Departamento de Matemática – Secção de Estatística e Aplicações

Probabilidades e Estatística

EXERCÍCIOS

Edição de Setembro de 2008

Formulário

$P(X=x)=\binom{n}{x}p^x(1-p)^{n-x}$ $x=0,1,\ldots,n$ $E(X)=np \quad Var(X)=np(1-p)$	$P(X=x)=\frac{e^{-\lambda}\lambda^x}{x!}$ $x=0,1,\ldots$ $E(X)=Var(X)=\lambda$	$P(X=x)=p(1-p)^{x-1}$ $x=1,2,\ldots$ $E(X)=\frac{1}{p} \quad Var(X)=\frac{(1-p)}{p^2}$

$P(X=x)=\binom{M}{x}\binom{N-M}{n-x}\Big/\binom{N}{n}$ $x=\max\{0,n-N+M\},\ldots,\min\{n,M\}$ $E(X)=n\frac{M}{N} \quad Var(X)=n\frac{M}{N}\frac{N-M}{N}\frac{N-n}{N-1}$	$f_X(x)=\frac{1}{b-a},\ a\le x\le b$ $E(X)=\frac{b+a}{2} \quad Var(X)=\frac{(b-a)^2}{12}$
$f_X(x)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\},\ x\in I\!R$ $E(X)=\mu \quad Var(X)=\sigma^2$	$f_X(x)=\lambda e^{-\lambda x},\ x\ge 0$ $E(X)=\frac{1}{\lambda} \quad Var(X)=\frac{1}{\lambda^2}$

$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$	$\frac{\bar{X}-\mu}{S/\sqrt{n}}\sim t_{(n-1)}$	$\frac{\left(\bar{X}_1-\bar{X}_2\right)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\sim N(0,1)$
$\frac{\bar{X}-\mu}{S/\sqrt{n}}\overset{a}{\sim} N(0,1)$	$S^2=\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2$	

$\frac{\left(\bar{X}_1-\bar{X}_2\right)-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}\overset{a}{\sim} N(0,1)$	$\frac{\left(\bar{X}_1-\bar{X}_2\right)-(\mu_1-\mu_2)}{\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}\sim t_{(n_1+n_2-2)}$

$\frac{(n-1)S^2}{\sigma^2}\sim\chi^2_{(n-1)}$	$\sum_{i=1}^{k}\frac{(O_i-E_i)^2}{E_i}\overset{a}{\sim}\chi^2_{(k-\beta-1)}$	$\sum_{i=1}^{r}\sum_{j=1}^{s}\frac{(O_{ij}-E_{ij})^2}{E_{ij}}\overset{a}{\sim}\chi^2_{(r-1)(s-1)}$

$Y_i=\beta_0+\beta_1x_i+\varepsilon_i$	$\hat{\beta}_0=\bar{Y}-\hat{\beta}_1\bar{x}$	$\hat{\beta}_1=\frac{\sum_{i=1}^{n}x_iY_i-n\bar{x}\bar{Y}}{\sum_{i=1}^{n}x_i^2-n\bar{x}^2}$

$\hat{\sigma}^2=\frac{1}{n-2}\sum_{i=1}^{n}\left(Y_i-\hat{Y}_i\right)^2,\ \hat{Y}_i=\hat{\beta}_0+\hat{\beta}_1x_i$	$\hat{\sigma}^2=\frac{1}{n-2}\left[\left(\sum_{i=1}^{n}Y_i^2-n\bar{Y}^2\right)-\left(\hat{\beta}_1\right)^2\left(\sum_{i=1}^{n}x_i^2-n\bar{x}^2\right)\right]$

$\frac{\hat{\beta}_0-\beta_0}{\sqrt{\left(\frac{1}{n}+\frac{\bar{x}^2}{\sum x_i^2-n\bar{x}^2}\right)\hat{\sigma}^2}}\sim t_{(n-2)}$	$\frac{\hat{\beta}_1-\beta_1}{\sqrt{\frac{\hat{\sigma}^2}{\sum x_i^2-n\bar{x}^2}}}\sim t_{(n-2)}$	$\frac{\left(\hat{\beta}_0+\hat{\beta}_1x_0\right)-(\beta_0+\beta_1x_0)}{\sqrt{\left(\frac{1}{n}+\frac{(\bar{x}-x_0)^2}{\sum x_i^2-n\bar{x}^2}\right)\hat{\sigma}^2}}\sim t_{(n-2)}$

$$R^2=\frac{\left(\sum_{i=1}^{n}x_iY_i-n\bar{x}\bar{Y}\right)^2}{\left(\sum_{i=1}^{n}x_i^2-n\bar{x}^2\right)\times\left(\sum_{i=1}^{n}Y_i^2-n\bar{Y}^2\right)}$$

Capítulo 1

Estatística descritiva

1.1 Uma escola avalia o seu curso através de um questionário com 50 perguntas sobre diversos aspectos de interesse. Cada pergunta tem uma resposta numa escala de 1 a 5, onde a maior nota significa melhor desempenho. Para cada aluno é então encontrada a nota média. Na última avaliação recorreu-se a uma amostra de 42 alunos, e os resultados estão em baixo.

4.2	2.7	4.6	2.5	3.3	4.7
4.0	2.4	3.9	1.2	4.1	4.0
3.1	2.4	3.8	3.8	1.8	4.5
2.7	2.2	3.7	2.2	4.4	2.8
2.3	1.9	3.6	3.9	2.3	3.4
3.3	1.8	3.5	4.1	2.2	3.0
4.1	3.4	3.2	2.2	3.0	2.8

(a) Proceda à organização dos dados construindo um quadro de frequências onde figurem as frequências absolutas, absolutas acumuladas e relativas acumuladas.

(b) Desenhe o respectivo histograma.

(c) Identifique as classes modal e mediana.

(d) Calcule a média e o desvio padrão usando os dados agrupados e também usando os dados não agrupados. Compare os resultados.

(e) Calcule a mediana e os 1º e 3º quartis.

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

nº de sementes germinadas por vaso	0	1	2	3	4	5
nº de vasos	16	32	89	137	98	25

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

(b) Represente graficamente os resultados.

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

1.3 Realizou-se uma experiência com uma perfuradora hidráulica a fim de conhecer a sua capacidade de perfuração em estruturas rochosas. Para tal foi observada a profundidade (em polegadas) de perfuração em 10 locais, cujos dados se encontram abaixo:

10.6	10.7	10.1	10.9	10.8
10.2	11.0	10.3	10.5	10.9

Apresente três medidas de localização e de dispersão para os dados observados, interpretando-as e sugerindo qual a melhor, dentro de cada um dos grupos de medidas.

1.4 As notas finais obtidas em 3 turmas na disciplina de Probabilidades e Estatística foram as seguintes:

Turma	1	2	3
nº alunos	30	35	40
média	13	10	9
desvio padrão	2	2.2	2.1

(a) Calcule a média e o desvio padrão das notas obtidas no conjunto de todos os alunos.

(b) No final o professor entendeu alterar linearmente as notas de forma que a média e o desvio padrão das notas de todos os alunos fossem 12 e 2 respectivamente. Sabendo que um aluno da turma 1 obteve 10 valores, calcule a sua nota na nova escala adoptada pelo professor.

1.5 O departamento de pessoal de uma certa firma fez um levantamento dos salários dos 120 funcionários do sector administrativo, tendo obtido os seguintes resultados.

Faixa salarial	Frequência Relativa
$[0, 2]$	0.25
$]2, 4]$	0.40
$]4, 6]$	0.20
$]6, 10]$	0.15

(a) Esboçe o histograma correspondente.

(b) Calcule aproximadamente a média, a variância e o desvio padrão dos salários.

(c) Se for concedido um aumento de 100% a todos os funcionários, haverá alteração na média dos salários? E na variância dos salários? Justifique.

(d) Responda à questão anterior para o caso de ser concedido um aumento de 2 unidades a todos os funcionários.

Capítulo 2

Noções de probabilidade

2.1 Admita que um lote contém peças pesando 5, 10, 15, 20 g e que existem pelo menos 2 peças de cada peso. Retiram-se 2 peças do lote. Seja X o peso da 1ª peça retirada e Y o peso da 2ª peça retirada. Utilizando o plano xy marque:

(a) O espaço de resultados.

(b) O acontecimento $A = \{(x, y) : x = y\}$.

(c) O acontecimento $B = \{(x, y) : y > x\}$.

(d) O acontecimento $C =$ "A 2ª peça é duas vezes mais pesada do que a 1ª ".

(e) O acontecimento $D =$ "A 1ª peça pesa menos 10g do que a 2ª ".

(f) O acontecimento $E =$ "O peso médio das duas peças é menor que 15 g".

2.2 Sejam A e B acontecimentos 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) Não se realizar nenhum dos dois acontecimentos.

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

(c) Que se realize pelo menos um dos dois acontecimentos.

(d) Que se realize quanto muito um único acontecimento.

2.3 Mostre que:

(a) Se A e B são acontecimentos tais que $A \subset B$ então $P(A) \leq P(B)$.

(b) Para quaisquer acontecimentos C e D tem-se

$$P(C \cap D) \leq P(C) \leq P(C \cup D).$$

(c) $P\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} P(A_i), \quad \forall n \in I\!N.$

2.4 Uma colecção de 100 programas de computador foi examinada para detectar erros de "sintaxe", "input/output" e de "outro tipo" diferente dos anteriores. Desses 100 programas, 20 tinham erros de "sintaxe", 10 tinham erros de "input/output" e 5 tinham erros de "outro tipo", 6 tinham erros de "sintaxe" e de "input/output", 3 tinham erros de "sintaxe"e de "outro tipo", 3 tinham erros de "input/output"e de "outro tipo”e 2 tinham os três tipos de erros considerados. Um programa é seleccionado ao acaso desta colecção. Determine a probabilidade de que o programa seleccionado tenha:

(a) Exclusivamente erros de “sintaxe”.

(b) Pelo menos um dos três tipos de erros.

2.5 Num lançamento de um dado viciado, a probabilidade de ocorrer cada número ímpar é o dobro da probabilidade de ocorrer cada número par.

(a) Indique qual o espaço de resultados e calcule a probabilidade de cada acontecimento elementar.

(b) Calcule a probabilidade de que o número de pontos obtido no lançamento do dado seja superior a 3.

(c) Calcule a probabilidade de que o número de pontos obtido no lançamento do dado seja um quadrado perfeito.

2.6 Uma lotaria tem 10000 bilhetes numerados de 0000 a 9999. O número do primeiro prémio é o número do bilhete saído numa extracção ao acaso.

(a) Um jogador comprou um bilhete com o número 6789. Qual a probabilidade de lhe sair o primeiro prémio?

(b) Se o jogador comprar todos os bilhetes cujos números têm todos os algarismos iguais, qual a probabilidade de lhe sair o primeiro prémio?

(c) Qual a probabilidade do número premiado ter todos os algarismos diferentes?

2.7 Numa fila de espera de autocarro estão 4 homens, 3 mulheres e 2 crianças. Qual a probabilidade de:

(a) As pessoas, dentro de cada um daqueles três grupos, estarem de seguida?

(b) As 2 crianças estarem juntas?

2.8 Considere o lançamento de 3 dados perfeitos, sendo um branco, outro preto e outro verde. Determine a probabilidade de obter uma soma de pontos igual a 10.

2.9 De um grupo de 50 alunos do IST (10 alunos por ano) é escolhida ao acaso uma comissão coordenadora de 4 pessoas. Qual a probabilidade de:

(a) Ser escolhido um e um só aluno do 1º ano?

(b) Serem escolhidos um aluno (e só um) do 1º ano e um aluno (e só um) do 5º ano?

(c) Serem escolhidos no máximo dois alunos do 1º ano?

(d) Serem todos do mesmo ano?

2.10 Um grupo de apostadores do totobola decidiu jogar todas as apostas possíveis contendo 7 vitórias em casa, 4 empates e 2 vitórias fora. Calcule a probabilidade desse grupo ganhar o totobola.

2.11 Suponha que uma cidade tem $n+1$ habitantes e que um deles conta um boato a outro, que por sua vez o repete a um terceiro, e assim sucessivamente. Em cada passo, a pessoa que ouve o boato é escolhida ao acaso de entre as n restantes. Determine a probabilidade de que um boato seja contado r vezes:

(a) Sem antes voltar a ser contado à pessoa que lhe deu início.

(b) Sem que ninguém o ouça mais do que uma vez.

2.12 Considere um dado equipamento que é constituído por 10 transístores dos quais dois são defeituosos. Suponha que dois transístores são seleccionados ao acaso, com reposição.

(a) Escreva o espaço de resultados correspondente a esta experiência aleatória e calcule as respectivas probabilidades.

(b) Calcule as probabilidades dos seguintes acontecimentos:

A_1 – Sair um transístor defeituoso na 1ª tiragem.

A_2 – Sair um transístor defeituoso na 2ª tiragem.

A_3 – Sair pelo menos um transístor defeituoso.

A_4 – Sair exactamente um transístor defeituoso.

(c) Responda às mesmas questões de (a) e (b) mas agora considerando que não houve reposição.

2.13 Uma bolsa contém moedas de prata e cobre em igual número. Extrai-se ao acaso e sem reposição duas moedas. Calcule a probabilidade de que:

(a) A segunda moeda extraída seja de prata, sabendo que a primeira era de cobre.

(b) Saia uma moeda de prata na 2ª tiragem.

(c) Uma e uma só das moedas seja de prata.

(d) Pelo menos uma das moedas seja de cobre.

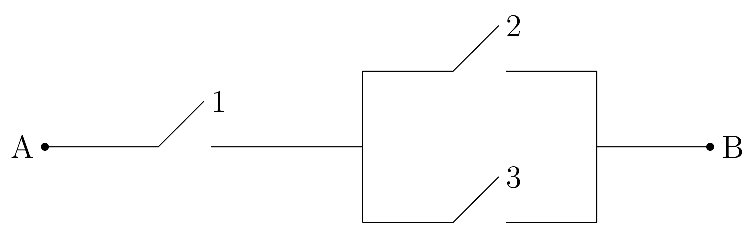
2.14 Uma urna contém 5 bolas brancas e 5 bolas pretas. Dois jogadores, A e B, tiram alternadamente e um de cada de vez uma bola da urna. O jogador que tirar a primeira bola branca ganha a partida.

(a) Considere a experiência aleatória associada a este jogo e escreva o correspondente espaço de resultados.

(b) Calcule a probabilidade de cada jogador ganhar a partida sabendo que o jogador A é o primeiro a tirar a bola de urna.

(c) Responda novamente às alíneas (a) e (b) mas agora considerando que as bolas são extraídas com reposição.

2.15 Considere o seguinte troço de um circuito eléctrico



e designe por F_i o acontecimento "o interruptor i está fechado" ($i = 1, 2, 3$). Suponha que F_1 e F_2 são independentes, com probabilidades iguais a $1/2$ e que F_3 tem uma probabilidade condicional de $1/8$ quando os interruptores 1 e 2 estão fechados e uma probabilidade condicional de $1/10$ quando apenas o interruptor 1 está fechado.

