IPEIO

Introdução às Probabilidades e Estatística

Luís Ramos
Pedro Mota

2019/20

Observação:

Estas folhas servem de apoio às aulas de IPEIO, Módulo de Probabilidades e Estatística. Para uma melhor compreensão dos assuntos abordados, aconselha-se a leitura de alguns dos livros indicados nas referências bibliográficas.

Conteúdo

Parte I

Exercícios Propostos

Capítulo 1

Introdução à Teoria da Probabilidade

1. Vinte e cinco membros de uma sociedade devem eleger um presidente, um secretário e um tesoureiro. Supondo que qualquer dos vinte e cinco membros é elegível para qualquer dos cargos, quantas são as hipóteses de um resultado final?

2. Considere o problema anterior. Suponha que não há diferenciação dos cargos. De quantas maneiras distintas se podia formar uma comissão, com três elementos escolhidos entre os vinte e cinco elementos?

3. Quantas palavras diferentes, com ou sem significado, se podem formar com as letras da palavra ROMA?

4. Quatro livros de Matemática, seis de Física e dois de Química, todos diferentes, devem ser arrumados numa prateleira. Quantas arrumações diferentes são possíveis se:

   (a) os livros de cada matéria ficarem todos juntos?

   (b) os livros de matemática devem ficar juntos?

5. Numa sala de cinema, de quantas maneiras diferentes se podem sentar numa fila de 12 lugares, 7 amigos?

6. De quantas maneiras 10 pessoas podem sentar-se num banco, se houver apenas 4 lugares?

7. De quantas formas diferentes se podem sentar 12 pessoas numa mesa redonda?

8. Um homem tem 3 camisas e 2 gravatas. De quantas maneiras pode vestir-se (com uma camisa e uma gravata)?

9. Num conjunto de 10 lâmpadas para árvore de natal, 2 são defeituosas. Quantas amostras de 6 lâmpadas podem ser escolhidas, de entre aquelas 10, de modo que:

   (a) as 6 lâmpadas escolhidas sejam todas boas?

   (b) entre a 6 escolhidas haja uma, e uma só defeituosa?

10. Dados 12 pontos num plano, não havendo 3 deles sobre a mesma recta,

    (a) quantas rectas são determinadas pelos pontos?

    (b) quantas dessas rectas passam pelo ponto A?

    (c) quantos triângulos são determinados pelos pontos?

    (d) Quantos desses triângulos contêm o ponto A como vértice?

    (e) Quantos desses triângulos contêm o lado AB?

11. De um baralho de 52 cartas são retiradas 10 cartas. Em quantos casos aparecem:
    (a) exactamente um ás?
    (b) pelo menos um ás?
    (c) exactamente dois ases?
    (d) pelo menos dois ases?

12. Determine o valor n que seja solução de:
    (a) $\binom{n-2}{2} = 6$
    (b) $\binom{n+1}{2} - \binom{n-1}{2} = \binom{n}{2} - 1$
    (c) $5 \times \binom{n}{3} = \binom{n+2}{4}$

13. Temos 15 livros para arrumar numa estante, em que 5 são de matemática. Qual a probabilidade de ficarem pelo menos 2 livros de matemática juntos?

14. Os atletas A, B, e C vão participar numa corrida e todos estão preparados para a ganhar. O sistema de cronometragem é suficientemente preciso de modo que não se admitem empates.
    (a) Qual a probabilidade de A terminar a corrida à frente de C?
    (b) Qual a probabilidade de A ganhar a corrida?

15. Por engano misturaram-se quatro pilhas novas com três usadas. Escolhendo, ao acaso e sem reposição, duas dessas pilhas, determine a probabilidade de:
    (a) Ambas serem novas
    (b) Nenhuma ser nova
    (c) Pelo menos uma ser nova

16. Num grupo de 20 congressistas, 8 só falam inglês, 5 só falam francês e 7 falam os dois idiomas. Qual a probabilidade de dois congressistas, escolhidos ao acaso, poderem conversar sem auxílio de um intérprete?

17. Uma urna contém quatro bolas amarelas, cinco bolas verdes, três bolas brancas e cinco bolas pretas. Extraem-se sucessivamente, ao acaso e sem reposição, quatro bolas. Qual a probabilidade de:
    (a) Obter na primeira extracção uma bola amarela, na segunda uma verde, depois uma branca e finalmente uma preta?
    (b) Obter o mesmo conjunto de cores independentemente da sua ordem?

18. (Teste de P.E. 2006/07) Considere os acontecimentos A e B de um espaço de resultados tais que $P(A \cup B) = 0.8$, e $P(A - B) = 0.3$. Qual o valor da $P(B)$?

19. Sejam A, B e C acontecimentos tais que $P(A) = P(B) = P(C) = \frac{1}{4}$, $P(A \cap B) = P(B \cap C) = 0$ e $P(A \cap C) = \frac{1}{8}$. Qual a probabilidade de se verificar pelo menos um dos 3 acontecimentos?

20. Sabendo que A e B são acontecimentos tais que $P(A) = \frac{2}{3}$, $P(B) = \frac{1}{2}$ e $P(A \cap B) = \frac{1}{3}$, determine $P(A - B)$, $P(A \cup B)$, $P(\bar{A} \cup \bar{B})$, $P(\bar{A} \cap B)$ e $P(A \cup \bar{B})$.

21. De 100 agricultores, 50 produzem vinho, 30 produzem milho e 10 produzem vinho e milho. Escolhendo um deste agricultores ao acaso qual a probabilidade de:
    (a) Ele produza vinho ou milho?
    (b) Ele não produza vinho nem milho?

22. A probabilidade de um homem estar vivo daqui a 25 anos é $\frac{3}{5}$ e a probabilidade da sua mulher ainda viver na mesma ocasião é de $\frac{2}{3}$. Determine a probabilidade de daqui a 25 anos:

    (a) Ambos estarem vivos.

    (b) Apenas o homem estar vivo.

    (c) Apenas a mulher estar viva.

    (d) Apenas um estar vivo.

23. Em determinada gelataria 40% dos clientes escolhem o sabor chocolate, 30% escolhem o sabor limão e 15% escolhem os dois. Seleccionou-se ao acaso um cliente dessa gelataria.

    (a) Se escolheu o sabor limão, qual a probabilidade de ter escolhido também o sabor chocolate? E vice-versa?

    (b) Qual a probabilidade de escolher limão ou chocolate?

24. Suponha que 10% da população de certo país sofre de problemas cardíacos e que, de entre estes, 70% são fumadores. De entre os que não sofrem de problemas cardíacos 45% fumam. Seleccionada ao acaso uma pessoa desta população:

    (a) Qual a probabilidade de ser fumadora?

    (b) Se for fumadora, qual a probabilidade de sofrer de problemas cardíacos?

25. Num clube de futebol treinam regularmente 30 jogadores, dos quais 8 são atacantes, 12 são médios e os restantes são defesas. Independentemente dos resultados dos restantes jogadores, cada atacante tem uma probabilidade de $3/4$ de marcar golo de penalty, cada médio tem uma probabilidade de $1/2$ de marcar golo por penalty e cada defesa consegue-o com probabilidade $1/5$.

    (a) Qual a probabilidade de que um jogador, escolhido ao acaso, marque golo devido a penalty?

    (b) Dado que, num jogo, um qualquer jogador marcou um golo de penalty, qual a probabilidade de esse jogador ser médio?

26. Sejam A e B acontecimentos independentes. Mostre que A e $\overline{B}$ são também acontecimentos independentes.

27. Um aluno conhece bem 60% da matéria dada. Num exame com cinco perguntas, sorteadas ao acaso, sobre toda a matéria, qual a probabilidade de vir a responder correctamente a duas perguntas?

28. Numa certa rua existem duas caixas Multibanco - A e B. A probabilidade de as máquinas avariarem é, independentemente uma da outra, de 0.05 para a A e 0.01 para a B. Determine a probabilidade de, num dia qualquer:

    (a) Ambas as máquinas estarem avariadas.

    (b) Apenas a máquina A estar avariada.

    (c) Pelo menos uma das máquinas estar avariada.

29. (Teste de P.E. 2006/07) Uma urna tem oito moedas, seis honestas e duas viciadas. O resultado do lançamento de uma moeda viciada é sempre "cara".

    (a) Escolhendo duas das oito moedas disponíveis, ao acaso e sem reposição, qual a probabilidade de seleccionar as duas moedas viciadas.

    (b) Escolhendo uma moeda ao acaso, qual a probabilidade de obter três caras em três lançamentos sucessivos dessa moeda?

    (c) Se em três lançamentos, da mesma moeda, o resultado foi sempre "cara", qual a probabilidade de ter escolhido a moeda viciada?

30. (Teste de Estatística - 2008/09) Um animal de uma dada população, quando exposto à temperatura de 35°C e humidade relativa de 75%, tem a probabilidade $2/5$ de morrer se for "são", mas de $4/5$ se estiver infectado com um vírus. Num laboratório, após se atingirem essas condições climatéricas, verificou-se que 50% dos animais estavam mortos. Qual é a proporção de animais infectados que existia na população inicial?

Capítulo 2

Variáveis aleatórias

1. A variável aleatória (v.a.) X representa o número de doentes com gripe que procuram, por dia, o Dr. Remédios. Em 50% dos dias, pelo menos 2 pacientes com gripe procuram o Dr. Remédios. A sua função de probabilidade é dada por:

$$X \left\{ \begin{array}{cccc} 0 & 1 & 2 & 3 \\ p & 0.2 & q & 0.3 \end{array} \right.$$

   (a) Determine p e q.

   (b) Determine a função de probabilidade das v.a.'s $Y = 40X$ e $W = \max(X, 1)$.

2. A v.a. X representa o número de pontos que saem no lançamento de um determinado dado. A sua função de distribuição segue-se:

$$F(x) = \left\{ \begin{array}{cc} 0, & x < 1 \\ 1/6, & 1 \leq x < 2 \\ 1/4, & 2 \leq x < 4 \\ 1/2, & 4 \leq x < 5 \\ 7/12, & 5 \leq x < 6 \\ 1, & x \geq 6 \end{array} \right.$$

   (a) Calcule as seguintes probabilidades, usando a função de distribuição:

      i) A probabilidade de o número de pontos saídos ser no máximo 3.

      ii) $P(1 < X \leq 2)$.

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

      iv) A probabilidade de o número de pontos saídos não distar de 2 pontos por mais de 1 ponto.

   (b) Determine a função de probabilidade de X e confirme os resultados acima obtidos.

   (c) Pode afirmar que o dado é equilibrado? Justifique.

   (d) Sabendo que o número de pontos saído é pelo menos 4, calcule a probabilidade de saírem 6 pontos.

3. O Sr. Matias possui um café nas vizinhanças de um estádio de futebol. Da sua experiência, o Sr. Matias sabe que, em dias de futebol, costuma vender ou 50, ou 100, ou 150 ou 200 sandes, com probabilidades 0.2, 0.4, 0.3 e 0.1, respectivamente.
O Sr. Matias costuma fazer 100 sandes e quando estas se esgotam recorre a um fornecedor da terra que lhe garante o envio atempado de mais sandes.

   (a) Qual a probabilidade de as sandes preparadas pelo Sr. Matias serem insuficientes para satisfazer a procura?

   (b) Calcule a probabilidade de vender 200 sandes, num dia em que as sandes por ele feitas não satisfazem a procura.

4. Seja X uma v.a. com a seguinte função densidade probabilidade:

$$f(x) = \begin{cases} k+x, & -1 \leq x < 0 \\ k-x, & 0 \leq x < 1 \\ 0, & c.c. \end{cases}$$

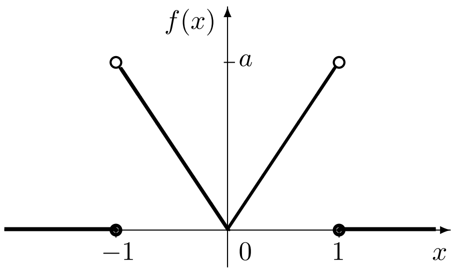
   (a) Determine o valor da constante k.

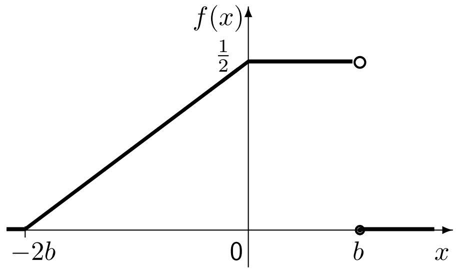
   (b) Determine $P(X > 0)$.

   (c) Determine $P(X > 0.5 | X > 0)$.

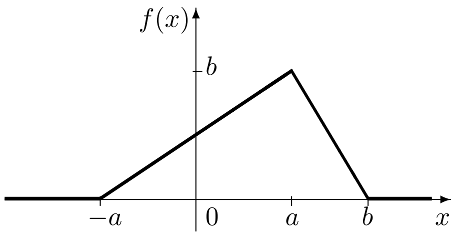
5. Considere funções densidade de probabilidade, representadas nos seguintes gráficos.

   (a) Determine o valor das constantes a e b.





   (b) Qual a relação entre a e b?



6. Seja X uma v.a. com a seguinte função densidade probabilidade:

$$f(x) = \begin{cases} 4x, & 0 < x < k \\ 0, & c.c. \end{cases}$$

   (a) Esboce o gráfico da função densidade e determine o valor da constante k.

   (b) Calcule $P(1/4 \leq X \leq 1/3)$, a mediana e o quantil de ordem 0.95.

7. A quantidade de tempo, em horas, que um computador funciona até avariar é uma v.a. com a seguinte função densidade probabilidade:

$$f(x) = \begin{cases} k\, e^{-\frac{x}{100}}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

   (a) Qual a probabilidade de o computador trabalhar entre 50 e 150 horas antes de avariar?

   (b) Qual a probabilidade de o computador funcionar menos de 100 horas até avariar? E exactamente 100 horas?

   (c) Qual a probabilidade de o computador avariar após 200 horas de funcionamento, sabendo que já funcionou mais de 100 horas?

8. (Exame de P.E. 2006/07) Seja X uma variável aleatória com função densidade

$$f(x) = \begin{cases} c(1+x), & -1 < x \leq 0; \\ c(1-\frac{x}{2}), & 0 < x < 2; \\ 0, & \text{outros valores de } x; \end{cases}$$

   (a) Mostre que $c = 2/3$.
   (b) Calcule $P(X \leq 0 | X \leq 1)$.

9. Determine o valor médio e a variância da variável aleatória discreta X com função de probabilidade: $P(X=0) = \frac{1}{8}$ $P(X=1) = \frac{3}{8}$ $f(2) = \frac{3}{8}$ $P(X=3) = \frac{1}{8}$. Calcule ainda:

$$E(g(X)), \text{ com } g(X) = X^3, \; E\left(\frac{1}{1+X}\right) \text{ e } E(X^2).$$

10. Seja X uma v.a. tal que $P(X=0) = \frac{1}{4}$, $P(X=1) = \frac{p}{2}$, $P(X=2) = \frac{5}{8} - \frac{p}{2}$ e $P(X=3) = \frac{1}{8}$, com $0 \leq p \leq \frac{1}{2}$. Determine p de forma a que $V(X)$ seja mínima.

11. Numa lotaria foram emitidos 10000 bilhetes. Sorteia-se 1 prémio de 25000 unidades monetárias (u.m.) e 10 prémios de 2500 u.m.. Seja X a v.a. que representa o valor do prémio de um bilhete qualquer.

   (a) Determine a função de probabilidade de X.
   (b) Qual a probabilidade de um bilhete não ter qualquer prémio?
   (c) Qual a probabilidade de um bilhete ter pelo menos 2500 u.m.?
   (d) Determine o $E(X)$, $V(X)$ e $CV(X)$.

12. Uma comissão de alunos está a organizar uma festa da faculdade. Os alunos vão comprar 200 litros de cerveja. Um fornecedor deste líquido (A) cobra 1 unidade monetária (u.m.) por litro permitindo a devolução da cerveja que sobrar (e que não tem de ser paga) e um outro fornecedor (B) cobra 0.5 u.m. por litro, não admitindo devoluções. Os alunos, independentemente de quanto lhes custe a cerveja, cobram 1.5 u.m. por litro.

    Sabendo que, se estiver bom tempo - o que acontecerá com probabilidade 0.8 - os alunos conseguem vender os 200 litros de cerveja, mas se estiver mau tempo só vendem metade, a quem devem comprar?

13. Seja X uma v.a. com a seguinte função de distribuição:

$$F(x) = \begin{cases} 0, & x < 0 \\ 1 - (x+1)e^{-x}, & x \geq 0 \end{cases}$$

   (a) Calcule $P(X < 1)$.
   (b) Calcule $P(X < 2 | X \geq 1)$.
   (c) Determine a função densidade probabilidade de X.
   (d) Dado que $\int_0^{+\infty} x^2 e^{-x} dx = 2 \int_0^{+\infty} x e^{-x} dx$ e $\int_0^{+\infty} x^3 e^{-x} dx = 3 \int_0^{+\infty} x^2 e^{-x} dx$, determine $E(X)$ e $V(X)$.

14. Determine $E(X)$, $E(X-1)$, $V(X)$ e $E(X(X-1))$ da v.a. X, que tem a seguinte função densidade probabilidade:

$$f(x) = \begin{cases} \frac{x}{2}, & 0 \leq x \leq 1 \\ \frac{1}{2}, & 1 < x \leq 2 \\ \frac{3-x}{2}, & 2 < x \leq 3 \\ 0, & x < 0 \vee x > 3 \end{cases}$$

15. A v.a. X tem a seguinte função densidade probabilidade:

$$f(x) = \begin{cases} k\sin(x), & 0 \leq x \leq \pi \\ 0, & c.c. \end{cases}$$

   (a) Determine o valor da constante k.
   (b) Dado que $\int_0^\pi x\sin(x)dx = \pi$ e $\int \cos(x)\sin(x)dx = \sin^2(x)/2 + c$, determine $E(X)$ e $E(\cos(X))$.
   (c) Sabendo que $V(X) = \frac{\pi^2}{4} - 2$, determine $E(X^2)$.
   (d) Determine $V(5X - 4)$.

16. (Teste de P.E. D - 2008/09) Uma empresa de produtos químicos fabrica um composto que vende em doses unitárias de 1 litro. Suponha que a fracção de álcool numa dose unitária do composto é uma variável aleatória X, com função densidade de probabilidade dada por:

$$f(x) = \begin{cases} kx^3(1-x), & 0 \leq x \leq 1; \\ 0, & \text{outros valores de } x; \end{cases}$$

   (a) Mostre que $k = 20$.
   (b) Sabendo que $E(X^2) = 10/21$, calcule $V(5X - 1)$.
   (c) Sabe-se que o custo de produção de uma dose de composto é sempre 0.8€/l, mas o seu preço de venda em €/l, V, depende do valor de X que lhe corresponde, sendo definido da seguinte forma:

$$V = \begin{cases} 1.2, & 1/3 \leq X \leq 2/3; \\ 1.0, & \text{outros valores de } X; \end{cases}$$

   Obtenha a função de probabilidade da variável aleatória, $L = V - 0.8$, o lucro obtido por cada dose daquele composto.

Capítulo 3

Principais distribuições

1. Um consumidor queixou-se às autoridades que no supermercado do Sr. Manuel se vendiam latas de cogumelos com o prazo de validade ultrapassado. No seguimento desta denúncia um inspector das actividades económicas dirigiu-se ao referido supermercado e seleccionou, ao acaso e sem reposição, 6 latas - do total de 50 que o Sr. Manuel ainda tinha para vender.
   Como na realidade ainda restavam 7 latas com o prazo de validade ultrapassado, qual a probabilidade de o Sr. Manuel ser multado (isto é, de o inspector descobrir pelo menos uma lata com o prazo ultrapassado)?

2. De forma a proceder a uma classificação geral do estado das praias Portuguesas, uma comissão Europeia vai inspeccionar 10 praias, seleccionadas ao acaso de entre as 100 existentes. A comissão atribui a classificação de Bom se pelo menos 8 das 10 praias inspeccionadas estiverem em bom estado. Sabendo que, da totalidade das 100 praias, 15 não apresentam boas condições, qual a probabilidade de Portugal:

   (a) Obter uma classificação de Suficiente, pelo facto da comissão só ter encontrado 7 praias em bom estado?
   (b) Obter uma boa classificação?
   (c) Se a comissão só inspeccionasse 5 praias, qual a probabilidade de não encontrar nenhuma em mau estado?
   (d) Nas praias inspeccionadas quantas se esperam que estejam em bom estado?

3. O senhor Sousa tem uma empresa que compra e vende selos e outros artigos de colecionismo. Ele guarda 20 selos dentro de uma bolsa preta, estando ainda cada um deles metido num envelope opaco. 6 destes selos valem 100 euros cada um e os restantes nada valem. O senhor Sousa, para promover a venda, cobra 20 euros por cada selo, mas não permitindo que o cliente veja o conteúdo do envelope. Suponha que um cliente compra 5 selos.

   (a) Qual a probabilidade dos cinco selos nada valerem?
   (b) Qual a probabilidade do cliente não perder nem ganhar dinheiro com a compra?

4. Num determinado percurso de avião, a probabilidade de uma pessoa qualquer que aí viaje pedir uma refeição vegetariana é de 0.2. Supondo que em determinado dia viajam 10 pessoas no avião, calcule a probabilidade de:

   (a) Ninguém pedir refeição vegetariana.
   (b) Todos pedirem refeição vegetariana.
   (c) Pelo menos uma pedir refeição vegetariana.

5. Determinado exame é constituído por 5 questões de escolha múltipla, em que cada questão tem 4 opções de resposta possíveis - apenas uma sendo a correcta. Supondo que um aluno que vai fazer

o exame responde a tudo ao acaso, qual é a probabilidade de ele acertar a mais de metade das questões? Qual é o número médio de respostas correctas? E o seu desvio padrão?

6. Sabe-se que 5% dos copos produzidos em determinada fábrica apresentam pequenos defeitos. Seleccionando-se da produção da fábrica, ao acaso, 50 copos, qual a probabilidade de:
   (a) Nenhum ser defeituoso?
   (b) Um ser defeituoso?
   (c) No máximo 1 ser defeituoso?
   (d) Calcule o número médio de copos defeituosos nesta amostra e o seu desvio padrão.
7. Na sala de aula de uma escola, 2 meninos lançam ao ar moedas equilibradas. O João faz 10 lançamentos e o Pedro 15. Qual a probabilidade de, no total dos lançamentos, saírem exactamente 12 caras?
8. O número de chamadas de emergência que um serviço de ambulâncias recebe por dia é uma v.a. de Poisson. Sabendo que a probabilidade de não haver nenhuma chamada num dia é de 0.15, calcule:
   (a) a probabilidade de haver apenas uma chamada num dia.
   (b) a probabilidade de haver 2 chamadas num dia.
   (c) a probabilidade de haver no máximo 3 chamadas num dia.
   (d) a probabilidade de haver pelo menos 4 chamadas num dia.
   (e) o número médio de chamadas por dia, o seu desvio padrão e coeficiente de variação.
9. Suponha que X é uma v.a. com distribuição de Poisson. Se $P(X = 2) = \frac{2}{3}P(X = 1)$, calcule $P(X = 0)$ e $P(X = 3)$.
10. Suponha que, o número de pessoas que utilizam uma caixa multibanco é um processo de Poisson de taxa $\lambda = 10$/hora. Calcule;
    (a) a probabilidade de não ir ninguém à caixa multibanco durante 1 hora.
    (b) a probabilidade de irem 20 pessoas à referida caixa durante 4 horas.
    (c) O número médio de visitas à caixa multibanco durante 4 horas e o seu coeficiente de variação.
11. Na portagem da ponte 25 de Abril o número de veículos automóveis que passa em cada cabine de pagamento da portagem, por minuto, segue uma distribuição de Poisson com valor médio 1 veículo. Supondo que em determinado minuto estão abertas 10 cabines, qual a probabilidade de serem, no total, atendidos 11 condutores nesse minuto?
12. Suponha que num livro de 500 páginas existem 300 erros tipográficos, distribuídos aleatoriamente por todo o livro. Assumindo que o número de erros segue uma distribuição de Poisson, determine a probabilidade de uma dada página conter:
    (a) 2 erros tipográficos.
    (b) Pelo menos 2 erros tipográficos
13. Um grande armazém de venda de material de vidro de laboratório emprega 100 pessoas. Tem-se verificado que o número de peças quebradas, por empregado e por mês, segue uma distribuição de Poisson de valor médio 1.5. Cada peça partida representa um prejuízo de 40 cêntimos, pelo que o armazém só arca com a despesa de um máximo de 3 peças por mês e por empregado. A partir deste valor é no salário do empregado que se desconta a despesa.
    (a) Qual a probabilidade de um empregado escolhido ao acaso ter de pagar do seu bolso algum prejuízo num qualquer mês?

(b) Considere agora a variável aleatória que representa o prejuízo do armazém, por mês e por empregado. Determine a sua função de probabilidade, qual é esse prejuízo médio e o seu desvio padrão.

14. Num posto dos correios o tempo (minutos) que a D. Hermínia demora a atender cada um dos seus clientes é uma v.a. exponencial de valor médio 3 minutos. Determine:

    (a) A função de distribuição de X.

    (b) A probabilidade de um cliente demorar mais de 5 minutos a ser atendido.

    (c) A probabilidade de um cliente demorar mais de 3 minutos a ser atendido.

    (d) A probabilidade de um cliente demorar mais de 5 minutos a ser atendido, sabendo que já está a ser atendido há pelo menos 2 minutos. Compare com a probabilidade anterior e comente.

    (e) O coeficiente de variação do tempo de atendimento.

15. Admita que os clientes chegam a uma loja de acordo com um processo de Poisson de média $\lambda = 2/minuto$. Calcule a probabilidade:

    (a) do tempo entre chegadas consecutivas ser superior a um minuto.

    (b) do tempo entre chegadas consecutivas ser inferior a quatro minutos.

    (c) do tempo entre chegadas consecutivas estar entre um e dois minutos.

