FACULDADE DE
CIÊNCIAS E TECNOLOGIA

PROBABILIDADES E ESTATÍSTICA

Exercícios sobre Estatística

Edição 2024

Propriedades estocásticas de uma amostra aleatória simples (a.a)

Considere uma população X. O vector aleatório $(X_1, X_2, \ldots, X_n)$ é uma amostra aleatória simples (a.a.) de dimensão $n \in \mathbb{N}$, $n \geq 2$, se e só se:

- $X_1, X_2, \ldots, X_n$ são v.a.'s independentes;
- $X_1, X_2, \ldots, X_n$ são v.a.'s identicamente distribuídas com distribuição igual à da população X.

Consequências imediatas

Se a população X tem valor médio $\mu \equiv E(X)$ e variância $\sigma^2 \equiv V(X)$, então:

- $E(X_1) = E(X_2) = \ldots = E(X_n) = E(X) \equiv \mu$;
- $V(X_1) = V(X_2) = \ldots = V(X_n) = V(X) \equiv \sigma^2$

Estatísticas (estimadores) importantes e suas propriedades

Seja X uma população com valor médio $\mu \equiv E(X)$ e variância $\sigma^2 \equiv V(X)$

Para uma a.a. $(X_1, X_2 \ldots, X_n)$, $n \in \mathbb{N}$, $n \geq 2$, da população X

Parâmetro	Estimador	Propriedades	
$\mu \equiv E(X)$	$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$	$E(\overline{X}) = E(X) \equiv \mu$	$V(\overline{X}) = \frac{V(X)}{n} \equiv \frac{\sigma^2}{n}$
$\sigma^2 \equiv V(X)$	$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$	$E(S^2) = V(X) \equiv \sigma^2$	

- Momento populacional centrado de ordem $j \in \mathbb{N}$: $\mu_j = E\left[(X - E(X))^j\right] = E\left[(X-\mu)^j\right]$

  Observações:

  - $\mu_1 = E(X - \mu) = E(X) - \mu = \mu - \mu = 0$
  - $\mu_2 = E\left[(X-\mu)^2\right] = V(X) \equiv \sigma^2$

- Momento amostral centrado de ordem $j \in \mathbb{N}$: $M_j = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^j$

  Observações:

  - $M_2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = \frac{n-1}{n}S^2$
  - $\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = \sum_{i=1}^{n} X_i^2 - n\overline{X}^2$

Estimação Pontual

1. O fluxo de tráfego num cruzamento, define-se pelo número de veículos X que chegam a este cruzamento por minuto. Efetuou-se um conjunto de 20 observações independentes de X, com os seguintes resultados:

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

   (a) Determine a média, a variância e o desvio padrão amostrais.

   (b) Apresente o mínimo, o máximo e a moda amostrais.

   Solução: a) 1.7, 2.221052632, $\sqrt{2.221052632} \approx 1.490319641$ b) 0, 4, 0

2. Considere $(X_1, X_2, \ldots, X_n)$, $n \geq 2$, uma amostra aleatória de uma população X.

   (a) Se $X \sim P(\lambda)$, deduza o estimador dos momentos para o parâmetro λ.

   (b) Se $X \sim B(m, p)$, deduza o estimador dos momentos para o parâmetro p.

   (c) Se Se $X \sim G(p)$, deduza o estimador dos momentos para o parâmetro p.

   (d) Se $X \sim E(\lambda, \delta)$, deduza os estimadores dos momentos para os parâmetros λ e δ.

   (e) Se $X \sim E(\lambda, \delta_0)$, sendo δ_0 o valor conhecido para o parâmetro δ, deduza o estimador dos momentos para o parâmetro λ.

   (f) Se $X \sim E(\lambda_0, \delta)$,sendo λ_0 o valor conhecido para o parâmetro λ, deduza o estimador dos momentos para o parâmetro δ.

   (g) Se $X \sim U(a, b)$, deduza os estimadores dos momentos para os parâmetros a e b.

   (h) Se $X \sim N\left(\mu, \sigma^2\right)$, deduza os estimadores dos momentos para os parâmetros μ e σ^2.
   Apresente o estimador dos momentos para o parâmetro σ.

   (i) Admita que X tem distribuição de Pareto de parâmetros (δ, α), com $\delta \in \mathbb{R}^+$ e $\alpha \in]2, +\infty[$.
   Sabendo que $E(X) = \dfrac{\alpha\delta}{\alpha - 1}$ e que $V(X) = \dfrac{\alpha\delta^2}{(\alpha - 2)(\alpha - 1)^2}$, deduza os estimadores dos momentos para os parâmetros δ e α.