(a) Prove que F_1 e $\overline{F}_2$ são independentes.

(b) Calcule a probabilidade de o interruptor 2 estar fechado dado que há corrente entre os terminais A e B.

2.16 A execução de um projecto de construção de um edifício no tempo programado está relacionada com os seguintes acontecimentos:

$E =$ "escavação executada a tempo"

$F =$ "fundações executadas a tempo"

$S =$ "superestrutura executada a tempo"

supostos independentes e com probabilidades iguais a, respectivamente, 0.8, 0.7 e 0.9. Calcule a probabilidade de:

(a) O edifício ser terminado no tempo previsto, devido ao cumprimento dos prazos nas três actividades referidas.

(b) O prazo de execução ser cumprido para a escavação e não ser cumprido em pelo menos uma das outras actividades.

2.17 Um certo tipo de motor eléctrico quando avariado pode apresentar quatro tipos de falhas, denotadas por F_1, F_2, F_3 e F_4, cujas probabilidades de ocorrência são iguais. Seja $A = \{F_1, F_2\}$, $B = \{F_1, F_3\}$, $C = \{F_1, F_4\}$ e $D = \{F_2, F_3\}$.

(a) Mostre que os acontecimentos A, B e C são independentes aos pares.

(b) Mostre que $P(C|A \cap B)$ é diferente de $P(C)$.

(c) Comente a afirmação: "Como a ocorrência simultânea de C e D é impossível, C e D são necessariamente dependentes".

2.18 Um geólogo crê que existe petróleo numa certa região com probabilidade 0.8 e que, caso haja petróleo, a probabilidade de sair petróleo na primeira perfuração é de 0.5.

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

(b) Tendo-se procedido à primeira perfuração da qual não resultou petróleo, qual é a nova probabilidade atribuída à existência de petróleo na região?

2.19 Suponha que 5% da população portuguesa sofre de hipertensão e que de entre estes, 75% ingerem bebidas alcoólicas. De entre os que não são hipertensos 50% ingerem bebidas alcoólicas.

(a) Qual a percentagem de pessoas que bebem álcool?

(b) Qual a percentagem de pessoas que bebendo álcool sofrem de hipertensão?

2.20 Para um certo tipo de cancro a taxa de prevalência (proporção de doentes na população em geral) é 0.005. Um teste diagnóstico para esta doença é tal que:

– a probabilidade do teste resultar positivo quando aplicado a um indivíduo com cancro (sensibilidade do teste) é 0.99;

– a probabilidade do teste resultar negativo quando o indivíduo não tem cancro (especificidade do teste) é 0.95.

(a) Calcule o valor preditivo do teste, isto é, a probabilidade de um indivíduo ter cancro sabendo que o teste resultou positivo.

(b) Supondo que o teste foi aplicado duas vezes consecutivas ao mesmo doente e que das duas vezes o resultado foi positivo, calcule a probabilidade do doente ter cancro (admita que, dado o estado do indivíduo, os resultados do teste em sucessivas aplicações, em qualquer indivíduo, são independentes). O que pode concluir quanto ao valor preditivo da aplicação do teste duas vezes consecutivas?

2.21 Um teste é constituído por uma pergunta com n alternativas. O indivíduo que o faz ou conhece a resposta ou responde ao acaso. Seja p a probabilidade de um indivíduo conhecer a resposta. Admitindo que a probabilidade de um indivíduo responder correctamente à questão dado que conhece a resposta é 1 e que a probabilidade de responder correctamente dado que responde ao acaso é $1/n$:

(a) Verifique que a probabilidade de um indivíduo não ter respondido ao acaso dado que respondeu correctamente é $\dfrac{np}{1+(n-1)p}$.

(b) Calcule a probabilidade de uma pessoa escolhida ao acaso não responder correctamente à questão, supondo $n = 5$ e $p = 0.2$.

2.22 Registos efectuados levaram a concluir que os motoristas que circulam em determinada estrada podem cometer dois e só dois tipos de transgressões ditas do tipo I e do tipo II, não se notando nenhum caso em que o motorista cometa ambas as transgressões. De entre 500 motoristas multados verificou-se serem 100 por transgressões do tipo I. Sabendo que 10% dos motoristas que cometem transgressões do tipo I são multados; que 1% cometem transgressões do tipo I e que 2% cometem transgressões do tipo II, calcule a probabilidade de que um motorista que circule nessa estrada e cometa uma transgressão do tipo II seja multado.

2.23 Um barco pesqueiro desapareceu e presume-se que o seu desaparecimento se deva a uma das três possíveis causas:

C_1 – afundou-se quando experimentava um sofisticado sistema de pesca para o qual não estava minimamente apetrechado;

C_2 – foi sequestrado por transportar um carregamento de material nuclear;

C_3 – foi destruido por um temporal.

Três brigadas de busca e salvamento, B_1 , B_2 *e* B_3 foram enviadas com a missão de procurar o barco, investigando cada uma delas uma das causas (i.e. a brigada B_i investiga a causa C_i). Suponha que:

1) as três causas do desaparecimento são igualmente prováveis;

2) a probabilidade da brigada B_i ser bem sucedida quando de facto o barco desapareceu devido à causa C_i é α_i ($\alpha_1 = 0.1$, $\alpha_2 = 0.7$, $\alpha_3 = 0.8$).

Sabendo que a investigação da brigada B_2 resultou infrutífera, calcule a probabilidade:

(a) Do barco ter sido sequestrado.

(b) Do barco ter sido destruido por um temporal.

Capítulo 3

Variáveis aleatórias e distribuições discretas

3.1 Uma caixa contém 6 iogurtes dos quais 2 estão estragados. Retiram-se ao acaso e sem reposição 3 iogurtes.

(a) i) Qual a probabilidade de obter quando muito um iogurte estragado?
   ii) Se nas 3 extracções apenas houve um iogurte estragado, qual a probabilidade de ter sido o segundo?

(b) Designe por X a variável aleatória que representa o número de iogurtes estragados nas 3 extracções. Determine:
   i) A função de probabilidade de X.
   ii) A função de distribuição de X.
   iii) O valor esperado e a variância de X.

(c) Responda novamente às alíneas (a) e (b), mas agora admitindo que as 3 extracções foram feitas com reposição.

3.2 Numa fábrica existem três máquinas iguais de uma mesma marca, que trabalham independentemente. A probabilidade de cada máquina avariar num dado espaço de tempo é 0.1. Seja X a variável aleatória que representa o número de máquinas que findo esse período de tempo estão a trabalhar. Determine:

(a) A função de probabilidade de X.

(b) A função de distribuição de X.

(c) O valor esperado, moda, mediana e variância de X.

3.3 Considere a variável aleatória discreta X com a seguinte função de probabilidade:

$$P(X = x) = \begin{cases} ax & , x = 1, 2, 3 \\ 0 & , \text{caso contrário} \end{cases}$$

sendo a uma constante real.

(a) Determine a.

(b) Determine a função de distribuição de X.

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

3.4 Seja X uma variável aleatória discreta com a seguinte função de probabilidade:

$$P(X=x)=\begin{cases}(1+3c)/4 & , x=1\\ (1-c)/4 & , x=2\\ (1+2c)/4 & , x=3\\ (1-4c)/4 & , x=4\\ 0 & , x\neq 1,2,3,4\end{cases}$$

(a) Determine o valor de c.

(b) Calcule o valor esperado e a variância de X.

3.5 Considere uma experiência aleatória associada a 5 acontecimentos elementares ω_i ($i=1,2,3,4,5$) com as seguintes probabilidades:

i	1	2	3	4	5
ω_i	0	1	2	3	4
$P(\omega_i)$	0.1	0.2	0.3	0.3	0.1

Considere a variável aleatória, definida à custa dos acontecimentos elementares,

$$X(\omega_i)=\begin{cases}2\omega_i & , \omega_i\geq 2\\ 6\omega_i-8 & , \omega_i<2\end{cases}$$

Determine o valor esperado de X e a probabilidade de X assumir um valor negativo.

3.6 Considere a variável aleatória discreta X com a seguinte função de distribuição:

$$F_X(x)=\begin{cases}0 & , x<0\\ 1/6 & , 0\leq x<2\\ 1/4 & , 2\leq x<4\\ 1/2 & , 4\leq x<6\\ 1 & , x\geq 6\end{cases}$$

(a) Determine a função de probabilidade de X.

(b) Calcule:

i) $P(X\leq 1)$.

ii) $P(X>5)$.

iii) $P(0<X\leq 2)$.

iv) $P(2\leq X<6)$.

3.7 Num armazém encontra-se um lote de 10000 latas de um certo produto alimentar que está a ser preparado para ser distribuído. 500 dessas latas já ultrapassaram o prazo de validade. É efectuada uma inspecção sobre uma amostra de 15 embalagens escolhidas ao acaso com reposição. A inspecção rejeita o lote se forem encontradas mais do que duas latas fora do prazo de validade nessa amostra.

(a) Qual a probabilidade de rejeição do lote?

(b) Qual o número esperado de latas fora do prazo de validade?

(c) Suponha que as latas são inspeccionadas sucessivamente (com reposição) até ser encontrada uma fora do prazo de validade.

i) Qual a probabilidade de ser necessário inspeccionar 4 ou mais latas?

ii) Qual o número esperado de latas inspeccionadas?

3.8 Num lote de 500 peças existem 50 defeituosas. Desse lote retira-se ao acaso e com reposição uma amostra. O lote é rejeitado se tal amostra incluir mais do que duas peças defeituosas. Calcule:

(a) A probabilidade de rejeição do lote se a amostra tiver dimensão 10.

(b) A dimensão que a amostra deve ter para que a probabilidade de rejeição seja inferior a 0.05.

(c) Nas condições da alínea (a) e se existirem 100 lotes nas condições indicadas, qual o número esperado de lotes em que se pode esperar que haja rejeição?

3.9 2000 pessoas de entre as 60000 que constituem a população de uma cidade estão a assistir a um programa de televisão. Escreva a expressão que lhe permitiria calcular a probabilidade exacta de que, entre 250 pessoas seleccionadas ao acaso e sem reposição da população da cidade, menos de 5 estejam a ver esse programa.

3.10 O número de partículas emitidas por uma fonte radioactiva, num dado período de tempo, é uma variável aleatória com distribuição de Poisson. Sabendo que a probabilidade de não ser emitida qualquer partícula nesse período de tempo é 1/3, calcule a probabilidade de que nesse período de tempo a fonte emita pelo menos 2 partículas.

3.11 Uma máquina electrónica de venda de chocolates e bebidas dá um lucro de 12 dezenas de euros por semana se não tiver avarias durante a semana. Se a máquina tiver x ($x \geq 1$) avarias durante a semana o custo da reparação é de $(x+1)^2$ dezenas de euros. Suponha que o número de avarias numa semana, X, é uma variável aleatória de Poisson de parâmetro $\lambda = 3/2$.

(a) Calcule a probabilidade de numa semana

i) não haver avarias.

ii) haver uma avaria, sabendo que de facto ocorreram avarias nessa semana.

(b) Determine, em dezenas de euros, o lucro esperado por semana.

3.12 Indique uma expressão que lhe permita calcular a probabilidade exacta de que pelo menos 2 pessoas de um grupo de 500 façam anos no dia de Natal (considere o ano com 365 dias). Obtenha um valor aproximado para esta probabilidade com base na distribuição de Poisson.

3.13 Um processo de fabrico de placas de vidro produz, em média, 4 bolhas de ar espalhadas aleatoriamente por $10\,m^2$ de placa. Sabendo que a distribuição do número de bolhas de ar pode ser modelada por uma distribuição de Poisson, calcule a probabilidade de:

(a) Uma placa de $2.5m \times 2m$ ter mais de 2 bolhas de ar.

(b) Obter, num lote de 10 placas de vidro com $1m \times 2.5m$, 6 placas perfeitas.

Capítulo 4

Variáveis aleatórias e distribuições contínuas

4.1 Suponha que o desvio da medida das peças produzidas por uma máquina em relação à norma especificada pelo mercado é uma variável aleatória X com a seguinte função de densidade de probabilidade:

$$f_X(x) = \begin{cases} 1+k+x & , -1 \leq x < 0 \\ 1+k-x & , 0 \leq x \leq 1 \\ 0 & , \text{restantes valores de } x \end{cases}$$

(a) Calcule o valor de k.

(b) Determine a função de distribuição de X.

(c) Calcule o valor esperado e a variância de X.

(d) Calcule a moda, a mediana e o 1º quartil de X.

(e) Calcule a probabilidade de que seja necessário extrair exactamente duas peças da produção da máquina para que apareça uma peça com um desvio positivo em relação à norma.

4.2 Seja $Y = 100\ X$ a variavel aleatória que representa a percentagem de álcool num certo composto, onde X é uma variável aleatória com a seguinte função de densidade de probabilidade:

$$f_X(x) = \begin{cases} 20\, x^3\,(1-x) & , 0 < x < 1 \\ 0 & , \text{caso contrário} \end{cases}$$

(a) Determine a função de distribuição de X e esboce o seu gráfico.

(b) Calcule a probabilidade de X ser inferior a $2/3$.

(c) Suponha que o preço de venda do composto depende do conteúdo em álcool: se $1/3 < X < 2/3$ o preço é de C_1 euros por litro; caso contrário o preço é de $C_2 < C_1$ euros por litro. Supondo o custo de produção igual a C_3 euros por litro:

i) Calcule a função de distribuição do lucro líquido por litro.

ii) Determine o valor esperado do lucro líquido por litro.

4.3 Uma empresa vende peças cuja duração em centenas de horas é uma variável aleatória contínua com a seguinte função de distribuição:

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x} & , x > 0 \\ 0 & , \text{caso contrário} \end{cases}$$

A empresa dispõe de um stock de peças dos tipos A e B. Ao tipo A está associado um parâmetro $\lambda = 1/2$ e ao tipo B um parâmetro $\lambda = 1$. De um lote formado por 100 peças do tipo A e 50 peças do tipo B, retirou-se ao acaso uma peça, cuja duração foi ensaiada. Em relação ao resultado desse ensaio sabe-se apenas que a duração da peça foi inferior a 90h. Calcule a probabilidade de que a peça escolhida seja do tipo B.

4.4 Considere uma variável aleatória contínua cuja função densidade de probabilidade é simétrica em relação ao seu valor esperado. Sabendo que $E(X) = 10$ e $V(X) = 25$ e que a variável aleatória Y se define por $Y = \beta\, X - \alpha$ com $\alpha, \beta > 0$, determine:

(a) α e β de modo que o valor esperado de Y seja nulo e a variância de Y seja unitária.

(b) $P(Y \leq 0)$.

4.5 Uma certa liga metálica contém uma percentagem de chumbo X, que pode ser considerada como uma variável aleatória com função de densidade de probabilidade dada por

$$f_X(x) = \begin{cases} \frac{3}{5}10^{-5}x(100 - x) & , 0 \leq x \leq 100 \\ 0 & , \text{caso contrário} \end{cases}$$

Suponha que L, o lucro líquido obtido na venda desta liga (por unidade de peso), depende da percentagem de chumbo através da relação:

$$L = C_1 + C_2 X$$

Calcule o valor esperado do lucro líquido por unidade de peso.