16. Seja X uma v.a. com distribuição $N(100, 20^2)$. Calcule:

    (a) $P(X < 125)$.

    (b) $P(X > 115)$.

    (c) $P(60 < X < 140)$.

17. Seja X uma v.a. normal com média 12 e variância 2. Determine c tal que:

    (a) $P(X < c) = 0.1$.

    (b) $P(X > c) = 0.25$.

    (c) $P(12 - c < X < 12 + c) = 0.95$.

18. Admita que o Q.I. das pessoas de determinado país é uma v.a. X com distribuição normal de média 90 e desvio padrão 12. Determine:

    (a) A percentagem da população com Q.I. entre 85 e 95.

    (b) O valor $c > 0$ tal que a percentagem da população com Q.I. entre $90 - c$ e $90 + c$ seja de 95%.

    (c) 10000 pessoas desta população concorreram ao selecto clube SMART, que apenas admite indivíduos com Q.I. superior a 120. Quantas destas pessoas espera o clube vir a admitir?

19. A altura (metros) a que crescem os pinheiros é uma v.a. X normalmente distribuída com desvio padrão igual a 1 metro. Supondo que 90% dos pinheiros atingem uma altura de pelo menos 16 metros, qual a altura média dos pinheiros?

20. Numa fábrica de embalar arroz este trabalho é executado por uma máquina. A quantidade de arroz (Kg) que entra nos pacotes é uma v.a. X seguindo uma distribuição normal de valor médio μ e desvio padrão σ.

    (a) Determine σ sabendo que a quantidade embalada difere da sua média por menos de 100g, em 95 % dos casos.

    (b) Supondo que $\mu = 1Kg$, determine a probabilidade de, em 10 pacotes de arroz embalados por esta máquina, 2 terem menos de 0.9Kg.

21. Considere X uma v.a. Normal de valor médio 2 e variância 9. Seja I um intervalo do tipo $[4-a, a]$. Determine o valor de a de modo a que $P(X \in I) = 0.90$.

22. A altura (metros) a que um atleta salta é uma v.a. Normal de média 1.8m e desvio padrão 20cm. Sabendo que 20% das vezes o atleta consegue saltar acima de h, determine h.

23. Num jardim zoológico existem um leão e um tigre que consomem, independentemente um do outro, o mesmo tipo de alimentação - carne de 2ª. A quantidade de carne (Kg) que cada um deles come por dia são variáveis aleatórias, representadas por X_1 para o leão e X_2 para o tigre, respectivamente, normalmente distribuídas com média 4Kg e desvio padrão 0.5Kg. Determine a probabilidade de, num determinado dia:

    (a) Ambos os animais comerem menos de 3Kg de carne cada.
    (b) O leão comer mais do que o tigre.
    (c) Metade do que o leão come juntamente com $\frac{3}{4}$ do que o tigre come, exceder os 4Kg.

24. (Teste de PE 2006/07) Um elevador está preparado para suportar uma carga até 450 kg. Sempre que este valor é ultrapassado o elevador não funciona. Um estudo recente indica que o peso, das pessoas que utilizam esse elevador, é uma variável aleatória com distribuição Normal de valor médio 70 kg.

    (a) Sabendo que a probabilidade de uma pessoa (que utiliza o elevador) pesar menos de 60 kg é 0.0228, determine o desvio padrão desta variável aleatória.
    (b) Se entrarem 6 pessoas no elevador, qual a probabilidade de o elevador não funcionar devido ao excesso de peso?

25. (Teste de P.E. D 2005/06) Um foguete espacial é constituído por 3 partes distintas, *cápsula*, *corpo* e *depósitos*. Representem as v.a.'s X, Y e W o peso da cápsula, o peso do corpo do foguete e o peso dos depósitos, respectivamente, em toneladas. Sabe-se que $X \sim N(5, 1)$, $Y \sim N(10, 2^2)$ e $W \sim N(7, 2^2)$, sendo as três variáveis independentes entre si.

    (a) Qual a probabilidade de o peso da cápsula estar compreendido entre 3 e 7 toneladas?
    (b) Qual o peso h que o corpo do foguete ultrapassa em 2.5% das vezes?
    (c) Qual a probabilidade de o peso da cápsula mais o peso dos depósitos excederem o peso do corpo do foguete?

26. Suponha que X é uma v.a. com distribuição χ^2 com 10 graus de liberdade, $X \sim \chi^2_{10}$. Determine o valor de c, tal que:

    (a) $P(X \leq c) = 0.95$;
    (b) $P(X \leq c) = 0.05$.

27. Admita que X é uma v.a. com distribuição t com 14 graus de liberdade, $X \sim t_{14}$. Determine o valor de c, tal que:

    (a) $P(X \leq c) = 0.75$;
    (b) $P(X \leq c) = 0.05$;
    (c) $P(|X| > c) = 0.4$.

Capítulo 4

Teorema Limite Central

1. Numa loja de conveniência cada pessoa gasta, em média, 10€, com um desvio padrão de 3.75€. Qual a probabilidade de 100 clientes gastarem mais de 1100€, admitindo que os gastos são independentes de pessoa para pessoa?

2. O número de sismos no Japão, por mês, é uma v.a. com média 5 sismos e desvio padrão 2 sismos. Admitindo que os sismos são independentes entre si, determine a probabilidade de nos próximos 40 anos haver no máximo 2300 sismos.

3. Uma empresa vende caixas com biscoitos e, quando lhe é solicitado, envia-as pelo correio. Para evitar pesar estas caixas, cobra sempre o valor de portes de correio correspondente a admitir que qualquer caixa pesa 2508g.
   Cada caixa leva 100 biscoitos e o peso da embalagem plástica é desprezável.
   Se soubermos que o peso de cada biscoito é variável mas que em média pesa 25g com um desvio padrão de 8g, determine a probabilidade do valor pago em portes de correio com o envio de uma caixa ser inferior ao valor que pagaria, caso a caixa fosse pesada.

4. Ao adicionar números, um computador arredonda cada número para o inteiro mais próximo. Admita que os erros cometidos são v.a.'s independentes e identicamente distribuídas (i.i.d.) com valor médio igual a 0 e variância igual a $1/12$.
   Se 1200 números forem adicionados, qual a probabilidade aproximada de que o erro total cometido não ultrapasse 15.4?

5. Envelopes de avião são empacotados em grupos de 100, sendo depois pesados. Supondo que o peso de cada envelope é uma v.a. com valor médio igual a 1 grama e desvio padrão de 0.05 g, independentemente de envelope para envelope, determine:
   (a) a probabilidade de que um pacote, com exactamente 100 envelopes, pese mais de 100.5 g.
   (b) a probabilidade de que a média dos pesos dos 100 envelopes de um pacote, diste do respectivo valor médio por uma quantidade superior a 0.01g.

6. Numa determinada estufa de produção de tulipas vão-se semear 240 bolbos desta flor. Sabe-se que em média cada bolbo produz 4 flores, com um desvio padrão de 2 flores. Qual a probabilidade aproximada de se conseguir obter uma produção final de mais de 1000 tulipas? Justifique.

Capítulo 5

Estimação Pontual

1. Considere a população formada pelo número de filhos por família (X) num determinado país, em que $X = 0, 1, 2, 3, 4$ (não há famílias com mais de 4 filhos). Suponha que se conhece a sua distribuição de probabilidade:

$$X \left\{ \begin{array}{ccccc} 0 & 1 & 2 & 3 & 4 \\ 0.15 & 0.25 & 0.30 & 0.20 & 0.10 \end{array} \right.$$

   (a) Quais os valores populacionais de μ e σ^2?

   (b) Desta população recolhe-se uma amostra aleatória constituída por 2 famílias - (X_1, X_2). Qual a distribuição de probabilidade de X_1 e X_2 e os respectivos parâmetros μ e σ^2?

   (c) Suponha que recolheu a seguinte amostra aleatória de 10 famílias:

$$(1, 3, 0, 0, 2, 3, 0, 2, 4, 1).$$

   Com base nesta amostra estime pontualmente μ e σ^2. Estime ainda o erro padrão da estimativa de μ.

2. Considere que se seleccionou uma amostra aleatória $(X_1, X_2, \ldots, X_n)$ de uma população com valor médio μ e variância σ^2.

   (a) Mostre que $\overline{X} = \dfrac{\sum_{i=1}^{n} X_i}{n}$ é estimador centrado e consistente da média populacional.

   (b) Mostre que $\hat{\theta}_1 = \dfrac{X_1 + X_n}{2}$ e $\hat{\theta}_2 = \dfrac{2X_1 + 3X_2 + 5X_3}{10}$ também são estimadores centrados de μ. Qual é melhor? São consistentes?

   (c) Mostre que $(\overline{X})^2$ não é estimador centrado de μ^2.

3. Suponha que seleccionou uma amostra aleatória $(X_1, X_2, \ldots, X_n)$ de uma população com distribuição $U(0, \theta)$, isto é, com função densidade:

$$f(x) = \left\{ \begin{array}{ll} \frac{1}{\theta}, & 0 \leq x \leq \theta \\ 0, & c.c. \end{array} \right.$$

   (a) Verifique se o estimador $2\overline{X}$ é centrado e consistente.

   (b) Dada a amostra (1.215, 1.580, 0.726, 2.843, 3.394, 0.612, 2.621, 1.181, 2.930, 0.317), estime o valor de θ. Nota: $\sum x_i = 17.42$.

4. Considere a amostra aleatória $(X_1, X_2, \ldots, X_n)$ de uma população com distribuição $B(r, p)$, com r conhecido. Verifique se o estimador $\hat{p} = \frac{\overline{X}}{r}$ é centrado e consistente.

5. Sabe-se que a idade de determinada camada do subsolo segue uma distribuição Normal com média de 0.5 milhões de anos e um desvio padrão de 20000 anos. Seleccionadas ao acaso 10 amostras de subsolo calcule a probabilidade de a média amostral das suas idades ser superior a 490000 anos.

6. Considere uma amostra aleatória de dimensão 25, extraída de uma população Normal de média 100 e desvio padrão 10.

   (a) Qual a probabilidade de a média amostral cair no intervalo de $E(\overline{X}) - 1.96 \times SE(\overline{X})$ a $E(\overline{X}) + 1.96 \times SE(\overline{X})$?

   (b) Quanto deverá ser o tamanho amostral tal que a amplitude do intervalo definido em (a) diminua para 2.

7. O tempo de espera em pista para a descolagem de cada avião no aeroporto de Lisboa é uma v.a. com valor médio 4 minutos e desvio padrão 2.5 minutos. Suponha que se selecciona ao acaso 50 aviões, para se registarem os seus correspondentes tempos de espera. Calcule a probabilidade de a média dos tempos de espera exceder os 5 minutos.

8. Assuma que o número de ovos que as tartarugas verdes depositam nas praias, em cada desova, é uma v.a. de $Poisson$, com valor médio 15 ovos. Seleccionando ao acaso uma amostra de 100 tartarugas verdes, qual a probabilidade de que a média do número de ovos destas esteja compreendido entre o seu valor médio e $\pm$ 3 vezes o seu erro padrão.

9. Suponha que o tempo de vida de determinada espécie de burros é uma v.a. com distribuição exponencial, de valor médio 25 anos. Seleccionando ao acaso uma amostra de 40 burros desta espécie, qual a probabilidade de que a média dos seus tempos de vida seja inferior a 20 anos?

10. No país das Maravilhas a proporção de loucos é de 0.45. Suponha que se pretende seleccionar uma amostra aleatória de 500 habitantes deste país. Qual a probabilidade de a proporção de loucos que vão calhar na amostra exceder 0.5?

11. Numa população Normal de média desconhecida e desvio padrão 5 calcule a probabilidade de a variância de uma amostra aleatória de dimensão 20 dessa população estar compreendida entre 26 e 58.

Capítulo 6

Estimação por Intervalo de Confiança

1. Para avaliar o peso médio das maçãs produzidas por um determinado agricultor analisaram-se 20 maçãs seleccionadas ao acaso da produção. Estas resultaram num peso médio de $\bar{x} = 320\text{g}$. Assuma que os pesos das maçãs têm distribuição Normal com desvio padrão $\sigma = 20\text{g}$.

   (a) Construa um intervalo de confiança a 90% para a média do peso.

   (b) Qual deve ser o tamanho da amostra de forma a que a amplitude do correspondente intervalo de confiança a 90% para a média seja de 1g? E 5g? Comente.

2. A quantidade de combustível dispendido num percurso de Lisboa a Faro (em litros) é uma variável aleatória normal.

   (a) Assuma que em 8 viagens Lisboa-Faro seleccionadas ao acaso se verificou um gasto médio de combustível de 36 litros e um desvio padrão de 10 litros. Construa intervalos de confiança para a média a 90% e a 95% e compare-os.

   (b) Assuma agora que foi em 50 viagens Lisboa-Faro, seleccionadas ao acaso, que se verificou um gasto médio de combustível de 36 litros e um desvio padrão de 10 litros. Construa intervalos de confiança para a média a 90% e a 95% e compare com os anteriores. Comente.

3. O nível de poluição do ar de determinada cidade (medido em concentração de monóxido de carbono no ar) distribui-se normalmente. Recolheram-se os seguintes valores da referida concentração em 10 dias diferentes (em ppm): 0.09, 0.33, 0.01, 0.25, 0.20, 0.05, 0.03, 0.18, 0.13, 0.24. Com base nesta amostra determine um intervalo de confiança a 99% para a concentração média de monóxido de carbono na atmosfera.

4. A quantidade de gordura em 100g de carne de determinado tipo de vacas, medido em gramas, tem desvio padrão 8g. Qual deve ser o tamanho de uma amostra aleatória a seleccionar de forma a que a amplitude de um intervalo de confiança a 95% para a gordura média por 100g de carne seja inferior a 2.5g? Refira eventuais pressupostos que teve de fazer.

5. Construa um intervalo de confiança a 95% para a temperatura média de uma determinada sala de espera, com base numa amostra de temperaturas recolhidas em 35 dias diferentes que resultaram nos valores $\bar{x} = 22.1°C$ e $s = 3.2°C$.

6. A tensão (MegaPascal) suportada por uma determinada barra de aço é uma variável aleatória com desvio padrão igual a 30 MPa. Com base numa amostra aleatória de n tensões observadas, para as quais se verificou que $\sum x_i = 10000\text{MPa}$, construiu-se um intervalo de confiança a 95% para a tensão média suportada, cujo extremo superior era de 208.3MPa. Determine o extremo inferior do referido intervalo e diga quanto vale o n, assumindo que $n > 30$.

7. (Exame de P.E. D - 2008/09) A população das estaturas dos alunos da FCT, em metros, segue uma distribuição Normal. Recolheu-se a seguinte amostra aleatória de estaturas de 40 alunos desta faculdade:

1.79	1.80	1.72	1.82	1.57	1.78	1.78	1.66	1.78	1.80
1.75	1.74	1.60	1.77	1.82	1.82	1.75	1.66	1.84	1.77
1.78	1.78	1.69	1.78	1.52	1.72	1.84	1.65	1.71	1.79
1.76	1.70	1.63	1.71	1.70	1.64	1.59	1.63	1.74	1.71

correspondendo a uma média amostral de 1.73 e a um desvio padrão amostral de 0.08.

(a) Indique uma estimativa pontual, com base nesta amostra, para a verdadeira estatura média populacional.

(b) Deduza e calcule um intervalo de confiança a 92% para a estatura média populacional.

8. O tempo médio (segundos) de reacção de uma determinada raça de cães a um certo estímulo tem interesse para um determinado treinador. Assim ele resolveu testar 32 cães escolhidos aleatoriamente tendo observado $\bar{x} = 1.2s$ e $\sum (x_i - \bar{x})^2 = 15.5s^2$.

   (a) Construa um intervalo de confiança a 95% para o tempo médio de reacção dos cães.

   (b) Suponha que só se conseguiu obter uma amostra de 15 cães, tendo resultado em $\bar{x} = 1.1s$ e $\sum (x_i - \bar{x})^2 = 15.9s^2$. Construa, para este caso, um intervalo de confiança a 95% para o tempo médio de reacção dos cães, referindo eventuais pressupostos que tenha tido de fazer.

9. Numa fábrica de embalagem de queijo em fatias seleccionaram-se aleatoriamente 100 embalagens, das quais se verificaram que 18 tinham peso inferior ao suposto - sendo por isso inadequadas. Construa um intervalo de confiança a 98%para a verdadeira proporção de pacotes inadequados na produção total.

10. De 200 casos de pessoas com cancro do cólon, aleatoriamente detectadas, 12 morreram após 5 anos da detecção.

    (a) Estime pontualmente a probabilidade de uma pessoa que contraia o cancro do cólon morrer após 5 anos da sua detecção.

    (b) Quanto deveria aumentar ao tamanho da sua amostra aleatória de forma a que a largura do intervalo de confiança a 90% para a probabilidade considerada na alínea anterior fosse inferior a 0.01?

11. O tempo (horas) que o Pedro dispende em filas de trânsito, por dia, é uma v.a. Normal. Seleccionando aleatoriamente 15 dias registaram-se os seguintes valores de espera:

    $$1.5 \quad 1.0 \quad 1.0 \quad 2.0 \quad 1.5 \quad 1.25 \quad 1.0 \quad 2.0 \quad 1.5 \quad 1.25 \quad 1.75 \quad 0.5 \quad 1.0 \quad 1.5 \quad 1.25$$

    Determine um intervalo de confiança a 99% para a variância do tempo de espera.

12. Um profissional de bowling jogou 8 partidas num torneio, tendo obtido as seguintes pontuações:

    $$117.0 \quad 220.2 \quad 199.5 \quad 237.2 \quad 249.5 \quad 179.8 \quad 259.2 \quad 248.5$$

    Admitindo a normalidade das pontuações, construa um intervalo de confiança a 95% para a variância e para o desvio padrão (este último fornece uma medida da consistência da prestação do jogador).

Capítulo 7

Teste de Hipóteses

1. Uma fábrica de gelados afirma que a procura do gelado de chocolate no verão, por dia e em euros, é uma v.a. Normalmente distribuída com valor médio €200 e desvio padrão €40. Numa amostra aleatória constituída por 10 dias seleccionados ao acaso do período de verão verificou-se que $\bar{x} = 216$.

   (a) Teste, ao nível de significância 5%, se de facto o consumo médio de gelado de chocolate no verão é de €200 por dia.

   (b) Teste, ao ao nível de significância 5%, se de facto o consumo médio de gelado de chocolate no verão é menor do que €200 por dia.

   (c) Qual a potência do teste, da alínea anterior, se $\mu = 190$.

   (d) Resolva as duas primeiras alíneas usando o valor-p.

2. Um produtor de azeite afirma que a acidez média do seu azeite é de 0.9°. De forma a confirmar tal facto recolheu-se uma amostra aleatória da sua produção de azeite, tendo-se medido os seguintes valores de acidez: $0.9 \quad 0.8 \quad 0.7 \quad 1.1 \quad 0.9 \quad 0.9 \quad 1.0 \quad 0.7 \quad 1.5 \quad 1.1$. Admitindo a Normalidade da acidez do azeite:

   (a) Teste, ao nível de significância 1%, se o produtor tem razão.

   (b) Teste, ao ao nível de significância 1%, se a acidez média é superior a 0.9°.

3. Um biólogo pretende demonstrar que o peso médio de uma determinada espécie de coelhos - coelhos anões - é superior a 250g. Para tal seleccionou aleatoriamente 40 coelhos, tendo obtido uma média dos pesos de 255.3g e um desvio padrão de 30g. Teste ao nível de significância 10% se o biólogo está certo, assumindo a Normalidade dos pesos dos coelhos.

4. A Inês recebe, para além do seu salário, vencimento correspondente a 2 horas extra que devia fazer todos os dias. Contudo ela está desconfiada que tem andado a trabalhar, em média, mais do que 2 horas extra. Como a empresa onde trabalha regista sempre a hora de entrada e de saída dos seus funcionários, ela seleccionou aleatoriamente 12 dias de trabalho passados e registou os seguintes valores relativos ao horário extra: $\bar{x} = 2.3h$ e $s = 0.5h$. Admitindo a Normalidade do tempo extra de trabalho, teste a um nível de significância de 5%, se as suas suspeitas se confirmam.

5. Uma companhia de seguros tem previsto no seu orçamento um total de 5000€/dia para pagar as indemnizações dos seus segurados. De forma a confirmar se o valor médio das indemnizações pagas por dia está bem previsto seleccionaram-se, de anos anteriores, 100 dias, tendo-se verificado $\bar{x} = 5625$€ e $\sum(x_i - \bar{x})^2 = 6187500$€^{2}. Teste, ao nível de significância 5%, se a previsão se adequa.

6. Numa fábrica de massas embalam-se pacotes de esparguete que deveriam ter peso médio de 500g. O peso dos pacotes é uma v.a. Normal com variância $\sigma^2 = 225g^2$. De forma a confirmar o peso médio destes pacotes, seleccionaram-se ao acaso 40 embalagens que tinham um peso médio de 495g. Teste, ao nível de significância 1%, se o peso médio das embalagens é menor do que as 500g indicadas.

7. Seja X uma v.a. com distribuição Normal de valor médio μ e desvio padrão σ. A partir de uma amostra de dimensão 30, retirada da população, obtiveram-se os seguintes resultados:

$$\sum_{i=1}^{30} x_i = 64.0 \qquad \sum_{i=1}^{30} (x_i - \bar{x})^2 = 84.4$$

   (a) Teste, ao nível de significância 1%, as hipóteses $H_0 : \mu = 2 \ vs \ H_1 : \mu > 2$.
   (b) Admitindo que, para outra amostra de igual dimensão, o valor observado da estatística do teste da alínea anterior foi de 1.699, calcule o correspondente valor-p e tome uma decisão ao nível de significância de 10%.
   (c) Suponha que está a testar a hipótese $H_0 : \mu = 2$ contra a hipótese $H_1 : \mu = 2.5$ e que rejeita a hipótese nula se $\bar{X}_{30} > 2.3$. Calcule as probabilidades dos erros de 1ª e 2ª espécie do teste, se $\sigma = 1$.
   (d) Suponha agora que pretende testar as hipóteses $H_0 : \sigma^2 = 3 \ \ vs \ \ H_1 : \sigma^2 < 3$ e que se rejeita a hipótese nula se $S^2 < 1.3552$ para uma amostra de dimensão 30. Calcule a probabilidade do erro de 1ª espécie.

8. Numa operação stop da brigada de trânsito, de 120 camiões TIR que foram parados, 42 iam com excesso de peso. Com base nesta amostra aleatória, teste a hipótese de que a proporção deste tipo de camiões, que circulam nas nossas estradas em situação ilegal, ultrapassa os 30%. Use um nível de significância de 10%.

9. (Exame de P.E. D - 2008/09) A população das estaturas dos alunos da FCT, em metros, segue uma distribuição Normal. Recolheu-se a seguinte amostra aleatória de estaturas de 40 alunos desta faculdade:

1.79	1.80	1.72	1.82	1.57	1.78	1.78	1.66	1.78	1.80
1.75	1.74	1.60	1.77	1.82	1.82	1.75	1.66	1.84	1.77
1.78	1.78	1.69	1.78	1.52	1.72	1.84	1.65	1.71	1.79
1.76	1.70	1.63	1.71	1.70	1.64	1.59	1.63	1.74	1.71

   correspondendo a uma média amostral de 1.73 e a um desvio padrão amostral de 0.08. Teste a hipótese de que a verdadeira proporção de alunos com estatura superior ou igual a 1.82m nesta população é maior que 0.2. Use um nível de significância de 5%.

10. Determinada desordem genética no sangue pode ser prevista com base num teste de sangue muito simples. De forma a ter uma noção da proporção de pessoas que na população possam vir a ter esta desordem, testaram-se 100 pessoas, seleccionadas ao acaso, para as quais 14 testes deram positivo. Efectue um teste de hipóteses, usando um nível de significância 5%, sobre se percentagem de pessoas com tal desordem é inferior a 10%.

11. No fabrico de parafusos admite-se, relativamente aos seus comprimentos, uma variabilidade máxima de 0.5mm^2. Recolheu-se uma amostra aleatória de 20 parafusos que se verificou terem $s^2 = 0.3$. Admitindo a Normalidade do comprimento dos parafusos, teste, ao nível de significância de 5% se a especificação sobre a variabilidade do comprimento dos parafusos está a ser respeitada.

12. Com base na amostra aleatória seguinte, teste $H_0 : \sigma = 1.3 \ \ vs \ \ H_1 : \sigma \neq 1.3$, a um nível de significância de 1%: $2.0 \quad 3.2 \quad 5.0 \quad 1.8 \quad 3.4 \quad 2.6$

13. A resistência de um determinado metal é dito ter uma variabilidade inferior a 0.01 ohm^2. Teste esta hipótese, a um nível de significância 10%, usando a seguinte amostra aleatória de resistências medidas para este metal:

$$0.14,\ 0.138,\ 0.143,\ 0.142,\ 0.144,\ 0.137$$

14. A altura (em mm) da espuma de sabão numa bacia é importante para os fabricantes de detergentes e supõe-se que tem distribuição Normal. Foi efetuada uma experiência, colocando a mesma quantidade de detergente em 10 bacias de tamanho standard e, depois de uma certa agitação da água, mediu-se a altura da espuma, obtendo-se $\sum_{i=1}^{10}(x_i - \bar{x})^2 = 1332$.

(a) Teste, ao nível de significância 1%, a hipótese de o verdadeiro desvio padrão da população ser inferior a 20 mm.

(b) Calcule o valor-p do teste da alínea anterior e tome uma decisão ao nível de significância 10%.