3. Seja $(X_1, X_2, \ldots, X_n)$, $n \geq 2$, uma amostra aleatória de uma população X. Admita que $E(X)$ e $V(X)$ existem.

   (a) Considere $\overline{X} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$ a média amostral.

      i. Mostre que $E\left(\overline{X}\right) = E(X)$.

      ii. Mostre que $V\left(\overline{X}\right) = \dfrac{V(X)}{n}$.

   (b) Considere $S^2 = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$. Mostre que $E\left(S^2\right) = V(X)$.
   Para tal, resolva as alíneas seguintes, .

      i. Verifique que $(n-1)S^2 = \displaystyle\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = \sum_{i=1}^{n} X_i^2 - n\overline{X}^2$

      ii. Demonstre que $E\left(\displaystyle\sum_{i=1}^{n} X_i^2\right) = n\left[V(X) + E^2(X)\right]$ e que $E\left(\overline{X}^2\right) = V\left(\overline{X}\right) + E^2\left(\overline{X}\right)$

      iii. Conclua por demonstração que $E\left(S^2\right) = V(X)$.

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

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

(a) Considere $\lambda^* = \overline{X}$, o estimador dos momentos para λ. Verifique que se trata de um estimador centrado. Deduza a sua variância e o seu erro padrão.

(b) Sabendo que a média e a variância amostrais são, respectivamente, $\overline{x} = 2.16$ e $s^2 = \dfrac{241.44}{99}$, apresente a estimativa dos momentos para λ assim como a estimativa do seu erro padrão.

Solução: a) –, λ/n, $SE(\lambda^*) = \sqrt{\lambda/n}$, b) $\lambda^* = \bar{x} = 2.16$, $SE^*(\lambda^*) = \sqrt{2.16/100}$

5. Seja X uma população com distribuição Exponencial de parâmetros (λ, δ_0) e uma amostra aleatória $(X_1, X_2, \ldots, X_n)$ desta população.

(a) Considere o estimador dos momentos para o parâmetro λ, $\lambda^* = \overline{X} - \delta_0$ (ver o exercício 2e) Mostre que se trata de um estimador centrado e determine a respectiva variância.

(b) Considere $\hat{\lambda} = \min(X_1, \ldots, X_n)$ um outro estimador para o parâmetro λ.

Sabemos que $\hat{\lambda}$ tem distribuição exponencial de parâmetros $\left(\lambda, \frac{\delta_0}{n}\right)$.

i. Mostre que $\hat{\lambda}$ não é um estimador centrado para λ e indique o respectivo enviesamento.

ii. Determine o erro quadrático médio do estimador $\hat{\lambda}$.

iii. Compare o erro quadrático médio destes estimadores para λ e indique qual o "melhor".

(c) Considere $\tilde{\lambda} = \min(X_1, \ldots, X_n) - \frac{\delta_0}{n} = \hat{\lambda} - \frac{\delta_0}{n}$

i. Confirme que $\tilde{\lambda}$ é um estimador centrado para o parâmetro λ.

ii. Averigue se o estimador λ^* é mais, menos ou tão eficiente que o estimador $\tilde{\lambda}$.

(d) Analise a consistência destes três estimadores para λ.

Solução: a) –, $\frac{\delta_0^2}{n}$ b)i. –, δ_0/n b)ii. $2\frac{\delta_0^2}{n^2}$, b)iii. $\hat{\lambda}$

6. Considere X o tempo (em anos) entre avarias consecutivas de um certo equipamento e a seguinte amostra de registos de tempos entre avarias consecutivas.

6.0	6.2	11.0	9.1	8.5	5.7	12.0	5.1	5.1	5.1
5.8	5.4	6.3	6.1	6.5	6.3	5.7	5.3	5.5	5.0
5.5	7.1	6.8	5.5	7.1	9.5	7.3	7.8	9.8	9.6
5.9	7.6	9.6	6.5	6.6	5.3	6.8	6.2	5.2	6.9
6.5	6.9	9.5	9.4	6.8	5.3	7.0	7.6	6.1	5.6

(a) Admitindo que X tem distribuição $E(\lambda, \delta)$, apresente as estimativas dos momentos para os seus parâmetros (ver o exercício 2d).

