das análises, em g/l, foram:

garrafa	1	2	3	···	18	19	20
método clássico	4.8	3.4	2.5	···	2.9	5.4	2.1
titulação automática	6.1	5.1	2.1	···	4.5	3.9	1.5

Os dados foram introduzidos no *software* R. Abaixo apresentam-se resultados de comandos, alguns inadequados. Responda às seguintes questões utilizando os resultados apresentados abaixo.

a) De acordo com a legislação em vigor um vinho de mesa deverá ter uma acidez total superior a 3.5 g/l. Com base nos resultados das análises efectuadas pelo <u>método clássico</u>, o enólogo poderá concluir que o seu vinho cumpre os requisitos de acidez impostos pela legislação? Explicite e valide os pressupostos necessários à resolução do problema.

b) Com base nos valores obtidos poder-se-á concluir que os dois métodos de análise da acidez total do vinho têm resultados significativamente diferentes? Explicite e valide os pressupostos necessários à resolução do problema.

```
> classico
 [1] 4.8 3.4 2.5 3.8 4.3 3.6 3.5 3.5 4.0 3.6 6.3 3.0 3.1 3.7 2.8 5.1 4.0 2.9 5.4 2.1
> automatico
 [1] 6.1 5.1 2.1 4.9 6.6 4.0 4.5 1.0 4.7 3.5 8.2 3.9 3.6 4.5 3.5 6.3 4.7 4.5 3.9 1.5

> var(classico)        > var(automatico)        > var(classico-automatico)
[1] 1.040105            [1] 2.887868                [1] 1.326605
```

```
> shapiro.test(classico)
        Shapiro-Wilk normality test
data:  classico
W = 0.951, p-value = 0.3827
```

```
> shapiro.test(automatico)
        Shapiro-Wilk normality test
data:  automatico
W = 0.9625, p-value = 0.5959
```

```
> shapiro.test(classico-automatico)
        Shapiro-Wilk normality test
data:  classico - automatico
W = 0.9163, p-value = 0.08413


> t.test(classico,mu=3.5)
        One Sample t-test
data:  classico
t = 1.184, df = 19, p-value = 0.2510
alternative hypothesis: true mean is not equal to 3.5
95 percent confidence interval:
 3.292693 4.247307
sample estimates:
mean of x
     3.77

> t.test(classico,mu=3.5,alternative="greater")
        One Sample t-test
data:  classico
t = 1.184, df = 19, p-value = 0.1255
alternative hypothesis: true mean is greater than 3.5
95 percent confidence interval:
 3.375677      Inf
sample estimates:
mean of x
     3.77

> t.test(classico,mu=3.5,alternative="less")
        One Sample t-test
data:  classico
t = 1.184, df = 19, p-value = 0.8745
alternative hypothesis: true mean is less than 3.5
95 percent confidence interval:
     -Inf 4.164323
sample estimates:
mean of x
     3.77
```

```
> t.test(classico,automatico,paired=FALSE, var.equal=TRUE)
        Two Sample t-test
data:  classico and automatico
t = -1.32, df = 38, p-value = 0.1947
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.4821486  0.3121486
sample estimates:
mean of x mean of y
    3.770     4.355

> t.test(classico,automatico,paired=TRUE)
        Paired t-test
data:  classico and automatico
t = -2.2714, df = 19, p-value = 0.03493
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.12405128 -0.04594872
sample estimates:
mean of the differences
                 -0.585
```

3.34. 10 pessoas asmáticas participaram numa experiência destinada a estudar o efeito de um novo tratamento da função pulmonar. Durante 1 segundo registou-se o volume de ar expirado (em litros) antes e depois da aplicação do tratamento. Os resultados obtidos foram os seguintes:

	1	2	3	4	5	6	7	8	9	10
Antes	1.69	2.77	1.00	1.66	0.85	1.42	2.82	2.58	1.98	2.02
Depois	2.69	2.22	3.07	3.35	2.74	3.61	5.14	2.44	2.25	3.41

Nota: Para responder a esta pergunta utilize, sempre que possível, resultados que lhe são apresentados no Anexo.

a) Determine um intervalo a 99% de confiança para o volume médio de ar expirado por uma pessoa antes da aplicação do tratamento. Indique e valide os pressupostos necessários à sua resolução.

b) Qual é o erro máximo cometido ao usar a média da amostra para estimar o volume médio de ar expirado por pessoa, com o nível de confiança da alínea anterior?

c) Os dados recolhidos dão indicação de que o volume de ar expirado é superior após o tratamento? Justifique, indicando e validando os pressupostos necessários à resolução desta questão.