Capítulo 8

Regressão Linear

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

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

$$\sum Y_i^2 = 1976.54 \quad \bar{Y} = 11.27143 \quad \sum x_i^2 = 12848 \quad \bar{x} = 30.14286 \quad \sum Y_i x_i = 4906.1$$

Considere o seguinte output do R para responder às questões:

```
> x<- c(29,28,29,35,26,25,32,31,34,27,33,33,32,28) #temperaturas
> Y<- c(10.5,9.0,10.4,18.6,5.5,5.2,11.6,10.4,17.8,9.9,13.7,14.2,12.3,8.7) #horas de utilização
> summary(lm(Y~x)) #sumário da regressão linear

Call:
lm(formula = Y~x)

 Residuals:
      Min         1Q     Median         3Q        Max
 -1.87517   -0.91932   -0.05554    0.54189    2.30895

Coefficients:
              Estimate   Std. Error   t value   Pr(>|t|)
 (Intercept)  -24.0267   3.6928       -6.506    2.91e-05   ***
 x            1.1710     0.1219       9.607     5.52e-07   ***
- - -
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.378 on 12 degrees of freedom
Multiple R-squared: 0.8849, Adjusted R-squared: 0.8753
F-statistic: 92.29 on 1 and 12 DF, p-value: 5.516e-07
```

(a) Disponha os dados em gráfico.

(b) Estime a recta de regressão linear simples. Refira quais os pressupostos efectuados. Desenhe-a no gráfico anterior.

(c) Estime a variância dos erros do modelo de regressão linear simples.

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

(e) Teste a hipótese de o verdadeiro declive da recta de regressão ser nulo. Comente o resultado à luz da alínea anterior.

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

(g) Deduza e calcule um intervalo de confiança a 90% para β_0.

(h) Determine um intervalo de confiança a 95% para a variância dos erros do modelo de regressão linear simples.

2. Pretende-se modelar a velocidade do vento Y, medida em Km/h, com a altitude x a que se faz a medição (m). Para tal registaram-se, para 9 valores de altitude, os correspondentes valores da velocidade do vento:

i	1	2	3	4	5	6	7	8	9
x_i	100	250	500	750	1000	1250	1500	1750	2000
Y_i	4	9	15	16	20	46	54	59	72

$$\sum Y_i^2 = 14675 \quad \bar{Y} = 32.78 \quad \sum x_i^2 = 12760000 \quad \bar{x} = 1011.11 \quad \sum Y_i x_i = 427900$$

Considere o seguinte output do R para responder às questões:

```
> x<-c(100,250,500,750,1000,1250,1500,1750,2000) #altitude
> Y<-c(4,9,15,16,20,46,54,59,72) #velocidade do vento
> summary(lm(Y~x)) #sumário da regressão linear

Call:
lm(formula = Y~x)

 Residuals:
     Min        1Q    Median      3Q      Max
 -12.3731   -0.6897   3.2048   3.9435   4.5214

Coefficients:
              Estimate    Std. Error   t value   Pr(> |t|)
 (Intercept)  -4.049017   4.023410     -1.006    0.348
 x            0.036422    0.003379     10.779    1.3e-05     ***
- - -
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.375 on 7 degrees of freedom
Multiple R-squared: 0.9432, Adjusted R-squared: 0.9351
F-statistic: 116.2 on 1 and 7 DF, p-value: 1.302e-05
```

(a) Ajuste um modelo de regressão linear simples aos dados. O que pode dizer sobre a qualidade do ajuste?

(b) Determine um intervalo de confiança a 95% para o verdadeiro declive da recta de regressão.

(c) Use o resultado da alínea anterior para testar a hipótese de que o verdadeiro declive da recta de regressão é nulo.

(d) Determine um intervalo de confiança a 90% para a variância dos erros do modelo de regressão linear simples.

(e) Determine o valor do resíduo da observação $(x_i; Y_i) = (100; 4)$.

3. (Exame de P.E. D - 2005/06) Pretende-se averiguar se existe uma relação directa entre a proximidade com campos de futebol da residência de casais e a taxa de divórcio. Assim registaram-se, em 5 locais seleccionados ao acaso, o correspondente número de estádios de futebol num raio de 50Km (x) e a respectiva taxa de divórcio por 1000 habitantes registada nessas localidades (Y):

Nº de campos de futebol, x_i	0	1	2	5	6
Taxa de divórcio (por 1000 habitantes), Y_i	2.2	2.5	3.5	4.1	4.8

$$\sum_{i=1}^{5} x_i = 14; \quad \sum_{i=1}^{5} x_i^2 = 66; \quad \sum_{i=1}^{5} Y_i = 17.1; \quad \sum_{i=1}^{5} Y_i^2 = 63.19; \quad \sum_{i=1}^{5} Y_i x_i = 58.8; \quad SQ_R = 0.2585075.$$

Considere o seguinte output do R para responder às questões:

```
> x<-c(0,1,2,5,6) #nº de campos de futebol
> Y<-c(2.2,2.5,3.5,4.1,4.8) #taxa de divórcios
> summary(lm(Y~x)) #sumário da regressão linear

Call:
lm(formula = Y~x)

 Residuals:
      Min         1Q    Median        3Q       Max
 -0.07910   -0.18657   0.40597  -0.21642   0.07612

Coefficients:
             Estimate  Std. Error  t value  Pr(>|t|)
 (Intercept) 2.2791    0.2060      11.063   0.00158    **
 x           0.4075    0.0567      7.186    0.00555    **
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2935 on 3 degrees of freedom
Multiple R-squared: 0.9451, Adjusted R-squared: 0.9268
F-statistic: 51.64 on 1 and 3 DF, p-value: 0.005553
```

(a) Ajuste uma recta de regressão linear a estes dados. Que pode dizer da qualidade do ajuste?

(b) Diga por suas palavras como interpreta o valor de $\hat{\beta}_1$ obtido.

(c) Teste a hipótese do verdadeiro valor declive da recta de regressão, β_1, ser nulo, a um nível de significância 10%. O resultado está de acordo com a qualidade do ajuste discutida em (a)?

(d) Numa localidade com 3 estádios de futebol na sua proximidade (menos de 50Km) quanto prevê que valha a correspondente taxa de divórcio?

4. (Exame de P.E. - 2006/07) Com o objectivo de estudar a qualidade do ar na região de Lisboa, pretende-se modelar a quantidade Y de Ozono troposférico (O_3), com a quantidade x de partículas em suspensão com diâmetro aerodinâmico inferior a 10 μm (PM_{10}). Para tal, registaram-se os seguintes dados:

x_i	60.5	78.8	89.8	80.9	74.8	49.9	97.5	92.5	36.5	18.1	29.6	15.9
y_i	124.2	158	177.1	185.6	179.2	145.7	163.7	188.8	122.2	75.4	94.8	80.3

$$\sum x_i = 724.8; \quad \sum x_i^2 = 53414.92; \quad S_{YY} = 18620.05; \quad \sum x_i y_i = 114890.35; \quad \hat{\sigma}^2 = 237.44;$$

Considere o seguinte output do R para responder às questões:

```
> x<-c(60.5, 78.8, 89.8, 80.9, 74.8, 49.9, 97.5, 92.5, 36.5, 18.1, 29.6,15.9) #particulas em suspensão
> Y<-c(124.2,158,177.1,185.6,179.2,145.7,163.7,188.8,122.2,75.4,94.8,80.3) #ozono
> summary(lm(Y~x)) #sumário da regressão linear

Call:
lm(formula = Y~x)

 Residuals:
      Min       1Q   Median       3Q      Max
  -25.719   -8.087   -2.747   13.419   19.253

Coefficients:
              Estimate   Std. Error   t value   Pr(>|t|)
 (Intercept)  62.829     10.473       5.999     0.000132   ***
 x            1.298      0.157        8.272     8.78e-06   ***
- - -
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 15.41 on 10 degrees of freedom
Multiple R-squared: 0.8725, Adjusted R-squared: 0.8597
F-statistic: 68.42 on 1 and 10 DF, p-value: 8.779e-06
```

(a) Ajuste um modelo de regressão linear simples aos dados. Refira quais os pressupostos do modelo.

(b) Comente a qualidade do modelo.

(c) Teste, ao nível de significância de 5%, a hipótese de o declive da recta de regressão ser nulo.

(d) Prove que qualquer recta dos mínimos quadrados passa por $(\overline{x}, \overline{y})$.

Capítulo 9

Soluções dos exercícios propostos

Introdução à Teoria da Probabilidade

1. 13800
2. 2300
3. 24
4. (a) 207360
   (b) 8709120
5. 3991680
6. 5040
7. 39916800
8. 6
9. (a) 28 (b) 112
10. (a) 66 (b) 11 (c) 220
    (d) 55 (e) 10
11. (a) 6708426560
    (b) 9279308324
    (c) 2264093964
    (d) 2570881764
12. (a) $n = 6$ (b) $n = 5$
    (c) $n = 3 \ \vee \ n = 14$
13. $1 - \frac{A_5^{11} \times 10!}{15!} = \frac{11}{13}$
14. (a)1/2 (b)1/3
15. (a) 2/7 (b) 1/7 (c) 6/7
16. 15/19
17. (a) 0.0053
    (b) 0.1261
18. $P(B) = 0.5$
19. $P(A \cup B \cup C) = 5/8$
20. $P(A - B) = 1/3$;
    $P(A \cup B) = 5/6$;
    $P(\bar{A} \cup \bar{B}) = 2/3$;
    $P(\bar{A} \cap B) = 1/6$;
    $P(A \cup \bar{B}) = 5/6$
21. (a) 0.7
    (b) 0.3
22. (a) 2/5
    (b) 1/5
    (c) 4/15
    (d) 7/15
23. (a) 0.5; 0.375
    (b) 0.55
24. (a) 0.475
    (b) 0.1474
25. (a) 7/15
    (b) 3/7
26.
27. 0.2304
28. (a) 0.0005
    (b) 0.0495
    (c) 0.0595
29. (a) 1/28
    (b) 11/32
    (c) 8/11
30. 25%

Variáveis aleatórias

1. (a) $p = 0.3$; $q = 0.2$

   (b)
   $$Y \left\{ \begin{array}{cccc} 0 & 40 & 80 & 120 \\ 0.3 & 0.2 & 0.2 & 0.3 \end{array} \right.$$

   $$W \left\{ \begin{array}{ccc} 1 & 2 & 3 \\ 0.5 & 0.2 & 0.3 \end{array} \right.$$

2. (a) **i)** 1/4
   ii) 1/12
   iii) 5/12

iv) $1/4$

(b)

$$X\left\{\begin{array}{ccccc} 1 & 2 & 4 & 5 & 6 \\ 1/6 & 1/12 & 1/4 & 1/12 & 5/12 \end{array}\right.$$

(c) Não.

(d) 5/9

3. (a) 0.4
   (b) 0.25
4. (a) k=1
   (b) 1/2
   (c) 1/4
5. (a) $a = 1$; $b = 1$
   (b) $(a + b)b = 2$
6. (a) $k = \sqrt{2}/2$
   (b) 7/72; $m_e = \chi_{0.5} = 0.5$; $\chi_{0.95} = 0.6892$
7. (a) 0.3834
   (b) 0.6321
   (c) 0.3679
8. (a)
   (b) 2/5
9. $E(X) = 3/2$; $V(X) = 3/4$; $E(X^3) = 27/4$; $E(\frac{1}{1+X}) = 15/32$; $E(X^2) = 3$
10. $V_p(X) = -\frac{p^2}{4} + \frac{p}{8} + \frac{63}{64}$, mínima para $p = 0$ ou $p = 1/2$
11. (a)

$$X\left\{\begin{array}{ccc} 0 & 2500 & 25000 \\ 0.9989 & 0.001 & 0.0001 \end{array}\right.$$

   (b) 0.9989
   (c) 0.0011
   (d) $E(X) = 5u.m.$, $V(X) = 68725u.m.^2$ e $CV(X) = 5243.09\%$
12. Sejam as v.a's L_A e L_B o lucro, em u.m., obtido através do fornecedor A e B respectivamente.
    $L_A\left\{\begin{array}{cc} 50 & 100 \\ 0.2 & 0.8 \end{array}\right.$ $L_B\left\{\begin{array}{cc} 50 & 200 \\ 0.2 & 0.8 \end{array}\right.$
    $E(L_A) = 90\ u.m.$ e $E(L_B) = 170\ u.m.$
    Devem comprar a B.
13. (a) 0.2642
    (b) 0.4482
    (c)

$$f(x) = \left\{\begin{array}{ll} 0, & x < 0 \\ xe^{-x}, & x \geq 0 \end{array}\right.$$

    (d) $E(X) = 2$; $V(X) = 2$.
14. $E(X) = 3/2$; $E(X - 1) = 1/2$; $V(X) = 5/12$; $E(X(X - 1)) = 7/6$
15. (a) $k = 1/2$
    (b) $E(X) = \pi/2$; $E(\cos(X)) = 0$
    (c) $\pi^2/2 - 2$
    (d) $25\pi^2/4 - 50$
16. (a)
    (b) $E(X) = 2/3$; $V(5X - 1) = 50/63$
    (c) $L\left\{\begin{array}{cc} 0.2 & 0.4 \\ 142/243 & 101/243 \end{array}\right.$

Principais Distribuições

1. 0.6164
2. (a) 0.1297
   (b) 0.8295
   (c) 0.4357
   (d) 8.5 praias
3. (a) 0.1291
   (b) 0.3874
4. (a) 0.1074.
   (b) 1.024×10^{-7}.
   (c) 0.8926.
5. 0.1035; 1.25 respostas; 0.9682 respostas
6. (a) 0.0769
   (b) 0.2025
   (c) 0.2794
   (d) 2.5 copos; 1.5411 copos
7. 0.1550 (soma de Binomiais independentes)
8. (a) 0.2846
   (b) 0.2699
   (c) 0.8752
   (d) 0.1248

(e) 1.8971 chamadas; 1.3774 chamadas

9. 0.2636; 0.1041

10. (a) 4.54×10^{-5}
    (b) 0.000192
    (c) 40; 15.81%

11. 0.1137

12. Assumindo distribuição Poisson:
    (a) 0.0988
    (b) 0.1219

13. (a) 0.0656
    (b) Seja Y o prejuízo do armazém, por mês e por empregado (€).
    $Y \left\{ \begin{array}{cccc} 0 & 0.4 & 0.8 & 1.2 \\ 0.2231 & 0.3347 & 0.2510 & 0.1912 \end{array} \right.$
    $E(Y) = 0.5641€$; $\sigma(Y) = 0.4139€$

14. (a)
$$F(x) = \left\{ \begin{array}{cc} 0, & x \leq 0 \\ 1 - e^{-x/3}, & x > 0 \end{array} \right.$$
    (b) 0.1889
    (c) 0.3679
    (d)] $0.3679 = P(X > 3)$ - Propriedade da falta de memória da exponencial
    (e) 100%

15. (a) 0.1353
    (b) 0.9997
    (c) 0.1170

16. (a) 0.8944
    (b) 0.2266
    (c) 0.9544

17. (a) 10.1898
    (b) 12.9475
    (c) 2.7719

18. (a) 32.56%
    (b) 23.52
    (c) 62 pessoas

19. 17.28m

20. (a) 0.051Kg
    (b) 0.023

21. 6.92

22. 1.968m

23. (a) 5.1984×10^{-4}
    (b) 0.5
    (c) 0.9868

24. (a) $\sigma = 5$
    (b) 0.0071

25. (a) 0.9544
    (b) $h = 13.92$
    (c) 0.7486

26. (a) 18.3
    (b) 3.94

27. (a) 0.692
    (b) -1.76
    (c) 0.868

Teorema Limite Central

1. 0.0038
2. 0.0113
3. 0.4602
4. 0.9382
5. (a) 0.1587
   (b) 0.0456
6. 0.0985

Estimação Pontual

1. (a) $\mu = 1.85$, $\sigma^2 = 1.4275$
   (b) X_1 e X_2 têm a mesma distribuição de probabilidade de X. Ambos têm média $\mu = 1.85$ e variância $\sigma^2 = 1.4275$.
   (c) $\hat{\mu} = \bar{x} = 1.6$, $\hat{\sigma}^2 = s^2 = 2.0(4)$, $\hat{SE}_{\bar{X}} = 0.4522$.

2. (a)

(b) $\hat{\theta}_2$ é melhor, pois tem menor Erro Quadrático Médio $(EQM(\hat{\theta}_2) = 0.38\sigma^2)$ que $\hat{\theta}_1$ $(EQM(\hat{\theta_1}) = 0.5\sigma^2)$. Não são consistentes.

(c) $E(\bar{X}^2) = \frac{\sigma^2}{n} + \mu^2 \neq \mu^2$.

3. (a) O estimador é centrado e consistente.

   (b) $\hat{\theta} = 3.484$.

4. É centrado $(E(\widehat{p}) = p)$ e consistente $(EQM(\hat{p}) = V(\hat{p}) \to 0$, quando $n \to \infty)$.

5. 0.9429.

6. (a) 0.95 .

   (b) 385.

7. 0.0023.

8. 0.9974.

9. 0.1038.

10. 0.0122 (distribuição aproximada).

11. $P(26 < S^2 < 58) = \cdots = F_{\chi^2_{19}}(44.08) - F_{\chi^2_{19}}(19.76) = 0.9991 - 0.5909 = 0.4082$, calculado com o software R, ou o valor aproximado 0.495, obtido com base na tabela da χ^2.

Estimação por Intervalo de Confiança

1. (a) $IC_{90\%}(\mu) \equiv (312.67; 327.33)$.

   (b) 4304; 173.

2. (a) $IC_{90\%}(\mu) \equiv (29.3002; 42.6998)$; $IC_{95\%}(\mu) \equiv (27.638; 44.362)$.

   (b) $IC_{90\%}(\mu) \equiv (33.6284; 38.3716)$; $IC_{95\%}(\mu) \equiv (33.1574; 38.8426)$.

3. $IC_{99\%}(\mu) \equiv (0.042; 0.26)$.

4. Assumo $n > 30$ - $n \geq 158$.

5. $IC_{95\%}(\mu) \equiv (21.04; 23.16)$.

6. 191.7; $n = 50$.

7. (a) $\widehat{\mu} = \overline{x} = 1.73$.

   (b) $IC_{92\%}(\mu) \equiv (1.708; 1.752)$.

8. (a) $IC_{95\%}(\mu) \equiv (0.955; 1.445)$

   (b) Assumo normalidade. $IC_{95\%}(\mu) \equiv (0.51; 1.69)$.

9. $IC_{98\%}(p) \equiv (0.09; 0.27)$.

10. **(a)** $\hat{p} = 0.06$.

    (b) $n = 6068$ (usando $\hat{p}$ da alínea anterior para estimar a variância de $\hat{P}$).

11. $IC_{99\%}(\sigma^2) \equiv (0.075; 0.573)$.

12. $IC_{95\%}(\sigma^2) \equiv (991.18; 9384.00)$; $IC_{95\%}(\sigma) \equiv (31.48; 96.87)$.

Teste de Hipóteses

1. (a) $H_0 : \mu = 200 \;\; vs \;\; H_1 : \mu \neq 200$; $R_{0.05} \equiv (-\infty; -1.96) \cup (1.96; +\infty)$; $z_{obs} = 1.26$; Não rejeitar H_0 a 5%.

   (b) $H_0 : \mu \geq 200 \;\; vs \;\; H_1 : \mu < 200$; $R_{0.05} \equiv (-\infty; -1.64)$; $z_{obs} = 1.26$; Não rejeitar H_0 a 5%. Os dados não evidenciam que o consumo médio de gelado de chocolate seja menor que 200€/dia.

   (c) 0.1977

   (d) (a) valor-p=0.2076 (b) valor-p=0.8962

2. **(a)** $H_0 : \mu = 0.9 \;\; vs \;\; H_1 : \mu \neq 0.9$; $R_{0.01} \equiv (-\infty; -3.25) \cup (3.25; +\infty)$; $t_{obs} = 0.802$; Não rejeitar H_0 a 1%.

   (b) $H_0 : \mu \leq 0.9 \;\; vs \;\; H_1 : \mu > 0.9$; $R_{0.01} \equiv (2.821; +\infty)$; $t_{obs} = 0.802$; Não rejeitar H_0 a 1%, pelo que os dados não evidenciam que a acidez média seja superior a 0.9.

3. $H_0 : \mu \leq 250 \;\; vs \;\; H_1 : \mu > 250$; $R_{0.10} \equiv (1.28; +\infty)$; $z_{obs} = 1.12$; Não rejeitar H_0 a 10%, pelo que os dados não indiciam que o biólogo tenha razão.

4. $H_0 : \mu \leq 2 \quad vs \quad H_1 : \mu > 2$; $R_{0.05} \equiv (1.796; +\infty)$; $t_{obs} = 2.08$; Rejeitar H_0 a 5%, pelo que os dados indicam que Inês parece ter razão.
5. $H_0 : \mu = 5000 \quad vs \quad H_1 : \mu \neq 5000$; $R_{0.05} \equiv (-\infty; -1.96) \cup (1.96; +\infty)$; $z_{obs} = 25$; Rejeitar H_0 a 5%, indicando os dados um desajuste no referido prémio médio.
6. $H_0 : \mu \geq 500 \quad vs \quad H_1 : \mu < 500$; $R_{0.05} \equiv (-\infty; -2.33)$; $z_{obs} = -2.108$; Não rejeitar H_0 a 5%, pelo que os dados não evidenciam que o peso médio dos pacotes seja inferior a 500g.
7. (a) $R_{0.01} \equiv (2.462; +\infty)$; $t_{obs} = 0.428$; Não rejeitar H_0 a 1%.
   (b) valor-$p = 0.05$; Rejeitar H_0 a 10%.
   (c) $\alpha =$P(erro 1º espécie)=0.0505; β=P(erro 2º espécie)=0.1357.
   (d) $\alpha =$0.005.
8. $H_0 : p \leq 0.3 \quad vs \quad H_1 : p > 0.3$; $R_{0.10} \equiv (1.28; +\infty)$; $z_{obs} = 1.20$; Não rejeitar H_0 a 10%, significando que os dados evidenciam que a proporção de camiões infractores ultrapasse os 30%.
9. $H_0 : p \leq 0.2 \quad vs \quad H_1 : p > 0.2$; $R_{0.05} \equiv (1.64; +\infty)$; $z_{obs} = -1.19$; Não rejeitar H_0 a 5%.
10. $H_0 : p \geq 0.1 \quad vs \quad H_1 : p < 0.1$; $R_{0.05} \equiv (-\infty; -1.64)$; $z_{obs} = 1.33$; Não rejeitar H_0 a 5%, significando que os dados evidenciam que a percentagem de possuidores desta desordem na população não é inferior a 10%.
11. $H_0 : \sigma^2 \leq 0.5 \quad vs \quad H_1 : \sigma^2 > 0.5$; $R_{0.05} \equiv (30.14; +\infty)$; $x^2_{obs} = 11.4$; Não rejeitar H_0 a 5%, pelo que a especificação parece estar a ser cumprida.
12. $H_0 : \sigma^2 = 1.3^2 \quad vs \quad H_1 : \sigma^2 \neq 1.3^2$; $R_{0.01} \equiv (0; 0.412) \cup (16.75; +\infty)$; $x^2_{obs} = 4.02$; Não rejeitar H_0 a 1%.
13. $H_0 : \sigma^2 \geq 0.01 \quad vs \quad H_1 : \sigma^2 < 0.01$; $R_{0.10} \equiv (0; 1.610)$; $x^2_{obs} = 0.0039$; Rejeitar H_0 a 10%, pelo que os dados evidenciam uma variabilidade inferior a 0.01.
14. (a) $H_0 : \sigma \geq 20$ vs $H_1 : \sigma < 20$; $R_{0.01} \equiv]0, 2.09[$; $x^2_{obs} = 3.33$; Não rejeitar H_0 a 1%.
    (b) valor-p=0.05; Rejeitar H_0 a 10%.

Regressão linear

1. (a)
   (b) $S_{xx} = 127.7143$; $S_{xy} = 149.5571$; $\hat{Y} = -24.027 + 1.171x$.
   (c) $S_{YY} = 197.9086$; $SQ_R = 22.77281$; $\hat{\sigma}^2 = 1.897734$
   (d) $R^2 = 0.8849$. Ajuste razoável.
   (e) $H_0 : \beta_1 = 0 \quad vs \quad H_1 : \beta_1 \neq 0$; $R_{0.05} \equiv (-\infty; -2.179) \cup (2.179; +\infty)$; $t_{obs} = 9.61$; Rejeitar H_0 a 5%, o que está de acordo com (d).
   (f) $\hat{Y}_{30} = 11.103h$. Para 40º não é possível estimar.
   (g) $t_{12:0.05} = 1.78$; $IC_{90\%}(\beta_0) \equiv (-30.60012; -17.45388)$.
   (h) $\chi^2_{12:0.025} = 23.3$; $\chi^2_{12:0.975} = 4.40$; $IC_{95\%}(\sigma^2) \equiv (0.9773737; 5.175638)$.
2. (a) $S_{xx} = 3558889$; $S_{xY} = 129622.2$; $\hat{Y} = -4.04902 + 0.03642x$; $S_{YY} = 5005.556$; $R^2 = 0.9432$. Ajuste razoável.
   (b) $IC_{95\%}(\beta_1) \equiv (0.0284; 0.0444)$.
   (c) Rejeitar H_0 a 5%.
   (d) $IC_{90\%}(\sigma^2) \equiv (20.17315; 131.079)$.
   (e) $Y_i - \widehat{Y}_i = 4.40702$.
3. (a) $\hat{Y} = 2.28 + 0.407x$; $R^2 = 0.9451$. Ajuste razoável.
   (b)
   (c) $H_0 : \beta_1 = 0 \quad vs \quad H_1 : \beta_1 \neq 0$; $R_{0.1} \equiv (-\infty; -2.35) \cup (2.35; +\infty)$; $\widehat{\sigma}^2 = 0.08616915$; $t_{obs} = 7.1778$; Rejeitar H_0 a 10%, o que está de acordo com (a).

(d) $\hat{Y}_3 = 3.501.$

4. (a) $\hat{Y} = 62.83 + 1.298x$
$(S_{xx} = 9637), \quad S_{xy} = 12512.35)$

(b) $R^2 = 0.87$

(c) $t_{obs} = 8.27$. Rejeitar H_0 a 5%.