(b) Admita agora que X tem distribuição $E(\lambda, 2)$. Apresente as estimativas para λ, resultantes da utilização do estimador dos momentos λ^* (ver o exercício 2e) e dos estimadores $\hat{\lambda}$ e $\tilde{\lambda}$ (ver o exercício 5b e o exercício 5c). Para cada um destes estimadores, determine ainda a estimativa do seu erro quadrático médio.

Informação amostral: $\bar{x} = 6.9$, $\sum_{i=1}^{50} (x_i - \bar{x})^2 = 138.22$

Solução: a) $\delta^* = 1.66264849$, $\lambda^* = 5.23735151$, b) $\lambda^* = 4.9$, $\hat{\lambda} = 5$, $\tilde{\lambda} = 4.96$, $\widehat{EQM}(\lambda^*) = 0.08$, $\widehat{EQM}\left(\hat{\lambda}\right) = 0.0032$, $\widehat{EQM}\left(\tilde{\lambda}\right) = 0.0016$

7. Seja X uma população cuja distribuição depende do conhecimento do valor de um parâmetro $\delta \in \mathbb{R}^+$. Considere $(X_1, X_2, \ldots, X_n)$ uma amostra aleatória de X, $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ a média amostral e $\delta^* = \overline{X}/4$ um estimador de δ.
   Sabendo que $E(X) = 4\delta$ e que $V(X) = 4\delta^2$, verifique que δ^* é um estimador centrado com desvio padrão (erro padrão) igual a $\frac{\delta}{2\sqrt{n}}$.

8. Se (X_1, X_2, X_3) constitui uma amostra aleatória de uma população X com valor esperado μ e variância σ^2:
   (a) Verifique que $\hat{\mu} = \overline{X}$ e $\mu^* = \frac{1}{4}(X_1 + 2X_2 + X_3)$ são estimadores centrados do parâmetro μ.
   (b) Qual o estimador mais eficiente?

   Solução: b) $V(\hat{\mu}) = \sigma^2/3$, $V(\mu^*) = 3\sigma^2/8$, $\hat{\mu}$ é mais eficiente que μ^*

9. Seja X uma população com distribuição $N(\mu, \sigma^2)$.
   (a) Determine os estimadores de momentos para os parâmetros μ e σ^2. Indique um estimador de momentos para σ.
   (b) Mostre que o estimador de momentos para μ é centrado e indique a respectiva variância. Mostre que o estimador para σ^2 não é centrado. Determine o seu enviesamento.

   Solução: a) $\mu^* = \overline{X}$, ${\sigma^2}^* = M_2 = \frac{n-1}{n}S^2$, $\sigma^* = \sqrt{M_2} = \sqrt{\frac{n-1}{n}}S$ b) $V(\mu^*) = \sigma^2/n$, *bias* $\left({\sigma^2}^*\right) = -\frac{\sigma^2}{n}$

10. Considere X o volume de água por garrafa (em mililitros) e a seguinte amostra observada de volumes de 50 garrafas (em mililitros):

1503.50	1501.55	1503.37	1513.31	1497.78
1498.61	1496.55	1503.15	1496.94	1500.76
1506.22	1495.77	1511.71	1507.69	1497.39
1511.38	1500.11	1504.57	1509.51	1503.19
1510.99	1501.13	1504.07	1514.59	1504.84
1513.67	1494.41	1502.43	1504.58	1505.14
1494.08	1502.16	1514.86	1502.38	1503.39
1503.83	1502.98	1509.33	1508.38	1515.97
1510.48	1505.67	1516.88	1503.09	1496.29
1499.57	1503.17	1501.73	1508.79	1501.32

Admitindo que X tem distribuição Normal, apresente a estimativa de momentos para o volume médio de água por garrafa. Determine ainda uma estimativa centrada para a variância do volume de água por garrafa.

Solução: $\overline{x} = 1504.4652$, $s^2 = 33.83483771$

11. Suponha que a voltagem suportada por um cabo eléctrico (com um certo isolamento), sofre variações aleatórias que se distribuem de acordo com uma distribuição $N\left(\mu, \sigma^2\right)$. Para a seguinte amostra de voltagens suportadas no momento da falha de 12 cabos:

$$53 \quad 64 \quad 38 \quad 68 \quad 67 \quad 52 \quad 60 \quad 44 \quad 48 \quad 46 \quad 70 \quad 62$$

e considerando os seguintes estimadores para os parâmetros μ e σ^2,

$$\hat{\mu} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{e} \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = \frac{n-1}{n}S^2$$

determine as estimativas do valor esperado, variância e desvio padrão da população, bem como a estimativa da probabilidade de um cabo suportar voltagens superiores à voltagem máxima registada nesta amostra.