ANEXO

```
>antes<-c(1.69,2.77,1.00,1.66,.85,1.42,2.82,2.58,1.98,2.02)
>depois<-c(2.69,2.22,3.07,3.35,2.74,3.61,5.14,2.44,2.25,3.41)

> var(antes);var(depois)
[1] 0.4807433
[1] 0.7567511

> var(antes-depois)
[1] 1.056112

> shapiro.test(antes)

        Shapiro-Wilk normality test

data:  antes
W = 0.9402, p-value = 0.5558

> shapiro.test(depois)

        Shapiro-Wilk normality test

data:  depois
W = 0.8633, p-value = 0.08348

> shapiro.test(antes-depois)

        Shapiro-Wilk normality test

data:  antes - depois
W = 0.8972, p-value = 0.2039

> t.test(antes,depois,paired=F,var.equal=T)

        Two Sample t-test

data:  antes and depois
t = -3.4482, df = 18, p-value = 0.002868
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.9520632 -0.4739368
```

```
sample estimates:
mean of x mean of y 
    1.879     3.092 

> t.test(antes,depois,paired=F,var.equal=T,alternative="less")

        Two Sample t-test

data:  antes and depois 
t = -3.4482, df = 18, p-value = 0.001434
alternative hypothesis: true difference in means is less than 0 
95 percent confidence interval:
       -Inf -0.6029904 
sample estimates:
mean of x mean of y 
    1.879     3.092 

> t.test(antes,depois,paired=F,var.equal=T,alternative="greater")

        Two Sample t-test

data:  antes and depois 
t = -3.4482, df = 18, p-value = 0.9986
alternative hypothesis: true difference in means is greater than 0 
95 percent confidence interval:
 -1.823010       Inf 
sample estimates:
mean of x mean of y 
    1.879     3.092 

> t.test(antes,depois,paired=T)

        Paired t-test

data:  antes and depois 
t = -3.7326, df = 9, p-value = 0.004679
alternative hypothesis: true difference in means is not equal to 0 
95 percent confidence interval:
 -1.9481531 -0.4778469 
sample estimates:
mean of the differences 
                 -1.213 
```

```
> t.test(antes,depois,paired=T,alternative="less")

        Paired t-test

data:  antes and depois
t = -3.7326, df = 9, p-value = 0.002339
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
       -Inf -0.6172772
sample estimates:
mean of the differences
                 -1.213

> t.test(antes,depois,paired=T,alternative="greater")

        Paired t-test

data:  antes and depois
t = -3.7326, df = 9, p-value = 0.9977
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 -1.808723       Inf
sample estimates:
mean of the differences
                 -1.213
```

3.35 Uma associação de defesa de consumidores pretende publicar um estudo comparativo dos preços de dois supermercados. Para isso seleccionou 20 produtos e registou os preços de cada um nos supermercados A e B (em cêntimos):

Produto	1	2	3	4	5	6	···	15	16	17	18	19	20
Super. A	202	201	560	253	384	332	···	624	851	742	501	476	765
Super. B	185	187	516	239	349	295	···	613	851	731	546	490	762

Utilize os resultados apresentados abaixo para responder às seguintes questões.

a) Determine uma estimativa para a diferença dos preços médios nos dois supermercados.

b) Com base nos dados recolhidos, o que pode afirmar relativamente aos preços médios praticados pelos dois supermercados? Explicite e valide as hipóteses necessárias à resolução do problema.

```
> A
 [1] 202 201 560 253 384 332 549 722 153 676 804 535 472 335 624 851 742 501 476 765
> B
 [1] 185 187 516 239 349 295 552 667 132 676 745 529 460 316 613 851 731 546 490 762
```

```
> mean(A)
[1] 506.85
> mean(B)
[1] 492.05
> var(A)
[1] 46116.77
> var(B)
[1] 46454.68
> var(A-B)
[1] 568.379
```

```
> shapiro.test(A)
  Shapiro-Wilk normality test
data:  A
W = 0.957, p-value = 0.4864
```

```
> shapiro.test(B)
 Shapiro-Wilk ormality test
data:  B
W = 0.9552, p-value = 0.4521
```

```
> shapiro.test(A-B)
    Shapiro-Wilk normality test
data:  A - B
W = 0.9514, p-value = 0.3889
```