(d)

Parte II

Exercícios Resolvidos

Capítulo 1

Introdução à Teoria da Probabilidade

1. Um centro comercial tem 8 portas. De quantas maneiras distintas se pode:

   (a) entrar e sair do centro comercial?

   (b) entrar por uma porta e sair por outra?

   Resolução:

   (a) Como temos 8 formas distintas de entrar (8 portas) e 8 formas distintas de sair temos $64 = 8 \times 8$, formas distintas de entrar e sair do centro comercial.

   (b) Para entrar continuamos a ter as 8 possibilidades de escolha (8 portas) mas quando se quiser sair apenas podemos optar por uma das 7 portas que não a utilizada para entrar, ou seja, temos $56 = 8 \times 7$ formas distintas.

2. Vinte e cinco membros de uma sociedade devem eleger um presidente, um secretário e um tesoureiro. Supondo que qualquer um dos 25 membros é elegível para qualquer dos cargos, quantas são as hipóteses distintas de eleição?

   Resolução:

   - Neste tipo de eleições não é usual a acumulação de cargos, pelo que teremos $25 \times 24 \times 23$ hipóteses de eleição, uma vez que para a eleição do presidente teremos 25 elementos disponíveis mas para a eleição do secretário já só teremos 24 membros disponíveis e finalmente quando for para eleger o tesoureiro a escolha está limitada aos 23 elementos que sobraram. Claro que $25 \times 24 \times 23 = \frac{25!}{22!} = \frac{25!}{(25-3)!} = A_3^{25}$ (arranjos) "= número de formas distintas de escolher 3 elementos de um universo de 25 quando a ordem pela qual são escolhidos faz diferença"(origina resultados diferente).
   - Para quem tiver curiosidade sobre o resultado se puder haver acumulação de cargos, teremos 25^3 hipóteses distintas uma vez que para cada cargo pode ser eleito qualquer dos 25 membros.

3. Quatro livros de Matemática, seis de Física e dois de Química, todos diferentes, devem ser arrumados numa prateleira. Quantas arrumações diferentes são possíveis, se:

   (a) os livros de cada matéria ficarem todos juntos?

   (b) os livros de Matemática devem ficar juntos?

   Resolução:

   (a) Podemos considerar primeiro cada matéria como um elemento pelo que o número de formas distintas de permutar as três matérias entre si é $3!$ (permutações de 3 elementos). De seguida temos de considerar que dentro de cada matéria os livros podem trocar de posições tendo-se que o número de formas distintas de colocar os 4 livros de Matemática é permutações de 4 elementos, ou seja, $4!$ e de modo análogo para a Física temos permutações de 6 elementos ($6!$) e para a Química $2!$. Portanto a resposta é $3! \times 4! \times 6! \times 2! = 207\,360$.

(b) Neste caso podemos começar por considerar que temos 9 elementos distintos, Matemática como um elemento e os outros 8 livros (Física e Química) como os restantes. Temos então permutações de 9 elementos originando $9!$ sequências distintas mas teremos ainda de considerar que os 4 livros de Matemática ao permutarem entre si (mantendo o bloco) originam mais $4!$ possibilidades, obtendo-se um total de $9! \times 4! = 8\,709\,120$ arrumações distintas.

4. Três amigos vão ao bar Pentagunus e pedem ao empregado de mesa que lhes sirva três bebidas diferentes. Este, na hora de servir as referidas bebidas, esquece-se de quem pediu o quê e decide colocar em frente de cada amigo uma bebida ao acaso. Qual a probabilidade de:

   (a) Todos receberem a bebida que efectivamente escolheram?

   (b) Ninguém receber a bebida correta?

   (c) Apenas um dos amigos receber a bebida que efectivamente escolheu?

   Resolução:

   (a) Repare-se que o número de formas diferentes do empregado dispor as três bebidas é $3!$ (permutações de 3 elementos) enquanto que apenas há uma forma de servir os pedidos corretamente. Logo a probabilidade pedida, p_1, segundo a lei de Laplace é: $p_1 = \frac{\text{número de casos favoráveis}}{\text{números de casos possíveis}} = \frac{1}{3!} = \frac{1}{6}$.

   (b) Continuamos com $3!$ casos possíveis mas agora temos 2 casos favoráveis à ocorrência do acontecimento indicado. O primeiro cliente tem duas formas de receber a bebida errada mas para os seus dois amigos apenas resta uma forma de distribuir as duas bebidas que faltam de modo a que nenhum fique com a bebida certa. Logo a probabilidade pedida, p_2, é: $p_2 = \frac{2}{3!} = \frac{1}{3}$.

   (c) Nesta situação o número de casos favoráveis é de 3 pois pode ser qualquer um dos três amigos aquele que recebe a bebida certa. Logo a probabilidade pedida, p_3, é: $p_3 = \frac{3}{3!} = \frac{1}{2}$.

5. Uma gaveta contém 10 meias verdes e 6 meias azuis. Tiram-se 2 meias ao acaso.

   (a) Se a extração for feita sem reposição, qual a probabilidade de se formar:

      i. um par verde?

      ii. um par de meias da mesma cor?

      iii. um par de meias de cores diferentes?

   (b) Se a extração for feita com reposição, qual a probabilidade de se formar:

      i. um par verde?

      ii. um par de meias da mesma cor?

      iii. um par de meias de cores diferentes?

   Resolução:

   (a) i. Repare-se que neste caso (e nos seguintes) o número de casos possíveis de tirar 2 meias da gaveta é C_2^{16} (combinações). Já o número de casos favoráveis é de C_2^{10}, que são as formas diferentes de tirar 2 meias verdes de entre as 10 que estão na gaveta. Vindo a probabilidade pedida $p_1 = \frac{C_2^{10}}{C_2^{16}} = 0.375$

      ii. Neste caso a probabilidade pedida será a de tirar um par de meias verdes ou azuis, logo como C_2^6 que são as formas diferentes de tirar 2 meias azuis de entre as 6 e na alínea anterior temos C_2^{10} para as meias verdes. A probabilidade pedida é $p_2 = \frac{C_2^{10}+C_2^6}{C_2^{16}} = 0.5$

      iii. Como o número de casos favoráveis é $C_1^{10}C_1^6$ (formas diferentes de tirar 1 meia verde de entre 10 e 1 meia azul de entre 6), teremos, $p_3 = \frac{C_1^{10}C_1^6}{C_2^{16}} = 0.5$

   (b) i. Repare-se que havendo reposição o número de casos favoráveis é de 10 em cada extração e o número de casos possíveis é de 16 em cada extração. Pelo que a probabilidade pedida será de $p_1 = \frac{10\times 10}{16\times 16} = 0.390625$.

ii. De forma análoga à alínea anterior, o número de casos favoráveis à obtenção de um par de meias azuis é de 6×6 pelo que a probabilidade pedida é igual à probabilidade de obtenção de um par verde ou um par azul, logo a probabilidade pedida será $p_2 = \frac{10\times10+6\times6}{16\times16} = 0.53125$.

iii. Como anteriormente, é fácil perceber que a probabilidade de se tirar uma meia verde seguida de uma azul é de $\frac{10\times6}{16\times16}$ que é igual à probabilidade de sair primeiro a azul e só depois a verde. Logo a probabilidade pedida é $p_3 = 2 \times \frac{10\times6}{16\times16} = 0.46875$.

6. Temos 15 livros para arrumar numa estante, em que 5 são de matemática. Qual a probabilidade de ficarem pelo menos 2 livros de matemática juntos?

Resolução:

Seja o acontecimento A-"ficarem pelo menos 2 livros de matemática juntos". Pretende-se calcular $P(A) = 1 - P(\bar{A})$ onde o complementar de A corresponde ao acontecimento "não haver livros de matemática juntos".

Para calcular o número de casos favoráveis ao acontecimento $\bar{A}$ vamos em primeiro lugar contabilizar o numero de disposições distintas dos livros de matemática de forma a não haver livros de matemática juntos. Assim, os 5 livros de matemática terão de ser colocados nas posições entre os outros livros ou nas duas posições dos extremos, tal como indicado em baixo:

$$\_\ \underline{O}\ \_\ \underline{O}\ \_\ \underline{O}\ \_\ \underline{O}\ \_\ \underline{O}\ \_\ \underline{O}\ \_\ \underline{O}\ \_\ \underline{O}\ \_\ \underline{O}\ \_\ \underline{O}\ \_$$

Neste caso, atribuímos os 11 lugares (posições) aos 5 livros de matemática, tendo em conta que não se repetem livros e que a ordem interessa. Tem-se assim A_5^{11} formas distintas de colocar os 5 livros de matemática de forma a não ficarem juntos, e como para cada organização dos livros de matemática temos todas as permutações dos restantes 10 livros, o número de casos favoráveis ao acontecimento $\bar{A}$ é dado por $A_5^{11} \times 10!$

Para os casos possíveis temos as permutações dos 15 livros, $15!$. Assim

$$P(A) = 1 - \frac{A_5^{11} \times 10!}{15!} = \frac{11}{13}$$

7. Considere os acontecimentos $A, B \in (\Omega, \mathcal{S}, P)$ tais que $P(A \cup B) = 0.8$ e $P(A - B) = 0.3$. Qual o valor da $P(B)$?

Resolução:

Repare-se que por

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \text{ e } P(A - B) = P(A) - P(A \cap B)$$

virá que

$$P(B) = P(A \cup B) - P(A) + P(A \cap B) = P(A \cup B) - P(A - B) = 0.8 - 0.3 = 0.5.$$

8. A execução de um projeto 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". Considerando os acontecimentos independentes e com probabilidades iguais a 0.8, 0.7 e 0.9, respetivamente, calcule a probabilidade de:

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

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

Resolução:

(a) Pretendemos calcular a probabilidade do acontecimento $(E \cap F \cap S)$, ou seja, a probabilidade das três atividades serem realizadas a tempo. Como os acontecimentos E, F e S são independentes, temos

$$P(E \cap F \cap S) = P(E)P(F)P(S) = 0.8 \times 0.7 \times 0.9 = 0.504.$$

(b) Tendo em conta os acontecimentos definidos no enunciado, o acontecimento "o prazo de execução ser cumprido para a escavação e não ser cumprido em pelo menos uma das outras atividades" pode traduzir-se por $\left(E \cap (\bar{F} \cup \bar{S})\right)$. Assim, por aplicação, em primeiro, da lei de Morgan e de seguida, o resultado da probabilidade para a diferença de acontecimentos, temos

$$P\left(E \cap (\bar{F} \cup \bar{S})\right) = P\left(E \cap \overline{F \cap S}\right) = P(E) - P(E \cap F \cap S) = 0.8 - 0.504 = 0.296.$$

9. Uma linha de produção em série é formada por três máquinas A, B e C, colocadas nesta ordem. O produto final resulta do processamento destas três máquinas. O funcionamento das máquinas B e C está dependente do funcionamento da(das) máquinas que a(as) antecedem. Sabemos que a probabilidade da máquina A sofrer uma avaria é de 0.1. Quando esta avaria, a máquina B pode sofrer uma avaria com probabilidade 0.3, e por sua vez, quando avariam em simultâneo as máquinas A e B, a máquina C deixa de operar com 0.5 de probabilidade. Determine a probabilidade de num determinado momento, se encontrarem avariadas as três máquinas.

   Resolução:

   Considerem-se respetivamente, A, B e C os acontecimentos: as máquinas A, B, C avariarem. Queremos determinar $P(A \cap B \cap C)$. Sabemos que

   $$P(A) = 0.1, P(B|A) = 0.3 \text{ e } P(C|A \cap B) = 0.5$$

   pelo que

   $$P(A \cap B \cap C) = P(C|A \cap B)P(A \cap B) = P(C|A \cap B)P(B|A)P(A) = 0.5 \times 0.3 \times 0.1 = 0.015.$$

10. Uma comissão investiga a queda de um avião, a qual só pode ser atribuída a uma das três seguintes causas:

    - C_1 - ataque deliberado com bomba ou míssil;
    - C_2 - falha técnica com explosão de combustível;
    - C_3 - ataque com míssil devido a falha humana.

    A experiência passada indica que a causa C_1 é cinco vezes mais provável do que a causa C_3 e que a causa C_2 é quatro vezes mais frequente que a causa C_3. Ao examinar os destroços, **foi confirmada** a presença de uma substância química que pode estar presente em qualquer das causas; mas, de acidentes anteriormente investigados, é sabido que a probabilidade de encontrar essa substância nos destroços é de 0.8 no caso C_1, de 0.9 no caso C_2 e de 0.95 no caso de C_3. Calcule a probabilidade de que o acidente tenha sido devido à causa C_2.

    Resolução:

    Consideremos o acontecimento E-"encontrar a substância química", e repare-se que temos:

    $$P(C_1) + P(C_2) + P(C_3) = 1, \quad P(C_1) = 5P(C_3), \quad P(C_2) = 4P(C_3)$$

    donde por substituição obtemos

    $$5P(C_3) + 4P(C_3) + P(C_3) = 1 \Leftrightarrow P(C_3) = 0.1,$$

    e portanto

    $$P(C_1) = 0.5, \quad P(C_2) = 0.4.$$

Por outro lado sabe-se que

$$P(E|C_1) = 0.8, \quad P(E|C_2) = 0.9, \quad P(E|C_3) = 0.95$$

como o pretendido é a probabilidade da causa do acidente ter sido C_2, sem esquecer que os destroços já foram examinados e a substância química foi encontrada, ou seja, o que se pretende é $P(C_2|E)$. Como

$$P(C_2|E) = \frac{P(C_2 \cap E)}{P(E)} = \frac{P(E|C_2)P(C_2)}{P(E)}$$

basta calcular $P(E)$ para termos todos os dados necessários. Como por utilização do teorema da probabilidade total podemos escrever

$$P(E) = P(E|C_1)P(C_1)+P(E|C_2)P(C_2)+P(E|C_3)P(C_3) = 0.8\times0.5+0.9\times0.4+0.95\times0.1 = 0.855.$$

Finalmente,

$$P(C_2|E) = \frac{0.9 \times 0.4}{0.855} = 0.421053.$$

Claro que podíamos ter utilizado diretamente o teorema de Bayes e escrever

$$P(C_2|E) = \frac{P(E|C_2)P(C_2)}{P(E|C_1)P(C_1) + P(E|C_2)P(C_2) + P(E|C_3)P(C_3)}.$$

11. Um supermercado tem para venda embalagens de café de três tipos diferentes: A, B e C. O número de embalagens de cada um dos tipos referidos é, respetivamente, 400, 500 e 100. Além disso o dono do supermercado sabe que:

    - das embalagens de café do tipo A, 5% estão fora do prazo de validade;
    - das embalagens de café que estão fora do prazo de validade, 1/5 contém café do tipo B;
    - não há embalagens de café do tipo C fora do prazo de validade.

    Escolhida ao acaso uma embalagem de café:

    (a) determine a probabilidade de que esteja fora do prazo de validade.

    (b) sabendo que é café do tipo B, determine a probabilidade de que esteja fora do prazo de validade.

    Resolução:

    (a) Considerem-se A, B, C-tipos de café e F-embalagem fora do prazo de validade. A partir dos dados temos que:

    $$P(A) = \frac{400}{1000} = 0.4; \quad P(B) = \frac{500}{1000} = 0.5; \quad P(C) = \frac{100}{1000} = 0.1$$

    e

    $$P(F|A) = 0.05; \quad P(B|F) = 1/5 = 0.2; \quad P(F|C) = 0.$$

    Pretende-se calcular $P(F)$, que pelo teorema da probabilidade total é dado por

    $$P(F) = P(F|A)P(A) + P(F|B)P(B) + P(F|C)P(C)$$

    mas não conhecemos $P(F|B)$, no entanto,

    $$P(F|B) = \frac{P(F \cap B)}{P(B)} = \frac{P(B|F)P(F)}{P(B)}$$

    pelo que

    $$\begin{aligned} P(F) &= P(F|A)P(A) + \frac{P(B|F)P(F)}{P(B)}P(B) + P(F|C)P(C) \\ &= P(F|A)P(A) + P(B|F)P(F) + P(F|C)P(C) = 0.05 \times 0.4 + 0.2 \times P(F) + 0 \times 0.1 \end{aligned}$$

    tendo-se finalmente

    $$P(F) = 0.02 + 0.2P(F) \Leftrightarrow P(F) = \frac{0.02}{0.8} = 0.025.$$

(b) Queremos $P(F|B)$, mas pelo que foi feito na alínea anterior teremos

$$P(F|B) = \frac{P(B|F)}{P(B)}P(F) = \frac{0.2}{0.5} \times 0.025 = 0.01.$$

12. Num escritório existem duas linhas de atendimento de chamadas, aqui designadas por linha A e linha B. A linha A é utilizada 60% das vezes. As probabilidades de serem atendidas 0,1 ou 2 chamadas por minuto na linha A são de 0.2, 0.4 e 0.4, respetivamente. A linha B, só tem capacidade de atendimento de uma chamada por minuto e a probabilidade de não serem atendidas chamadas num minuto é de 0.6.

    (a) Determine a probabilidade de, num minuto, ser atendida uma chamada neste escritório.

    (b) Se num minuto foi atendida uma chamada, qual a probabilidade de ter sido recebida pela linha A?

    (c) Os acontecimentos referidos na alínea anterior são independentes? Justifique.

    Resolução:

    (a) Considerem-se os acontecimentos A, B-chamada atendida pela linha A, B (respetivamente) e C_i-"i chamadas atendidas por minuto", $i = 0, 1, 2$. Dos dados temos

    $P(A) = 0.6$; $P(C_0|A) = 0.2$; $P(C_1|A) = 0.4$; $P(C_2|A) = 0.4$; $P(C_2|B) = 0$; $P(C_0|B) = 0.6$

    a partir dos dados anteriores também facilmente se conclui que $P(B) = 1 - P(A) = 0.4$ e $P(C_1|B) = 1 - (P(C_0|B) + P(C_2|B)) = 0.4$. Pretende-se calcular $P(C_1)$, que pelo teorema da probabilidade total será

    $$P(C_1) = P(C_1|A)P(A) + P(C_1|B)P(B) = 0.4 \times 0.6 + 0.4 \times 0.4 = 0.4.$$

    (b) Neste caso queremos

    $$P(A|C_1) = \frac{P(A \cap C_1)}{P(C_1)} = \frac{P(C_1|A)P(A)}{P(C_1)} = \frac{0.4 \times 0.6}{0.4} = 0.6.$$

    (c) Dois acontecimentos E e F dizem-se independentes se, e só se, $P(E \cap F) = P(E)P(F)$, mas como também é verdade que se $P(E|F) = P(E)$ então E e F são independentes, uma vez que

    $$P(E|F) = P(E) \Leftrightarrow \frac{P(E \cap F)}{P(F)} = P(E) \Rightarrow P(E \cap F) = P(E)P(F).$$

    Podemos agora concluir que A e C_1 são independentes pois pela alínea anterior verifica-se que $P(A|C_1) = 0.6 = P(A)$.

13. As probabilidades de dois corredores de velocidade percorrerem 100 metros em menos de 10 segundos são respetivamente: 1/2 e 1/4. Considerando que os tempos dos dois atletas são independentes, calcule a probabilidade de, uma corrida em que participam apenas os dois atletas, ser ganha em menos de 10 segundos.

    Resolução:

    Admitindo que os corredores são designados por atleta 1 e 2 respetivamente, considere os acontecimentos: A- o atleta 1 correr 100m em menos de 10 segundos; B- o atleta 2 correr 100m em menos de 10 segundos. Temos: $P(A) = 1/2$ e $P(B) = 1/4$. O acontecimento "a corrida ser ganha em menos de 10 segundos" verifica-se se e só se pelo menos um dos 2 atletas correr em menos de 10 segundos, ou seja, se e só se o acontecimento $A \cup B$ se verifica. Como os acontecimentos A e B são independentes, logo $P(A \cap B) = P(A)P(B) = 1/8$. Assim

    $$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.5 + 0.25 - 0.125 = 0.625$$

    ou

    $$P(A \cup B) = 1 - P(\overline{A} \cap \overline{B}) \overset{A(i)B}{=} 1 - P(\overline{A})P(\overline{B}) = 1 - 0.5 \times 0.75 = 0.625.$$

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

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

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

    (b) Se no lote A houver 50 bolbos dos quais 40 germinarão, qual a probabilidade de em 3 bolbos, escolhidos ao acaso desse lote, pelo menos um germinar?

    Resolução:

    Considere os acontecimentos: G - o bolbo germinar; A - o bolbo pertencer ao lote A; B - o bolbo pertencer ao lote B; C - o bolbo pertencer ao lote C. Temos: $P(G|A) = 0.8$; $P(G|B) = 0.85$ e $P(G|C) = 0.9$.

    (a) Escolhendo um lote ao acaso, tem-se $P(A) = P(B) = P(C) = 1/3$.

    i. Aplicando o teorema da probabilidade total, tem-se

    $$P(G) = P(G|A)P(A)+P(G|B)P(B)+P(G|C)P(C) = 0.8\times\frac{1}{3}+0.85\times\frac{1}{3}+0.9\times\frac{1}{3} = 0.85.$$

    ii. Aplicando o Teorema de Bayes

    $$P(C|\overline{G}) = \frac{P(\overline{G}|C).P(C)}{P(\overline{G})} = \frac{0.1\times\frac{1}{3}}{0.15} = \frac{2}{9}.$$

    (b) Considere o acontecimento: D - germinar pelo menos um dos três bolbos. Como não há repetição de elementos e a ordem não interessa, então pela lei de Laplace, tem-se

    $$P(D) = 1 - P(\overline{D}) = 1 - \frac{\binom{10}{3}}{\binom{50}{3}} = 0.9939.$$

15. O José entrou agora na Universidade e foi informado de que há 0.3 de probabilidade de vir a receber uma bolsa de estudo. No caso de a receber, a probabilidade de acabar o curso é de 0.85. Ainda se sabe que a probabilidade de não receber a bolsa e acabar o curso é de 0.32. Qual a probabilidade de que o José acabe o curso?

    Resolução:

    Considere os acontecimentos: R - vir a receber uma bolsa; A - acabar o curso. Temos: $P(R) = 0.3$; $P(A|R) = 0.85$ e $P(\overline{R}\cap A) = 0.32$.
    Aplicando o teorema da probabilidade total e de seguida o teorema da probabilidade composta, tem-se

    $$P(A) = P(A|R)P(R)+P(A|\overline{R})P(\overline{R}) = P(A|R)P(R)+P(A\cap\overline{R}) = 0.85*0.3+0.32 = 0.575.$$

16. Um fabricante de caixas de fósforo usualmente fornecidas em quartos de hotel, recebeu a dada altura um elevado número de queixas quanto à qualidade dos fósforos que fabrica. Numa rápida análise às condições de produção, verificou-se que 10% dos fósforos saem defeituosos tendo como consequência não se acenderem. Sabe-se ainda que cada caixa contém 15 fósforos.

    (a) Calcule a probabilidade de numa caixa acabada de se formar

    i. apenas o primeiro fósforo ser defeituoso;

    ii. haver apenas um fósforo defeituoso.

(b) Suponha agora que as caixas de fósforos são provenientes, em igual proporção, de duas linhas de fabrico, A e B, e que a percentagem de caixas aproveitáveis em cada linha é de 80% e 90%, respetivamente. Qual é a probabilidade de uma caixa escolhida ao acaso, da produção total das duas linhas, seja aproveitável?

Resolução:

(a) Considere os acontecimentos: D_i - que se verificam se e só se o fósforo i for defeituoso, $i = 1, ..., 15$.

i. Como os acontecimentos $D_i,\ i = 1, ..., 15$, são independentes, então

$$P(D_1 \cap \bar{D}_2 \cap \cdots \cap \bar{D}_{15}) = P(D_1)P(\bar{D}_2) \cdots P(\bar{D}_{15}) = 0.1 \times 0.9^{14} = 0.02287679.$$

ii.

$$P\left((D_1 \cap \bar{D}_2 \cap \cdots \cap \bar{D}_{15}) \cup (\bar{D}_1 \cap D_2 \cap \bar{D}_3 \cap \cdots \cap \bar{D}_{15}) \cup \cdots \cup (\bar{D}_1 \cap \cdots \cap \bar{D}_{14} \cap D_{15})\right) =$$
$$\stackrel{(1)}{=} P(D_1 \cap \bar{D}_2 \cap \cdots \cap \bar{D}_{15}) + P(\bar{D}_1 \cap D_2 \cap \bar{D}_3 \cap \cdots \cap \bar{D}_{15}) + \cdots + P(\bar{D}_1 \cap \cdots \cap \bar{D}_{14} \cap D_{15}) =$$
$$\stackrel{(2)}{=} P(D_1)P(\bar{D}_2) \cdots P(\bar{D}_{15}) + P(\bar{D}_1)P(D_2)P(\bar{D}_3) \cdots P(\bar{D}_{15}) + \cdots + P(\bar{D}_1) \cdots P(\bar{D}_{14})P(D_{15}) =$$
$$= 15 \times 0.1 \times 0.9^{14} = 0.3431519.$$

(1) Acontecimentos disjuntos.
(2) Acontecimentos independentes.

(b) Considere os acontecimentos: A - proveniente da linha A; B - proveniente da linha B; AP - caixa aproveitável. Temos: $P(A) = P(B) = 0.5$; $P(AP|A) = 0.8$ e $P(AP|B) = 0.9$. Aplicando o teorema da probabilidade total, temos

$$P(AP) = P(AP|A)P(A) + P(AP|B)P(B) = 0.8 \times 0.5 + 0.9 \times 0.5 = 0.85.$$

17. É possível ter acontecimentos A, B, e C disjuntos, dois a dois, tais que $P(A) = 0.3$, $P(B) = 0.4$ e $P(C) = 0.5$? Justifique.

Resolução:

Como A, B e C são acontecimentos disjuntos dois a dois, então pelo axioma 3 da teoria das probabilidades tem-se $P(A \cup B \cup C) = P(A) + P(B) + P(C) = 1.2 > 1$. Assim, não é possível ter aquelas probabilidades para acontecimentos disjuntos dois a dois.