4.6 A procura diária de arroz num supermercado, em centenas de quilos, é uma variável aleatória com função densidade de probabilidade:

$$f_X(x) = \begin{cases} (2x)/3 & , 0 \leq \ x \ < 1 \\ -x/3 + 1 & , 1 \leq \ x \ \leq 3 \\ 0 & , \text{restantes valores de } x \end{cases}$$

(a) Qual a probabilidade da procura exceder 150 Kg de arroz num dia escolhido ao acaso?

(b) Calcule o valor esperado da procura diária de arroz, assim como uma medida da variabilidade dessa procura.

(c) Qual a quantidade de arroz que deve ser deixada diariamente à disposição do público para que não falte arroz em 95% dos dias?

4.7 Seja X uma variável aleatória com distribuição normal de valor esperado 10 e variância 4, que representa o comprimento de uma barra de ferro. Suponha que a barra é considerada não defeituosa se $8 \leq X \leq 12$ e defeituosa caso contrário.

(a) Qual a probabilidade de que uma barra seja não defeituosa?

(b) Qual a probabilidade de que, em 10 barras escolhidas ao acaso e com reposição do fabrico diário, pelo menos 2 sejam defeituosas?

4.8 O comprimento das peças produzidas por uma máquina é uma variável aleatória normal com valor esperado μ (mm) e variância σ^2 (mm^2). Uma peça é defeituosa se o seu comprimento diferir do valor esperado mais do que σ. Sabe-se que 50% das peças produzidas têm comprimento inferior a 2.5 mm e 47.5% das peças produzidas têm comprimento entre 2.5 mm e 3.42 mm.

(a) Calcule μ e σ.

(b) Determine a probabilidade de que uma peça seja não defeituosa.

4.9 O tempo de vida de um laser tem distribuição normal com média igual a 7000 horas e desvio padrão igual a 600 horas.

(a) Qual é a probabilidade de um desses lasers falhar até 5300 horas?

(b) Qual é a duração que 90% desses lasers excede?

(c) Um produto inclui três lasers e falha se algum deles falhar. Se os tempos de vida dos três lasers forem independentes, qual é a probabilidade desse produto durar mais do que 7000 horas? (*Teste B 13 Mai 2000*)

4.10 Uma componente electrónica tem uma duração de vida, em centenas de horas, que é uma variável aleatória com distribuição exponencial de valor esperado 0.5.

(a) Calcule a função de distribuição da variável aleatória X.

(b) Calcule a probabilidade de que a componente electrónica tenha uma duração de vida superior a 150h, sabendo que já funcionou pelo menos durante 100 h.

4.11 O número de mensagens electrónicas recebidas por dia ($24h$) numa pequena empresa de entregas rápidas tem distribuição de Poisson com média igual a 10.

(a) Calcule a probabilidade de num dia a empresa não receber mais do que 7 mensagens.

(b) Qual é a probabilidade do intervalo entre duas mensagens consecutivas exceder 1 hora? (*Exame 5 Fev 2002*)

Capítulo 5

Distribuições conjuntas de probabilidade e complementos

5.1 Uma loja de electrodomésticos vende televisores da marca X e da marca Y. A função de probabilidade conjunta do número de televisores vendidos diariamente é a seguinte:

$Y \backslash X$	0	1	2
0	0.12	0.25	0.13
1	0.05	0.30	0.01
2	0.03	0.10	0.01

(a) Calcule as funções de probabilidade marginais de X e de Y.

(b) Calcule a função de distribuição marginal de X.

(c) Calcule a probabilidade de que num dia a marca Y seja mais vendida do que a marca X.

(d) Determine o valor esperado e a variância do número total de televisores vendidos diariamente.

5.2 Durante um treino de basquetebol um jogador efectua três lançamentos da linha de lançamento livre. A probabilidade que ele tem de encestar em cada lançamento é de 0.6 e os lançamentos podem ser considerados independentes.

(a) Descreva o espaço de resultados.

(b) Seja X a variável aleatória que representa o número de vezes que o jogador encesta nos dois primeiros lançamentos e Y a variável aleatória que representa o número de vezes que o jogador encesta nos dois últimos lançamentos.

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

ii) Determine as funções de probabilidade marginais de X e de Y.

5.3 Sejam X e Y duas variáveis aleatórias discretas com função de probabilidade conjunta dada por:

$Y \backslash X$	1	2	3
1	1/9	0	1/18
2	0	1/3	1/9
3	1/9	1/6	1/9

(a) Determine:

i) A função de probabilidade marginal de X.

ii) A função de distribuição marginal de Y.

iii) $P(X + Y \leq 4)$.

iv) As funções de probabilidade de X condicionais a $Y = 1$ e $Y = 3$.

v) $E(X|Y = 1)$.

(b) Defina $E(X|Y)$.

(c) Diga, justificando, se X e Y são variáveis aleatórias independentes.

(d) Calcule a $V(X + Y)$.

5.4 Para ser admitido num certo curso um aluno tem que realizar duas provas, A e B, independentes. A classificação em cada uma das provas será de insuficiente (0), suficiente (1) ou bom (2). A probabilidade do aluno obter 0, 1 ou 2 nas provas A e B é apresentada em seguida:

Classificação	Prova A	Prova B
0	0.2	0.2
1	0.5	0.6
2	0.3	0.2

Considere o par aleatório (X, Y) onde:

X = "diferença (em módulo) das classificações nas provas A e B";
Y = "soma das classificações das provas A e B".

(a) Determine:

i) A função de probabilidade conjunta do par aleatório (X, Y).

ii) As funções de probabilidade marginais de X e de Y.

iii) A função de distribuição marginal de X.

iv) A função de probabilidade de X condicional a $Y = 2$.

(b) Diga, justificando, se X e Y são independentes.

(c) Calcule:

i) Todas as funções de probabilidade de Y condicionais a X.

ii) $E(Y|X = 2)$ e $V(Y|X = 2)$.

iii) $F_{Y|X=0}(y)$.

iv) $P(Y = 2|X.Y = 0)$.

v) $P(X + Y$ ser ímpar$)$.

5.5 A função de probabilidade conjunta de duas variáveis aleatórias, X e Y, é tal que:

$$P(X = x, Y = y) = \begin{cases} 1/10 & , \; x = 1, 2, 3, 4, \;\; y = 1, 2, 3, 4 \text{ e } y \leq x \\ 0 & , \text{ caso contrário} \end{cases}$$

(a) Calcule o coeficiente de correlação de X e Y e diga, justificando, se as variáveis aleatórias são ou não independentes.

(b) Calcule $E(X|Y = 3)$.

5.6 Sejam X e Y variáveis aleatórias com função de probabilidade conjunta dada por:

$X \backslash Y$	-1	0	1
-1	0	1/4	0
0	1/4	0	1/4
1	0	1/4	0

Mostre que $Cov(X,Y) = 0$ mas que X e Y não são independentes.

5.7 Considere o par aleatório (X,Y) cuja função de probabilidade é

$$P(X = x, Y = y) = \begin{cases} p^{2-x-y}q^{x+y} & , \; x,y = 0,1, \quad 0 < p < 1, \quad q = 1-p \\ 0 & , \text{ caso contrário} \end{cases}$$

(a) Calcule $V(Z)$, onde $Z = X + Y$.

(b) Defina a variável aleatória $E(X|Y)$.

(c) Apresente um exemplo dum par aleatório discreto (U,V) com as mesmas funções de probabilidade marginais que (X,Y), mas tal que $P(U = x, V = y) \neq P(X = x, Y = y)$.

5.8 A emissão de uma fonte radioactiva é tal que o número de partículas emitidas em cada período de 10 segundos, X, tem distribuição de Poisson com $E(X^2) = 6$.

(a) Observada a emissão durante 7 períodos consecutivos de 10 segundos, qual a probabilidade de, em pelo menos um desses períodos, serem emitidas 4 ou mais partículas?

(b) Um contador Geiger-Muller, que vai registando as emissões sucessivas, tem uma probabilidade 0.9 de registar cada partícula que é emitida.

i) Sabendo que o número de partículas registadas em x ($x \geq 1$) partículas emitidas por período tem uma distribuição binomial, mostre que o número de partículas registadas por período tem uma distribuição de Poisson com parâmetro $\lambda = 0.9 \times 2$.

ii) Determine o valor esperado e a mediana do número de partículas registadas por período.

5.9 Sejam X e Y duas variáveis aleatórias contínuas com função de densidade de probabilidade conjunta

$$f_{X,Y}(x,y) = \begin{cases} 1/2 & , \; -a \leq x \leq a, \; -a \leq y \leq a \, , \; a \in \mathbb{R}^+ \\ 0 & , \text{ caso contrário} \end{cases}$$

(a) Determine o valor de a.

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

(c) Calcule a função de distribuição da variável aleatória Y.

5.10 Considere o par aleatório com densidade conjunta

$$f_{X,Y}(x,y) = \begin{cases} 6(1-x-y) & , \; 0 < y < 1-x \, , \; x > 0 \\ 0 & , \text{ caso contrário} \end{cases}$$

(a) Serão X e Y variáveis aleatórias independentes? Justifique.

(b) Calcule a função de distribuição da variável aleatória X.

(c) Determine $f_{X|Y=y}(x)$.

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

(e) Calcule $P(X < 3/4|Y > 1/2)$.

5.11 Considere para origem do eixo do tempo o horário de partida de certo comboio e para unidade um intervalo de 10 minutos. Sejam X e Y o momento de chegada do passageiro à estação e o momento de partida do comboio, respectivamente. A função de densidade de probabilidade conjunta do par aleatório (X, Y) é dada por

$$f_{X,Y}(x, y) = \begin{cases} \{1 + x(y-1)[x^2 - (y-1)^2]\}/4 & , |x| < 1 \ , \ 0 < y < 2 \\ 0 & , \text{caso contrário} \end{cases}$$

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

(b) Calcule a probabilidade de o passageiro apanhar o comboio.

5.12 Duas pessoas combinam encontrar-se entre as 14 e as 15 horas ficando entendido que nenhuma delas esperará mais do que 15 minutos pela outra. Assuma que iguais intervalos de tempo têm associadas iguais probabilidades de chegada. Qual a probabilidade de as duas pessoas se encontrarem?

5.13 Considere a variável aleatória bidimensional contínua (X, Y) com função densidade de probabilidade conjunta:

$$f_{X,Y}(x, y) = \begin{cases} 2 & , 0 < x < y < 1 \\ 0 & , \text{caso contrário} \end{cases}$$

(a) Calcule o coeficiente de correlação entre X e Y.

(b) Calcule a $V(X|Y = y)$.

(c) Verifique que $E(X) = E[E(X|Y)]$.

5.14 O diâmetro interior de um tubo cilíndrico é uma variável aleatória X com distribuição normal de valor esperado 3 cm e desvio padrão 0.02 cm e a espessura Y do mesmo tubo é uma variável com distribuição normal de valor esperado 0.3 cm e desvio padrão 0.005 cm, independente de X.

(a) Calcule o valor esperado e o desvio padrão do diâmetro exterior do tubo.

(b) Calcule a probabilidade de que o diâmetro exterior do tubo exceda 3.62 cm.

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

(a) Calcule a probabilidade do peso destes 20 utilizadores exceder a carga máxima.

(b) Sabendo que estão 15 pessoas no elevador com um peso de 950 Kg e que se espera a entrada de mais 5 pessoas para completar a lotação e iniciar a viagem, determine a probabilidade do peso total destes 20 passageiros exceder a carga máxima.

(c) Qual a probabilidade de haver nas 20 pessoas, que em certo momento viajam no elevador,

    i) quando muito 2 com peso superior a 85 Kg?

    ii) pelo menos 1 com peso inferior a 40 Kg?

(d) Acha que, em face do tipo de população que utiliza o elevador, a carga máxima indicada é adequada? Explique a sua opinião.

5.16 Um posto de transformação permite uma carga total de 2800KW. Sabe-se que esse posto de transformação alimenta uma fábrica com consumo permanente de 2500KW e além disso o mesmo posto de transformação alimenta 100 consumidores domésticos. Estes gastam em média 2KW em electrodomésticos (sendo o desvio padrão igual a 0.5KW) e 0.5KW com a iluminação (sendo o desvio padrão de 0.25KW). Determine a probabilidade do transformador disparar por excesso de carga, admitindo que os vários tipo de consumos domésticos são independentes e normalmente distribuídos.

5.17 O número de itens dum certo tipo procurados num armazém durante uma semana segue uma distribuição de Poisson com $\lambda = 50$. Calcule a dimensão mínima do *stock* a adquirir de modo a que a probabilidade de satisfazer a procura seja de 98% (use a aproximação à normal).

5.18 Um atirador acerta num alvo com probabilidade 1/3. Numa sequência de 30 tiros calcule aproximadamente a probabilidade do atirador acertar pelo menos 15 vezes no alvo.

5.19 O tempo de produção de uma certa peça de porcelana é uma variável aleatória com distribuição exponencial de valor esperado 2 horas.

(a) Qual a probabilidade duma peça levar pelo menos 1h 45m a ser produzida?

(b) Verificando-se que em certo momento uma peça já está a ser produzida há 45m, qual a probabilidade de ser necessário esperar pelo menos 1h 45m para concluir a peça? Compare este resultado com o da alínea (a) e comente.

(c) Num dia em que a fábrica não tinha qualquer peça em stock foi aceite uma encomenda de 100 peças, tendo a fábrica assumido o compromisso de fornecer as peças no prazo máximo de 30 dias (o que corresponde a 240 horas de trabalho). Acha que a fábrica tem boas possibilidades de cumprir o seu compromisso? Justifique.

(d) A fábrica mantém os registos do tempo de execução de cada peça. Seis peças foram escolhidas ao acaso. Qual a probabilidade de 4 delas terem sido executadas no máximo em 1h 45m cada uma?

5.20 Um estudante decidiu amealhar diariamente uma pequena quantia para comprar uma bicicleta. As probabilidades do estudante amealhar 50, 100 e 250 cêntimos em cada dia são respectivamente 0.3, 0.6 e 0.1. Calcule, justificando, a probabilidade do estudante amealhar mais do que 350 euros durante o ano (365 dias).

5.21 O intervalo de tempo, em minutos, entre a passagem de dois comboios numa estação de metropolitano tem, em horas de ponta, distribuição uniforme no intervalo de $(5, 15)$.

(a) Determine a probabilidade de se ter de esperar mais de 8 minutos entre dois comboios.

(b) Sabendo que o último comboio passou há oito minutos, qual é a probabilidade de se ter de esperar pelo menos mais cinco minutos pelo próximo comboio? Calcule o valor esperado desse tempo de espera adicional.