3.36 Uma empresa agrícola pretende adquirir um gerador. Face a dois tipos de geradores (que designamos por ger1 e ger2), de preços semelhantes, quer decidir comparando a produção de energia eléctrica, em kWh, de cada um dos geradores. Recolheu uma amostra de 25 observações da quantidade de energia eléctrica produzida por cada gerador, em diferentes situações. Os dados foram introduzidos no *software* R. Considere os seguintes resultados referentes ao gerador tipo 1:

```
>var(ger1); mean(ger1)
[1] 18.90528
[1] 9.6056
```

```
>shapiro.test(ger1)
Shapiro-Wilk normality test
 data:  ger1
 W = 0.9514, p-value = 0.2698
```

a) Indique uma estimativa da variância da produção de energia eléctrica do gerador 1. Explicite o estimador que lhe está associado.

b) Poder-se-á considerar que a variância da produção de energia eléctrica no gerador 1 é superior a $15(\text{kWh})^2$? Justifique convenientemente a sua resposta.

c) Para obter um intervalo de confiança a 95% para a diferença entre as produções médias de energia eléctrica dos geradores 1 e 2, usou-se o comando do R que se apresenta de seguida e respectivo resultado (no qual três valores foram substituídos por **????**):

```
>t.test(ger1,ger2,paired=????,var.equal=T)

        Two Sample t-test

data:  ger1 and ger2
t = -0.6632, df = ???? , p-value = 0.5104
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -3.009280  ????
sample estimates:
mean of x mean of y
   9.6056   10.3520
```

i) Qual a diferença entre as médias observadas para a produção de energia eléctrica dos dois geradores? Qual o erro máximo que se cometeria, a 95% de confiança, ao identificar esta diferença com a verdadeira diferença entre as produções médias de energia eléctrica dos geradores?

ii) Complete o *output* indicando a informação em falta nos três locais assinalados por **????**.

iii) Que pressupostos foi necessário verificar para construir o intervalo de confiança dado no *output*?

– O que pode afirmar, a 95% de confiança, sobre a diferença entre as produções médias de energia eléctrica dos geradores?

– A conclusão da alínea anterior seria alterada se considerasse uma confiança de 99%? Justifique.

Exercícios de Revisão de Inferência Estatística

R3.1. Recolheu-se uma amostra aleatória de dimensão 5 de uma população normal. Determine a probabilidade de o desvio padrão da amostra ser inferior ao desvio padrão da população.

R3.2. Seja X uma variável aleatória com distribuição binomial, $X \cap B(n, p)$. Considere os estimadores para p assim definidos $\widehat{P_1} = \frac{X+1}{n+1}$ $\widehat{P_2} = \frac{X}{n+1}$

a) Determine, em função dos parâmetros da distribuição de X, o valor médio e a variância dos estimadores $\widehat{P}_1$ e $\widehat{P}_2$. Qual deles preferiria? Justifique.

b) Suponha que numa amostra de dimensão 20, se observou a ocorrência de 5 sucessos. Determine uma estimativa para p.

R3.3. A quantidade (em *ppm*) de um poluente no solo de uma certa região é uma v.a. X que se admite ter distribuição normal com valor médio μ e variância σ^2, desconhecidos.

Recolheu-se uma amostra de dimensão 10 tendo-se obtido o desvio padrão de 1.3943 *ppm*. Com base na amostra construiu-se o seguinte intervalo de confiança para μ

$$]22.447\ ;\ 25.313[$$

a) Indique estimativas do valor médio e da variância da quantidade daquele poluente no solo.

b) Qual o grau de confiança daquele intervalo?

c) Construa um intervalo a 95% de confiança para σ^2.

R3.4. Num certo processo químico é muito importante que uma dada solução tenha um pH de exactamente 8.20. O método utilizado na determinação do pH fornece medições que se admite terem distribuição normal de valor médio igual ao verdadeiro valor do pH da solução e desvio padrão de 0.02.