18. A probabilidade de uma conexão elétrica, que se mantém seca, falhar durante o período de garantia de um computador portátil é de 1%. Se a conexão humedecer, a probabilidade de falha durante o período de garantia é de 5%. Sabemos ainda que 90% das conexões são mantidas secas e 10% humedecem.

(a) Determine a probabilidade de falha da conexão durante o período de garantia.

(b) Se uma conexão falhar durante o período de garantia, qual a probabilidade de ela se ter mantido seca.

Resolução:

Considere os acontecimentos: F - conexão elétrica falhar; S - conexão elétrica se manter seca. Temos:

$$P(S) = 0.9; \quad P(F|S) = 0.01; \quad P(F|\overline{S}) = 0.05.$$

(a) Então, aplicando o Teorema da Probabilidade Total

$$P(F) = P(F|S).P(S) + P(F|\overline{S}).P(\overline{S}) = 0.014.$$

(b) Aplicando o Teorema de Bayes, temos

$$P(S|F) = \frac{P(F|S).P(S)}{P(F)} = 0.6429.$$

19. Seja $\{E_1, E_2, E_3\}$ uma partição do espaço de resultados, Ω, com $P(E_i) > 0,\ i = 1, 2, 3$. Mostre que, para um qualquer acontecimento B, com $B \in \Omega$, se tem $P(B) = \sum_{i=1}^{3} P(B|E_i)P(E_i)$.

Resolução:

Como $\{E_1, E_2, E_3\}$ é partição de Ω, então $\Omega = E_1 \cup E_2 \cup E_3$ e $E_i \cap E_j = \emptyset,\ i \neq j,\ i, j = 1, 2, 3.$ Assim

$$P(B) = P(B \cap \Omega) = P(B \cap (E_1 \cup E_2 \cup E_3)) = P((B \cap E_1) \cup (B \cap E_2) \cup (B \cap E_3)) =$$
$$\stackrel{(1)}{=} P(B \cap E_1) + P(B \cap E_2) + P(B \cap E_3) \stackrel{(2)}{=} P(B|E_1)P(E_1) + P(B|E_2)P(E_2) + P(B|E_3)P(E_3).$$

(1) pelo axioma 3, pois $(B \cap E_i) \cap (B \cap E_j) = B \cap E_i \cap E_j = B \cap \emptyset = \emptyset,\ i \neq j,\ i, j = 1, 2, 3.$

(2) pelo teorema da probabilidade composta.

20. Um aluno tem um despertador que toca numa determinada hora, permitindo-lhe chegar a horas às aulas. O despertador toca na hora pretendida com probabilidade 0.7. Se tocar, a probabilidade do aluno chegar a horas à escola é 0.8, se não tocar, a probabilidade do aluno chegar a horas à escola é 0.3. Qual a probabilidade do aluno chegar a horas à escola?

Resolução:

Sejam os acontecimentos: C - chegar a horas à escola; T - o despertador tocar. Temos: $P(T) = 0.7; P(C|T) = 0.8; P(C|\bar{T}) = 0.3$. Aplicando o Teorema da Probabilidade Total, temos

$$P(C) = P(C|T)P(T) + P(C|\bar{T})P(\bar{T}) = 0.8 \times 0.7 + 0.3 \times 0.3 = 0.65.$$

Capítulo 2

Variáveis Aleatórias

1. De uma v.a. X, sabe-se que:

   - Tem por suporte o conjunto $\mathcal{D} = \{0, 2, 4\}$;
   - $P((X = 0) \cup (X = 2)) = 0.8$;
   - $P(X = 0) = \frac{3}{2}P(X = 4)$.

   (a) Indique a função de probabilidade da v.a. X.

   (b) Calcule a $P(0 < X < 4))$.

   (c) Determine a função de probabilidade da v.a. $Y = \min(X, 2)$.

   Resolução:

   (a) Repare-se que $P((X = 0) \cup (X = 2)) = 0.8 \Rightarrow P(X = 4) = 0.2$ e $P(X = 0) = \frac{3}{2}P(X = 4) \Rightarrow P(X = 0) = 0.3$ e portanto $P(X = 2) = 0.8 - 0.3 = 0.5$. Vindo a função de probabilidade de X

   $$X \left\{ \begin{array}{ccc} 0 & 2 & 4 \\ 0.3 & 0.5 & 0.2 \end{array} \right.$$

   (b) Basta ter em atenção que $P(0 < X < 4)) = P(X = 2) = 0.5$.

   (c) Tendo em atenção o suporte de X é fácil de concluir que o suporte de Y é $\mathcal{E} = \{0, 2\}$, como

   $$P(Y = 0) = P(\min(X, 2) = 0) = P(X = 0) = 0.3$$

   e

   $$P(Y = 2) = P(\min(X, 2) = 2) = P((X = 2) \cup (X = 4)) = 0.5 + 0.2 = 0.7.$$

   Teremos a função de probabilidade de Y,

   $$Y \left\{ \begin{array}{cc} 0 & 2 \\ 0.3 & 0.7 \end{array} \right.$$

2. O Sr. Matias possui um café nas vizinhanças de um estádio de futebol. Da sua experiência, o Sr. Matias sabe que, em dias de futebol, costuma vender 50, ou 100, ou 150 ou 200 sandes, com probabilidades 0.2, 0.4, 0.3 e 0.1, respetivamente. O Sr. Matias costuma fazer 100 sandes e quando estas se esgotam recorre a um fornecedor da terra que lhe garante o envio atempado de mais sandes. Todas as sandes são vendidas a 1 euro. Cada sandes feita pelo Sr. Matias custa 0.25 euros e as que são encomendadas ao fornecedor custam 0.65 euros.

   (a) Qual a probabilidade de as sandes preparadas pelo Sr. Matias serem insuficientes para satisfazer a procura?

(b) Calcule a probabilidade de vender 200 sandes, num dia em que as sandes por ele feitas não satisfazerem a procura.
(c) Qual o número médio de sandes vendidas num dia de futebol? E o desvio padrão?
(d) Deduza a função de probabilidade do lucro diário obtido pelo Sr. Matias.
(e) Determine o lucro médio diário. Expresse através do desvio padrão, a dispersão do lucro diário.

Resolução:

(a) Seja X a v.a. que representa a procura de sandes num dia de futebol. Então X tem a seguinte função de probabilidade

$$X \left\{ \begin{array}{cccc} 50 & 100 & 150 & 200 \\ 0.2 & 0.4 & 0.3 & 0.1 \end{array} \right.$$

A probabilidade pedida é $P(X > 100) = P(X = 150) + P(X = 200) = 0.4$.

(b) Queremos

$$P(X = 200|X > 100) = \frac{P((X = 200) \cap (X > 100))}{P(X > 100)} = \frac{P(X = 200)}{P(X > 100)} = \frac{0.1}{0.4} = 0.25.$$

(c) O valor médio é

$$E(X) = \sum_{i=1}^{4} x_i P(X = x_i) = 50 \times 0.2 + 100 \times 0.4 + 150 \times 0.3 + 200 \times 0.1 = 115 \text{ sandes.}$$

Sabemos que a variância é, por definição, $V(X) = E[(X - E(X))^2]$, mas podemos calcular a variância utilizando a fórmula

$$V(X) = E(X^2) - (E(X))^2.$$

Como,

$$E(X^2) = \sum_{i=1}^{4} x_i^2 P(X = x_i) = 50^2 \times 0.2 + 100^2 \times 0.4 + 150^2 \times 0.3 + 200^2 \times 0.1 = 15250,$$

o desvio padrão, $\sigma(X) = \sqrt{V(X)}$ será

$$\sigma(X) = \sqrt{E(X^2) - (E(X))^2} = \sqrt{15250 - 115^2} = 45 \text{ sandes.}$$

(d) Repare-se que o Sr. Matias tem um custo fixo de produzir 100 sandes, custo esse que é de $100 \times 0.25 = 25$euros. Se vender 50 sandes terá um lucro de $L = 50 - 25 = 25$; se vender 100 sandes terá um lucro de $L = 100 - 25 = 75$; se vender 150 sandes terá um custo extra de $50 \times 0.65 = 32.5$ pelo que o lucro será de $L = 150 - (25 + 32.5) = 92.5$; se vender 200 sandes terá um custo extra de $100 \times 0.65 = 65$ pelo que o lucro será de $L = 200 - (25 + 65) = 110$euros. Vindo a função de probabilidade da variável L

$$L \left\{ \begin{array}{cccc} 25 & 75 & 92.5 & 110 \\ 0.2 & 0.4 & 0.3 & 0.1 \end{array} \right.$$

(e) O lucro médio é

$$E(L) = \sum_{i} l_i P(L = l_i) = 25 \times 0.2 + 75 \times 0.4 + 92.5 \times 0.3 + 110 \times 0.1 = 73.75 \text{ euros/dia.}$$

Como,

$$E(L^2) = \sum_{i} l_i^2 P(L = l_i) = 25^2 \times 0.2 + 75^2 \times 0.4 + 92.5^2 \times 0.3 + 110^2 \times 0.1 = 6151.88,$$

e o desvio padrão

$$\sigma(L) = \sqrt{V(L)} = \sqrt{E(L^2) - (E(L))^2} = 26.6985 \text{ euros/dia.}$$

3. Um vendedor ambulante tem 8 relógios para vender, dos quais 3 estão avariados. Um cliente resolve comprar-lhe 4 relógios.

   (a) Determine a função de probabilidade de X, o número de relógios avariados comprados.

   (b) Qual a probabilidade do comprador adquirir relógios avariados e em número não superior a 2.

   (c) Determine o valor médio e a variância de X.

   Resolução:

   (a) Repare-se que X só pode tomar os valores $0, 1, 2$ ou 3, e que tendo em atenção a lei de Laplace

   $$P(X = k) = \frac{C_k^3 C_{4-k}^5}{C_4^8}, k = 0, 1, 2, 3.$$

   Pelo que a função de probabilidade da variável X será

   $$X \left\{ \begin{array}{cccc} 0 & 1 & 2 & 3 \\ \frac{C_0^3 C_4^5}{C_4^8} & \frac{C_1^3 C_3^5}{C_4^8} & \frac{C_2^3 C_2^5}{C_4^8} & \frac{C_3^3 C_1^5}{C_4^8} \end{array} \right. \quad \Leftrightarrow \quad X \left\{ \begin{array}{cccc} 0 & 1 & 2 & 3 \\ \frac{1}{14} & \frac{3}{7} & \frac{3}{7} & \frac{1}{14} \end{array} \right.$$

   (b) Note-se que o que é pedido é

   $$P(0 < X \leq 2) = P(1 \leq X \leq 2) = P(X = 1) + P(X = 2) = 6/7.$$

   (c) O valor médio é

   $$E(X) = \sum_i x_i P(X = x_i) = 0 \times \frac{1}{14} + 1 \times \frac{3}{7} + 2 \times \frac{3}{7} + 3 \times \frac{1}{14} = \frac{21}{14}.$$

   Como

   $$E(X^2) = \sum_i x_i^2 P(X = x_i) = 0^2 \times \frac{1}{14} + 1^2 \times \frac{3}{7} + 2^2 \times \frac{3}{7} + 3^2 \times \frac{1}{14} = \frac{39}{14},$$

   teremos a variância

   $$V(X) = E(X^2) - (E(X))^2 = 0.535714.$$

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

   $$P(X = x) = ax, \quad x = 1, 2, 3.$$

   (a) Mostre que $a = 1/6$.

   (b) Calcule a moda, a mediana, o valor médio e a variância de X.

   Resolução:

   (a) Como a função de probabilidade de X

   $$X \left\{ \begin{array}{ccc} 1 & 2 & 3 \\ a & 2a & 3a \end{array} \right.$$

   verifica as propriedades $P(X = x_i) \geq 0, \ \forall x_i \in D$ e $\sum_i P(X = x_i) = 1$, temos

   $$\sum_i P(X = x_i) = 1 \Leftrightarrow a + 2a + 3a = 1 \Leftrightarrow a = 1/6.$$

(b) A moda, m_o é o valor que maximiza a função de probabilidade desde que seja único, logo, tendo em conta a função de probabilidade de X, $m_o = 3$.

A mediana é o quantil de ordem $1/2$. Ora, $P(X \leq 2) = P(X = 1) + P(X = 2) = 1/2$, logo a mediana é $m_e = 2$.

O valor médio de X é

$$E(X) = \sum_i x_i P(X = x_i) = 1 \times \frac{1}{6} + 2 \times \frac{2}{6} + 3 \times \frac{3}{6} = \frac{7}{3}.$$

A variância de X é

$$V(X) = E(X^2) - E^2(X) = 6 - \left(\frac{7}{3}\right)^2 = \frac{5}{9},$$

onde

$$E(X^2) = \sum_i x_i^2 P(X = x_i) = 1^2 \times \frac{1}{6} + 2^2 \times \frac{2}{6} + 3^2 \times \frac{3}{6} = 6.$$

5. Seja X uma variável aleatória com função densidade:

$$f(x) = \begin{cases} x, & 0 < x < 1; \\ 2 - x, & 1 < x < 2; \\ 0, & \text{outros valores de } x. \end{cases}$$

(a) Mostre que se trata efetivamente de uma função densidade.
(b) Calcule $P(X < 1)$, $P(X = 1)$ e $P(X > 0.5|X \leq 1)$.
(c) Determine m tal que $P(X > m) = \frac{7}{8}$.
(d) Calcule o valor médio de X.

Resolução:

(a) f é uma função densidade se $f(x) \geq 0,\ \forall x \in \Re$ e $\int_{-\infty}^{+\infty} f(x)dx = 1$.
Como $x > 0$ para $0 < x < 1$, e $2 - x > 0$ para $1 < x < 2$ então $f(x) \geq 0,\ \forall x \in \Re$, e dado que

$$\int_{-\infty}^{+\infty} f(x)dx = \int_0^1 xdx + \int_1^2 2 - xdx = \left[\frac{x^2}{2}\right]_0^1 + \left[2x - \frac{x^2}{2}\right]_1^2 = \frac{1}{2} + \left(2 - \frac{3}{2}\right) = 1$$

logo, f é uma função densidade.

(b) $P(X < 1) = \int_{-\infty}^{1} f(x)dx = \left[\frac{x^2}{2}\right]_0^1 = \frac{1}{2}$; $P(X = 1) = 0$ pois X é uma v.a. contínua;

$$P(X > 0.5|X \leq 1) = \frac{P\left((X > 0.5) \cap (X \leq 1)\right)}{P(X \leq 1)} = \frac{P\left(0.5 < X \leq 1\right)}{P(X \leq 1)} = \frac{\int_{0.5}^{1} f(x)dx}{0.5} =$$

$$= \frac{\int_{0.5}^{1} xdx}{0.5} = \frac{\left[\frac{x^2}{2}\right]_{0.5}^1}{0.5} = \frac{0.375}{0.5} = 0.75.$$

(c) Como $P(X < 1) = 1/2$ e $P(X \leq m) = 1/8$ então $m < 1$, logo

$$P(X > m) = \frac{7}{8} \Leftrightarrow P(X \leq m) = \frac{1}{8} \Leftrightarrow \int_{-\infty}^{m} f(x)dx = \frac{1}{8} \Leftrightarrow$$

$$\Leftrightarrow \int_0^m xdx = \frac{1}{8} \Leftrightarrow \left[\frac{x^2}{2}\right]_0^m = \frac{1}{8} \Leftrightarrow \frac{m^2}{2} = \frac{1}{8} \Leftrightarrow m = -\frac{1}{2} \vee m = \frac{1}{2},$$

e dado que $P(X < -1/2) = \int_{-\infty}^{-1/2} f(x)dx = \int_{-\infty}^{-1/2} 0dx = 0$ então $m = 1/2$.

(d)

$$E(X) = \int_{-\infty}^{+\infty} xf(x)dx = \int_0^1 x^2 dx + \int_1^2 2x - x^2 dx = \left[\frac{x^3}{3}\right]_0^1 + \left[x^2 - \frac{x^3}{3}\right]_1^2 = \frac{1}{3} + \left(\frac{4}{3} - \frac{2}{3}\right) = 1.$$

6. Seja X uma variável aleatória com a seguinte função de distribuição:

$$F(x) = \begin{cases} 0 & x < -2 \\ 0.25x + 0.5 & -2 \leq x < 1 \\ 0.5x + 0.25 & 1 \leq x < 1.5 \\ 1 & 1.5 \leq x \end{cases}$$

(a) Calcule $P(X > 1)$ e $P(X \leq 0|X < 1)$.

(b) Determine o valor de a de forma a que $P(X > a) = 0.7$.

(c) Determine a função densidade de probabilidade de X. Calcule também o valor médio de X.

Resolução:

(a)

$$P(X > 1) = 1 - P(X \leq 1) = 1 - F(1) = 1 - (0.5 \times 1 + 0.25) = 0.25.$$

$$P(X \leq 0|X < 1) = \frac{P((X \leq 0) \cap (X < 1))}{P(X < 1)} = \frac{P(X \leq 0)}{P(X < 1)} = \frac{F(0)}{F(1^-)} = \frac{0.25 \times 0 + 0.5}{0.25 \times 1 + 0.5} = 0.6667.$$

(b) $P(X > a) = 0.7 \Leftrightarrow P(X \leq a) = 0.3 \Leftrightarrow F(a) = 0.3$.
Sabe-se que $F(x)$ é não decrescente, e como $F(-2) = 0$ e $F(1^-) = 0.75 > 0.3$ então $-2 \leq a < 1$, logo $F(a) = 0.3 \Leftrightarrow 0.25 \times a + 0.5 = 0.3 \Leftrightarrow a = -0.8$.

(c) Como $f(x) = F'(x)$, então

$$f(x) = \begin{cases} 0.25, & -2 \leq x < 1 \\ 0.5, & 1 \leq x < 1.5 \\ 0, & c.c. \end{cases}$$

e

$$E(X) = \int_{-\infty}^{+\infty} xf(x)dx = \int_{-2}^{1} x0.25dx + \int_1^{1.5} x0.5dx = -0.0625.$$

7. Seja X uma variável aleatória contínua com função densidade:

$$f(x) = \begin{cases} \frac{x^2}{9}, & 0 < x < 3 \\ 0, & \text{outros valores} \end{cases}$$

(a) Calcule o valor médio e a mediana de X.

(b) Seja Y a variável aleatória que representa a parte inteira de X. Indique, justificando, a função de probabilidade de Y.

Resolução:

(a) $E(X) = \int_{-\infty}^{+\infty} xf(x)dx = \int_0^3 \frac{x^3}{9}dx = \left[\frac{x^4}{36}\right]_0^3 = \frac{9}{4}$;
$m_e:\ P(X \leq m_e) = 0.5 \Leftrightarrow \int_0^{m_e} \frac{x^2}{9}dx = 0.5 \Leftrightarrow \left[\frac{x^3}{27}\right]_0^{m_e} = 0.5 \Leftrightarrow m_e \approx 2.3811$

(b) A função de probabilidade de Y é dada por:

$$Y \left\{ \begin{array}{ccc} 0 & 1 & 2 \\ \frac{1}{27} & \frac{7}{27} & \frac{19}{27} \end{array} \right.$$

onde:

$P(Y=0) = P(0 \leq X < 1) = \int_0^1 \frac{x^2}{9}dx = \left[\frac{x^3}{27}\right]_0^1 = \frac{1}{27}$;

$P(Y=1) = P(1 \leq X < 2) = \int_1^2 \frac{x^2}{9}dx = \left[\frac{x^3}{27}\right]_1^2 = \frac{7}{27}$;

$P(Y=2) = P(2 \leq X < 3) = \int_2^3 \frac{x^2}{9}dx = \left[\frac{x^3}{27}\right]_2^3 = \frac{19}{27}$.

8. O diretor de compras da empresa SOL, pretende definir uma política de aquisição de matéria prima. A necessidade de matéria prima por dia, expressa em toneladas, é representada por uma v.a. X contínua com função densidade de probabilidade:

$$f(x) = \left\{ \begin{array}{cc} 1 - \frac{x}{c} & 0 < x < c \\ 0 & \text{outros valores de } x \end{array} \right. , c > 0.$$

(a) Determine o valor de c.

(b) Calcule a probabilidade de num determinado dia, a necessidade de matéria prima não ultrapassar 1.5 toneladas?

(c) Qual a necessidade diária esperada de matéria prima?

(d) O diretor da fábrica considera que quando a necessidade diária de matéria prima é superior a 1.5 toneladas o sistema está em ruptura. A administração propôs dar-lhe um prémio de 20 u.m. por cada dia em que não houvesse ruptura, mas cobrar-lhe uma multa de 100 u.m. caso houvesse ruptura. Deduza a função de probabilidade do prémio diário.

Resolução:

(a) Sabemos que toda a função densidade de probabilidade deve verificar $f(x) \geq 0$ e também $\int_{-\infty}^{\infty} f(x)dx = 1$, pelo que

$$\int_{-\infty}^{\infty} f(x)dx = 1 \Leftrightarrow \int_0^c \left(1 - \frac{x}{c}\right) dx = 1 \Leftrightarrow \left[x - \frac{x^2}{2c}\right]_0^c = 1$$

$$c - \frac{c^2}{2c} = 1 \Leftrightarrow c - \frac{c}{2} = 1 \Leftrightarrow c = 2,$$

e repare-se que com $c = 2$ também se verifica $f(x) \geq 0$.

(b) Queremos determinar, $P(X \leq 1.5)$, mas

$$P(X \leq 1.5) = \int_{-\infty}^{1.5} f(x)dx = \int_0^{1.5} \left(1 - \frac{x}{2}\right) dx = \left[x - \frac{x^2}{4}\right]_0^{1.5} = 1.5 - \frac{1.5^2}{4} = 0.9375.$$

(c) Queremos $E(X)$ e sabemos que $E(X) = \int_{-\infty}^{\infty} x.f(x)dx$, donde

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_0^2 x\left(1 - \frac{x}{2}\right) dx = \left[\frac{x^2}{2} - \frac{x^3}{6}\right]_0^2 = \frac{4}{2} - \frac{8}{6} = \frac{2}{3}.$$

(d) Represente-se a v.a. prémio diário por Y. Como o suporte de Y é $\mathcal{D} = \{-100, 20\}$ virá a função de probabilidade de Y

$$Y \left\{ \begin{array}{cc} -100 & 20 \\ P(Y=-100) & P(Y=20) \end{array} \right. \Leftrightarrow Y \left\{ \begin{array}{cc} -100 & 20 \\ 0.0625 & 0.9375 \end{array} \right. ,$$

uma vez que

$$P(Y=20) = P(X \leq 1.5) = 0.9375 \text{ e } P(Y=-100) = 1 - P(Y=20).$$

9. A procura semanal (em toneladas) de um determinado produto é uma v.a. X contínua com função densidade de probabilidade:

$$f(x) = \begin{cases} \frac{x-1}{8}, & 4 < x < 6 \\ 0, & \text{restantes valores de } x \end{cases}.$$

   (a) Qual a probabilidade de numa semana a procura ser inferior a 5 toneladas?

   (b) Qual a quantidade mínima de produto a ter em stock de modo a que a procura seja satisfeita em pelo menos 90% das semanas?

   (c) Determine a procura semanal média do produto, assim como o respetivo desvio padrão, sabendo que $E\left[\left(X - \frac{61}{12}\right)^2\right] = \frac{47}{144}$.

Resolução:

(a) Queremos determinar, $P(X < 5)$, mas

$$P(X < 5) = \int_{-\infty}^{5} f(x)dx = \int_{4}^{5} \frac{x-1}{8}dx = \left[\frac{(x-1)^2}{16}\right]_4^5 = 1 - \frac{9}{16} = \frac{7}{16}.$$

(b) Queremos determinar q de modo que $P(X \leq q) = 0.9$, logo

$$P(X \leq q) = 0.9 \Leftrightarrow \int_4^q \frac{x-1}{8}dx = 0.9 \Leftrightarrow \left[\frac{(x-1)^2}{16}\right]_4^q = 0.9 \Leftrightarrow \frac{(q-1)^2 - 9}{16} = 0.9$$
$$\Leftrightarrow q^2 - 2q - 8 = 0.9 \times 16 \Leftrightarrow 10q^2 - 20q - 224 = 0 \Rightarrow q = 5.83735.$$

Observação: Note-se que o valor de q é o quantil de probabilidade 90% de X.

(c) Queremos $E(X)$ e sabemos que $E(X) = \int_{-\infty}^{\infty} xf(x)dx$, donde

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_4^6 x\left(\frac{x-1}{8}\right)dx = \frac{1}{8}\left[\frac{x^3}{3} - \frac{x^2}{2}\right]_4^6 = \frac{61}{12}.$$

Mas com este resultado podemos escrever que

$$E\left[\left(X - \frac{61}{12}\right)^2\right] = \frac{47}{144} \Leftrightarrow E\left[(X - E(X))^2\right] = \frac{47}{144} \Leftrightarrow V(X) = \frac{47}{144},$$

pelo que o desvio padrão de X será

$$\sigma(X) = \sqrt{V(X)} = \sqrt{\frac{47}{144}}.$$

Capítulo 3

Principais Distribuições

1. A Rádio Eletrão quer vender rapidamente os 30 computadores portáteis que tem em armazém, pelo que realizou uma promoção com descontos atrativos e oferecendo a pré-instalação do sistema operativo. Infelizmente, o processo de instalação do sistema operativo não é completamente fiável e 10 dos portáteis necessitarão de assistência complementar. Suponha que uma empresa comprou 20 portáteis e considere X o número de portáteis com problemas, de entre os comprados.

   (a) Qual a distribuição de X?

   (b) Determine a probabilidade de menos de 3 portáteis necessitarem de assistência complementar.

   (c) Determine a probabilidade de mais de 6 portáteis necessitarem de assistência complementar.