(c) Admitindo que os intervalos de tempo entre passagens sucessivas dos comboios são variáveis aleatórias independentes, calcule um valor aproximado para a probabilidade da média dos intervalos de tempo entre 100 passagens exceder 9 minutos. (*Exame 19 Jan 2002*)

5.22 O tempo (em horas) que João Pestana dorme por noite é uma variável aleatória com distribuição uniforme no intervalo (7,12).

(a) Calcule a probabilidade de João Pestana dormir mais de 11 horas numa noite.

(b) Calcule a probabilidade de, em 20 noites, João Pestana dormir mais de 11 horas em pelo menos 3 dessas noites.

(c) Qual a probabilidade de João Pestana dormir mais de 1100 horas em 100 noites?

Capítulo 6

Amostragem e estimação pontual

6.1 Considere a população X com função densidade de probabilidade

$$f(x) = \begin{cases} |x|, & |x| < 1 \\ 0, & |x| \geq 1 \end{cases}$$

e a amostra aleatória $(X_1, \ldots, X_5)$.

(a) Diga o que entende por amostra aleatória. Determine a função densidade de probabilidade da amostra aleatória $(X_1, \ldots, X_5)$.

(b) Determine o valor esperado e a variância da média da amostra aleatória, e a variância da amostra (-0.9; 0.8; 0.95; -0.5; 0.75) que representa um valor particular de $(X_1, \ldots, X_5)$.

(c) Calcule a probabilidade do menor valor da amostra aleatória, considerada em (a), ser inferior a 1/7 e ainda a probabilidade do maior valor da amostra aleatória ser superior a 1/7.

6.2 (a) Mostre que se $\hat{\theta}$ é um estimador centrado do parâmetro θ e $V(\hat{\theta}) > 0$ então $(\hat{\theta})^2$ não é um estimador centrado de θ^2.

(b) Se $\hat{\theta}$ é um estimador de θ, o seu enviesamento é dado por $b = [E(\hat{\theta}) - \theta]$. Mostre que $E[(\hat{\theta} - \theta)^2] = \mathrm{V}\,(\hat{\theta}) + b^2$.

6.3 Seja $\overline{X}_1$, a média de uma amostra aleatória de dimensão n extraída de uma população normal de valor esperado μ e variância σ_1^2 e $\overline{X}_2$ a média de uma amostra aleatória de dimensão n, independente da primeira, extraída de uma população normal de valor esperado μ e variância σ_2^2. Mostre que:

(a) $[w\overline{X}_1 + (1 - w)\overline{X}_2]$, em que $0 \leq w \leq 1$, é um estimador centrado de μ.

(b) A variância do estimador indicado em a) é mínima quando

$$w = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$$

6.4 Se (X_1, X_2, X_3) constitui uma amostra aleatória de dimensão 3 extraída de uma população normal com valor esperado μ e variância σ^2, qual a eficiência de $\hat{\mu} = (X_1 + 2X_2 + X_3)/4$ relativamente a $\overline{X}$?

6.5 Considere uma população X, com função densidade de probabilidade $f(x)$ e valor desconhecido da mediana, ξ. A mediana da amostra aleatória $\widetilde{X}$ é estimador centrado de ξ e o seu desvio padrão é aproximadamente $[2\sqrt{n}\, f(\xi)]^{-1}$. Calcule a eficiência relativa da média da amostra aleatória $\overline{X}$ em relação à mediana da amostra aleatória $\widetilde{X}$, como estimadores do parâmetro μ,

(a) para o caso duma população normal com valor esperado μ e desvio padrão σ.

(b) para o caso da variável aleatória cuja função densidade de probabilidade é:

$$f(x) = \frac{1}{\sqrt{2}\sigma} e^{-\sqrt{2}\left|\frac{x-\mu}{\sigma}\right|}$$

em que μ e σ representam, respectivamente, o valor esperado e o desvio padrão.

(c) O que pode concluir, na sequência dos resultados obtidos em (a) e (b)?

6.6 T_1 *e* T_2 são estimadores de um parâmetro θ, tais que:

$$\begin{array}{ll} E(T_1) = \theta & V(T_1) = 9 \\ E(T_2) = 3\theta & V(T_2) = 3 \end{array}$$

Diga, justificando, qual destes estimadores é melhor estimador de θ.

6.7 Considere uma urna com bolas brancas e pretas na proporção de 3/1 desconhecendo-se, no entanto, qual a cor dominante. Seja p a probabilidade de sair uma bola preta numa extracção.

Qual a estimativa de máxima verosimilhança de p se, ao extraírmos com reposição 3 bolas da urna, encontrássemos

(a) 1 bola preta?

(b) 2 bolas pretas?

(c) Suponha agora que desconhecíamos qualquer relação entre o número de bolas brancas e pretas. Qual a estimativa de máxima verosimilhança de p, se ao extrairmos 3 bolas com reposição encontrássemos 2 bolas pretas?

6.8 Uma urna contém N bolas, umas brancas e outras pretas. Seja R a razão (desconhecida) entre o número de bolas brancas e o número de bolas pretas. Supondo que dessa urna foram extraídas, com reposição, n bolas e que se observaram k bolas brancas, determine a estimativa de máxima verosimilhança para R.

(*Sugestão:* exprima as probabilidades de extrair uma bola branca ou uma bola preta em termos de R).

6.9 Num trabalho de rotina de controlo de qualidade da produção duma fábrica de pneus foram analisados 4 lotes de 80 pneus cada, tendo-se obtido 2.5%, 3.75%, 5% e 6.25% de pneus defeituosos, respectivamente. Considere a distribuição do número de pneus defeituosos por lote e deduza o estimador de máxima verosimilhança da probabilidade de um pneu ser defeituoso. Calcule a estimativa de máxima verosimilhança com base na amostra de 4 lotes.

6.10 O número de andares vendidos em cada dia por uma empresa imobiliária segue uma distribuição de Poisson de parâmetro λ.

(a) Com base numa amostra aleatória proveniente dessa população, deduza o estimador de máxima verosimilhança do parâmetro λ. Diga, justificando, se é ou não centrado.

(b) Indique um estimador centrado para a variância da variável aleatória em estudo.

(c) Sabendo que durante 20 dias consecutivos são vendidos 8 andares, calcule a estimativa da máxima verosimilhança de λ .

(d) Sabendo que durante 15 dias consecutivos não foram vendidos andares e que nos dois dias seguintes a empresa vendeu pelo menos um andar em cada dia, calcule a estimativa da máxima verosimilhança de λ .

6.11 Suponha que X é uma variável aleatória normal de valor esperado μ e desvio padrão $\sigma = 2$. Calcule a partir de uma amostra aleatória de dimensão n dessa população o estimador de máxima verosimilhança para μ. Será um estimador centrado?

6.12 Suponha que a voltagem que um cabo eléctrico com um certo isolamento pode suportar varia de acordo com uma distribuição Normal. Para uma amostra de 12 cabos as falhas ocorreram nos seguintes níveis de voltagem:

$$52 \quad 64 \quad 38 \quad 68 \quad 66 \quad 52 \quad 60 \quad 44 \quad 48 \quad 46 \quad 70 \quad 62$$

Determine as estimativas de máxima verosimilhança dos seguintes parâmetros: valor esperado, variância, desvio padrão, bem como da probabilidade de um cabo suportar níveis superiores a voltagem máxima registada na amostra acima.

6.13 A tensão de rotura de uma "amostra" de betão é uma variável aleatória X com valor esperado μ e variância σ^2, finitos mas desconhecidos. Cem determinações independentes desta variável originaram os seguintes valores:

$$\sum_{i=1}^{100} x_i^2 = 5706 \qquad \sum_{i=1}^{100} x_i^2 - \frac{1}{100}\left(\sum_{i=1}^{100} x_i\right)^2 = 81$$

(a) Justifique a afirmação:
"A estatística $T_1 = \frac{1}{n^2(n-1)}\sum_{i=1}^{n}(nX_i - \sum_{j=1}^{n} X_j)^2$ é um estimador centrado para σ^2 enquanto que o estimador $T_2 = \frac{1}{n^3}\sum_{i=1}^{n}(nX_i - \sum_{j=1}^{n} X_j)^2$ subestima, em valor esperado, σ^2, sendo centrado apenas assintoticamente".

(b) Indique, justificando detalhadamente, qual dos dois estimadores T_1 e T_2 é o estimador de máxima verosimilhança de σ^2, caso X possua uma distribuição normal.

6.14 Certo tipo de pilhas tem uma duração (em horas) que se distribui exponencialmente com valor esperado μ. A duração global de 10 pilhas tomadas aleatoriamente foi de 1740 horas. Qual a estimativa de máxima verosimilhança da probabilidade de uma pilha durar mais de 200 horas?

6.15 Tem sido sugerido que em certos locais e certas condições climatéricas, a altura X das ondas do mar segue aproximadamente a distribuição de Rayleigh cuja função densidade de probabilidade é

$$f(x;\alpha) = \begin{cases} \dfrac{x}{\alpha^2}e^{-\frac{1}{2}(\frac{x}{\alpha})^2}, & x \geq 0 \\ 0, & x < 0 \end{cases} \qquad (\alpha > 0)$$

Relativamente a variável aleatória X sabe-se que $E(X) = \alpha\sqrt{\frac{\pi}{2}}$ e $V(X) = (2-\frac{\pi}{2})\alpha^2$.

(a) Suponha que se observaram ondas com as seguintes alturas (em metros):

1.4 3.5 2.4 1.9 3.1 2.7 2.5 3.1 4.1 2.8 2.5 3.3

Obtenha a estimativa de máxima verosimilhança do valor esperado e da variância de X.

(b) Faça um esboço gráfico da função densidade de probabilidade $f(x; \hat{\alpha})$ correspondente à população especificada pelas observações referidas em a). Marque no eixo x os valores de $\hat{\alpha}$ e $\widehat{E(X)}$. Como se designa habitualmente o valor $\hat{\alpha}$?

6.16 Uma amostra aleatória de tamanho 5 é obtida de uma população normal com valor médio 12 e desvio padrão 2.

(a) Qual é a probabilidade de a média da amostra aleatória exceder 13?

(b) Qual é a probabilidade de o mínimo da amostra aleatória ser inferior a 10?

(c) Qual é a probabilidade de o máximo da amostra aleatória ser superior a 15?

6.17 Seja $(X_1, \ldots, X_n)$ uma amostra aleatória de tamanho n proveniente da população X com distribuição $U(0, 1)$. Calcule a probabilidade de $\overline{X}$ ser pelo menos 0.9.

6.18 Uma amostra de dimensão 40, $(X_1, \ldots, X_{40})$, é extraída duma população de Poisson com $\lambda = 10$. Recorra à distribuição normal para calcular um valor aproximado de $P(\overline{X} < 9)$.

6.19 Um processo de fabrico é delineado para produzir unidades com um máximo de 2% de defeituosas. A sua verificação é feita diariamente testando 10 unidades selecionadas aleatoriamente da produção diária. Se se encontrar pelo menos uma defeituosa, o processo é parado momentaneamente e examinado. Se a probabilidade de ser produzida uma unidade defeituosa é efectivamente 0.01:

(a) Qual a probabilidade de o processo ser interrompido?

(b) Qual a probabilidade de, num dado teste, não se obter nenhuma defeituosa?

(c) Qual o valor esperado e o desvio padrão da proporção de unidades defeituosas em amostras de 10 unidades?

6.20 Suponha que o diâmetro de um certo tipo de tubo tem uma distribuição Normal de valor médio μ e desvio padrão 0.01 cm.

(a) Qual a probabilidade de um tubo ter um diâmetro que se desvie do seu valor esperado por mais de $\pm\, 0.02\, cm$?

(b) Em 1000 tubos produzidos, quantos esperaria rejeitar se os limites de especificação fossem $2.77 \pm 0.03\, cm$ e o valor esperado da distribuição fosse de 2.79 cm?

(c) Qual o tamanho da amostra a obter para que não seja superior a 5% a probabilidade de a média da amostra aleatória diferir do valor esperado da população por mais de $\pm\, 0.01\, cm$?

Capítulo 7

Estimação por intervalos

7.1 Medições do comprimento de 25 peças produzidas por uma máquina conduziram a uma média $\overline{x} = 140$ mm. Admita que cada peça tem comprimento aleatório com distribuição normal de valor esperado μ e desvio padrão $\sigma = 10$ mm, e que o comprimento de cada peça é independente das restantes. Construa um intervalo de confiança a 95% para o valor esperado da população.

7.2 Admita que a densidade de construção, X, num projecto de urbanização tem distribuição normal. Uma amostra aleatória de 50 lotes desse projecto conduziu a

$$\sum_{i=1}^{50} x_i = 227.2 \; ; \; \sum_{i=1}^{50} x_i^2 = 2242.6$$

Assumindo que o desvio padrão de X é igual a 4, construa um intervalo de confiança a 95% para a densidade média de construção. Que dimensão deveria ter a amostra para que a amplitude desse intervalo fosse reduzida a metade? (*Exame 19 Jan 2002*)

7.3 Foram efectuados estudos em Los Angeles com o objectivo de determinar a concentração de monóxido de carbono perto de vias rápidas. Para isso recolheram-se amostras de ar, para as quais se determinou a respectiva concentração (usando um espectrómetro). Os resultados das medições em ppm (partes por milhão) foram os seguintes (para um período de um ano):

102.2	98.4	104.1	101.0	102.2	100.4	98.6	88.2	78.8	83.0
84.7	94.8	105.1	106.2	111.2	108.3	105.2	103.2	99.0	98.8

Determine um intervalo de confiança a 95% para a concentração esperada de monóxido de carbono, assim como para a sua variância. Indique as hipóteses consideradas.

7.4 Suponha que a intensidade da corrente, em amperes, num certo circuito é uma variável aleatória com distribuição normal. Uma amostra de dimensão 12 desta variável aleatória conduziu aos seguintes resultados:

2.3	1.9	2.1	2.8	2.3	3.6	1.4	1.8	2.1	3.2	2.0	1.9

Construa um intervalo de confiança de 99% para:

(a) O valor esperado da intensidade da corrente.

(b) O desvio padrão da intensidade da corrente.

7.5 Um engenheiro civil, tencionando comparar a resistência a forças compressivas de dois tipos de betão, seleccionou aleatoriamente 10 elementos de cada tipo de betão e registou as seguintes medições.

Tipo I	3250	3268	4302	3184	3266	3297	3332	3502	3064	3116
Tipo II	3094	3268	4302	3184	3266	3124	3316	3212	3380	3018

Se se assumir que as amostras provêm de populações normais com desvio padrão igual a 353 e 133, respectivamente, determine um intervalo de confiança a 95% para a diferença entre os valores esperados das duas populações.