Solução: 56, ≈ 102.8333, ≈ 10.1407, $\hat{P}(X > 70) = P(Z > 1.38) = 0.0838$

12. Seja X uma população absolutamente contínua com função densidade de probabilidade

$$f_X(x) = \begin{cases} \dfrac{2x}{\theta^2}, &]0, \theta[\\ 0, & x \notin]0, \theta[\end{cases}$$

cujo $E(X) = \dfrac{2}{3}\theta$ e cuja $V(X) = \theta^2/18$.

Considere $(X_1, X_2, \ldots, X_n)$ uma amostra aleatória desta população, de dimensão $n \geq 2$, e as estatísticas

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \ \text{ e } \ M_n = \max(X_1, X_2, \ldots, X_n)$$

Tenha ainda em conta que $E(M_n) = \dfrac{2n}{2n+1}\theta$ e que $V(M_n) = \dfrac{n}{(n+1)(2n+1)^2}\theta^2$

(a) O estimador dos momentos para o parâmetro θ é $\theta^* = \dfrac{2}{3}\overline{X}$. Verifique que é um estimador centrado com variância $\dfrac{\theta^2}{18n}$.

(b) Seja $\tilde{\theta} = aM_n + b$ com $a \neq 0, b$ constantes reais não dependentes de θ, um estimador do parâmetro θ. Este estimador é centrado para o parâmetro θ se e só se $a = \dfrac{2n+1}{2n}$ e $b = 0$.

(c) O estimador $\hat{\theta} = M_n = \max(X_1, X_2, \ldots, X_n)$ tem um erro quadrático médio com expressão: $EQM\left(\hat{\theta}\right) = \dfrac{\theta^2}{(n+1)(2n+1)^2}$?

(d) Considere os estimadores $\theta^* = \dfrac{2}{3}\overline{X}$ e $\hat{\theta}$. No que diz respeito à sua dispersão como estimadores do parâmetro θ, qual é o melhor?

(e) Considere os estimadores θ^* e $\tilde{\theta} = \dfrac{2n+1}{2n}M_n$. Qual é o mais eficiente?

13. Considere a amostra aleatória $(X_1, X_2, \ldots, X_n)$ de uma população X com distribuição $B(1,p)$, em que $p \in]0,1[$ tem valor desconhecido.

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

(a) Mostre que $\hat{P} = \bar{X}$ é um estimador de centrado para p e determine a sua variância.

(b) Qual a distribuição exacta de $n\hat{P} = n\bar{X}$?

(c) Invoque o T.L.C. para justificar o seguinte resultado: Para $n \geq 30$,

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

Solução: a) $p(1-p)/n$ b) $B(n,p)$

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

Tipo	Duração	Custo	Tipo	Duração	Custo
L	55	0.041	M	68	0.496
M	60	0.248	LOP	20	0.07
LOP	116	0.096	M	20	0.248
M	70	0.496	L	97	0.075
L	91	0.083	LOP	36	0.07
M	72	0.496	M	33	0.248
L	27	0.041	LOP	93	0.085
L	49	0.07	M	171	0.744
L	200	0.131	INT	5	0.1
M	53	0.248	INT	502	2.992
L	58	0.07	M	27	0.248
L	27	0.07	L	137	0.081
M	394	1.735	M	190	0.992
L	395	0.216	M	62	0.496
LOP	39	0.07	L	85	0.074
M	26	0.248	M	41	0.248

Estime a proporção p, de chamadas locais (L) e o respectivo desvio padrão.
Determine o valor aproximado da probabilidade de obter uma estimativa de p com um erro absoluto superior a 5%.

Solução: $\hat{p} = 11/32$, $\sqrt{\hat{V}(\hat{p})} = \sqrt{\hat{p}(1-\hat{p})/n} = 0.083961661$, $P\left(\left|\hat{P} - p\right| > 0.05\right) \approx P(|Z| > 0.60) = 0.5486$

15. Seja uma população X com valor médio μ e variância σ^2. Considere $(X_1, X_2, \ldots, X_n)$ uma amostra aleatória de X, $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ a sua média aritmética e $S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$ um estimador da variância da população (isto é, de σ^2).