   Resolução:

   (a) Repare-se que X contabiliza o número de computadores com problemas numa amostra de 20 unidades, retirada (sem reposição) de uma população de 30 computadores dos quais 10 têm problemas. Uma v.a. com estas características terá distribuição Hipergeométrica de parâmetros 30 (dimensão da população), 10 (número de elementos da população que têm a característica em análise) e 20 (dimensão da amostra). Pelo que pode-se escrever

   $$X \sim H(N = 30, M = 10, n = 20) \Leftrightarrow P(X = k) = \frac{C_k^{10} C_{20-k}^{20}}{C_{20}^{30}}, k = 0, 1, 2, ..., 10.$$

   (b) Queremos

   $$P(X < 3) = P((X = 0) \cup (X = 1) \cup (X = 2)) = \frac{C_0^{10} C_{20}^{20}}{C_{20}^{30}} + \frac{C_1^{10} C_{19}^{20}}{C_{20}^{30}} + \frac{C_2^{10} C_{18}^{20}}{C_{20}^{30}} = 0.000291263.$$

   (c) Queremos

   $$P(X > 6) = P(X = 7) + P(X = 8) + P(X = 9) + P(X = 10) =$$
   $$= \frac{C_7^{10} C_{13}^{20}}{C_{20}^{30}} + \frac{C_8^{10} C_{12}^{20}}{C_{20}^{30}} + \frac{C_9^{10} C_{11}^{20}}{C_{20}^{30}} + \frac{C_{10}^{10} C_{10}^{20}}{C_{20}^{30}} = 0.560339.$$

   (d) Sabe-se que (ver formulário)

   $$E(X) = \frac{n.M}{N} = \frac{20 \times 10}{30} = 6.666(7),$$

   e que

   $$\sigma(X) = \sqrt{V(X)} = \sqrt{\frac{n.M.(N - M)(N - n)}{N^2(N - 1)}} = \sqrt{\frac{20 \times 10 \times 20 \times 10}{30^2 \times 29}} = 1.23797.$$

2. Uma determinada praga atacou uma unidade agrícola tendo contaminado três quartos da sua produção de maçãs. Considere 4 maçãs escolhidas ao acaso desta produção.

   (a) Identifique a distribuição da v.a. X que contabiliza o número de maçãs contaminadas, entre as 4 escolhidas, e apresente a respetiva função de probabilidade.

   (b) Determine:

      i. A probabilidade de todas elas terem sido contaminadas.

      ii. A probabilidade de nenhuma ter sido contaminada.

      iii. A probabilidade de terem sido contaminadas menos de 3 maçãs.

   (c) Determine o valor médio e a variância de X.

   (d) Suponha que se recolheram ao acaso, duas amostras de maçãs, uma com 4 e outra com 3 maçãs. Determine a probabilidade de, no conjunto das duas amostras, se encontrarem 2 maçãs contaminadas. Identifique a distribuição para o total de maçãs contaminadas, no conjunto das duas amostras.

   Resolução:

   (a) Repare-se que X contabiliza o número de maçãs contaminadas numa amostra de 4 unidades, onde a probabilidade de cada maça estar contaminada é de $3/4$ e pode-se assumir que o resultado (contaminada ou não) de cada maçã é independente dos restantes. Assim sendo, estamos perante a contabilização do número de sucessos (maça contaminada) na realização de 4 provas independentes (extração da maça) todas com a mesma probabilidade de sucesso $(3/4)$. Uma v.a. com estas características terá distribuição Binomial de parâmetros $n = 4$ (número de provas) e $p = 3/4$ (probabilidade de sucesso). Pelo que pode-se escrever,

   $$X \sim B(n = 4, p = 0.75) \Leftrightarrow P(X = k) = C_k^4 0.75^k (1 - 0.75)^{4-k}, k = 0, 1, 2, 3, 4.$$

   (b) i. O pedido é

   $$P(X = 4) = C_4^4 0.75^4 (1 - 0.75)^0 = 0.316406.$$

      ii. Neste caso

   $$P(X = 0) = C_0^4 0.75^0 (1 - 0.75)^4 = 0.00390625.$$

      iii. Queremos

   $$\begin{aligned} P(X < 3) = P(X \le 2) = P(X = 0) + P(X = 1) + P(X = 2) = \\ = C_0^4 0.75^0 0.25^4 + C_1^4 0.75^1 0.25^3 + C_2^4 0.75^2 0.25^2 = 0.261719. \end{aligned}$$

   (c) Sabe-se que (ver formulário)

   $$E(X) = n.p = 4 \times 0.75 = 3,$$

   e que

   $$V(X) = n.p.(1 - p) = 4 \times 0.75 \times 0.25 = 0.75.$$

   (d) Sejam X a v.a. anterior e Y o número de maçãs contaminadas numa segunda amostra, então $Y \sim B(n_Y = 3, p = 0.75)$. Como X e Y são independentes e ambas têm distribuição Binomial com a mesma probabilidade de sucesso, sabe-se que

   $$X + Y \sim B(n + n_Y = 7, p = 0.75),$$

   pelo que

   $$P(X + Y = 2) = C_2^7 0.75^2 (1 - 0.75)^5 = 0.0115356.$$

3. O número de automóveis, X, que chegam por dia a uma pequena oficina para serem reparados é uma variável aleatória de Poisson com parâmetro igual a 2. Devido à reduzida dimensão da oficina, só podem ser atendidos, no máximo, 3 automóveis por dia. Se chegarem mais de 3 automóveis, os excedentes são encaminhados para outras oficinas.

(a) Determine a probabilidade de, num dia qualquer, serem encaminhados automóveis para outras oficinas.

(b) Em quanto deverão ser alargadas as instalações, de modo a que a oficina possa reparar, em aproximadamente 95% dos dias, todos os automóveis que chegam?

(c) Qual o número esperado de automóveis que chegam por dia?

(d) Determine a função de probabilidade para o número de automóveis atendidos diariamente.

(e) Qual o número médio de automóveis atendidos diariamente?

(f) Qual é o número médio de automóveis que são diariamente encaminhados para outras oficinas?

(g) Determine a probabilidade de, em cinco dias, chegarem 9 automóveis.

Resolução:

(a) Como
$$X \sim P(2) \Leftrightarrow P(X=k) = \frac{e^{-2} \times 2^k}{k!}, k = 0, 1, 2, ...$$
e serão encaminhados automóveis para outra oficina se chegarem mais de 3, o que se pretende é
$$P(X > 3) = 1 - P(X \leq 3) = 1 - (P(X=0) + P(X=1) + P(X=2) + P(X=3)) =$$
$$= 1 - \left(e^{-2} + 2e^{-2} + \frac{e^{-2}2^2}{2!} + \frac{e^{-2}2^3}{3!}\right) = 0.142877.$$

(b) Queremos determinar k de modo que $P(X \leq k) \approx 0.95$. Como pela alínea anterior temos que $P(X \leq 3) = 0.857123$ e por outro lado $P(X=4) = \frac{e^{-2}2^4}{4!} = 0.0902235$ virá que
$$P(X \leq 4) = P(X \leq 3) + P(X = 4) = 0.947347 \approx 0.95$$
pelo que bastará aumentar a capacidade para permitir um atendimento diário de 4 automóveis.

(c) Sabe-se que (ver formulário)
$$X \sim P(2) \Rightarrow E(X) = 2.$$

(d) Seja Y o número de automóveis atendidos diariamente, então Y tem suporte $\mathcal{D} = \{0, 1, 2, 3\}$ e como
$$P(Y=0) = P(X=0), P(Y=1) = P(X=1), P(Y=2) = P(X=2), P(Y=3) = P(X \geq 3)$$
uma vez que o número de automóveis atendidos coincide com o número de automóveis que chegam à oficina nos dias em que esse número é menor que 3, nos dias em que chegam 3 ou mais automóveis só 3 serão atendidos. Vindo a função de probabilidade de Y
$$Y \left\{ \begin{array}{cccc} 0 & 1 & 2 & 3 \\ 0.1353 & 0.2707 & 0.2707 & 0.3233 \end{array} \right.$$
Repare-se que ao arredondar as probabilidades é necessário ter em atenção que a soma final deve valer 1.

(e) O valor médio de Y é
$$E(Y) = 0 \times 0.1353 + 1 \times 0.2707 + 2 \times 0.2707 + 3 \times 0.3233 = 1.782.$$

(f) Seja Z o número de automóveis encaminhados para outras oficinas, é fácil de perceber que, $Z = X - Y$ donde, utilizando propriedades do valor médio, virá
$$E(Z) = E(X - Y) = E(X) - E(Y) = 2 - 1.782 = 0.218.$$

(g) Sejam X_i o número de automóveis que chegam à oficina no dia i, com $i = 1, 2, ..., 5$ então $X_i \sim P(2)$, $i = 1, ..., 5$. Fazendo $S_5 = \sum_{i=1}^{5} X_i$, podemos escrever a probabilidade pretendida como $P(S_5 = 9)$ mas sabemos que $S_5 \sim P(5 \times 2) = P(10)$. Portanto

$$P(S_5 = 9) = \frac{e^{-10}10^9}{9!} = 0.12511.$$

4. Entre os 64 programadores de uma empresa, 48 são do sexo masculino. A empresa decide sortear 10 programadores para fazer um curso avançado de programação. Qual o número esperado de programadores do sexo masculino que são sorteados? Justifique.

Resolução:

Seja X - número de programadores do sexo masculino em 10. Dado que os 10 programadores são selecionados sem reposição, então $X \sim H(N, M, n)$ com $N = 64$ (dimensão da população), $M = 48$ (número de programadores do sexo masculino) e $n = 10$ (dimensão da amostra). Logo, $E(X) = n\frac{M}{N} = 10 \times \frac{48}{64} = 7.5$.

5. Uma aposta simples do concurso Euromilhões consiste em inscrever cinco números do conjunto $N = \{1, 2, ..., 50\}$ e duas estrelas do conjunto $E = \{1, 2, ..., 9\}$ num boletim de apostas. Em cada concurso são sorteados 5 números do conjunto N e duas estrelas do conjunto E, ganhando o 1º Prémio qualquer aposta que contenha estes números e estas estrelas.

(a) Qual a probabilidade de ganhar um 1º Prémio com uma aposta simples?

(b) Qual a probabilidade de ganhar um 1º Prémio com uma aposta múltipla, em que se inscrevem 8 números e 4 estrelas?

(c) Se 200 pessoas fizerem uma aposta simples, qual a probabilidade de 2 delas ganharem um 1º Prémio?

Resolução:

(a) Como a ordem dos elementos não interessa e não há repetições então o número de casos possíveis é dado por $\binom{50}{5}\binom{9}{2} = 76275360$. Assim, a probabilidade de ganhar um 1º prémio com uma aposta simples é $1/76275360 = 1.311039 \times 10^{-08}$.

(b) A probabilidade de ganhar um 1º prémio com uma aposta múltipla de 8 números e 4 estrelas é igual a probabilidade de ganhar um 1º Prémio com o número de apostas simples distintas que é possível formar com 8 números e 4 estrelas.

O número de casos favoráveis, correspondente ao número de apostas simples que é possível formar com 8 números e 4 estrelas, é $\binom{8}{5}\binom{4}{2} = 336$

Assim, tendo em conta os casos possíveis obtidos na alínea anterior, a probabilidade de ganhar um 1º Prémio com uma aposta múltipla de 8 números e 4 estrelas é $336/76275360 = 4.405092 \times 10^{-06}$.

(c) Seja X a v.a. que representa o número de pessoas, em 200, que ganham um 1º Prémio tendo feito uma aposta simples. Como se tratam de provas de Bernoulli independentes, dado que as pessoas apostam de forma independente, e sendo a probabilidade de sucesso para cada pessoa (prova) igual a probabilidade de ganhar um 1º Prémio com uma aposta simples, calculada na alínea a), então pode-se concluir que $X \sim Bin(200, 1/76275360)$. Assim

$$P(X = 2) = \binom{200}{2}\left(\frac{1}{76275360}\right)^2\left(1 - \frac{1}{76275360}\right)^{200-2} \approx 3.42 \times 10^{-12}.$$

6. Um estudante de Engenharia que vai passar férias ao Algarve decidiu deslocar-se à boleia. Para esse efeito, colocou-se à entrada da auto-estrada. Sabe-se que o número de automóveis que entram na auto-estrada num intervalo de t minutos, $N(t)$, se distribui segundo um processo de Poisson de taxa $\lambda = 2$ por minuto.

(a) Qual a probabilidade de não entrar qualquer automóvel durante 1 minuto?

(b) Será demasiado grande o valor de 0.9 para a probabilidade de em 2 minutos entrar algum automóvel?

(c) Indique a distribuição da v.a. T que expressa o tempo (em minutos) entre passagens dos automóveis.

(d) Calcule a probabilidade do estudante ter de esperar menos de 2 minutos até à chegada do primeiro automóvel.

(e) A probabilidade de qualquer automóvel parar para dar boleia ao estudante é 0.2. Se X representar o número de automóveis que entram na auto-estrada até que o primeiro pare para dar boleia ao estudante:

Resolução:

(a) Sabe-se que se $N(t)$ é um processo de Poisson de parâmetro λ então teremos

$$P(N(t)=k)=e^{-\lambda t}\frac{(\lambda t)^k}{k!}, k=0,1,...$$

No nosso caso temos $\lambda=2$ e queremos nesta alínea

$$P(N(1)=0)=e^{-2\times 1}\frac{(2\times 1)^0}{0!}=e^{-2}=0.135335.$$

(b) Como a probabilidade de em 2 minutos entrar algum automóvel é

$$P(N(2)>0)=1-P(N(2)=0)=1-e^{-2\times 2}\frac{(2\times 2)^0}{0!}=1-e^{-4}=0.981684$$

pelo que 0.9 não será demasiado grande.

(c) Sabe-se que por a entrada dos automóveis (por minuto) na auto-estrada seguir um processo de Poisson de parâmetro $\lambda=2$, então a v.a. T que expressa o tempo entre passagens consecutivas de automóveis terá distribuição Exponencial de parâmetros $\lambda=2$, ou seja, $T\sim Exp(2)$.

(d) Queremos calcular $P(T<2)$, pelo que foi dito na alínea anterior sobre T e pelo que conhecemos da distribuição Exponencial (ver formulário), teremos

$$P(T<2)=\int_0^2 2e^{-2t}dt=\left[-e^{-2t}\right]_0^2=1-e^{-4}=0.981684.$$

Observação: Repare-se que esta probabilidade é igual à obtida no cálculo realizado na alínea b), pois são equivalentes os acontecimentos; "em 2 minutos entrar algum automóvel"e "esperar menos de 2 minutos até à chegada do primeiro automóvel".

7. Uma máquina de encher garrafas de água mineral foi calibrada para deitar uma média de 1.5 litros em garrafas com uma capacidade nominal de 1.55 litros. Sabe-se ainda que o volume de água despejado é normalmente distribuído com um desvio padrão de 30 ml. Determine:

(a) A percentagem de garrafas que contêm menos de 1.52 litros.

(b) A probabilidade de uma garrafa conter entre 1.48 e 1.52 litros.

(c) O valor de c tal que a percentagem de garrafas com um volume de água entre $1.5-c$ e $1.5+c$ litros, seja de 0.95.

(d) O número esperado de garrafas, das próximas 1000 a serem enchidas, em que a máquina consegue engarrafar uma quantidade de água superior á capacidade.

(e) O volume de água abaixo do qual se encontra a fração de 25% das garrafas menos cheias.

Resolução:

(a) Repare-se que tendo-se, $X \sim N(1.5, 0.03^2)$, para o cálculo de probabilidades associadas com esta v.a. teremos de centrar e reduzir a v.a. (ver teorema 4.32, pag. 34), uma vez que só temos tabela para a Normal reduzida. Queremos então $P(X < 1.52)$, pelo que

$$P(X < 1.52) = P\left(\frac{X - 1.5}{0.03} < \frac{1.52 - 1.5}{0.03}\right) = P\left(Z < \frac{2}{3}\right) = \Phi\left(\frac{2}{3}\right) = 0.7475,$$

onde $Z = \frac{X-1.5}{0.03} \sim N(0,1)$ e Φ representa a função distribuição da $N(0,1)$.

(b) Queremos

$$P(1.48 < X < 1.52) = P\left(-\frac{0.02}{0.03} < \frac{X - 1.5}{0.03} < \frac{0.02}{0.03}\right) = P\left(Z < \frac{2}{3}\right) - P\left(Z \leq -\frac{2}{3}\right)$$
$$= \Phi\left(\frac{2}{3}\right) - \Phi\left(-\frac{2}{3}\right) = 2\Phi\left(\frac{2}{3}\right) - 1 = 0.4950,$$

onde usámos a simetria em torno do ponto zero da $N(0,1)$ para deduzir que $\Phi(-z) = 1 - \Phi(z)$.

(c) Queremos determinar c tal que, $P(1.5 - c < X < 1.5 + c) = 0.95$, como

$$P(1.5 - c < X < 1.5 + c) = 0.95 \Leftrightarrow P\left(-\frac{c}{0.03} < \frac{X - 1.5}{0.03} < \frac{c}{0.03}\right) = 0.95$$
$$\Leftrightarrow \Phi\left(\frac{c}{0.03}\right) - \Phi\left(-\frac{c}{0.03}\right) = 0.95 \Leftrightarrow 2\Phi\left(\frac{c}{0.03}\right) - 1 = 0.95 \Leftrightarrow \Phi\left(\frac{c}{0.03}\right) = 0.975$$
$$\Leftrightarrow \frac{c}{0.03} = \Phi^{-1}(0.975) \Leftrightarrow \frac{c}{0.03} = 1.96 \Leftrightarrow c = 0.0588.$$

Onde Φ^{-1} representa a função inversa da função distribuição de $N(0,1)$, ou seja, $\Phi^{-1}(p)$ é o quantil de probabilidade p da $N(0,1)$.

(d) Seja Y a v.a. que contabiliza o número de garrafas com mais de 1.55 litros em 1000 garrafas enchidas. Repare-se que temos 1000 provas independentes (quantidade de água a ser engarrafada) e cada prova tem dois resultados possíveis (quantidade superior à capacidade ou não), considerando-se como "sucesso"a ocorrência de "quantidade de água superior à capacidade"e cuja probabilidade de ocorrer é $P(X > 1.55) = 1 - P(X \leq 1.55) = 1 - \Phi(5/3) = 0.0478$. É fácil de perceber que esta v.a. terá distribuição Binomial de parâmetros $n = 1000$ e $p = P(X > 1.55)$, pelo que podemos escrever, $Y \sim B(1000, 0.0478)$ e o que pretendemos é $E(Y)$, vindo

$$E(Y) = np = 47.8.$$

(e) Queremos q tal que $P(X < q) = 0.25$, como

$$P(X < q) = 0.25 \Leftrightarrow P\left(\frac{X - 1.5}{0.03} < \frac{q - 1.5}{0.03}\right) = 0.25 \Leftrightarrow \Phi\left(\frac{q - 1.5}{0.03}\right) = 0.25$$
$$\Leftrightarrow 1 - \Phi\left(\frac{1.5 - q}{0.03}\right) = 0.25 \Leftrightarrow \Phi\left(\frac{1.5 - q}{0.03}\right) = 0.75 \Leftrightarrow \frac{1.5 - q}{0.03} = \Phi^{-1}(0.75)$$
$$\Leftrightarrow \frac{1.5 - q}{0.03} = 0.67 \Leftrightarrow q = 1.4799 \text{ litros.}$$

8. A resistência à compressão de amostras de cimento é uma variável aleatória X com distribuição normal de valor médio 6000Kg/cm^2 e desvio padrão 100Kg/cm^2.

(a) Qual é a probabilidade de uma amostra de cimento ter resistência à compressão inferior a 6250Kg/cm^2?

(b) Que resistência à compressão é excedida em 95% das amostras?

(c) Supondo que as resistências à compressão de amostras de cimento são variáveis aleatórias independentes, determine o número médio de amostras com resistência inferior a 6000Kg/cm^2 em 100 amostras de cimento.

Resolução:
Seja X a v.a. que representa a resistência, em Kg/cm^2, de uma amostra de cimento.

Dado que $X \sim N(6000, 100^2)$ então $Z = \frac{X-6000}{100} \sim N(0,1)$.

(a) $P(X < 6250) = P\left(Z < \frac{6250-6000}{100}\right) = P(Z < 2.5) = \Phi(2.5) = 0.9938$.

(b) Pretendemos determinar t de forma que

$$P(X > t) = 0.95 \Leftrightarrow P(X \leq t) = 0.05 \Leftrightarrow P(Z \leq \frac{t-6000}{100}) = 0.05 \Leftrightarrow \Phi(\frac{t-6000}{100}) = 0.05$$
$$\Leftrightarrow \Phi(\frac{6000-t}{100}) = 0.95 \Leftrightarrow \frac{6000-t}{100} = 1.64 \Leftrightarrow t = 5836.$$

(c) Seja Y a v.a. que representa o número de amostras de cimento cuja a resistência à compressão é inferior a 6000Kg/cm^2 em 100 amostras. Como a resistência de cada amostra é independente das resistências das outras amostras, temos 100 provas de Bernoulli independentes com a mesma probabilidade de "sucesso", $p = P(X < 6000) = P(Z < 0) = \Phi(0) = 0.5$, logo $Y \sim Bin(100, 0.5)$. Assim, o número médio de amostras com resistência inferior a 6000Kg/cm^2 é $E[Y] = 100 \times 0.5 = 50$ amostras.

9. Sabe-se que o tempo gasto pelos indivíduos para resolver um teste é normalmente distribuído com média de 20 minutos e desvio padrão de 4 minutos.

(a) Determine a probabilidade de que uma pessoa gaste entre 16 e 22 minutos para resolver o teste.

(b) Qual o tempo de resolução, excedido com probabilidade 0.8?

(c) Num teste realizado por 20 pessoas qual a probabilidade de apenas 2 destas pessoas gastarem no máximo 20 minutos para resolver o teste?

(d) Determine a probabilidade do tempo total gasto por 10 pessoas, na resolução do teste, ser inferior a 220 minutos.

Resolução:
Seja X o tempo gasto, em minutos, por um indivíduo na realização do teste. Ora, $X \sim N(20, 16)$.

(a) Como $Z = \frac{X-20}{4} \sim N(0,1)$ então,

$$P(16 < X < 22) = P\left(\frac{16-20}{4} < \frac{X-20}{4} < \frac{22-20}{4}\right) = P(-1 < Z < 0.5) =$$
$$\overset{v.a.\ cont.}{=} \Phi(0.5) - \Phi(-1) = \Phi(0.5) - (1 - \Phi(1)) = 0.6915 - (1 - 0.8413) = 0.5328.$$

(b) Pretende-se saber o tempo t, em minutos, tal que $P(X > t) = 0.8$. Assim,

$$P(X > t) = 0.8 \Leftrightarrow P(X \leq t) = 0.2 \Leftrightarrow P\left(\frac{X-20}{4} \leq \frac{t-20}{4}\right) = 0.2 \Leftrightarrow P\left(Z \leq \frac{t-20}{4}\right) = 0.2$$
$$\Leftrightarrow \Phi\left(\frac{t-20}{4}\right) = 0.2 \overset{(1)}{\Leftrightarrow} \Phi\left(-\frac{t-20}{4}\right) = 0.8 \Leftrightarrow -\frac{t-20}{4} = 0.84 \Leftrightarrow t = 16.64,$$

(1) $\Phi(-c) = 1 - \Phi(c),\ \forall c \in \Re$.

(c) Seja Y - número de pessoas, em 20, que gastam no máximo 20 minutos na realização do teste.
Como o tempo de realização do teste é independente de pessoa para pessoa então temos 20 provas de Bernoulli independentes com probabilidade de sucesso, p, constante. Assim, $Y \sim Bin(20, p)$ com

$$p = P(X \leq 20) = P\left(\frac{X-20}{4} \leq \frac{20-20}{4}\right) = P(Z \leq 0) = \Phi(0) = 0.5,$$

logo

$$P(Y=2)=\binom{20}{2}0.5^2(1-0.5)^{20-2}=0.0002.$$

(d) Sejam as variáveis aleatórias X_i - tempo gasto pela pessoa i na realização do teste, $i=1,...,10$. As variáveis $X_1,...,X_{10}$ são independentes e têm distribuição normal de valor médio 20 e desvio padrão 4.
Tendo em conta a definição das variáveis $X_i,\ i=1,...,10$, o tempo total gasto na realização do teste para as 10 pessoas é dado por $T=\sum_{i=1}^{10}X_i$.

Como T é uma combinação linear de variáveis aleatórias independentes com distribuição normal, tem-se

$$T\sim N(E(T),V(T))\equiv N(200,160)$$

onde

$$E(T)=E\left(\sum_{i=1}^{10}X_i\right)=\sum_{i=1}^{10}E(X_i)=\sum_{i=1}^{10}20=200$$

e

$$V(T)=V\left(\sum_{i=1}^{10}X_i\right)\overset{v.a.'s\ ind.}{=}\sum_{i=1}^{10}V(X_i)=\sum_{i=1}^{10}16=160.$$

Assim,

$$P(T<220)=P\left(\frac{T-200}{\sqrt{160}}<\frac{220-200}{\sqrt{160}}\right)\overset{(2)}{=}\Phi\left(\frac{20}{\sqrt{160}}\right)\approx\Phi(1.58)=0.9429.$$

(2) Se $T\sim N(200,160)$ então $\frac{T-200}{\sqrt{160}}\sim N(0,1)$.

Capítulo 4

Teorema Limite Central

1. Uma empresa vende caixas com biscoitos e, quando lhe é solicitado, envia-as pelo correio. Para evitar pesar estas caixas, cobra sempre o valor de portes de correio correspondente a admitir que qualquer caixa pesa 2508g. Cada caixa leva 100 biscoitos e o peso da embalagem plástica é desprezável. Se soubermos que o peso de cada biscoito é variável mas que em média pesa 25g com um desvio padrão de 8g, determine a probabilidade do valor pago em portes de correio com o envio de uma caixa ser inferior ao valor que pagaria, caso a caixa fosse pesada.