7.6 Um fabricante de cigarros enviou a dois laboratórios amostras de tabaco supostamente idênticas. Cada laboratório efectuou cinco determinações do conteúdo em nicotina (em mg). Os resultados foram os seguintes:

Laboratório 1 (x_1)	24	27	26	21	24
Laboratório 2 (x_2)	27	28	23	31	26

$$\bar{x}_1 = 24.4 \quad \bar{x}_2 = 27.0 \quad \sum_i x_{1i}^2 = 2998 \quad \sum_i x_{2i}^2 = 3679$$

Admite-se que os resultados de cada laboratório seguem distribuições normais independentes com variância comum. Determine um intervalo de confiança a 99% para a diferença das médias entre os resultados fornecidos pelos dois laboratórios. Acha que se pode concluir que as médias das duas populações são iguais? (*Exame 5 Fev 2002*)

7.7 Para comparar a eficiência de dois métodos de ensino, uma turma de 24 alunos foi dividida aleatoriamente em dois grupos. Cada grupo é ensinado de acordo com um método diferente. Os resultados no fim de semestre, numa escala de 0 a 100, são os seguintes:

1º grupo	$n_1 = 13$	$\overline{x}_1 = 74.5$	$s_1^2 = 82.6$
2º grupo	$n_2 = 11$	$\overline{x}_2 = 71.8$	$s_2^2 = 112.6$

Assumindo que as populações são normais e com variâncias iguais e desconhecidas obtenha um intervalo de confiança a 95% para a diferença entre os valores esperados das duas populações.

7.8 Para estimar a diferença de tempos esperados de vida entre fumadores e não fumadores, numa grande cidade dos E.U.A., foram recolhidos duas amostras independentes de, respectivamente, 36 não fumadores e 44 fumadores tendo-se obtido os seguintes resultados:

	Dimensão	Média	Desvio padrão corrigido
Não fumadores	36	72	9
Fumadores	44	62	11

Calcule um intervalo de confiança a 90% para a diferença dos valores esperados dos tempos de vida.

7.9 Uma amostra de 100 peças de uma linha de produção revelou 17 peças defeituosas.

(a) Determine um intervalo de confiança a 95% para a verdadeira proporção p de peças defeituosas produzidas.

(b) Quantas peças adicionais devemos recolher para estarmos confiantes a 98% que o erro de estimação de p seja menor que 2%?

7.10 Num trabalho realizado há já algum tempo concluiu-se que 62% dos passageiros que entram na estação A do metro tem como destino o centro da cidade. Esse valor tem vindo a ser utilizado em todos os estudos de transportes realizados deste então.

O Engenheiro Vivaço começou a ter dúvidas sobre a actualidade daquele valor, acreditando que ele tem vindo a diminuir, acompanhando o declínio do centro. Resolveu, portanto, realizar um inquérito na estação A, tendo sido inquiridos 240 passageiros dos quais 126 indicaram o centro como destino.

(a) Com base nestes resultados construa um intervalo de confiança a 90% para a percentagem de passageiros entrados em A e que saiem no centro, e interprete-o, admitindo que tem como interlocutor um leigo em Estatística.

(b) Quantos passageiros deveriam ser inquiridos caso se pretendesse estimar aquela percentagem com margem de erro não superior a 2% e com um grau de confiança de pelo menos 90%?

7.11 Estudos efectuados ao longo do tempo pela secção de Controlo de Qualidade de uma dada empresa permitiram constatar que o número de artigos defeituosos (isto é, fora dos padrões de especificação) produzidos por lote é bem modelado por uma distribuição de Poisson com um valor esperado λ, que tem girado em torno de 80%.

Tendo-se criado uma certa desconfiança quanto ao funcionamento adequado do processo de produção, a secção verificou 465 lotes idênticos de artigos, com os seguintes resultados:

nº de artigos defeituosos por lote	0	1	2	3	4	5
nº de lotes	216	156	71	15	5	2

(a) Derive o estimador de máxima verosimilhança de λ, T, e o estimador de máxima verosimilhança da probabilidade de o número de artigos defeituosos por lote não ser superior a 1.

(b) Indique, justificando, a distribuição amostral aproximada de T e, com base nela, construa um intervalo de confiança a 90% para λ.

7.12 Considere uma população X com distribuição exponencial com valor esperado α^{-1}, $\alpha > 0$, isto é, com função densidade de probabilidade

$$f(x) = \begin{cases} \alpha e^{-\alpha x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

Observada uma amostra de dimensão 100 obteve-se $\overline{x} = 2.5$. Deduza, com base nesta amostra, um intervalo de confiança a 95% para o parâmetro α.

Capítulo 8

Testes de hipóteses

8.1 Seja $X \sim N(\mu, 4)$. Para testar a hipótese $H_0 : \mu = 1$ contra a alternativa $H_1 : \mu = 2$ usa-se a seguinte região crítica: $\overline{x} > c$.

(a) Para uma amostra de dimensão 25 determine c de modo que $\alpha = 0.1$.

(b) Determine a dimensão da amostra n e c de modo que $\alpha = 0.05$ e $\beta = 0.10$.

(c) Suponha que para amostras de dimensão 2 dessa população se fixa o seguinte teste: rejeita-se H_0 se $\overline{x} > 1.5$. Calcule as probabilidades dos erros de 1ª e 2ª espécie.

8.2 Uma fábrica de adubos tem um novo adubo que se diz produzir, em valor esperado, 20 quintais de um determinado cereal por hectare. O desvio padrão da produção deste cereal é conhecido como sendo de 4 quintais por hectare.

Para testar a hipótese $H_0 : \mu = 20$ contra a hipótese $H_1 : \mu \neq 20$ é extraída uma amostra aleatória de 16 hectares numa área agrícola experimental. Considerando que a produção do cereal pode ser representada por uma variável aleatória X, normalmente distríbuida de valor esperado μ e que, se $18 < \overline{x} < 22$ aceita-se H_0 e caso contrário rejeita-se H_0:

(a) Identifique a estatística do teste.

(b) Calcule a probabilidade de aceitar H_0 quando $\mu = 17$; 18; 19; 20; 21; 22; 23.

(c) Com base nos resultados de (b) faça um gráfico aproximado da função potência do teste.

Note que a função potência do teste, para testar a hipótese H_0 contra uma hipótese alternativa H_1, referente a um dado parâmetro θ é dada pela seguinte função de θ: $\beta(\theta) = P(\text{rejeitar } H_0|\theta)$

8.3 Para controlar a qualidade de lotes que vão sendo produzidos relativamente ao peso das embalagens decidiu-se usar o seguinte esquema: recolher uma amostra de dimensão n de cada lote, e calcular a média amostral $\overline{x}$ dos pesos das embalagens e:

se $\overline{x} \leq c$ rejeita-se o lote
se $\overline{x} > c$ aceita-se o lote

Acordou-se ainda que se o valor esperado do peso das embalagens no lote (μ) for inferior ou igual a 5.3, a probabilidade de rejeitar o lote deve ser pelo menos 99% e se μ for superior ou igual a 5.5 a probabilidade de aceitar o lote deve ser pelo menos 90%. Admita que os pesos das embalagens têm distribuição normal com desvio padrão, em cada lote, igual a 0.2.

Calcule o valor de c e o menor valor de n requerido por este esquema de amostragem. Justifique.

8.4 Para testar a hipótese $H_0 : p = 1/2$ contra $H_1 : p = 3/4$ (p é a probabilidade de obter cara no lançamento duma moeda), com base no número de caras saídas com o lançamento de uma moeda 4 vezes consecutivas, consideram-se as seguintes regiões críticas:

$$\begin{aligned} C_1 &= \{\, 2, 3, 4 \,\} \\ C_2 &= \{\, 3, 4 \,\} \\ C_3 &= \{\, 4 \,\} \end{aligned}$$

Calcule, com base nos valores da tabela seguinte, as probabilidades dos erros de $1^{\underline{a}}$ *e* $2^{\underline{a}}$ espécie associados a cada uma das regiões críticas.

Caras saídas	$H_0 : p = 1/2$	$H_1 : p = 3/4$
0	0.0625	0.0039
1	0.2500	0.0469
2	0.3750	0.2109
3	0.2500	0.4219
4	0.0625	0.3164

Escolha justificando, uma região crítica para definir o teste.

8.5 Da produção diária de determinado fertilizante tiraram-se seis pequenas porções que se analisaram para calcular a percentagem de nitrogénio. Os resultados foram os seguintes:

6.2 5.7 5.8 5.8 6.1 5.9

Sabe-se, por experiência, que o processo de análise fornece valores com distribuição que se pode considerar normal com $\sigma^2 = 0.25$.

(a) Suportam as observações a garantia de que a percentagem esperada de nitrogénio, μ, é igual a 6% ao nível de significância de 10%?

(b) Responda à alínea anterior usando o valor-p.

8.6 Uma máquina de ensacar açúcar está regulada para encher sacos de 16 quilos. Para controlar o funcionamento escolheram-se ao acaso 15 sacos da produção de determinado período, tendo-se obtido os pesos seguintes:

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	

Admitindo que o peso de cada saco possui distribuição normal:

a) Que conclusão pode tirar sobre a regulação da máquina?

b) Que evidência fornece a concretização de S^2 sobre a hipótese $H_0 : \sigma^2 = 0.25$?

8.7 Seja X uma variável aleatória com distribuição normal de valor esperado μ e desvio padrão σ. A partir de uma amostra de dimensão 30 dessa variável obtiveram-se os seguintes resultados:

$$\sum_{i=1}^{30} x_i = 64.0 \qquad \sum_{i=1}^{30} (x_i - \overline{x})^2 = 84.8$$

Teste ao nível de significância de 5% a hipótese $H_0 : \mu = 2.0$ contra a hipótese alternativa $H_1 : \mu > 2.0$.

8.8 Um ensaio de rotura a compressão efectuado sobre 12 provetes cúbicos de betão conduziu aos seguintes valores da tensão de rotura (kgf/cm^2).

263	254	261	236	228	253	249	262	250	252	257	258

Admita (como aliás é feito no Regulamento de Betões de Ligantes Hidráulicos) que a variável em estudo segue uma distribuição normal.

(a) Um engenheiro pretende saber se a tensão esperada de rotura não é inferior a 255 kgf/cm^2. Que evidência fornecem os dados acerca desta questão se se admitir um nível de significância menor ou igual a 5%? Justifique.

(b) Sabendo que o valor característico da tensão de rotura se define como o valor da variável que tem uma probabilidade de 95% de ser excedido, calcule uma estimativa do valor característico da tensão de rotura daquele betão, justificando o procedimento adoptado.

8.9 A cotação na bolsa de uma dada empresa está sujeita a flutuações em torno de um valor médio (2500) relativamente estável. Admite-se que a cotação desta empresa pode ser considerada uma variável aleatória com distribuição aproximadamente normal. O valor que se admite para a variância é tal que há 95% de probabilidade de a cotação pertencer ao intervalo (2300,2700).

(a) Observou-se durante 16 dias as cotações da empresa e obteve-se a média amostral de 2538 e um desvio padrão amostral corrigido de 91.5. Que conclusão pode tirar acerca da variabilidade da cotação dessa empresa?

(b) Após um período de remodelação da empresa observaram-se durante 13 dias a sua cotaçao na bolsa e obteve-se a média amostral de 2670 e o desvio padrão amostral corrigido igual a 86.3. Será que pode concluir pela eficácia das medidas introduzidas?

8.10 Dois alunos de estatística decidiram fazer uma aposta relativamente à nota da disciplina de Probabilidades e Estatística. O aluno A acredita que o valor esperado da nota é 8 e o aluno B afirma que será 10. Para decidir qual o vencedor fizeram um teste ao valor proposto pelo aluno A. Admitindo que A perde a aposta se a sua hipótese for rejeitada, selecionaram ao acaso 30 notas de Probabilidades e Estatística $(x_1, \ldots, x_{30})$ e verificaram que $\sum_{i=1}^{30} x_i = 270$. Acrescente-se que a variância divulgada pela secção de Estatística e Aplicações foi 16 e os alunos acordaram um nível de significância de 5%.

(a) Quem ganhou a aposta?

(b) Acha que a aposta foi justa (no sentido da probabilidade de cada um dos jogadores perder injustamente ser igual)? Identifique essas probabilidades.

8.11 O departamento de segurança de uma fábrica quer saber se o tempo esperado que o empregado nocturno da segurança leva a dar uma volta a fábrica é de 30 minutos. Em 32 voltas a média do tempo foi de 30.8 minutos com um desvio padrão corrigido de $s = 1.5$ minutos. Diga se, ao nível de significância de 1%, é de admitir a hipótese considerada.

8.12 Um mesmo tipo de material (em relação ao qual a temperatura de fusão é importante) pode ser adquirido a dois fabricantes (A e B). Uma amostra de 21 observações da temperatura de fusão de material de cada fabricante produziu os seguintes valores:

fabricante	A	B
média (^oC)	420	426

É sabido que o desvio padrão das temperaturas de fusão do material fornecido pelos dois fabricantes é de 4^oC.

(a) Acha que a temperatura esperada de fusão do material fornecido pelos dois fabricantes pode ser considerada igual? Use um teste de hipóteses conveniente e um nível de significância de 1%, não se esquecendo de indicar alguma hipótese de trabalho que seja necessária.

(b) Determine a probabilidade de o teste da alínea (a) detectar diferença entre as temperaturas esperadas de fusão do material produzido pelos fabricantes B e A quando existe uma diferença de $+3^oC$ entre essas temperaturas. (*Exame A 29 Jan 2000*)

8.13 Para confrontar dois tipos de máquina de ceifar (segadeiras) um trigal foi dividido em secções longitudinais e cada duas secções adjacentes tratadas por cada uma das máquinas, sendo a indicação da máquina obtida lançando uma moeda ao ar. As produtividades foram as seguintes:

Segadeira 1	8.0	8.4	8.0	6.4	8.6	7.7	7.7	5.6	6.2
Segadeira 2	5.6	7.4	7.3	6.4	7.5	6.1	6.6	6.0	5.5

Ao agricultor que experimenta as segadeiras interessa averiguar se a produtividade esperada das duas máquinas se pode considerar igual ou se existe diferença significativa que o leve a preferir uma delas.

Responda a esta questão admitindo que as produtividades possuem distribuição normal com:

(a) As variâncias conhecidas e iguais a 1.13 e 0.62, respectivamente.

(b) As variâncias iguais com valor comum desconhecido.

8.14 Um fabricante de pneus pretende comparar, através de ensaios piloto, 2 métodos de produção dos pneus. Selecionados 10 e 8 pneus produzidos, respectivamente segundo o 1º e 2º métodos, resolve-se testá-los. Os pneus da 1ª amostra foram testados numa zona A, os da 2ª numa zona B, com as durações (em unidades de 100 km):

Amostra 1	61.1	58.2	62.3	64	59.7	66.2	57.8	61.1	62	63.6
Amostra 2	62.2	56.6	66.4	56.2	57.4	58.4	57.6	65.4		

Sabe-se de estudo anteriores que a duração de um pneu varia segundo uma distribuição normal, em que o valor esperado é eventualmente influenciável pelo método de produção, e cujo desvio padrão é susceptível de ser fortemente afectado pelas características da zona onde se procede a rodagem.

(a) Será que se pode admitir que a duração esperada de um pneu do 1º tipo não excede 6000 km?

(b) Os dados são significativamente compatíveis com a conjectura do desvio padrão da duração de um pneu do 1º tipo ser igual a 400 km?