Antes de mais, recorde que, se Y é uma v.a. com valor médio e variância, então:

$$V(Y) = E\left(Y^2\right) - E^2(Y) \quad \text{e} \quad E\left(Y^2\right) = V(Y) + E^2(Y)$$

(a) Mostre que $(n-1)\, S^2 = \displaystyle\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = \left(\sum_{i=1}^{n} X_i^2\right) - n\overline{X}^2$.

(b) Verifique que $E\left[(n-1)\, S^2\right] = (n-1)\, \sigma^2$.

(c) Conclua que S^2 é um estimador centrado de σ^2. Conclua também que $\hat{\sigma}^2 = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$ é um estimador enviesado de σ^2, com um enviesamento *bias* $\left(\hat{\sigma}^2\right) = -\dfrac{\sigma^2}{n} < 0$.

16. (a) Considere $(X_1, X_2, \ldots, X_n)$ uma amostra aleatória de uma população X que se sabe ter distribuição Normal, com valor médio μ desconhecido e variância conhecida com valor σ_0^2. Adopte $\overline{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$ como estimador de μ.

i. Justifique a seguinte afirmação: $\sqrt{n}\dfrac{\overline{X} - \mu}{\sigma_0}$ tem distribuição Normal Reduzida.

ii. Se quisermos assegurar que o erro absoluto de amostragem na estimação de μ não excede $a \in \mathbb{R}^+$ com 0.95 de probabilidade, quais os valores admissíveis para μ?

(b) Considere $(X_1, X_2, \ldots, X_n)$ uma amostra aleatória de dimensão $n \geq 30$, de uma população X que se sabe ter valor médio μ desconhecido e variância conhecida com valor σ_0^2. Adopte $\overline{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$ como estimador de μ.

i. Invoque o T.L.C. para justificar que $\sqrt{n}\dfrac{\overline{X} - \mu}{\sigma_0}$ tem distribuição aproximada Normal Reduzida.

ii. Se quisermos assegurar que o erro absoluto de amostragem na estimação de μ não excede $a \in \mathbb{R}^+$ com probabilidade aproximada de 80%, quais os valores admissíveis para μ?

Solução: a) ii) $P\left(\left|\overline{X} - \mu\right| \leq a\right) = 0.95 \Leftrightarrow \mu \in \left[\overline{X} - 1.96\dfrac{\sigma_0}{\sqrt{n}}, \overline{X} + 1.96\dfrac{\sigma_0}{\sqrt{n}}\right]$,

b) ii) $P\left(\left|\overline{X} - \mu\right| \leq a\right) \approx 0.80 \Leftrightarrow \mu \in \left[\overline{X} - 1.28\dfrac{\sigma_0}{\sqrt{n}}, \overline{X} + 1.28\dfrac{\sigma_0}{\sqrt{n}}\right]$

Estimação por Intervalo de Confiança

17. Considere X uma população com distribuição Normal, de valor médio μ e variância igual a 9, e a seguinte amostra de valores desta população: $(9, 14, 10, 12, 7, 3, 11, 12, 3)$.

    (a) Determine as estimativas por intervalo de 80% e 90% de confiança para o valor médio da população.

    (b) Qual o nível de confiança associado ao intervalo para o valor médio, de amplitude igual a 2.7 ?

    Informação amostral: $n = 9$, $\overline{x} = 9$

    Solução: a) $IC_{80\%}(\mu) = [5.16, 12.84]$, $IC_{90\%}(\mu) = [4.05, 13.95]$ b) 82.3%

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

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

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

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

    985 992 1011 976 997 1000 1004 983 998

    (a) Determine estimativas por intervalo de confiança para o efectivo peso médio de uma embalagem, usando coeficientes de confiança de: 90%, 95% e 99%. Como varia a precisão do intervalo com o coeficiente de confiança escolhido?

    (b) Suponha que, em vez da amostra de pesos de 9 embalagens, tinha sido obtida uma amostra de pesos de 100 embalagens, e que esta fornecia uma média de 994 gramas. Estime o peso médio de uma embalagem por intervalo de 95% de confiança. Que ilação retira do aumento da dimensão da amostra?

    (c) Qual deverá ser o tamanho mínimo da amostra a recolher, de modo a que a amplitude do intervalo a 95% de confiança para o peso médio de uma embalagem seja de, no máximo, 2 gramas?

    (d) Admita agora que o desvio padrão σ do peso das embalagens tem um valor desconhecido. Determine estimativas por intervalo de confiança para o efectivo peso médio de uma embalagem, usando coeficientes de confiança de: 90%, 95% e 99%. Compare a amplitude destes intervalos com a amplitude dos intervalos obtidos na alínea a).