   Resolução:
   Considere as seguintes v.a., X_i-peso do biscoito i, com $i = 1, ..., 100$. Como $\mathbb{E}[X_i] = 25$ e $\sigma(X_i) = 8$, $i = 1, ..., 100$ e por $S_{100} = \sum_{i=1}^{100} X_i$ ser uma soma de v.a. independentes e identicamente distribuídas, sabe-se pelo teorema limite central que

   $$Z = \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{S_{100} - 100 \times 25}{8 \times \sqrt{100}} \overset{a}{\sim} N(0, 1).$$

   Repare-se que o peso da caixa com os biscoitos corresponde à variável S_{100} pelo que o pretendido pode ser representado por $P(S_{100} > 2508)$. Como $n = 100 \geq 30$, podemos utilizar o teorema limite central, para aproximar a probabilidade pedida:

   $$P(S_{100} > 2508) = 1 - P(S_{100} \leq 2508) = 1 - P\left(\frac{S_{100} - 2500}{80} \leq \frac{2508 - 2500}{80}\right) \overset{TLC}{\approx} 1 - \Phi(0.1) = 0.4602.$$

2. Uma empresa tem produção constante de 90 toneladas/mês do produto que fabrica. Sabe-se que a procura mensal desse produto (em toneladas) é uma v.a. de valor médio $E(X) = 80$ e desvio padrão $\sigma(X) = 10$ e é independente de mês para mês. Calcule a probabilidade da procura nos próximos 3 anos ser superior a 2950 toneladas.

   Resolução:
   Sejam as variáveis aleatórias X_i, procura em toneladas no mês i, $i = 1, ..., 36$. Assim $T = \sum_{i=1}^{36} X_i$ representa a procura total, em toneladas, durante os próximos 3 anos. Dado que T é uma soma de variáveis independentes e idênticamente distribuidas com $E(X_i) = 80$ e $\sigma(X_i) = 10$ ($\forall i = 1, ..., 36$), então como $n = 36 > 30$ estamos nas condições do teorema limite central, pelo que,

   $$Z = \frac{T - 36 \times 80}{\sqrt{36 \times 100}} \overset{a}{\sim} N(0, 1)$$

   e

   $$P(T > 2950) = 1 - P(T \leq 2950) = 1 - P\left(\frac{T - 36 \times 80}{\sqrt{36 \times 100}} \leq \frac{2950 - 36 \times 80}{\sqrt{36 \times 100}}\right) =$$
   $$= 1 - P\left(Z \leq \frac{2950 - 2880}{\sqrt{3600}}\right) \approx 1 - \Phi\left(\frac{2950 - 2880}{\sqrt{3600}}\right) = 1 - 0.8790 = 0.1210$$

3. A procura diária para certo tipo de artigo na loja A tem procura média e variância diária de 2 produtos. Considerando as variáveis aleatórias X_i, o número de produtos vendidos no dia i, $i = 1, 2, ..., 365$, calcule a probabilidade **aproximada** do número de produtos vendidos num ano (com 365 dias) ser no máximo 730.

   Resolução:

   Seja X a procura diária de certo artigo na loja A, sabe-se que $E(X) = 2$ e que $V(X) = 2$.
   Com X_i, o número de produtos vendidos no dia i, $i = 1, 2, ..., 365$ podemos definir $S_{365} = \sum_{i=1}^{365} X_i$ como o número de produtos vendidos num ano. Queremos,

$$P(S_{365} \leq 730)$$

   mas como estamos nas condições do teorema limite central $(X_i, i = 1, ..., n$ i.i.d. com $n = 365 > 30)$ e como $\forall i = 1, ..., 365, E(X_i) = 2$ e $V(X_i) = 2$, então

$$Z = \frac{S_{365} - 365 \times 2}{\sqrt{365 \times 2}} \overset{a}{\sim} N(0, 1)$$

   pelo que

$$P(S_{365} \leq 730) = P\left(\frac{S_{365} - 365 \times 2}{\sqrt{365 \times 2}} \leq \frac{730 - 365 \times 2}{\sqrt{365 \times 2}}\right) = P(Z \leq 0) \approx \Phi(0) = 0.5$$

Capítulo 5

Estimação Pontual

1. Considere a amostra aleatória $(X_1, X_2, \ldots, X_n)$ de uma população com distribuição Binomial de parâmetros 2 e p.

   (a) O estimador $T = 0.5\overline{X}$ é consistente de p? Justifique.

   (b) Dada a amostra $(0, 0, 2, 1, 2, 1, 0, 2, 2, 1)$, estime pontualmente a $P(X_1 = 2)$.

   Resolução:

   (a) Dado que

   $$E(T) = E\left(0.5\overline{X}\right) = E\left(\frac{1}{2n}\sum_{i=1}^{n} X_i\right) = \frac{1}{2n}\sum_{i=1}^{n} E(X_i) = \frac{1}{2n}\sum_{i=1}^{n} 2p = p,$$

   isto é, T é estimador centrado de p, então

   $$EQM(T) = V(T) = V\left(\frac{1}{2n}\sum_{i=1}^{n} X_i\right) = \frac{1}{4n^2} V\left(\sum_{i=1}^{n} X_i\right) \underset{X_i' s\, ind.}{=} \frac{1}{4n^2}\sum_{i=1}^{n} V(X_i) =$$
   $$= \frac{1}{4n^2}\sum_{i=1}^{n} 2p(1-p) = \frac{p(1-p)}{2n}.$$

   Como $\lim_{n\to\infty} EQM(T) = \lim_{n\to\infty} \frac{p(1-p)}{2n} = 0$ então o estimador T é consistente para p.

   (b) Como $P(X_1 = 2) = \binom{2}{2} p^2 (1-p)^{2-2} = p^2$ e $\hat{p} = 0.5\overline{X}$ é estimador de p, então a estimativa pontual de $P(X_1 = 2)$ é $P(\widehat{X_1 = 2}) = (0.5\overline{x})^2 = (0.5 \times 1.1)^2 = 0.3025$.

2. Considere a amostra aleatória $(X_1, X_2, \ldots, X_n)$ de uma população com distribuição Poisson de parâmetro λ. Considere os estimadores de λ, $\hat{\lambda}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i$ e $\hat{\lambda}_2 = \frac{1}{n-2}\sum_{i=1}^{n-2} X_i$, $n > 2$. Qual é o estimador mais eficiente? Justifique.

   Resolução:

   Como

   $$E\left(\hat{\lambda}_1\right) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n} \lambda = \lambda$$

   e

   $$E\left(\hat{\lambda}_2\right) = E\left(\frac{1}{n-2}\sum_{i=1}^{n-2} X_i\right) = \frac{1}{n-2}\sum_{i=1}^{n-2} E(X_i) = \frac{1}{n-2}\sum_{i=1}^{n-2} \lambda = \lambda$$

   então $\hat{\lambda}_1$ e $\hat{\lambda}_2$ são estimadores centrados de λ, logo

$$EQM(\hat{\lambda}_1) = V\left(\hat{\lambda}_1\right) = V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}V\left(\sum_{i=1}^{n} X_i\right) \underset{X_i's\,ind.}{=} \frac{1}{n^2}\sum_{i=1}^{n} V\left(X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\lambda = \frac{\lambda}{n}$$

e

$$EQM(\hat{\lambda}_2) = V\left(\hat{\lambda}_2\right) = \frac{1}{(n-2)^2}V\left(\sum_{i=1}^{n-2} X_i\right) \underset{X_i's\,ind.}{=} \frac{1}{(n-2)^2}\sum_{i=1}^{n-2} V\left(X_i\right) = \frac{1}{(n-2)^2}\sum_{i=1}^{n-2}\lambda = \frac{\lambda}{n-2}.$$

Assim, $\hat{\lambda}_1$ é mais eficiente do que $\hat{\lambda}_2$ uma vez que $EQM(\hat{\lambda}_1) < EQM(\hat{\lambda}_2)$.

3. O estimador da proporção, $\hat{P}$, é centrado e consistente de p? Justifique.

Resolução:
O estimador da proporção, p, é dado por $\hat{P} = X/n$ onde $X \sim Bin(n, p)$. Assim

$$E(\hat{P}) = E\left(\frac{X}{n}\right) = \frac{E(X)}{n} = \frac{np}{n} = p,$$

logo, o estimador $\hat{P}$ é centrado de p.

Como $\hat{P}$ é um estimador centrado, $b(\hat{P}) = E(\hat{P}) - p = 0$, então

$$EQM(\hat{P}) = V(\hat{P}) + b^2(\hat{P}) = V(\hat{P}) = V\left(\frac{X}{n}\right) = \frac{V(X)}{n^2} = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n},$$

logo, o estimador $\hat{P}$ é consistente de p, pois $\lim_{n\to\infty} EQM(\hat{P}) = \lim_{n\to\infty} \frac{p(1-p)}{n} = 0$.

4. No país das Maravilhas a proporção de loucos é de 0.45. Suponha que se pretende selecionar uma amostra aleatória de 500 habitantes deste país. Qual a probabilidade de a proporção de loucos que vão calhar na amostra exceder 0.5?

Resolução:
A proporção de loucos que vão calhar na amostra é representada pelo estimador $\hat{P}$ (da verdadeira proporção de loucos na população, $p = 0.45$).
Como $n = 500 \geq 30$, então

$$Z = \frac{\hat{P} - p}{\sqrt{p(1-p)/n}} = \frac{\hat{P} - 0.45}{\sqrt{0.45(1-0.45)/500}} \overset{a}{\sim} N(0, 1).$$

Assim,

$$P(\hat{P} > 0.5) = 1 - P(\hat{P} \leq 0.5) = 1 - P\left(\frac{\hat{P} - 0.45}{\sqrt{0.45(1-0.45)/500}} \leq \frac{0.5 - 0.45}{\sqrt{0.45(1-0.45)/500}}\right) =$$
$$= 1 - P\left(Z \leq 2.247333\right) \approx 1 - \Phi(2.247333) = 1 - 0.9878 = 0.0122.$$

5. Numa população Normal de média desconhecida e desvio padrão 5 calcule a probabilidade de a variância de uma amostra aleatória de dimensão 20 dessa população estar compreendida entre 26 e 58.

Resolução:
Numa população Normal de média desconhecida, desvio padrão 5 e para uma amostra de dimensão 20, tem-se

$$X^2 = \frac{(n-1)S^2}{\sigma^2} = \frac{19S^2}{25} \sim \chi^2_{19}.$$

Assim, a probabilidade de a variância de uma amostra aleatória de dimensão 20 estar compreendida entre 26 e 58 é

$$P\left(26 < S^2 < 58\right) = P\left(\frac{19 \times 26}{25} < \frac{19S^2}{25} < \frac{19 \times 58}{25}\right) = P\left(19.76 < X^2 < 44.08\right) =$$
$$= F_{\chi^2_{19}}(44.08) - F_{\chi^2_{19}}(19.76) = 0.9991 - 0.5909 = 0.4082^*.$$

(*) Valor calculado com o software R (http://www.r-project.org).

6. Seja $(X_1, X_2, ..., X_n)$ uma amostra aleatória de uma população com distribuição normal de valor médio μ e variância σ^2. Mostre que $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ é um estimador centrado para σ^2.

   Resolução:

   Sabemos que nas condições do enunciado $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$, logo $E\left(\frac{(n-1)S^2}{\sigma^2}\right) = n - 1$.
   Assim,

$$E(S^2) = \frac{\sigma^2}{n-1}E\left(\frac{(n-1)S^2}{\sigma^2}\right) = \frac{\sigma^2}{n-1}(n-1) = \sigma^2,$$

   concluindo-se que o estimador S^2 é centrado para σ^2.

Capítulo 6

Estimação por Intervalo de Confiança

1. A resistência à força de compressão (em psi) de um certo tipo de betão é modelado por uma variável aleatória normal de valor médio μ e variância 1000. Numa amostra de dimensão 12 observou-se uma resistência média à força de compressão de 3250 psi.

   (a) Deduza e calcule um intervalo a 95% de confiança para μ.

   (b) Se utilizássemos a mesma amostra para obter um intervalo com amplitude 30 psi, qual seria o nível de confiança desse intervalo?

   Resolução:

   (a) A população tem distribuição normal e a variância é conhecida. Vamos utilizar a variável pivot:

   $$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1).$$

   Com $1 - \alpha = 0.95$ temos o quantil de probabilidade:

   $$c : P(Z > c) = \frac{\alpha}{2} = 0.025 \Rightarrow c = z_{\alpha/2} = z_{0.025} = \Phi^{-1}(1 - 0.025) = 1.96.$$

   Assim $P(-c < Z < c) = 0.95$ e

   $$-c < Z < c \Leftrightarrow -c < \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} < c \Leftrightarrow -c\frac{\sigma}{\sqrt{n}} < \overline{X} - \mu < c\frac{\sigma}{\sqrt{n}} \Leftrightarrow$$
   $$\Leftrightarrow -\overline{X} - c\frac{\sigma}{\sqrt{n}} < -\mu < -\overline{X} + c\frac{\sigma}{\sqrt{n}} \Leftrightarrow \overline{X} - c\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + c\frac{\sigma}{\sqrt{n}}$$

   então, o intervalo de confiança aleatório a 95% é dado por:

   $$IC_{95\%}(\mu) \equiv \left]\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}; \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right[.$$

   Assim, temos o intervalo de confiança a 95% para μ:

   $$IC_{95\%}(\mu) \equiv \left]3250 - 1.96\frac{\sqrt{1000}}{\sqrt{12}}; 3250 + 1.96\frac{\sqrt{1000}}{\sqrt{12}}\right[\equiv]3232.108; 3267.892[.$$

   (b) Tendo em conta a alínea anterior, o intervalo aleatório a $(1 - \alpha) \times 100\%$ para μ é dado por:

   $$IC_{(1-\alpha)\times 100\%}(\mu) \equiv \left]\overline{X} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}; \overline{X} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right[,$$

   com amplitude $2z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} = 2z_{\frac{\alpha}{2}}\frac{\sqrt{1000}}{\sqrt{12}}$. Tem-se,

   $$2z_{\frac{\alpha}{2}}\frac{\sqrt{1000}}{\sqrt{12}} = 30 \Leftrightarrow z_{\frac{\alpha}{2}} = 1.643168 \Leftrightarrow \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) = 1.643168 \Rightarrow 1 - \frac{\alpha}{2} \approx 0.95,$$

   ou seja, $\alpha \approx 0.1$, logo o nível de confiança seria de aproximadamente 90%.

2. Um Engenheiro Químico necessita que o catalisador, que vai usar numa reação química, tenha pH próximo de 6.4. Como não consegue medir sem erro o pH do catalisador, analisou 15 amostras distintas (desse catalisador) e obteve $\sum_{i=1}^{15} x_i = 96.242$ e $\sum_{i=1}^{15} x_i^2 = 617.5785$. Assuma que o pH do catalisador tem distribuição Normal de valor médio μ e variância σ^2.

   (a) Estime pontualmente μ e σ^2.

   (b) Estime σ por intervalo de 90% de confiança. Apresente a dedução do intervalo.

   Resolução:

   (a) Dado que $\overline{x} = \frac{1}{15}\sum_{i=1}^{15} x_i = 6.416133$ e $s^2 = \frac{1}{15-1}\left(\sum_{i=1}^{15} x_i^2 - 15\overline{x}^2\right) = 0.0055$, tem-se $\hat{\mu} = 6.416133$ e $\hat{\sigma}^2 = 0.0055$, estimativas pontuais de μ e σ^2 respetivamente.

   (b) Como a população tem distribuição normal e valor médio desconhecido, vamos utilizar a variável pivot:

   $$X^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}.$$

   Com $\alpha = 0.1$ e $n = 15$ temos os quantis de probabilidade:

   $$a: \ P(X^2 > a) = 1 - \frac{\alpha}{2} \Rightarrow a = \chi^2_{n-1:1-\frac{\alpha}{2}} = \chi^2_{14:0.95} = 6.57$$

   e

   $$b: \ P(X^2 > b) = \frac{\alpha}{2} \Rightarrow b = \chi^2_{n-1:\frac{\alpha}{2}} = \chi^2_{14:0.05} = 23.7.$$

   Como $P(a < X^2 < b) = 1 - \alpha$ e

   $$a < X^2 < b \Leftrightarrow a < \frac{(n-1)S^2}{\sigma^2} < b \Leftrightarrow \frac{1}{b} < \frac{\sigma^2}{(n-1)S^2} < \frac{1}{a} \Leftrightarrow$$
   $$\Leftrightarrow \frac{(n-1)S^2}{b} < \sigma^2 < \frac{(n-1)S^2}{a} \Leftrightarrow \sqrt{\frac{(n-1)S^2}{b}} < \sigma < \sqrt{\frac{(n-1)S^2}{a}}$$

   então, o intervalo de confiança aleatório a $(1-\alpha)\times 100\%$ é dado por:

   $$IC_{(1-\alpha)\times 100\%}(\sigma) \equiv \left]\sqrt{\frac{(n-1)S^2}{\chi^2_{n-1:\frac{\alpha}{2}}}}; \sqrt{\frac{(n-1)S^2}{\chi^2_{n-1:1-\frac{\alpha}{2}}}}\right[.$$

   Assim, temos o intervalo de confiança a 90% para σ:

   $$IC_{90\%}(\sigma) \equiv \left]\sqrt{\frac{(15-1)\times 0.0055}{23.7}}; \sqrt{\frac{(15-1)\times 0.0055}{6.57}}\right[\equiv]0.057; 0.10826[.$$

3. Numa amostra de aparelhos de ar condicionado, fabricados por determinada empresa, foram registados os seguintes coeficientes de transferência de calor:

   $$65, \quad 63, \quad 60, \quad 68, \quad 62.$$

   Admitindo a normalidade da população:

   (a) Deduza e determine um intervalo de confiança a 99% para o coeficiente médio de transferência de calor, de todos os aparelhos de ar condicionado fabricados pela referida empresa.

   (b) Considere a variância da população igual a 11.64. Qual deverá ser a dimensão da amostra para garantir que a amplitude do intervalo de confiança a 95% para o valor médio seja inferior a 5. Não é necessário fazer a dedução do intervalo de confiança.

   Resolução:

(a) Como a população tem distribuição normal de variância desconhecida vamos utilizar a variável pivot:

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}.$$

Com $1 - \alpha = 0.99$ temos o quantil de probabilidade:

$$c : P(T > c) = \frac{\alpha}{2} = 0.005 \Rightarrow c = t_{n-1;\alpha/2} = t_{4;0.005} = 4.60.$$

Assim $P(-c < T < c) = 0.99$ e

$$-c < T < c \Leftrightarrow -c < \frac{\overline{X} - \mu}{S/\sqrt{n}} < c \Leftrightarrow -c\frac{S}{\sqrt{n}} < \overline{X} - \mu < c\frac{S}{\sqrt{n}} \Leftrightarrow$$
$$\Leftrightarrow -\overline{X} - c\frac{S}{\sqrt{n}} < -\mu < -\overline{X} + c\frac{S}{\sqrt{n}} \Leftrightarrow \overline{X} - c\frac{S}{\sqrt{n}} < \mu < \overline{X} + c\frac{S}{\sqrt{n}}$$

então, o intervalo de confiança aleatório a $(1 - \alpha) \times 100\%$ é dado por:

$$IC_{(1-\alpha)\times 100\%}(\mu) \equiv \left]\overline{X} - t_{n-1;\alpha/2}\frac{S}{\sqrt{n}}; \overline{X} + t_{n-1;\alpha/2}\frac{S}{\sqrt{n}}\right[.$$

Para esta amostra, $n = 5$; $\overline{x} = 63.6$; $s = 3.04959$, temos o intervalo de confiança a 99% para μ:

$$IC_{99\%}(\mu) \equiv \left]63.6 - 4.6\frac{3.04959}{\sqrt{5}}; 63.6 + 4.6\frac{3.04959}{\sqrt{5}}\right[\equiv]57.32644; 69.87356[.$$

(b) Como a população tem distribuição normal e $\sigma^2 = 11.64$ então, o intervalo de confiança aleatório a $(1 - \alpha) \times 100\%$ para o valor médio, μ, é dado por:

$$IC_{(1-\alpha)\times 100\%}(\mu) \equiv \left]\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}; \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right[.$$

com amplitude $\Delta = 2z_{\alpha/2}\sigma/\sqrt{n}$.
Com $\alpha = 0.05$, $z_{\alpha/2} = \Phi^{-1}(1 - \alpha/2) = 1.96$, tem-se

$$\Delta < 5 \Leftrightarrow 2 \times 1.96\sqrt{11.64}/\sqrt{n} < 5 \Leftrightarrow n > \left(\frac{2 \times 1.96\sqrt{11.64}}{5}\right)^2 \Leftrightarrow n > 7.154596.$$

Para garantir um intervalo de confiança a 95% para μ com amplitude inferior a 5 teríamos de considerar amostras com dimensão superior ou igual a 8.

4. A resistência, em ohms, dum certo tipo de componentes eletrónicos tem distribuição normal. Recolheu-se a seguinte amostra das resistências de 30 componentes eletrónicos:

10.11	10.20	10.17	10.55	10.30	10.18	10.03	10.99	10.03	10.24
10.41	10.52	10.49	10.69	10.68	10.68	10.10	10.43	10.32	10.09
10.07	10.99	10.23	10.94	10.38	10.64	10.25	10.00	10.56	10.93

para a qual se tem $\overline{x} = 10.4067$ e $s = 0.3043$.

(a) Obtenha uma estimativa pontual para a variância populacional.

(b) Deduza e calcule o intervalo de confiança a 95% para a variância populacional.

Resolução:

(a) A estimativa pontual da variância é dada por $\hat{\sigma}^2 = s^2 = 0.3043^2 = 0.0926$.

(b) Como a população tem distribuição normal e valor médio desconhecido, podemos utilizar a variável pivot:

$$X^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}.$$

Com $\alpha = 0.05$ e $n = 30$ temos os quantis de probabilidade:

$$a: \; P(X^2 > a) = 1 - \frac{\alpha}{2} \Rightarrow a = \chi^2_{n-1:1-\frac{\alpha}{2}} = \chi^2_{29:0.975} = 16.0$$

e

$$b: \; P(X^2 > b) = \frac{\alpha}{2} \Rightarrow b = \chi^2_{n-1:\frac{\alpha}{2}} = \chi^2_{29:0.025} = 45.7.$$

Como $P(a < X^2 < b) = 1 - \alpha$ e

$$a < X^2 < b \Leftrightarrow a < \frac{(n-1)S^2}{\sigma^2} < b \Leftrightarrow \frac{1}{b} < \frac{\sigma^2}{(n-1)S^2} < \frac{1}{a} \Leftrightarrow$$
$$\Leftrightarrow \frac{(n-1)S^2}{b} < \sigma^2 < \frac{(n-1)S^2}{a}$$

então, o intervalo de confiança aleatório a $(1-\alpha) \times 100\%$ é dado por:

$$IC_{(1-\alpha)\times 100\%}(\sigma^2) \equiv \left] \frac{(n-1)S^2}{\chi^2_{n-1:\frac{\alpha}{2}}}; \frac{(n-1)S^2}{\chi^2_{n-1:1-\frac{\alpha}{2}}} \right[.$$

Assim, temos o intervalo de confiança a 95% para σ^2:

$$IC_{95\%}(\sigma^2) \equiv \left] \frac{(30-1) \times 0.3043^2}{45.7}; \frac{(30-1) \times 0.3043^2}{16} \right[\equiv]0.0588; 0.1678[.$$

5. Seja X uma população com distribuição normal de média μ e desvio padrão σ. Uma amostra aleatória de dimensão 25 foi extraída desta população e revelou $\sum_{i=1}^{25} x_i = 1200.25$ e $\sum_{i=1}^{25} x_i^2 = 62218.58$.

   (a) Indique estimativas pontuais, com base nesta amostra, para μ e σ.

   (b) Deduza e calcule um intervalo de confiança a 95% para a média populacional.

   (c) Admita agora que o intervalo de confiança para o valor médio da população é dado por:

   $$\left] \bar{X} - z_{\alpha/2} \times \frac{3}{\sqrt{n}}; \bar{X} + z_{\alpha/2} \times \frac{3}{\sqrt{n}} \right[.$$

      i. Qual deverá ser a dimensão da amostra de forma a que a amplitude deste intervalo de confiança a 92% seja inferior a 1?

      ii. Qual o efeito na amplitude deste intervalo de confiança quando o nível de confiança aumenta? Justifique.

Resolução:

(a) Dado que $\overline{x} = \frac{1}{25}\sum_{i=1}^{25} x_i = 48.01$ e $s^2 = \frac{1}{25-1}\left(\sum_{i=1}^{25} x_i^2 - 25\overline{x}^2\right) = 191.4407$, tem-se $\hat{\mu} = 48.01$ e $\hat{\sigma} = \sqrt{191.4407} = 13.83621$, estimativas pontuais de μ e σ respetivamente.