(c) Admita que as variâncias da duração dos dois tipos de pneus são iguais. Teste a hipótese de não haver uma diferença significativa na duração média dos dois tipos de pneus.

8.15 Dois grupos de 36 estudantes foram seleccionados ao acaso para participarem numa experiência que consiste em aprender o significado de palavras numa língua que não conhecem.

Durante 30 minutos os estudantes tentaram aprender o maior número de palavras. No grupo I os estudantes trabalharam isoladamente. No grupo II os estudantes trabalharam aos pares procurando certificar-se mutuamente que iam aprendendo as palavras. Em seguida foi efectuado um teste para determinar o número de palavras aprendidas por cada aluno, tendo-se obtido os seguintes resultados:

Grupo I	24	14	16	17	18	23	14	15	15	17
	18	16	17	19	20	21	20	19	19	18
	16	25	18	20	22	16	16	15	25	22
	16	20	20	19	25	18				
Grupo II	21	22	25	21	20	18	20	17	16	14
	17	15	18	23	17	19	15	23	19	20
	16	22	15	18	16	16	22	21	17	19
	15	18	23	20	20	18				

Acha que o segundo método de aprendizagem pode considerar-se significativamente superior ao primeiro?

8.16 Um laboratório lançou no mercado um novo medicamento para o tratamento de uma alergia, afirmando que a sua eficácia, num período de 8 horas, é de 90%. A sua aplicação a uma amostra de 200 indivíduos sofrendo de tal alergia revelou-se eficaz em 160 dos casos. Será a afirmação acima consistente com os dados obtidos? Indique o valor-p do teste efectuado.

8.17 Uma empresa fabricante de lâmpadas considera que a sua produção é eficaz se a probabilidade de se seleccionar ao acaso uma lâmpada não defeituosa for de pelo menos 90%. Para verificar a qualidade da produção das lâmpadas, foi efectuado um teste a 200 lâmpadas, tendo-se verificado que 24 tinham defeitos. A que conclusão deve chegar o estatístico da empresa? Justifique.

8.18 Um comerciante retalhista recebe carregamentos de roupa de homem normalmente com 10% de peças defeituosas. A fim de se certificar que a qualidade do produto não diminuiu, resolve verificar 100 peças, determinar a percentagem de peças defeituosas e conduzir um teste de hipóteses com nível de significância igual a 8%.

(a) Especifique a hipótese nula que está em causa assim como a hipótese alternativa.

(b) Indique a estatística e determine a região crítica do teste.

(c) Suponha que, nas peças verificadas, foram encontradas 12 defeituosas. O comerciante deve ou não rejeitar o carregamento?

8.19 Numa empresa recolheu-se uma amostra relativa à produção de energia eléctrica em kW/h de dois tipos de geradores. Admita que a distribuição de energia segue uma distribuição normal e que aos dois tipos de geradores está associado uma variância igual. Os resultados obtidos foram os seguintes:

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

(a) Teste se a produção média de energia eléctrica segundo os dois geradores é igual.

(b) Construa um intervalo de confiança a 95% para o valor esperado da produção de energia eléctrica.

(c) O fabricante afirma que a variância da produção de energia eléctrica é de $4\ (kW/h)^2$. Comente a afirmação do fabricante.

(d) Seja p a proporção desconhecida de geradores cuja produção se situa abaixo dos $5\ kW/h$. Estes são considerados defeituosos e o comprador será indemnizado. Teste a hipótese de a proporção de geradores defeituosos ser inferior ou igual a 10%.

8.20 Uma empresa agrícola tem uma estação agronómica experimental onde produz novas variedades de ervilhas. Uma amostra sobre as características das ervilhas resultou em 310 ervilhas amarelas e de casca macia, 109 ervilhas amarelas e de casca dura, 100 ervilhas verdes e de casca macia e 37 ervilhas verdes e de casca dura. Numa experiência semelhante, Mendel, através de um modelo matemático simples, previu que o resultado seria de 56.25% de ervilhas amarelas de casca macia, 18.75% de ervilhas amarelas de casca dura, 18.75% de ervilhas verdes de casca macia e 6.25% de ervilhas verdes de casca dura. Serão os resultados da estação agronómica compatíveis com os resultados de Mendel para os níveis de significância de 5% e 1%, respectivamente?

8.21 O recenseamento de 320 famílias com 5 filhos conduziu aos seguintes resultados:

Rapazes	5	4	3	2	1	0
Famílias	18	56	110	88	40	8

(a) Verifique se estes resultados são compatíveis com a hipótese do número de rapazes ser uma variável aleatória com distribuição binomial, admitindo a equiprobabilidade dos sexos, ao nível de significância de 0.1%.

(b) Indique um intervalo para o valor-p do teste efectuado para responder à alínea anterior.

8.22 Suponha que o departamento de defesa acredita que a distribuição de probabilidade do número de avarias, durante uma dada missão, ocorridas numa determinada zona do submarino Polaris segue uma distribuição de Poisson. Os dados relativos a 500 destas missões são os seguintes:

número de falhas por missão	0	1	2	3	4 ou mais
número de missões	185	180	95	30	10

(a) Teste ao nível de significância de 5% a hipótese da referida variável aleatória possuir uma distribuição de Poisson, com valor esperado igual a 1.

(b) A estimativa de máxima verosimilhança do valor esperado avaliada numericamente com base na amostra agrupada é igual a 0.9845. Será que o modelo de Poisson é uma boa escolha para descrever o conjunto de dados?

8.23 Numa experiência com tubos de vácuo foram observados os tempos de vida (em horas) de 100 tubos, tendo-se registado as seguintes frequências absolutas:

Intervalo	$]0, 30]$	$]30, 60]$	$]60, 90]$	$]90, +\infty[$
Frequências absolutas	41	31	13	15

Serão os dados consistentes com a hipótese de o tempo de vida de um tubo de vácuo ter distribuição *exponencial* com valor esperado igual a 50 horas? Calcule um intervalo para o valor-p e comente. (*Exame 13 Jul 2002*)

8.24 A altura, em metros, dos indivíduos de determinada população é uma variável aleatória X. Escolhidos aleatoriamente 100 desses indivíduos e medidas as suas alturas obtiveram-se os seguintes resultados:

Classes	F_i^0
$[1.595, 1.625[$	5
$[1.625, 1.655[$	18
$[1.655, 1.685[$	42
$[1.685, 1.715[$	27
$[1.715, 1.745[$	8

(a) Teste o ajustamento da distribuição normal com valor esperado 1.675 e variância 0.029^2.

(b) Teste ao nível de significancia de 1% a hipótese H_0 : "X é uma variável aleatória com distribuição normal", admitindo que as estimativas de máxima verosimilhança de μ e σ^2 são os respectivos momentos da amostra agrupada.

8.25 Mil indivíduos foram classificados segundo o sexo e o daltonismo tendo-se obtido o seguinte quadro:

	Homem	Mulher
Não daltónico	442	514
Daltónico	38	6

Acha que o daltonismo é independente do sexo? Justifique. Considere um nível de significância de 5%.

8.26 Uma importante empresa de equipamento desportivo pretende seleccionar um de três programas de treino de vendas A, B ou C. Os resultados do desempenho de vendas de 120 vendedores após o treino foram os seguintes:

	Resultados		
Programa	Medíocre	Suficiente	Bom
A	6	25	9
B	8	20	7
C	10	30	5

Teste se o desempenho dos vendedores não é influenciado pelo programa de treino, justificando o procedimento adoptado.

8.27 Num estudo sobre os efeitos da vacinação na mortalidade por contracção de varíola em Londres no ano de 1901, obteve-se o seguinte conjunto de resultados:

	Recuperaram	Morreram
Vacinados	847	153
Não vacinados	126	158

Conjectura-se que não existe associação entre a vacinação contra a varíola e a mortalidade devido a essa doença. Verifique se esta hipótese é apoiada pelos dados recolhidos ao nível de significância de 10%.

8.28 Num levantamento de opinião pública em 1982 nos Estados Unidos da América foram postas as duas seguintes questões a 1397 pessoas:

- É a favor da obrigatoriedade do registo de porte de arma?
- Concorda com a pena de morte?

tendo-se obtido o conjunto de resultados na tabela abaixo.

	Pena de morte	
Registo obrigatório	Sim	Não
Sim	784	236
Não	311	66

Formule e teste a hipótese de não existir associação entre as respostas às duas questões.

Capítulo 9

Introdução à regressão linear simples

9.1 Interessa estudar a relação entre a resistência de um determinado tipo de plástico (Y) e o tempo que decorre a partir da conclusão do processo de moldagem até ao momento de medição da resistência ($x\,[horas]$). As observações que se seguem foram efectuadas em 12 peças construídas com este plástico, escolhidas aleatoriamente.

i	1	2	3	4	5	6	7	8	9	10	11	12
x_i	32	48	72	64	48	16	40	48	48	24	80	56
y_i	230	262	323	298	255	199	248	279	267	214	359	305

(a) Represente graficamente as observações e desenhe a recta que, no seu entender, melhor se ajusta às observações.

(b) Considere um modelo de regressão linear simples para explicar as observações. Obtenha a estimativa dos mínimos quadrados dos coeficientes da recta de regressão e desenhe-a no gráfico.

(c) Calcule o coeficiente de determinação e comente o valor obtido.

(d) Proceda ao teste da hipótese "O coeficiente angular é nulo". Qual o interesse desta hipótese? Relacione-o com o resultado obtido em (c).

(e) Calcule o intervalo de confiança a 95% para o valor esperado da resistência obtida 48 horas depois de concluída a moldagem. Acha legítimo usar o mesmo procedimento tratando-se de um período de 10 horas em vez de 48 horas? Justifique a sua resposta.

9.2 Um estudo sobre a influência da velocidade do vento (X), em m/s, na quantidade de água (Y) que se evapora por dia, em centenas de litros, na albufeira de certa barragem, a temperaturas constantes, conduziu a:

x_i	20	50	30	100	70
y_i	3	5	3	10	8

(a) Adoptando um modelo de regressão linear simples, estime a recta de regressão de Y sobre X e obtenha uma estimativa da quantidade média de água evaporada quando a velocidade do vento é igual a $90m/s$. Faça uso dos seguintes valores:

$$\bar{x} = 54.0 \quad \bar{y} = 5.8 \quad \sum_i x_i^2 = 18700 \quad \sum_i y_i^2 = 207 \quad \sum_i x_i y_i = 1960$$

(b) Calcule o coeficiente de determinação do modelo estimado.

(c) Teste a significância da regressão. Indique o valor-p desse teste e comente o resultado face ao valor obtido na alínea anterior. (*Exame 5 Fev 2002*)

9.3 O modelo de regressão linear simples foi usado para estudar a relação entre a produção de uma variedade de trigo (Y) e a quantidade de adubo usada como fertilizante (x). Foram efectuadas 7 observações:

i	1	2	3	4	5	6	7
x_i	100	200	300	400	500	600	700
y_i	40	50	50	70	65	65	80

As observações foram tratadas em seguida usando o pacote estatístico R. Parte do output obtido é o seguinte:

```
> producao <- c(40,50,50,70,65,65,80)
> adubo <- c(100,200,300,400,500,600,700)
> mrl <- lm(producao~adubo)
> summary(mrl)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 36.42857    5.03812   7.231  0.00079 ***
adubo        0.05893    0.01127   5.231  0.00338 **

Residual standard error: 5.961 on 5 degrees of freedom
Multiple R-Squared: 0.8455,     Adjusted R-squared: 0.8146
```

(a) Proceda ao teste da hipótese de que a adubação não tem influência na produção.

(b) Acha que o modelo se ajusta adequadamente às observações? Justifique.

(c) Calcule uma estimativa do valor esperado da produção com uma quantidade de adubo à sua escolha e indique uma estimativa da variância associada.

9.4 Da análise do consumo médio de energia por agregado familiar durante 10 dias de um mês de Inverno numa cidade obtiveram-se os seguintes resultados:

i	1	2	3	4	5	6	7	8	9	10
x_i	15	14	12	14	12	11	11	10	12	13
y_i	4.3	4.4	5.3	4.6	5.5	5.9	5.7	6.2	5.2	5.0

X: Temperatura diária média (^oC), Y: Consumo médio de energia (kW)

$$\sum_{i=1}^{10} x_i = 124 \qquad \sum_{i=1}^{10} y_i = 52.1 \qquad \sum_{i=1}^{10} x_i y_i = 637.1$$

$$\sum_{i=1}^{10} x_i^2 = 1560 \qquad \sum_{i=1}^{10} y_i^2 = 275.13$$

O modelo de regressão linear simples foi usado para estudar a relação entre o consumo médio de energia por agregado familiar e a temperatura diária média.

(a) Escreva a equação da recta de regressão estimada e obtenha um intervalo de confiança a 90% para o verdadeiro valor do declive da recta de regressão.

(b) Qual o valor predito para o consumo médio num dia de temperatura média igual a 10^oC? Que responderia se lhe fosse pedida uma predição do consumo médio para um dia com temperatura média de 20^oC?

9.5 Uma amostra de alunos seleccionada ao acaso dum curso com as disciplinas de Matemática e Estatística produziu as seguintes classificações num teste efectuado no final do ano lectivo (escala 0-100):

i	1	2	3	4	5	6	7	8	9	10
x_i (Mat.)	56	50	72	67	31	50	65	40	80	61
y_i (Est.)	60	50	67	75	44	56	72	48	76	62

A partir destes dados, o professor resolveu determinar o valor de algumas quantidades:

$$\sum_{i=1}^{10} x_i = 572 \quad \sum_{i=1}^{10} x_i^2 = 34716 \quad \sum_{i=1}^{10} x_i y_i = 36335$$
$$\sum_{i=1}^{10} y_i = 610 \quad \sum_{i=1}^{10} y_i^2 = 38394 \quad \sum_{i=1}^{10} (x_i - \overline{x})(y_i - \overline{y}) = 1443$$

e a partir delas deduziu a equação de regressão estimada pelo método dos minimos quadrados:

$$\hat{E}(Est.|Mat. = x) = 19.7 + 0.722x.$$

(a) Qual o interesse no uso do modelo de regressão em geral e em particular no caso presente?

(b) A Joana, o António e a Maria obtiveram 60, 95 e 20 em Matemática, respectivamente, mas faltaram ao teste de Estatística.

Poderá sugerir valores para as notas esperadas no teste de Estatística dos alunos que faltaram? Justifique a sua resposta. Acha que os valores que sugere para as notas de Estatística são de confiança?

(c) Suponha que o João obteve 70 em Estatística e faltou a Matemática.

Obtenha uma nova recta de regressão que permita estimar uma nota para o teste de Matemática deste aluno e indique esse valor predito. Justifique a resposta.

9.6 Uma liga metálica é submetida a várias tensões ($x[10^3 Kgf/cm^2]$), tendo-se registado o tempo decorrido ($T\,[horas]$) até se atingir a rotura. Alguns dos resultados obtidos nesta experiência foram os seguintes:

i	1	2	3	4
x_i	15	20	25	30
t_i	2500	600	200	70

Admite-se que as duas variáveis estão relacionadas de acordo com o seguinte modelo de regressão linear: $\ln T = \beta_0 + \beta_1 X + \varepsilon$.