    (Robalo, 1994a)

    Solução: a) $IC_{90\%}(\mu) = [987.4, 1000.6]$, $IC_{95\%}(\mu) = [986.16, 1001.84]$, $IC_{99\%}(\mu) = [983.68, 1004.32]$,
    b) $IC_{95\%}(\mu) = [991.648, 996.352]$, c) 554,
    d) $IC_{90\%}(\mu) = [987.1378575, 1000.862143]$, $IC_{95\%}(\mu) = [985.4776617, 1002.522338]$, $IC_{99\%}(\mu) = [981.6038716, 1006.396128]$

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

Tipo A:	1405	1725	1610	1605	1950	1575	
Tipo B:	1615	1665	1730	1755	1632	1606	1792

Admita que o tempo de duração dos fatos de cada marca tem distribuição Normal.

(a) Estime por intervalos de 90% de confiança, o tempo médio de duração dos fatos de cada marca.

(b) Estime por intervalos de 99% de confiança, a variância do tempo de duração dos fatos de cada marca.

(c) Estime por intervalos de 99% de confiança, o desvio padrão do tempo de duração dos fatos de cada marca.

Solução: a) $IC_{90\%}(\mu_A) = [1495.306324, 1794.693676]$, $IC_{90\%}(\mu_B) = [1630.848527, 1739.151473]$
b) $IC_{99\%}(\sigma_A^2) = [9865.269461, 399878.6408]$, $IC_{99\%}(\sigma_B^2) = [1768.864865, 48408.28402]$,
c) $IC_{99\%}(\sigma_A) = [99.32406285, 632.3595819]$, $IC_{99\%}(\sigma_B) = [42.05787518, 220.0188265]$

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

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

Considere a evolução do montante dispendido em formação profissional no período decorrido de 5 anos, $D = X - Y$. (Admita que D tem distribuição normal).

(a) Estime por intervalo de 95% de confiança, a evolução média deste montante.

(b) Estime por intervalo de 90% de confiança, a variância da evolução D. Apresente também a estimativa por intervalo de 90% de confiança do desvio padrão de D.

Solução: a) $\overline{d} = -3.6$, $s_D^2 = 55.15555556$, $IC_{95\%}(\mu_D) = [-8.907659706, 1.707659706]$,
b) $IC_{90\%}(\sigma_D^2) = [29.37278107, 149.0690691]$, $IC_{90\%}(\sigma_D) = [5.41966614, 12.20938447]$

21. Com base numa amostra com n observações, retirada de uma população com distribuição $N(\mu, \sigma^2)$, obteve-se a seguinte estimativa por intervalo de confiança para o valor esperado da população: $[7.398, 12.602]$.

(a) Sabendo que $\sigma^2 = 64$, para um coeficiente de confiança de 80%, determine a média amostral.

(b) Sabendo que $\sigma = 4$ e que $n = 16$, determine o coeficiente de confiança associado ao intervalo. Sugestão: Apoie a sua resolução na amplitude do intervalo.

(c) Para uma amostra de dimensão $n = 11$ cuja variância amostral é $s^2 = 9$, determine a estimativa por intervalo de confiança a 80% para σ^2 e a estimativa por intervalo de confiança a 80% para σ.

(d) Para uma amostra de dimensão $n = 16$ com desvio padrão amostral $s = 4$, qual o valor do coeficiente de confiança associado à estimativa $[9.6, 33.05785124]$ por intervalo de confiança para σ^2?

Solução: a) $\overline{x} = 10$, b) 99.06% c) $IC_{80\%}(\sigma^2) = [5.625, 18.48049281]$, $IC_{80\%}(\sigma) = [2.371708245, 4.298894371]$ d) 90%

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

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

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

    Solução: a) $IC_{95\%}(\mu) = [2353.514518, 2446.485482]$, b) 86.12%

23. Seleccionaram-se aleatoriamente 100 clientes de um supermercado, tendo-se registado os respectivos tempos dispendidos a fazer as compras numa visita (em minutos). Verificou-se que a média e a variância amostrais desses tempos foram 35 m e 225 m^2, respectivamente, e que 62 dispenderam mais de meia hora a fazer as compras numa visita.

    (a) Estime por intervalo de 95% de confiança, o tempo médio dispendido/cliente ao fazer as compras numa visita, μ.

    (b) Determine uma estimativa por intervalo de 98% de confiança para a proporção de clientes que numa visita dispenderão mais de meia hora a fazer as compras.