Para avaliar o pH de uma solução, efectuaram-se 10 medições independentes tendo-se obtido os seguintes valores:

8.18 8.16 8.17 8.22 8.19 8.17 8.15 8.21 8.16 8.18

a) Indique uma estimativa do valor médio do pH da solução.

b) Com base nestas 10 medições, o que pode concluir relativamente à utilização desta solução no referido processo químico?

c) Pretende-se efectuar um novo conjunto de medições para diminuir o erro máximo cometido na estimativa do verdadeiro valor do pH da solução. Mantendo-se todas as condições referidas acima, qual deverá ser o tamanho da amostra para que aquele erro máximo não exceda 0.01, a 95% de confiança?

R3.5. Faz-se uma experiência para saber se dois regimes alimentares A e B produzem o mesmo aumento de peso nos animais, durante um período de tempo fixado. Tomam-se 20 animais e de entre eles 10 ao acaso aos quais é dado o alimento A. Aos outros 10 é dado o alimento B. Os aumentos de peso (expressos em kg) no mesmo intervalo de tempo são os seguintes:

Regime A	-2.0	0.0	4.2	6.3	9.6	4.3	10.2	11.0	12.4	13.1
Regime B	4.0	6.0	8.0	11.3	12.3	14.4	14.5	14.7	14.7	16.0

Diga se existe diferença significativa entre os dois regimes alimentares, justificando convenientemente todas as hipóteses necessárias à resolução do problema.

R3.6. Um estudo pretende comparar um tipo de semente melhorada com o tipo de semente usado anteriormente. A nova semente passará a ser utilizada se, em média, o crescimento das plantas após 20 dias fôr superior ao das obtidas das 'velhas' sementes. São criadas 15 diferentes situações laboratoriais, variando temperatura e humidade. Em cada situação semeia-se uma semente de cada tipo e obtêm-se os seguintes resultados para o crescimento (em cm) das plantas após 20 dias :

Situação	1	2	3	4	5	6	7	8
Sementes melhoradas	3.46	3.48	2.74	2.83	4.00	4.95	2.24	6.92
sementes tradicionais	3.18	3.67	2.92	3.10	4.10	4.86	2.21	6.91

Situação	9	10	11	12	13	14	15
sementes melhoradas	6.57	6.18	8.30	3.44	4.47	7.59	3.87
sementes tradicionais	6.83	6.19	8.05	3.46	4.18	7.43	3.85

Deverá passar a usar-se as sementes melhoradas? Responda justificando e explicitando quaisquer hipóteses adicionais que seja necessário impôr.

R3.7. Num estudo sobre a incidência de certa doença numa população de insectos, um grupo de biólogos registou ao longo de um ano o número de insectos contaminados em cada amostra de 5 insectos, tendo para tal recolhido 100 amostras. Os resultados obtidos foram:

Num. de insectos contaminados	0	1	2	3	4	5
Num. de amostras	9	26	34	22	8	1

a) Construa uma tabela de frequências absolutas, relativas e relativas acumuladas para o número de insectos contaminados por amostra.

b) Determine a média e a mediana do número de insectos contaminados por amostra.

c) Considerando agora a totalidade de insectos observados como uma amostra aleatória da população de insectos:

   i) Determine uma estimativa da verdadeira proporção de insectos contaminados.

ii) Qual é o erro máximo associado à estimativa obtida na alínea anterior, com uma confiança de 95% ?

R3.8. O dono de uma ervanária produz um chá relativamente ao qual afirma que é 90% eficaz para curar dores de cabeça. Num inquérito feito a 250 pessoas, 198 concordaram que o chá cura as dores de cabeça. Acha que o resultado do inquérito é compatível com a pretensão do produtor ?

R3.9. Um investigador pretende estudar a incidência a nível nacional, de uma doença que ataca os pinheiros. Observações efectuadas através do país resultaram em 1233 casos de pinheiros afectados (a nível nacional) num total de 4250 observações.

a) Estime a percentagem de pinheiros afectados a nível nacional.

b) Determine um intervalo a 95% de confiança para a verdadeira proporção de pinheiros afectados.

R3.10. Em várias espécies de animais, altos níveis de testosterona tornam os machos mais atractivos para as fêmeas. No entanto, conjectura-se que altos níveis de testosterona podem enfraquecer o sistema imunitário, i.e., reduzem o número de anticorpos. Para responder a esta questão consideraram-se 13 melros de asa-vermelha, em cada um dos quais foi implantado um tubo contendo testosterona. Em cada melro avaliou-se o número de anticorpos no sangue antes e após o implante. Os resultados obtidos, em unidades convenientes, foram introduzidos no *software* R.