(b) Como a população tem distribuição normal e variância desconhecida, vamos utilizar a variável pivot:

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}.$$

Com $\alpha = 0.05$ e $n = 25$ temos o quantil de probabilidade:

$$c: \; P(T > c) = \frac{\alpha}{2} \Leftrightarrow P(T \leq c) = 1 - \frac{\alpha}{2} \Rightarrow c = t_{n-1:\frac{\alpha}{2}} = t_{24:0.025} = 2.06.$$

Como $P(-c < T < c) = 0.95$ e

$$-c < T < c \Leftrightarrow -c < \frac{\overline{X} - \mu}{S/\sqrt{n}} < c \Leftrightarrow -c\frac{S}{\sqrt{n}} < \overline{X} - \mu < c\frac{S}{\sqrt{n}} \Leftrightarrow$$
$$\Leftrightarrow -\overline{X} - c\frac{S}{\sqrt{n}} < -\mu < -\overline{X} + c\frac{S}{\sqrt{n}} \Leftrightarrow \overline{X} - c\frac{S}{\sqrt{n}} < \mu < \overline{X} + c\frac{S}{\sqrt{n}}$$

então, o intervalo de confiança aleatório a 95% é dado por:

$$IC_{95\%}(\mu) \equiv \left]\overline{X} - t_{n-1;\alpha/2}\frac{S}{\sqrt{n}}; \overline{X} + t_{n-1;\alpha/2}\frac{S}{\sqrt{n}}\right[.$$

Assim, temos o intervalo de confiança a 95% para μ:

$$IC_{95\%}(\mu) \equiv \left]48.01 - 2.06\frac{13.83621}{\sqrt{25}}; 48.01 + 2.06\frac{13.83621}{\sqrt{25}}\right[\equiv]42.30948; 53.71052[.$$

(c) i. Para um nível de confiança 92%, isto é, $\alpha = 0.08$, tem-se

$$z_{\alpha/2} = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) = \Phi^{-1}(0.96) = 1.75$$

logo, a amplitude do intervalo de confiança é dada por:

$$\Delta = \bar{X} + z_{\alpha/2} \times \frac{3}{\sqrt{n}} - \left(\bar{X} - z_{\alpha/2} \times \frac{3}{\sqrt{n}}\right) = 2z_{\alpha/2} \times \frac{3}{\sqrt{n}} = \frac{10.5}{\sqrt{n}}.$$

Assim,

$$\Delta < 1 \Leftrightarrow \frac{10.5}{\sqrt{n}} < 1 \Leftrightarrow n > 10.5^2 \Leftrightarrow n > 110.25,$$

ou seja, a dimensão da amostra deverá ser superior ou igual a 111.

ii. Como $z_{\alpha/2} = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$ e Φ^{-1} é uma função estritamente crescente, então, quanto maior for o nível de confiança, $1 - \alpha$, maior será o valor de $z_{\alpha/2}$ e, consequentemente, a amplitude do intervalo de confiança será maior.

6. De uma população escolheram-se aleatoriamente 120 pessoas para serem inquiridas sobre programas que viam na televisão. Na amostra existem 35 pessoas que gastam mais de 15 horas por semana a ver televisão.

(a) Com base nesta amostra, estime a proporção de pessoas que gastam mais de 15 horas semanais a ver televisão.

(b) Determine um intervalo com nível de confiança 95%, para a proporção de pessoas que gastam mais de 15 horas de televisão por semana.

Resolução:

(a) A estimativa da proporção de pessoas que gastam mais de 15 horas semanais a ver televisão, p, é dada por $\hat{p} = 35/120 = 0.2917$.

(b) Como $n \geq 30$, vamos utilizar a variável pivot:

$$Z = \frac{\hat{P} - p}{\sqrt{p(1-p)/n}} \overset{a}{\sim} N(0,1).$$

Com $1-\alpha = 0.95$ temos o quantil de probabilidade:

$$c : P(Z > c) = \frac{\alpha}{2} = 0.025 \Rightarrow c = z_{\alpha/2} = z_{0.025} = \Phi^{-1}(1-0.025) = 1.96.$$

A dedução do intervalo em ordem a p torna-se muito mais simples se substituirmos $\sqrt{p(1-p)/n}$, pela sua estimativa, $\sqrt{\hat{P}(1-\hat{P})/n}$.
Assim $P(-c < Z < c) = 1-\alpha$ e

$$-c < Z < c \Leftrightarrow -c < \frac{\hat{P}-p}{\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}} < c \Leftrightarrow -c\sqrt{\frac{\hat{P}(1-\hat{P})}{n}} < \hat{P}-p < c\sqrt{\frac{\hat{P}(1-\hat{P})}{n}} \ \Leftrightarrow$$

$$-c\sqrt{\frac{\hat{P}(1-\hat{P})}{n}} - \hat{P} < -p < c\sqrt{\frac{\hat{P}(1-\hat{P})}{n}} - \hat{P} \ \Leftrightarrow \hat{P} - c\sqrt{\frac{\hat{P}(1-\hat{P})}{n}} < p < \hat{P} + c\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}$$

então, o intervalo de confiança aleatório a $(1-\alpha)\times 100\%$ é dado por:

$$IC_{(1-\alpha)\times 100\%}(p) = \left]\hat{P} - z_{\alpha/2}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}\ ;\ \hat{P} + z_{\alpha/2}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}\right[.$$

Assim, temos o intervalo de confiança a 95% para p:

$$IC_{95\%}(p) \equiv \left]\frac{35}{120} - 1.96\sqrt{\frac{\frac{35}{120}(1-\frac{35}{120})}{120}}; \frac{35}{120} + 1.96\sqrt{\frac{\frac{35}{120}(1-\frac{35}{120})}{120}}\right[\equiv]0.2103; 0.3730[.$$

Capítulo 7

Testes de Hipóteses

1. Seja X uma v.a. com distribuição Normal de valor médio μ e desvio padrão σ. A partir de uma amostra de dimensão 30, retirada da população, obtiveram-se os seguintes resultados:

$$\sum_{i=1}^{30} x_i = 64.0 \qquad \sum_{i=1}^{30}(x_i - \bar{x})^2 = 84.4$$

   (a) Teste, ao nível de significância 1%, as hipóteses $H_0 : \mu = 2 \ vs \ H_1 : \mu > 2$.

   (b) Suponha que está a testar a hipótese $H_0 : \mu = 2$ contra a hipótese $H_1 : \mu = 2.5$ e que rejeita a hipótese nula se $\bar{X}_{30} > 2.3$. Calcule as probabilidades dos erros de 1ª e 2ª espécie do teste, se $\sigma = 1$.

Resolução:

(a) As hipóteses a testar são $H_0 : \mu = 2$ vs $H_1 : \mu > 2$ (teste unilateral direito)

A população tem distribuição normal e a variância é desconhecida, logo a estatística de teste é:

$$T = \frac{\bar{X} - 2}{S/\sqrt{n}} \underset{\mu=2}{\sim} t_{n-1}.$$

A região de rejeição, para um nível de significância $\alpha \times 100\%$, é dada por:

$$R_\alpha =]c; +\infty[=]t_{n-1;\alpha}; +\infty[,$$

onde c é tal que:

$$P(T > c) = \alpha \Leftrightarrow c = t_{n-1:\alpha}.$$

Assim, tem-se para $\alpha = 0.01$, $c = t_{29;0.01} = 2.462$ e $R_{0.01} =]2.462; +\infty[$.
Dado que $s^2 = \frac{1}{30-1}\sum_{i=1}^{30}(x_i - \bar{x})^2 = 84.4/29$, o valor observado da estatística do teste é:

$$t_{obs} = \frac{\bar{x} - 2}{s/\sqrt{n}} = \frac{64/30 - 2}{\sqrt{84.4/29}/\sqrt{30}} = 0.4280822.$$

Como $t_{obs} \notin R_{0.05}$ não rejeitamos H_0 ao nível de significância de 1%.

(b) Uma vez que a população tem distribuição normal e variância conhecida ($\sigma^2 = 1$), tem-se:

$$Z = \frac{\bar{X}_{30} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1).$$

Assim, a probabilidade do erro de 1ª espécie é:

$$\alpha = P(\text{Rejeitar } H_0|H_0 \text{ verdadeira}) = P(\bar{X}_{30} > 2.3|\mu = 2) = 1 - P\left(\frac{\bar{X}_{30} - 2}{1/\sqrt{30}} \leq \frac{2.3 - 2}{1/\sqrt{30}}|\mu = 2\right)$$
$$= 1 - P\left(Z \leq 0.3\sqrt{30}|\mu = 2\right) = 1 - \Phi(0.3\sqrt{30}) = 1 - 0.9495 = 0.0505,$$

e a probabilidade do erro de 2ª espécie é:

$$\beta = P(\text{Não Rejeitar } H_0|H_0 \text{ falsa}) = P(\bar{X}_{30} \leq 2.3|\mu = 2.5) = P\left(\frac{\bar{X}_{30} - 2.5}{1/\sqrt{30}} \leq \frac{2.3 - 2.5}{1/\sqrt{30}}|\mu = 2.5\right)$$
$$= P\left(Z \leq -0.2\sqrt{30}|\mu = 2.5\right) = \Phi(-0.2\sqrt{30}) = 1 - \Phi(0.2\sqrt{30}) = 1 - 0.8643 = 0.1357.$$

2. A resistência, em ohms, dum certo tipo de componentes eletrónicos tem distribuição normal com desvio padrão 0.35. Recolheu-se a seguinte amostra das resistências de 30 componentes eletrónicos:

10.11	10.20	10.17	10.55	10.30	10.18	10.03	10.99	10.03	10.24
10.41	10.52	10.49	10.69	10.68	10.68	10.10	10.43	10.32	10.09
10.07	10.99	10.23	10.94	10.38	10.64	10.25	10.00	10.56	10.93

para a qual se tem $\overline{x} = 10.4067$ e $s = 0.3043$.

(a) Teste ao nível de significância 1% se o valor médio da população é inferior a 10.5.

(b) Admitindo que o valor observado da estatística do teste da alínea anterior foi de -1.46, calcule o valor-p do teste. Indique para que valores do nível de significância rejeitaria a hipótese nula do teste.

Resolução:

(a) As hipóteses a testar são $H_0 : \mu \geq 10.5$ vs $H_1 : \mu < 10.5$ (teste unilateral esquerdo)

A população tem distribuição normal e variância conhecida ($\sigma^2 = 0.35^2$), logo a estatística de teste é:

$$Z = \frac{\bar{X} - 10.5}{\sigma/\sqrt{n}} \underset{\mu=10.5}{\sim} N(0, 1).$$

Para $\alpha = 0.01$, tem-se $z_{0.01} = \Phi^{-1}(1 - 0.01) = 2.33$ e a região de rejeição é:

$$R_{0.01} =]-\infty; -z_{0.01}[=]-\infty; -2.33[.$$

Como $z_{obs} = \frac{10.4067-10.5}{0.35/\sqrt{30}} = -1.4601 \notin R_{0.01}$ não rejeitamos H_0 ao nível de significância 1%.

(b) Dado que se trata de um teste unilateral esquerdo, tem-se:

$$\text{valor}-p = P(Z < z_{obs}|H_0) = P(Z < -1.46|H_0) = \Phi(-1.46) = 1-\Phi(1.46) = 1-0.9279 = 0.0721.$$

Assim, rejeitamos H_0 para $\alpha > 0.0721$.

3. Considere uma população com distribuição normal de valor médio μ e desvio padrão σ. A partir de uma amostra de dimensão 30 dessa população obtiveram-se os seguintes resultados:

$$\sum_{i=1}^{30} x_i = 64 \quad \sum_{i=1}^{30}(x_i - \bar{x})^2 = 84.8$$

(a) Teste ao nível de significância de 5% a hipótese do valor médio da população ser superior a 2.

(b) Admitindo que, para outra amostra de igual dimensão, o valor observado da estatística do teste da alínea anterior foi de 1.311, calcule o correspondente valor-p e tome uma decisão ao nível de significância de 8%.

(c) Suponha agora que pretende testar as hipóteses $H_0 : \sigma^2 = 3 \;\; vs \;\; H_1 : \sigma^2 < 3$ e que se rejeita a hipótese nula se $S^2 < 1.3552$ para uma amostra de dimensão 30. Calcule a probabilidade do erro de 1ª espécie.

Resolução:

(a) As hipóteses a testar são $H_0 : \mu \leq 2$ vs $H_1 : \mu > 2$.
A população tem distribuição normal e variância desconhecida, logo a estatística de teste é:

$$T = \frac{\bar{X} - 2}{S/\sqrt{n}} \underset{\mu=2}{\sim} t_{n-1}.$$

A região de rejeição, para um nível de significância $\alpha \times 100\%$, é dada por:

$$R_\alpha =]c; +\infty[=]t_{n-1;\alpha}; +\infty[,$$

onde c é tal que:

$$P(T > c) = \alpha \Rightarrow c = t_{n-1;\alpha} = F^{-1}_{t_{n-1}}(1 - \alpha).$$

Assim, temos para $\alpha = 0.05$, $c = t_{29;0.05} = 1.699$ e $R_{0.05} =]1.699; +\infty[$.
Dado que $s^2 = \frac{1}{30-1}\sum_{i=1}^{30}(x_i - \bar{x})^2 = 84.8/29$, o valor observado da estatística do teste é:

$$t_{obs} = \frac{\bar{x} - 2}{s/\sqrt{n}} = \frac{64/30 - 2}{\sqrt{84.8/29}/\sqrt{30}} = 0.4270713.$$

Como $t_{obs} \notin R_{0.05}$ não rejeitamos H_0 ao nível de significância de 5%.

(b) Trata-se de um teste unilateral direito, logo

$$\text{valor} - p = P(T > t_{obs}|H_0) = P(T > 1.311|H_0) = 0.1$$

pois $t_{29:0.1} = 1.311$.
Assim, não rejeitamos H_0 a 8% se significância pois o valor-p é superior ao nível de significância 0.08.

(c) Como a população tem distribuição normal com valor médio desconhecido, temos:

$$X^2 = \frac{(n-1)S^2}{\sigma^2} = \frac{29S^2}{3} \underset{\sigma^2=3}{\sim} \chi^2_{29}.$$

Então, a probabilidade do erro de 1ª espécie é:

$$\alpha = P(\text{Rejeitar}H_0|H_0) = P(S^2 < 1.3552|H_0) = P\left(\frac{29S^2}{3} < \frac{29 \times 1.3552}{3}|H_0\right) =$$
$$= P\left(X^2 < 13.10027|H_0\right) = F_{\chi^2_{29}}(13.10027) \approx 0.005.$$

Obs: Como $\chi^2_{29:0.995} = 13.121$, $P\left(X^2 > 13.10027|H_0\right) \approx P\left(X^2 > 13.121|H_0\right) = 0.995$.

4. Numa fábrica, tem sido estudada qual a proporção de circuitos integrados defeituosos. Com base numa amostra aleatória de 300 circuitos verificou-se que 13 apresentavam defeitos.

(a) Utilize os dados para testar $H_0 : p = 0.05$ vs $H_1 : p \neq 0.05$. Considere $\alpha = 0.05$.

(b) Calcule o valor-p (aproximado) do teste e tome uma decisão para $\alpha = 0.1$.

Resolução:

(a) $H_0 : p = 0.05$ vs $H_1 : p \neq 0.05$ (teste bilateral)

Dado que $n \geq 30$, a estatística do teste é:

$$Z = \sqrt{n}\frac{\hat{P} - 0.05}{\sqrt{0.05(1-0.05)}} \underset{p=0.05}{\overset{a}{\sim}} N(0,1).$$

Para $\alpha = 0.05$, tem-se $z_{\alpha/2} = z_{0.025} = \Phi^{-1}(1-0.025) = 1.96$ e a região de rejeição é:

$$R_{0.05} =]-\infty; -z_{\alpha/2}[\cup]z_{\alpha/2}; +\infty[=]-\infty; -1.96[\cup]1.96; +\infty[.$$

Como $z_{obs} = \sqrt{300}\frac{13/300 - 0.05}{\sqrt{0.05(1-0.05)}} = -0.5298 \notin R_{0.05}$ não rejeitamos H_0 ao nível de significância 5%, isto é, não existe evidência, ao nível de significância 5%, para rejeitar a hipótese de que $p = 0.05$.

(b) O valor-p (aproximado) deste teste (teste bilateral) é dado por:

valor$-p = 2\min\{P(Z<z_{obs}|H_0), P(Z>z_{obs}|H_0)\} = 2P(Z>0.5298|H_0) \approx 2\,(1-\Phi(0.53)) = 0.5962$

Como o valor$-p \geq 0.1$ então não rejeitamos H_0 ao nível de significância de 10%.

5. Foram efetuados estudos em Lisboa com o objetivo de determinar a concentração de carbono (CO) perto das vias rápidas. Para o efeito recolheram-se amostras de ar para as quais se determinou a respetiva concentração de carbono. Os resultados (em ppm) foram os seguintes:

$$101.1, \quad 98.2, \quad 100.4, \quad 96.8, \quad 88.2.$$

Tendo conhecimento que tais concentrações se distribuem normalmente alguém com responsabilidade gostaria que a variância da concentração de carbono não excedesse o valor 23.0. Ao nível de significância de 5%, os dados revelam que ele pode estar tranquilo?

Resolução:

As hipóteses a testar são $H_0 : \sigma^2 \leq 23.0$ vs $H_1 : \sigma^2 > 23.0$ (teste unilateral direito)

A população tem distribuição normal e valor médio, μ, desconhecido, logo a estatística de teste é:

$$X^2 = \frac{(n-1)S^2}{23} \underset{\sigma^2=23.0}{\sim} \chi^2_{n-1}.$$

Para $\alpha = 0.05$, tem-se a região de rejeição $R_{0.05} =]\chi^2_{4;0.05}; +\infty[=]9.488; +\infty[$.

Com $s^2 = \frac{1}{5-1}\left(\sum_{i=1}^{30} x_i^2 - n\bar{x}^2\right) = (47094.09 - 5 \times 96.94^2)/4 = 26.818$, tem-se

$$\chi^2_{obs} = \frac{(5-1)s^2}{23} = 4.664.$$

Como $\chi^2_{obs} \notin R_{0.05}$ não rejeitamos H_0 ao nível de significância 5%, isto é, não existe evidência, ao nível de significância 5%, para rejeitar a hipótese de que a variância da concentração de carbono não excede o valor 23.0.

6. A altura (em mm) da espuma de sabão numa bacia é importante para os fabricantes de detergentes e supõe-se que tem distribuição Normal. Foi efetuada uma experiência, colocando a mesma quantidade de detergente em 10 bacias de tamanho standard e, depois de uma certa agitação da água, mediu-se a altura da espuma, obtendo-se $\sum_{i=1}^{10}(x_i - \bar{x})^2 = 1332$.

Teste, ao nível de significância 1%, a hipótese de o verdadeiro desvio padrão da população ser inferior a 20 mm.

Resolução:

$H_0 : \sigma \geq 20$ vs $H_1 : \sigma < 20$

A população tem distribuição normal e valor médio desconhecido, logo a estatística de teste é

$$X^2 = \frac{(n-1)S^2}{20^2} \underset{\sigma=20}{\sim} \chi^2_{n-1}.$$

Para $\alpha = 0.01$, tem-se a região de rejeição, para um teste unilateral esquerdo, $R_{0.01} =]0, \chi^2_{9;0.99}[=]0, 2.088[$.

Com $s^2 = \frac{1}{10-1} \sum_{i=1}^{10} (x_i - \bar{x})^2 = 1332/9$, tem-se

$$\chi^2_{obs} = \frac{(10-1)s^2}{20^2} = \frac{1332}{20^2} = 3.33.$$

Como $\chi^2_{obs} \notin R_{0.01}$ não rejeitamos H_0 ao nível de significância 1%, isto é, não existe evidência, ao nível de significância 1%, para rejeitar a hipótese de que o verdadeiro desvio padrão da população seja superior ou igual a 20 mm.

Capítulo 8

Regressão Linear

1. Pretende-se estabelecer uma relação entre os proventos anuais (x em milhares de Euros) das famílias de uma dada comunidade e a sua correspondente poupança anual (Y em milhares de Euros). Para tal foram obtidos os seguintes dados relativos a nove famílias:

x_i	12	12	14	15	16	17	19	19	20
y_i	0.0	0.1	0.2	0.2	0.4	0.6	0.6	0.7	0.8

Resolva as questões com base nos cálculos obtidos pelo R:

```
Coefficients:
              Estimate   Std. Error   t value   Pr(>|t|)
 (Intercept)  -1.0889    0.138324     -7.872    0.000101   ***
 x            0.0931     0.008513     10.931    1.19e-05   ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.07224 on 7 degrees of freedom
Multiple R-squared: 0.9447, Adjusted R-squared: 0.9367
```

(a) Estime a reta do modelo de regressão linear.

(b) Determine a estimativa do valor de poupança anual obtido pelo modelo para um provento anual familiar de 20 mil Euros.

(c) Usando o valor do coeficiente de determinação, o que pode concluir quanto à qualidade do ajustamento?

(d) Para um nível de significância de 5%, o que pode concluir quanto à contribuição de x na a explicação de Y.

(e) Determine um intervalo de confiança a 95% para a variância dos erros do modelo de regressão linear simples.

Resolução:

(a) A reta de regressão estimada é: $\hat{Y} = -1.0889 + 0.0931\,x$, uma vez que no output do R, a linha *intercept* corresponde à ordenada na origem (β_0), a linha **x** ao declive da reta (β_1) e a coluna *estimate* corresponde às estimativas desses parâmetros.

(b) Para se obter a estimativa do valor de poupança anual obtido pelo modelo para um provento anual familiar de 20 mil Euros, teremos de substituir x por 20 na reta de regressão estimada. Assim sendo o resultado será: $\hat{Y}\,(20) = -1.0889 + 0.0931 \times 20 = 0.7731$.

(c) Como o coeficiente de determinação, R^2, aparece no output do R em *Multiple R-squared*, temos que $R^2 = 0.9447 \geq 0.8$, pelo que a qualidade do ajustamento é "'razoável"'.

(d) Para se concluir se a variável x é relevante na explicação da variável Y devemos testar se o declive da reta de regressão (β_1) é ou não nulo. Assim sendo teremos de realizar o seguinte teste:

$$H_0 : \beta_1 = 0 \ vs \ H_1 : \beta_1 \neq 0$$

- Região de rejeição para $\alpha = 0.05$: $t_{7:\,0.025} = 2.365$ e $R_{0.05} =]-\infty, -2.365[\cup]2.365, +\infty[$.
- Valor observado da estatística de teste: $t_{obs} = 10.931$, que pode ser encontrado no output do R em *t value*.
- Decisão: $t_{obs} \in R_{0.05}$ logo rejeitamos H_0, isto é, x contribui para a explicação de Y ao nível de significância de 5%.

Observação: Tomaríamos a mesma decisão utilizando o $valor - p$, pois $valor - p = 1.19 \times 10^{-5} < 0.05$. Repare-se que este valor pode ser encontrado no output do R, em $Pr(>|t|)$.

(e) Queremos um intervalo de confiança a 95% para σ^2 e sabemos que esse intervalo é dado por:

$$IC_{95\%}(\sigma^2) \equiv \left]\frac{(n-2)\hat{\sigma}^2}{\chi^2_{n-2;0.025}} \, ; \, \frac{(n-2)\hat{\sigma}^2}{\chi^2_{n-2;0.975}}\right[.$$

Como, no output do R, $\hat{\sigma}$ =Residual standard error= 0.07224 e com $n = 9$, da tabela da distribuição Qui Quadrado virá $\chi^2_{7;0.975} = 1.690$ e $\chi^2_{7;0.025} = 16.013$, pelo que obteremos:

$$IC_{95\%}(\sigma^2) \equiv \left]\frac{7 \times 0.07224^2}{16.013} \, ; \, \frac{7 \times 0.07224^2}{1.690}\right[\equiv]0.0023 \, ; \, 0.0216[.$$

2. Pretende-se modelar o peso de pintos Y, em gramas, com a idade x, em semanas. Para tal registaram-se 10 valores da idade e o correspondente peso.

Idade, x	1	2	3	4	5	6	7	8	9	10
Peso, y	60	100	120	150	200	210	310	320	330	360

Resolva as questões com base nos cálculos obtidos pelo R:

Coefficients:

	Estimate	Std. Error	t value	$Pr(>\|t\|)$	
(Intercept)	22.6667	13.4487	1.6854	0.1304	***
x	35.1515	2.1675	16.2178	2.101e-07	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 19.687 on 8 degrees of freedom
Multiple R-squared: 0.97048, Adjusted R-squared: 0.96679

Assumindo que existe uma relação linear entre as variáveis X e Y:

(a) Escreva a reta estimada no modelo de regressão linear.

(b) Obtenha a estimativa do peso médio de um pinto a meio da 8ª semana, isto é, quando $x = 8.5$.

(c) Ao testar a hipótese de o verdadeiro declive da recta de regressão ser nulo, a 5% de significância, tem-se o $valor - p =$ ________________________, logo ________________________ a hipótese H_0 ao nível de significância de 5%.

Resolução:

(a) A reta estimada é $\hat{Y} = 22.6667 + 35.1515x$, uma vez que no output do R,a linha *intercept* corresponde à ordenada (β_0) na origem, a linha **x** ao declive (β_1) da reta e a coluna *estimate* corresponde às estimativas desses parâmetros.

(b) Para se obter a estimativa do peso médio de um pinto a meio da 8ª semana, basta substituir $x = 8.5$, na reta de regressão estimada $\hat{Y}(8.5) = 22.6667 + 35.1515 \times 8.5 = 321.4544$.

(c) Como testar se o declive da reta de regressão (β_1) é ou não nulo, corresponde a realizar o seguinte teste:

$$H_0 : \beta_1 = 0 \ vs \ H_1 : \beta_1 \neq 0$$

A afirmação ficará:
Ao testar a hipótese de o verdadeiro declive da recta de regressão ser nulo, a 5% de significância, tem-se o $valor - p = 2.101 \times 10^{-07} < 0.05$, logo rejeita-se a hipótese H_0 ao nível de significância de 5%.

Bibliografia sugerida

- Cordeiro e Magalhães(2004). *Introdução à Estatística. Uma perspectiva química*. Lidel-Edições Técnicas, Lda.

- Guimarães e Cabral(1997). *Estatística*. McGraw-Hill.

- Montgomery e Runger (2002). *Applied Statistics and Probability for Engineers*. Wiley.

- Mood, Graybill e Boes (1974). *Introduction to the Theory of Statistics*. McGraw-Hill.

- Murteira, B., Ribeiro, C., Silva, J. e Pimenta, C. (2007). Introdução à Estatística, 2ª edição. McGraw-Hill

- Paulino e Branco (2005). *Exercícios de Probabilidade e Estatística*. Escolar Editora.

- Pestana, D. e Velosa, S. (2002). *Introdução à Probabilidade e à Estatística*. Fundação Calouste Gulbenkian, Lisboa.

- Rohatgi (1976). *An Introduction to Probability Theory and Mathematical Statistics.* Wiley.

- Sokal e Rohlf (1995). *Biometry*. Freeman.

- Tiago de Oliveira (1990). *Probabilidades e Estatística: Conceitos, Métodos e Aplicações, vol. I, II*. McGraw-Hill.