(a) Assumindo as hipóteses que julgar convenientes, obtenha as estimativas dos mínimos quadrados de β_0 e β_1.

(b) O modelo foi utilizado para prever os tempos correspondentes às tensões de $25 \times 10^3\, Kgf/cm^2$ e $50 \times 10^3\, Kgf/cm^2$. Calcule as estimativas desses tempos. Diga, justificando, se concorda que o modelo adoptado seja usado para predizer aqueles tempos.

9.7 Numa fábrica deseja-se estimar o valor esperado do custo total para produzir um item, $E(Y)$, como função do número de unidades produzidas (x). Após um certo período de observação, foi possível obter os dados da tabela seguinte:

i	1	2	3	4	5	6	7
x_i	35	75	138	161	199	224	252
y_i	81	88	133	165	239	282	343

(a) Admitindo que as variáveis em causa estão relacionadas de acordo com o modelo $Y = \alpha e^{\beta x}\epsilon$, determine as estimativas dos parâmetros α e β.

(b) Acha que o custo total de produção do item é significativamente influenciado pelo número de unidades produzidas? Justifique.

(c) Construa um intervalo de confiança de 95% para α.

Soluções

(Edição de Fevereiro de 2008)

Capítulo 1

1.1 **d)** $\bar{x} = 3.167$; $s = 0.886$ (dados não agrupados)

e) mediana = $\tilde{x} = 3.25$; $q_1 = 2.4$; $q_3 = 3.9$

1.2 **a)** $\bar{x} = 2.866$; $\tilde{x} = 3$; moda = 3 **c)** 0.3098

1.3 localização: $\bar{x} = 10.6$; $\tilde{x} = 10.65$; moda = 10.9

dispersão: $s^2 = 0.1$; $s = 0.3162$; $r = 0.9$; $c.v. = 2.98 \times 10^{-2}$

1.4 **a)** $\bar{x} = 10.476$; $s = 2.665$ **b)** 11.64

1.5 **a)** $\bar{x} \cong 3.65$; $s^2 \cong 5.17$; $s \cong 2.27$ **b)** média — sim; variância — sim

c) média — sim; variância — não

Capítulo 2

2.2 **a)** $1 + y - x$ **b)** $x - 2y$ **c)** $x - y$ **d)** $1 - y$

2.4 **a)** 0.13 **b)** 0.25

2.5 **a)** $S = \{1,2,3,4,5,6\}$, $P(\{1\}) = P(\{3\}) = P(\{5\}) = \dfrac{2}{9}$, $P(\{2\}) = P(\{4\}) = P(\{6\}) = \dfrac{1}{9}$

b) $4/9$ **c)** $1/3$

2.6 **a)** 0.0001 **b)** 0.001 **c)** 0.504

2.7 **a)** $1/210$ **b)** $2/9$

2.8 0.125

2.9 **a)** 988/2303 **b)** 435/2303 **c)** 22529/23030 **d)** $3/658$

2.10 Admitindo equiprobabilidade, $25740/3^{13}$

2.11 **a)** $\left(\dfrac{n-1}{n}\right)^{r-1}$ **b)** $\dfrac{n!}{n^r(n-r)!}$

2.12 **a)** $S = \{(D,D),(F,D),(D,F),(F,F)\}$, onde $D \equiv$ defeituoso; $F \equiv$ não defeituoso

$P\{(D,D)\} = 0.04$; $P\{(F,D)\} = P\{(D,F)\} = 0.16$; $P\{(F,F)\} = 0.64$

b) $P(A_1) = P(A_2) = 0.2$; $P(A_3) = 0.36$; $P(A_4) = 0.32$

c) $P\{(D,D)\} = 2/90$; $P\{(F,D)\} = P\{(D,F)\} = 16/90$; $P\{(F,F)\} = 56/90$

$P(A_1) = P(A_2) = 0.2$; $P(A_3) = 34/90$; $P(A_4) = 32/90$

2.13 **a)** $n/(2n-1)$, onde n é o número de moedas de cada tipo

b) $1/2$ **c)** $n/(2n-1)$ **d)** $(3n-1)/(4n-2)$

2.14 **b)** P(A ganhar) = 0.6587; P(B ganhar)= 0.3413

c) P(A ganhar) = $2/3$; P(B ganhar)= $1/3$

2.15 **b)** $10/11$

2.16 **a)** 0.504 **b)** 0.296

2.18 **a)** 0.4 **b)** 2/3

2.19 **a)** 51.25% **b)** 7.32%

2.20 **a)** 0.0905 **b)** 0.6633

2.21 **b)** 0.64

2.22 0.2

2.23 **a)** 0.13043 **b)** 0.43478

Capítulo 3

3.1 **a)** i) 4/5 ii) 1/3

b) i) $f_X(x)=\begin{cases}0.2, & x=0\\ 0.6, & x=1\\ 0.2, & x=2\end{cases}$ ii) $F_X(x)=\begin{cases}0, & x<0\\ 0.2, & 0\le x<1\\ 0.8, & 1\le x<2\\ 1, & x\ge 2\end{cases}$

iii) $E(X)=1$ e $V(X)=0.4$

c) a) i) 20/27 ii) 1/3

b) i) $f_X(x)=\begin{cases}8/27, & x=0\\ 12/27, & x=1\\ 6/27, & x=2\\ 1/27, & x=3\end{cases}$ ii) $F_X(x)=\begin{cases}0, & x<0\\ 8/27, & 0\le x<1\\ 20/27, & 1\le x<2\\ 26/27, & 2\le x<3\\ 1, & x\ge 3\end{cases}$

iii) $E(X)=1$ iv) $V(X)=2/3$

3.2 **a)** $f_X(x)=\begin{cases}0.001, & x=0\\ 0.027, & x=1\\ 0.243, & x=2\\ 0.729, & x=3\end{cases}$ **b)** $F_X(x)=\begin{cases}0, & x<0\\ 0.001, & 0\le x<1\\ 0.028, & 1\le x<2\\ 0.271, & 2\le x<3\\ 1, & x\ge 3\end{cases}$

c) $E(X)=2.7$, moda=mediana=3 e $V(X)$=0.27

3.3 **a)** $a=1/6$ **b)** $F_X(x)=\begin{cases}0, & x<1\\ 1/6, & 1\le x<2\\ 1/2, & 2\le x<3\\ 1 & x\ge 3\end{cases}$

c) moda = 3; mediana = qualquer ponto $\in$ [2,3]; $E(X)=7/3$

3.4 **a)** $-1/3\le c\le 1/4$ **b)** $E(X)=\dfrac{10-9c}{4}$ **c)** $V(X)=\dfrac{20-8c-81c^2}{16}$

3.5 $E(X)=2.6$; $P(X<0)=0.3$

3.6 **a)** $f_X(x)=\begin{cases}1/6, & x=0\\ 1/12, & x=2\\ 1/4, & x=4\\ 1/2, & x=6\end{cases}$ **b)** i) 1/6 ii) 1/2 iii) 1/12 iv) 1/3

3.7 **a)** 0.0362 **b)** 0.75 **c)** i) 0.8574 ii) 20

3.8 **a)** 0.0702 **b)** ≤ 8 **c)** 7.02

3.9 $$\sum_{k=0}^{4}\frac{\binom{2000}{k}\binom{58000}{250-k}}{\binom{60000}{250}}$$

3.10 0.3005

3.11 **a)** 0.2231; 0.4308 **b)** 4.47

3.12 $$\sum_{x=2}^{500}\binom{500}{x}\left(\frac{1}{365}\right)^{x}\left(\frac{364}{365}\right)^{500-x}\cong 0.3977$$

3.13 **a)** 0.3233 **b)** 0.0831

Capítulo 4

4.1 **a)** $k=0$ **b)** $F_X(x)=\begin{cases}0, & x<-1\\ x^2/2+x+1/2, & -1\leq x<0\\ -x^2/2+x+1/2, & 0\leq x<1\\ 1, & x\geq 1\end{cases}$ **c)** $E(X)=0$; $V(X)=1/6$

d) moda = 0; mediana = 0; $F_X^{-1}(1/4)=-1+\sqrt{2}/2$ **e)** 1/4

4.2 **a)** $F_X(x)=\begin{cases}0, & x<0\\ 5x^4-4x^5, & 0\leq x<1\\ 1, & x\geq 1\end{cases}$ **b)** $\dfrac{112}{243}$

c) i) $P(L=C_1-C_3)=101/243$; $P(L=C_2-C_3)=142/243$:

$$F_L(l)=\begin{cases}0, & l<C_2-C_3\\ 142/243, & C_2-C_3\leq l<C_1-C_3\\ 1, & l\geq C_1-C_3\end{cases}$$

4.3 0.4502

4.4 **a)** $\alpha=2$; $\beta=1/5$ **b)** 0.5

4.5 C_1+50C_2

4.6 **a)** 0.375 **b)** $\mu_X=4/3=133.3Kg$; $\sigma_X=62.36Kg$ **c)** 245.23 Kg

4.7 **a)** 0.6826 **b)** 0.8759

4.8 **a)** $\mu=2.5$; $\sigma=0.469$ **b)** 0.6826

4.9 **a)** 0.0023 **b)** 6231.04 horas **c)** 1/8

4.10 a) $F_X(x)=\begin{cases}0, & x<0\\ 1-e^{-2x}, & x\geq 0\end{cases}$ **b)** 0.3679

4.11 a) 0.2202 **b)** 0.6592

Capítulo 5

5.1 a) $f_X(x)=\begin{cases}0.2, & x=0\\ 0.65, & x=1\\ 0.15, & x=2\end{cases}$; $f_Y(y)=\begin{cases}0.5, & y=0\\ 0.36, & y=1\\ 0.14, & y=2\end{cases}$

b) $F_X(x)=\begin{cases}0. & x<0\\ 0.2, & 0\leq x<1\\ 0.85, & 1\leq x<2\\ 1, & x\geq 2\end{cases}$ **c)** 0.18 **d)** 1.59; 0.7619

5.2 b)

Y\X	0	1	2	$f_Y(y)$
0	0.064	0.096	0	0.16
1	0.096	0.240	0.144	0.48
2	0	0.144	0.216	0.36
$f_X(x)$	0.16	0.48	0.36	1

5.3 a) i) $f_X(x)=\begin{cases}2/9, & x=1\\ 1/2, & x=2\\ 5/18, & x=3\end{cases}$ ii) $F_Y(y)=\begin{cases}0, & y<1\\ 1/6, & 1\leq y<2\\ 11/18, & 2\leq y<3\\ 1, & y\geq 3\end{cases}$ iii) 11/18

iv) $f_{X|Y=1}(x)=\begin{cases}2/3, & x=1\\ 1/3, & x=3\end{cases}$; $f_{X|Y=3}(x)=\begin{cases}2/7, & x=1\\ 3/7, & x=2\\ 2/7, & x=3\end{cases}$ v) 5/3

b) $P[E(X|Y)=z]=\begin{cases}P(Y=1), & z=5/3\\ P(Y=2), & z=9/4\\ P(Y=3), & z=2\end{cases}$

c) Não, porque $\exists_{(x,y)}: f_{X,Y}(x,y)\neq f_X(x)f_Y(y)$, por exemplo $(x,y)=(1,2)$

d) 1.0895

5.4 **a)** i) e ii):

X\Y	0	1	2	3	4	$f_X(x)$
0	0.04	0	0.3	0	0.06	0.4
1	0	0.22	0	0.28	0	0.5
2	0	0	0.1	0	0	0.1
$f_Y(y)$	0.04	0.22	0.4	0.28	0.06	1

iii) $F_X(x)=\begin{cases}0, & x<0\\ 0.4, & 0\le x<1\\ 0.9, & 1\le x<2\\ 1, & x\ge 2\end{cases}$ iv) $f_{X|Y=2}(x)=\begin{cases}0.75, & x=0\\ 0.25, & x=2\end{cases}$

b) Não, porque $\exists_{(x,y)}: f_{X,Y}(x,y)\neq f_X(x)f_Y(y)$, por exemplo $(x,y)=(1,0)$

c) i) $f_{Y|X=0}(y)=\begin{cases}0.1, & y=0\\ 0.75, & y=2\\ 0.15, & y=4\end{cases}$; $f_{Y|X=1}(y)=\begin{cases}0.44, & y=1\\ 0.56, & y=3\end{cases}$; $f_{Y|X=2}(y)=\begin{cases}1, & y=2\\ 0, & y\neq 2\end{cases}$

ii) $E(Y|X=2)=2$; $V(Y|X=2)=0$ iii) $F_{Y|X=0}(y)=\begin{cases}0, & y<0\\ 0.1, & 0\le y<2\\ 0.85, & 2\le y<4\\ 1, & y\ge 4\end{cases}$

iv) 0.75 v) 0

5.5 **a)** $\rho_{X,Y}=0.5\neq 0 \Rightarrow X$ e Y não são independentes **b)** 3.5

5.7 **a)** $V(Z)=2pq$ **b)** $P[E(X|Y)=q]=1$ **c)** Por exemplo:

$u \setminus v$	0	1
0	p	0
1	0	q

5.8 **a)** 0.6602 **b)** 1.8; 2

5.9 **a)** $a=\sqrt{2}/2$ **b)** Sim, porque $\forall_{(x,y)}: f_{X,Y}(x,y)=f_X(x)f_Y(y)$

c) $F_Y(y)=\begin{cases}0, & y<-a\\ \frac{\sqrt{2}}{2}\left(y+\frac{\sqrt{2}}{2}\right), & -a\le y\le a\\ 1, & y>a\end{cases}$

5.10 **a)** Não, porque $\exists_{(x,y)}: f_{X,Y}(x,y)\neq f_X(x)f_Y(y)$, por exemplo $(x,y)=\left(\frac{3}{4},\frac{3}{4}\right)$

b) $F_X(x)=\begin{cases}0, & x<0\\ x^3-3x^2+3x, & 0\le x\le 1\\ 1, & x>1\end{cases}$

c) $f_{X|Y=y}(x)=\begin{cases}\frac{2(1-x-y)}{(1-y)^2}, & 0<x<1-y\\ 0, & \text{caso contrário}\end{cases}$, $0<y<1$ **d)** 3/4 **e)** 1

5.11 a) $f_X(x)=\begin{cases}1/2, & -1<x<1\\ 0. & \text{c. c.}\end{cases}$; $f_Y(y)=\begin{cases}1/2, & 0<y<2\\ 0. & \text{c. c.}\end{cases}$ **b)** $P(X<Y)=7/8$