Utilize, sempre que possível, os resultados apresentados no **Anexo** para responder às seguintes questões (alguns dos procedimentos apresentados são inadequados).

a) Indique estimativas para o valor médio do número de anticorpos antes e depois do implante.

b) Determine um intervalo de confiança a 95% para a variabilidade do número de anticorpos antes da introdução do implante.

c) A conjectura de que o aumento do nível de testosterona enfraquece o sistema imunitário será compatível com os resultados obtidos? Explicite e valide os pressupostos necessários à resolução do problema.

ANEXO

```
antes <-c(4.65,4.21,4.91,4.50,4.80,5.01,4.88,4.78,4.98,4.41,4.75,4.70,4.93)
depois <-c(4.54,3.80,4.98,4.45,5.00,5.00,4.35,4.96,5.02,4.73,4.77,4.60,5.01)

> var(antes)                > var(depois)
[1] 0.05638077              [1] 0.1299974
```

```
> shapiro.test(antes)
        Shapiro-Wilk normality test
data:  antes
W = 0.9196, p-value = 0.2477

> shapiro.test(depois)
        Shapiro-Wilk normality test
data:  depois
W = 0.831, p-value = 0.01631

> shapiro.test(antes-depois)
        Shapiro-Wilk normality test
data:  antes - depois
W = 0.9169, p-value = 0.2275

> var.test(antes,depois)
        F test to compare two variances
data:  antes and depois
F = 0.4337, num df = 12, denom df = 12, p-value = 0.1622
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.1323375 1.4213773
sample estimates:
ratio of variances
         0.4337068

> t.test(antes,depois,paired=F,var.equal=T,alternative="two.sided")
        Two Sample t-test
data:  antes and depois
t = 0.1927, df = 24, p-value = 0.8488
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.2240467  0.2702005
sample estimates:
mean of x mean of y
 4.731538  4.708462

> t.test(antes,depois,paired=F,var.equal=T,alternative="less")
        Two Sample t-test
data:  antes and depois
t = 0.1927, df = 24, p-value = 0.5756
alternative hypothesis: true difference in means is less than 0
```

```
95 percent confidence interval:
      -Inf 0.2279316
sample estimates:
mean of x mean of y
 4.731538  4.708462

> t.test(antes,depois,paired=F,var.equal=T,alternative="greater")
        Two Sample t-test
data:  antes and depois
t = 0.1927, df = 24, p-value = 0.4244
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 -0.1817778        Inf
sample estimates:
mean of x mean of y
 4.731538  4.708462

> t.test(antes,depois,paired=T,alternative="two.sided")
        Paired t-test
data:  antes and depois
t = 0.3562, df = 12, p-value = 0.7279
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.1180915  0.1642453
sample estimates:
mean of the differences
             0.02307692

> t.test(antes,depois,paired=T,alternative="less")
        Paired t-test
data:  antes and depois
t = 0.3562, df = 12, p-value = 0.6361
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
      -Inf 0.1385539
sample estimates:
mean of the differences
             0.02307692

> t.test(antes,depois,paired=T,alternative="greater")
        Paired t-test
data:  antes and depois
```

```
t = 0.3562, df = 12, p-value = 0.3639
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 -0.09240003         Inf
sample estimates:
mean of the differences
             0.02307692
```

R3.11. Um agricultor pretende experimentar dois tipos (I e II) de semeador de milho para comparar a produtividade das duas máquinas.

Para isso, num campo agrícola marcaram-se aleatoriamente 10 parcelas de igual área sendo cada uma delas dividida em duas secções iguais. Em cada parcela sorteou-se a atribuição de um dos tipos de semeador a uma das secções.

A produtividade (em unidades adequadas) de cada um dos semeadores foi registada, e os dados foram introduzidos no *software* R. Utilizando os resultados apresentados abaixo (alguns desnecessários), poderá o agricultor admitir que a produtividade esperada das duas máquinas é igual? Justifique convenientemente a sua resposta.

```
>x<-c(5.6, 8.4, 8.0, 8.6, 6.4, 7.8, 6.2, 7.7, 8.0, 7.7)
observações para o semeador tipo I.
>y<-c(6.0, 7.4, 7.3, 7.5, 6.4, 6.0, 5.5, 6.6, 5.6, 6.1)
observações para o semeador tipo II.


>mean(x)       >mean(y)        >mean(x-y)
[1] 7.44        [1] 6.44         [1] 1

>var(x)        >var(y)         >var(x-y)
[1] 1.018222    [1] 0.5448889    [1] 0.68


> shapiro.test(x)

        Shapiro-Wilk normality test

data:  x
W = 0.8731, p-value = 0.1087

> shapiro.test(y)

        Shapiro-Wilk normality test
```

```
data:  y
W = 0.9026, p-value = 0.2341

> shapiro.test(x-y)

        Shapiro-Wilk normality test

data:  x - y
W = 0.9765, p-value = 0.9435
```