    Solução: a) $IC_{95\%}(\mu) = [32.06, 37.94]$, b) Intervalo mais preciso $IC_{98\%}(p) = [0.503502859, 0.724138708]$, Intervalo menos preciso $IC_{98\%}(p) = [0.506904959, 0.733095041]$

24. Considere o tema do exercício 4. Resumindo,

    - População: X - n.º de veículos que chegam ao cruzamento/minuto
    - $X \sim P(\lambda) \qquad E(X) = \lambda \qquad V(X) = \lambda$
    - Estimador de λ: $\lambda^* = \overline{X}$
    - Informação amostral: $n = 100$, $\overline{x} = 2.16$, $s^2 = 241.44/99$

    (a) Estime, por intervalo de 95% de confiança, o número médio de veículos que chegam ao cruzamento (por minuto), tendo em conta as seguintes variáveis pivot para λ:

    i. $W = \sqrt{n}\dfrac{\overline{X} - \lambda}{S} \overset{a}{\sim} N(0, 1)$

    ii. $W = \sqrt{n}\dfrac{\overline{X} - \lambda}{\sqrt{\lambda}} \overset{a}{\sim} N(0, 1)$

    iii. $W = \sqrt{n}\dfrac{\overline{X} - \lambda}{\sqrt{\overline{X}}} \overset{a}{\sim} N(0, 1)$

    (b) Estime, por intervalo de 95% de confiança, o desvio padrão do número de veículos que chegam ao cruzamento (por minuto), tendo em conta as seguintes variáveis pivot para λ:

    i. $W = \sqrt{n}\dfrac{\overline{X} - \lambda}{S} \overset{a}{\sim} N(0, 1)$

    ii. $W = \sqrt{n}\dfrac{\overline{X} - \lambda}{\sqrt{\lambda}} \overset{a}{\sim} N(0, 1)$

    iii. $W = \sqrt{n}\dfrac{\overline{X} - \lambda}{\sqrt{\overline{X}}} \overset{a}{\sim} N(0, 1)$

    Solução: a) i. $IC_{95\%}(\lambda) = [1.853914268, 2.466085732]$, a) ii. $IC_{95\%}(\lambda) = [1.890508316, 2.467907684]$, a) iii. $IC_{95\%}(\lambda) = [1.871940006, 2.448059994]$, b) i. $IC_{95\%}\left(\sqrt{\lambda}\right) = [1.361585204, 1.570377576]$, b) ii. $IC_{95\%}\left(\sqrt{\lambda}\right) = [1.374957569, 1.570957569]$, b) iii. $IC_{95\%}\left(\sqrt{\lambda}\right) = [1.368188586, 1.564627749]$

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

Solução: Intervalo mais preciso $IC_{95\%}(p) = [0.731572138, 0.81319591]$,
Intervalo menos preciso $IC_{95\%}(p) = [0.734076932, 0.815923068]$

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

Homens (H):	26	40	56	28	36	55	41	61	53	50	52	37	50
Mulheres (M):	58	52	52	60	55	53	59	50	53	55	58		

Supondo a normalidade das populações,

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

(b) Estime por intervalo de 50% de confiança, a idade média de aparecimento dos primeiros sintomas, para os homens e para as mulheres.

Solução: a) $IC_{90\%}(\sigma_H) = [8.383657572, 16.79934443]$, $IC_{90\%}(\sigma_M) = [2.451719593, 5.283822643]$,
b) $IC_{50\%}(\mu_H) = [42.86220684, 47.13779316]$, $IC_{50\%}(\mu_M) = [54.3, 55.7]$

27. Seja X uma população cuja distribuição é caracterizada por dois parâmetros $(\lambda, \delta) \in \mathbb{R} \times \mathbb{R}^+$ e tal que: $E(X) = \lambda + \delta$ e $V(X) = \delta^2$. Considere $(X_1, X_2, \ldots, X_n)$ uma amostra aleatória desta população.

Suponha que o parâmetro δ tem um valor conhecido δ_0 e considere os dois estimadores do parâmetro λ: $\lambda^* = \overline{X} - \delta_0$ e $\hat{\lambda} = \min(X_1, \ldots, X_n)$.