5.12 0.4375

5.13 a) $\rho_{X,Y}=0.5$ **b)** $V(X|Y=y)=y^2/12$

5.14 a) $\mu_D=3.6$ cm; $\sigma_D=0.0224$ cm **b)** 0.1867

5.15 a) 0.0367 **b)** 0.0222 **c)** 0.9994; 0.3032

5.16 $\cong 0$

5.17 ≥ 65

5.18 0.0409

5.19 a) 0.4169 **b)** 0.4169

c) A probabilidade de cumprir o compromisso é elevada (0.9772) **d)** 0.3014

5.20 0.9236

5.21 a) 0.7 **b)** 0.286 e 3.5 m **c)** 0.9997

5.22 a) 0.2 **b)** 0.7939 **c)** $\cong 0$

Capítulo 6

6.1 a) $f_{X_1,\ldots,X_5}(x_1,\ldots,x_5)=\begin{cases}\prod_{i=1}^{5}|x_i|, & |x_i|<1,\ i=1,\ldots,5\\ 0, & \text{c. c.}\end{cases}$

b) $E(\overline{X})=0$; $V(\overline{X})=0.1$; $s^2=\dfrac{\sum(x_i-\overline{x})^2}{n-1}=0.73075$ **c)** 0.9718; 0.9654

6.4 8/9, logo $\overline{X}$ é mais eficiente

6.5 a) $\pi/2>1$ **b)** $1/2<1$ **c)** Para a população normal $\overline{X}$ é mais eficiente, para a população da alínea b) verifica-se o contrário

6.6 T_1 é melhor para $|\theta|>\sqrt{3/2}$ e T_2 é melhor para $|\theta|<\sqrt{3/2}$

6.7 a) $\hat{p}=1/4$ **b)** $\hat{p}=3/4$ **c)** $\hat{p}=2/3$

6.8 $\hat{R}=\dfrac{k}{n-k}$

6.9 $\hat{p}=0.04375$

6.10 a) $\hat{\lambda}=\overline{X}$, é centrado **b)** $\hat{\sigma}^2=\overline{X}$ **c)** $\hat{\lambda}=0.4$ **d)** $\hat{\lambda}=0.1252$

6.11 $\hat{\mu}=\overline{X}$, é centrado

6.12 $\hat{\mu}=55.833$; $\hat{\sigma}^2=101.639$; $\hat{\sigma}=10.082$; $\hat{P}(X>70)=0.0793$

6.13 b) T_2

6.14 $\hat{P}(X>200)=0.3168$

6.15 a) $\hat{\alpha}=2.022$; $\hat{E}(X)=2.534$; $\hat{V}(X)=1.755$ **b)** moda

6.16 a) 0.1314 **b)** 0.5785 **c)** 0.2923

6.17 $\cong 1-\Phi\left(0.4\sqrt{12n}\right)$, n elevado

6.18 0.0212 (com correcção de continuidade), 0.0202 (sem correcção de continuidade)

6.19 **a)** 0.0956 **b)** 0.9044 **c)** 0.01; 0.0315

6.20 **a)** 0.0456 **b)** 158.7 **c)** $n \geq 4$

Capítulo 7

7.1 $IC_{95\%}(\mu) = (136.08;143.92)$

7.2 $IC_{95\%}(\mu) = (3.435;5.653)$ e n=200

7.3 $IC_{95\%}(\mu) = (94.604;102.736)$; $IC_{95\%}(\sigma^2) = (43.656;161.008)$

7.4 **a)** $IC_{99\%}(\mu) = (1.723;2.844)$ **b)** $IC_{99\%}(\sigma) = (0.4007;1.285)$

7.5 $IC_{95\%}(\mu_1 - \mu_2) = (-192.106;275.506)$

7.6 $IC_{99\%}(\mu_1 - \mu_2) = (-8.174;2.974)$. Como $0 \in IC_{99\%}$ não se rejeita a hipótese de que as médias são iguais para n. s. $\leq 1\%$.

7.7 $IC_{95\%}(\mu_1 - \mu_2) = (-5.635;11.035)$

7.8 $IC_{90\%}(\mu_1 - \mu_2) = (6.322;13.678)$

7.9 **a)** $IC_{95\%}(p) = (0.0964;0.2436)$ **b)** 1809

7.10 **a)** $IC_{90\%}(p) = (0.4720;0.5780)$ **b)** $n \geq 1688$

7.11 **a)** $T = \overline{X}$; $p = P(X \leq 1)$, $\hat{p} = e^{-T} + Te^{-T}$ **b)** $IC_{90\%}(\lambda) \cong (0.7338;0.8705)$

7.12 $IC_{95\%}(\alpha) \cong (0.3216;0.4784)$

Capítulo 8

8.1 **a)** $c = 1.5126$ **b)** $n = 35$; $c = 1.5621$ **c)** $\alpha = \beta = 0.3632$

8.2 **a)** $\overline{X} \underset{\text{sob}\, H_0}{\sim} N(20;1)$

b)

μ	17	18	19	20	21	22	23
$P(\text{aceitar } H_0 \mid \mu)$	.1587	.5000	.8400	.9544	.8400	.5000	.1587

c) $\beta(\mu) = 1 - P(\text{aceitar } H_0 \mid \mu)$

8.3 $n = 14$, $c = 5.429$

8.4
C_1: $\alpha = 0.6875$, $\beta = 0.0508$
C_2: $\alpha = 0.3125$, $\beta = 0.2617$
C_3: $\alpha = 0.0625$, $\beta = 0.6836$

8.5 **a)** Sim (-1.645 < -0.408 < 1.645) **b)** Valor-p = 0.6818 > 0.1

8.6 **a)** $H_0: \mu = 16$ versus $H_1: \mu \neq 16$, não rejeitar H_0 para $\alpha \leq 5\%$ e rejeitar para $\alpha \geq 10\%$ (-2.145 < -2.037 < -1.761)

b) $H_0{:}\sigma^2 = 0.25$ versus $H_1{:}\sigma^2 \neq 0.25$, rejeitar H_0 pelo menos para $\alpha \geq 0.1\%$, ou seja para os níveis de significância usuais $(1.296 < 2.697)$

8.7 Não se rejeita H_0 $(0.427 < 1.699)$

8.8 **a)** $H_0{:}\mu \geq 255$ versus $H_1{:}\mu < 255$ (ou $H_0{:}\mu = 255$ versus $H_1{:}\mu < 255$), não se rejeita H_0 para $\alpha \leq 5\%$ (-1.017 > -1.796)

b) $\hat{\xi}_{0.05} = \hat{\mu} - 1.6449\hat{\sigma} = 234.64$

8.9 **a)** Teste sobre μ: -2.131 < 1.661 < 2.131, aceitar a 5%

Teste sobre σ: 6.262 < 12.06 < 27.49, aceitar a 5%

b) Teste sobre $\mu_1 - \mu_2$ com $\sigma_1 = \sigma_2$: -3.962 < -1.703, rejeitar a 5%, isto é as medidas introduzidas provocaram um aumento significativo na cotação esperada.

8.10 **a)** ganha A **b)** P(A perder injustamente) = 0.05,

P(B perder injustamente) = 0.1379, logo a aposta não é justa.

8.11 $H_0{:}\mu = 30$ versus $H_1{:}\mu \neq 30$, rejeitar H_0 para $\alpha = 1\%$ $(3.02 > 2.576)$

8.12 **a)** Assumindo que as variáveis têm distribuições normais rejeita-se H_0 para $\alpha \geq 1\%$ (-4.86<-2.5758). **b)** 0.4404

8.13 $H_0{:}\mu_1 = \mu_2$ versus $H_1{:}\mu_1 \neq \mu_2$

a) Rejeitar H_0 para $\alpha > 3.94\%$ e não rejeitar para $\alpha < 3.94\%$

b) Rejeitar H_0 para $\alpha \geq 10\%$ e não rejeitar para $\alpha \leq 5\%$ (1.746 < 2.088 < 2.120)

8.14 **a)** $H_0{:}\mu_1 \leq 60$ versus $H_1{:}\mu_1 > 60$ (ou $H_0 : \mu_1 = 60$ versus $H_1 : \mu_1 > 60$), rejeitar H_0 para $\alpha \geq 5\%$ e não rejeitar para $\alpha \leq 2.5\%$ (1.833 < 1.93 < 2.262)

b) $H_0{:}\sigma_1^2 = 16$ versus $H_1{:}\sigma_1^2 \neq 16$, rejeitar H_0 para $\alpha \geq 20\%$ e não rejeitar para $\alpha \leq 15\%$ (3.785 < 3.855 < 4.168), ou seja para os níveis de significância usuais

c) $H_0{:}\mu_1 = \mu_2$ versus $H_1{:}\mu_1 \neq \mu_2$, rejeitar H_0 para $\alpha \geq 40\%$ e não rejeitar para $\alpha \leq 30\%$ (0.865 < 0.996 < 1.071), ou seja para os níveis de significância usuais

8.15 $H_0{:}\mu_1 = \mu_2$ versus $H_1{:}\mu_1 < \mu_2$ ou $H_0 : \mu_1 \geq \mu_2$ versus $H_1{:}\mu_1 < \mu_2$. Valor-p $\cong 0.4522$. Como não se rejeita H_0 para os níveis de significância usuais não se poderá concluir que o segundo método é significativamente superior ao primeiro.

8.16 $H_0{:}p = 0.9$ versus $H_1{:}p \neq 0.9$. Valor-p $\approx 2\times10^{-6}$. Rejeita-se H_0 para $\alpha \geq 2\times10^{-6}$, ou seja para os níveis de significância usuais, logo a afirmação não é consistente com os dados

8.17 $H_0{:}p = 0.1$ versus $H_1{:}p > 0.1$ (p é a proporção de lâmpadas com defeitos), não rejeitar H_0 para $\alpha \leq 17.36\%$, ou seja para os níveis de significância usuais

8.18 **a)** $H_0{:}p = 0.1$ versus $H_1{:}p > 0.1$ ou $H_0{:}p \leq 0.1$ versus $H_1{:}p > 0.1$

b) $T = \dfrac{\hat{P} - 0.1}{\sqrt{\dfrac{0.1 \times 0.9}{100}}} \underset{\text{sob}\, H_o}{\overset{a}{\sim}} N(0,1)$; Região crítica: $\{T > 1.4051\}$

c) $t_{obs} = 0.67 < 1.4051$, não deve rejeitar o carregamento (para o n. s. de 8%)

8.19 a) $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 \neq \mu_2$, não se rejeita H_0 para $\alpha \leq 80\%$, ou seja para os n. s. usuais (-0.255 < -0.02 < 0.255)

b) $IC_{95\%}(\mu) = (8.936;11.136)$

c) $H_0 : \sigma^2 = 4$ versus $H_1 : \sigma^2 \neq 4$, rejeita-se H_0 para $\alpha \geq 0.1\%$, ou seja para os n. s. usuais (183.65 > 89.56)

d) $H_0 : p = 0.1$ versus $H_1 : p > 0.1$ ou $H_0 : p \leq 0.1$ versus $H_1 : p > 0.1$. Não se rejeita H_0 para $\alpha \leq 31.9\%$, ou seja para os n. s. usuais

8.20 Sim, para ambos os níveis (0.5596 < 7.815 < 11.34)

8.21 a) 11.96 < 20.52, os resultados são compatíveis com distribuição binomial e equiprobabilidade dos sexos ao n. s. de 0.1%

b) 0.025 < valor-p < 0.05

8.22 a) A hipótese não é rejeitada (0.2308 < 9.488)

b) É, aos n. s. usuais (não se rejeita para $\alpha \leq 97.5\%$, 0.4214 < 0.472)

8.23 valor-p $\in (0.5,0.6)$ pelo que não se rejeita H_0 para os n. s. usuais

8.24 a) Não rejeitar para $\alpha \leq 92.5\%$ (0.4347 < 0.472), ou seja para os n. s. usuais

b) A hipótese não é rejeitada (0.4505 < 6.635)

8.25 Rejeita-se a hipótese de independência para $\alpha = 5\%$ (27.14>3.841)

8.26 A hipótese de independência entre o programa de treino e os resultados não é rejeitada para $\alpha \leq 60\%$, ou seja para os n. s. usuais (2.742 > 2.753)

8.27 A hipótese de não associação não é apoiada pelos dados ao n. s. de 10% (195.97 > 2.706)

8.28 Rejeita-se a hipótese de não associação para $\alpha \geq 2.5\%$, não se rejeita para $\alpha \leq 1\%$ (5.024 < 5.15 < 6.635)

Capítulo 9

9.1 b) $\hat{\beta}_0 = 153.917$; $\hat{\beta}_1 = 2.417$

c) $R^2 = 0.9593$. A recta estimada ajusta-se bem, 95.9% da variação de Y é explicada pela relação linear com x.

d) $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 \neq 0$, rejeita-se H_0 para $\alpha \geq 0.1\%$, ou seja para os n.s. usuais (15.35 > 4.587)

e) $IC_{95\%}\left(\mu_{Y|x=48}\right)=(263.64;276.20)$. Não é legítimo para x = 10 porque $10 \notin \left(\min_i x_i ; \max_i x_i\right)$

9.2 **a)** $\hat{\beta}_0 = 0.6359$; $\hat{\beta}_1 = 0.0965$; $\hat{E}(Y|x=90) \cong 9.2427$

b) $R^2 = 0.9711$.

c) $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 \neq 0$, rejeita-se H_0 para os n.s. usuais. $t_{obs} = 9.929$, valor-p $\in (0.002, 0.01)$.

9.3 **a)** $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 \neq 0$, $t_{obs} = 5.23$, valor-p = 0.003, rejeita-se H_0 para $\alpha \geq 0.3\%$, ou seja para os n.s. usuais. Os dados indicam que a adubação influencia significativamente a produção

b) $R^2 = 84.5\%$. O modelo ajusta-se bem às observações.

c) $x_0 = 400$; $\mu_{Y|x_0} = 60$; $s^2\left(\hat{\mu}_{Y|x_0}\right) = 5.08$

9.4 **a)** $\hat{\mu}_{Y|x} = 10.1589 - 0.3991x$; $IC_{90\%}(\beta_1) = (-0.4474; -0.3508)$

b) $\hat{\mu}_{Y|x=10} = 6.17$, não se deve usar para x = 20

9.5 **b)** Só para x = 60, as outras duas são extrapolações. $\hat{E}(Est|Mat = 60) \cong 63$

c) $\hat{E}(Mat|Est = y) = -17.144 + 1.2188y$; $\hat{E}(Mat|Est = 70) \cong 68$

9.6 **a)** $\hat{\alpha} = 11.2633$; $\hat{\beta} = -0.23651$

b) $\hat{t}_{|x=25} = 210.7$ horas, $\hat{t}_{|x=50} = 0.57$ não deve ser usado com x = 50 porque se trata de uma extrapolação

9.7 **a)** $\hat{\alpha} = 55.77$; $\hat{\beta} = 0.0070805$

b) $H_0 : \beta = 0$ versus $H_1 : \beta \neq 0$, rejeita-se H_0 para $\alpha \geq 0.1\%$, ou seja para os n.s. usuais (15.19 > 6.869), conclui-se que o custo total é significativamente influenciado pelo número de unidades produzidas

c) $IC_{95\%}(\alpha) = (45.43; 68.47)$