R3.12. Um estudante de agronomia pretende estudar a produção de uma certa variedade de milho numa dada região. Para esse efeito, feita a colheita, pesou 50 espigas e organizou os resultados na tabela abaixo:

Peso(g)	x'_i	n_i	N_i	f_i	F_i
] 180, 190]	185	3			
] 190, 200]			12		
] 200, 210]		9		0.18	
] 210, 220]			30		
] 220, 230]		12			0.84
] 230, 240]		8	50		

a) Identifique e classifique, justificando, a variável estatística em estudo.

b) Diga o que representa o símbolo no cabeçalho de cada coluna e complete a tabela. Esboce o histograma associado.

c) Calcule um valor aproximado da média e da mediana do peso das espigas.

d) Indique uma estimativa da percentagem de espigas com peso superior a 220 g.

e) O estudante tirou a seguinte conclusão: "Esta variedade de milho tem uma boa produção porque mais de 30% das espigas têm peso superior a 220 g". Está de acordo com o seu colega? Justifique convenientemente.

R3.13. Seja $X_1, X_2, \cdots, X_n$ uma amostra aleatória retirada de uma população com distribuição normal com valor médio μ e variância σ^2 conhecida.

a) Qual é a probabilidade de o intervalo aleatório

$$\left]\overline{X} - 1.96\frac{\sigma}{\sqrt{n}}, \overline{X} + 1.96\frac{\sigma}{\sqrt{n}}\right[$$

conter o valor μ?

b) Qual é o erro máximo cometido ao usar $\overline{x}$ para estimar μ?

c) Qual é a redução do erro máximo que se obtém quando a dimensão da amostra duplica?

d) Mostre que $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ é um estimador centrado de σ^2.

R3.14. Seja X uma variável aleatória com valor esperado μ e variância σ^2 desconhecidos. Recolheram-se aleatoriamente 15 dados da população X, que se introduziram no *software* no objecto `x`. Responda às seguintes questões utilizando, sempre que possível, o *output* apresentado em Anexo.

a) Pretende-se estimar o valor médio, μ, da população X.

i) Mostre que $\overline{X}$, média de uma amostra aleatória de dimensão n retirada de X, é um estimador centrado de μ. Exprima a variância de $\overline{X}$ como função dos parâmetros da população.

ii) Com base na amostra recolhida indique estimativas do valor médio de X e da variância de $\overline{X}$.

b) Pretende-se averiguar se a variância da população é superior a 1, pelo que se decide realizar um teste de hipóteses.

i) Formule as hipóteses do teste que permite responder a esta questão.

ii) Indique a estatística de teste e a sua distribuição. Valide os pressupostos necessários à existência dessa distribuição.

iii) Conclua o teste de hipóteses.

c) Recolheu-se uma amostra de dimensão 15 de outra população Y. Os dados foram introduzidos no objecto `y` do `R`.

i) Indique um intervalo de confiança a 95% para a diferença das médias das populações X e Y. Explicite e valide os pressupostos necessários.

ii) Os dados recolhidos evidenciam que as médias das duas populações são iguais? Justifique convenientemente.

ANEXO

```
> mean(x)                  > mean(y)
[1] 5.686667              [1] 4.646667

> var(x)                  > var(y)
[1] 1.084095              [1] 1.855524


> shapiro.test(x)                             > shapiro.test(y)
```

```
        Shapiro-Wilk normality test                    Shapiro-Wilk normality test
data:  x                                       data:  y
W = 0.9747, p-value = 0.6733                   W = 0.9577, p-value = 0.2704

> shapiro.test(x-y)
        Shapiro-Wilk normality test
data:  x - y
W = 0.924, p-value = 0.2214

> var.test(x,y)

        F test to compare two variances

data:  x and y
F = 0.5843, num df = 14, denom df = 14, p-value = 0.3261
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.196151 1.740249
sample estimates:
ratio of variances
         0.5842529

> t.test(x,y,paired=T)
        Paired t-test
data:  x and y
t = 2.1864, df = 14, p-value = 0.04627
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.01977868 2.06022132
sample estimates:
mean of the differences
                   1.04


> t.test(x,y,paired=F,var.equal=T)
        Two Sample t-test
data:  x and y
t = 2.3493, df = 28, p-value = 0.0261
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.1331903 1.9468097
sample estimates:
mean of x mean of y
 5.686667  4.646667
```

Soluções de alguns Exercícios

3.1. a) 0.2112

b) $\simeq 0.2112$

3.2. a) Como $n = 100$ é "grande", pelo Teorema Limite Central, verifica-se para a média amostral, $\overline{X} = \sum_{i=1}^{100} X_i/n$, $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1) \Longrightarrow \frac{\overline{X}-31}{0.24} \sim N(0,1)$. Logo, $P[\overline{X} < 30] \simeq \Phi(\frac{30-31}{0.24}) = \Phi(-4.17) = 1 - \Phi(4.17) = 1 - 1 = 0$

b) Seja $S_{100} = \sum_{i=1}^{100} X_i$ o peso total da amostra aleatória. Pelo Teorema Limite Central tem-se

$$\frac{S_{100} - 100\mu}{\sigma\sqrt{100}} \sim N(0,1) \Longrightarrow \frac{S_{100} - 3100}{24} \sim N(0,1).$$

Então, $P[S_{100} > 3150] = 1 - P[S_{100} \leq 3150] \simeq 1 - \Phi(\frac{3150-3100}{24}) = 1 - \Phi(2.08) = 1 - 0.9812 = 0.0188$.

3.3. a) $\overline{x} = 56.2$ cm e $s = 4$ cm

b) 0.42074

c) 4/5=0.8 cm.

3.4. b) p

d) $\hat{p}_1 = 1/3$ e $\hat{p}_2 = 1/2$

3.5. 1082

3.6. $\approx 1 - \Phi(0.4\sqrt{12n})$, para n elevado

3.7. ≈ 0.01267

3.12. a) $]77.27,\ 79.33[$

b) 1.03

c) 2654

3.13. 18.3 e $n = 12$

3.14. a) $\overline{x} = 0.959$ Kg e $s^2 = 0.0116$ Kg^2

c) $]0.00886,\ 0.0155[$

3.15. a) 99%

b) 0.185

3.16. a) i) 27 ii) 0.95

b)]14.76, 64.77[

3.23. a) Unidade estatística: planta de tabaco; variável: nº de folhas por planta

b) $\overline{x} = 2967/150 = 19.78$ e $\tilde{x} = 20$

3.25. a) $\overline{x}_{\mathrm{S}} = 100$, $\overline{x}_{\mathrm{N}} = 60$, $s^2_{\mathrm{S}} = 283.33$ e $s^2_{\mathrm{N}} = 207.89$

3.29. a) Verdadeira

b) Falsa

R3.1. 0.7127

R3.2. a) $E[\widehat{P_1}] = \frac{np+1}{n+1}$; $E[\widehat{P_2}] = \frac{np}{n+1}$;

$Var[\widehat{P_1}] = Var[\widehat{P_2}] = \frac{npq}{(n+1)^2}$.

$\widehat{P_2}$ é o que apresenta menor EQM portanto é o melhor.

b) Usando o melhor estimador, $\widehat{P_2}$, temos uma estimativa para p, $\hat{p}_2 = 5/21$.

R3.3. a) Estimativa do valor médio é $\overline{x} = 23.88\ ppm$ e de σ^2 é $s^2 = 1.9441\ ppm^2$

b) O grau de confiança é 99%.

R3.4. a) $\overline{x} = 8.179$

c) $n \geq 16$

R3.7. a)

x_i	Freq. Abs. n_i	Freq. Rel. f_i	Freq. Rel Acum. F_i
0	9	0.09	0.09
1	26	0.26	0.35
2	34	0.34	0.69
3	22	0.22	0.91
4	8	0.08	0.99
5	1	0.01	1.00

b) $\overline{x} = 197/100$ e $\tilde{x} = 2$

c) i) $\hat{p} = 197/500 = 0.394$ ii) 0.0428

R3.9. a) 29.01%

b)]0.2765, 0.3037[