(a) Para $n \geq 30$,

i. Invoque o T.L.C. para mostrar que $\sqrt{n}\dfrac{\lambda^* - \lambda}{\delta_0} \overset{a}{\sim} N(0,1)$.

ii. Justifique a seguinte afirmação: $W = \sqrt{n}\dfrac{\lambda^* - \lambda}{\delta_0}$ é uma variável pivot para λ.

iii. Recorra ao método pivotal para concluir que:

$$IC_{100(1-\alpha)\%}(\lambda) \equiv \left[\lambda^* - z_{\alpha/2}\frac{\delta_0}{\sqrt{n}}, \lambda^* + z_{\alpha/2}\frac{\delta_0}{\sqrt{n}}\right]$$

é um intervalo assintótico de confiança $100(1-\alpha)\%$ para λ.

(b) Relativamente ao estimador $\hat{\lambda}$, considere o seguinte resultado: A estatística $W = n\dfrac{\hat{\lambda} - \lambda}{\delta_0}$ tem uma distribuição por amostragem caracterizada pela função de distribuição,

$$F_W(w) = P(W \leq w) = \begin{cases} 0, & w > 0 \\ 1 - e^{-w}, & w \geq 0 \end{cases}$$

isto é, $W \sim E(0,1)$.

i. Justifique a seguinte afirmação: W é uma variável pivot para λ.

ii. Complete os seguintes passos do método pivotal para a dedução de um intervalo de confiança $100(1-\alpha)\%$ para λ.

- Variável pivot: $W = n\dfrac{\hat{\lambda} - \lambda}{\delta_0} \sim E(0,1)$
- Se $P(W \leq a) = \frac{\alpha}{2} \Leftrightarrow a =$ ____________ e, se $P(W > b) = \frac{\alpha}{2} \Leftrightarrow b =$ ____________

- $a \leq W \leq b \Leftrightarrow a \leq n\dfrac{\hat{\lambda}-\lambda}{\delta_0} \leq b \Leftrightarrow \hat{\lambda} - b\dfrac{\delta_0}{n} \leq \lambda \leq$ ______________
- $IC_{100(1-\alpha)\%}(\lambda) \equiv \left[\text{______________} \; , \; \hat{\lambda} + \ln\left(1-\frac{\alpha}{2}\right)\dfrac{\delta_0}{n}\right]$

(c) Considere a população X-tempo entre avarias consecutivas do equipamento (em anos) apresentada no exercício 6. Admita que X tem distribuição Exponencial com parâmetros (λ, δ). com as características inicialmente descritas neste exercício e que $\delta_0 = 2$. Usando os resultados deste exercício, apresente as duas estimativas por intervalo de 95% de confiança para λ.
Comente a qualidade da estimação por intervalo de 95% confiança do parâmetro λ.

Solução: b) ii) $a = -\ln(1-\alpha/2)$, $b = -\ln(\alpha/2)$, $\hat{\lambda} - a\frac{\delta_0}{n}$, $\hat{\lambda} + \ln(\alpha/2)\frac{\delta_0}{n}$ c) $IC_{100(1-\alpha)\%}(\lambda) = [4.345628284, 5.454371716]$, $IC_{100(1-\alpha)\%}(\lambda) = [4.852444822, 4.998987288]$

28. Considere X uma população com distribuição Uniforme no intervalo $]0, \theta]$, $\theta \in \mathbb{R}^+$ e $(X_1, X_2, \ldots, X_n)$ uma amostra aleatória desta população.

(a) Mostre que o estimador dos momentos de θ é $\theta^* = 2\overline{X}$. Verifique que é um estimador centrado com variância $V(\theta^*) = \dfrac{\theta^2}{3n}$.

Apresente também o estimador dos momentos para a mediana de X.

(b) Para uma amostra de dimensão $n \geq 30$, mostre que $\sqrt{3n}\,\dfrac{\theta^* - \theta}{\theta} \overset{a}{\sim} N(0,1)$.

(c) Deduza o intervalo de 80% de confiança para o parâmetro θ, usando a distribuição assintotica da variável pivot $W = \sqrt{3n}\dfrac{\theta^* - \theta}{\theta}$.

(d) Considere agora $\widehat{\theta} = \max(X_1, X_2, \ldots, X_n)$ outro estimador para o parâmetro θ. Sabendo que $W = \dfrac{\widehat{\theta}}{\theta}$ é uma variável pivot e que a sua função distribuição é

$$F_W(w) = \begin{cases} 0, & w < 0 \\ w^n, & 0 \leq w < 1 \\ 1, & w \geq 1 \end{cases}$$

complete os seguintes passos da metodologia pivotal para a obtenção de um intervalo de 80% de confiança para θ:

- Variável pivot: $W = \dfrac{\widehat{\theta}}{\theta}$
- Se $P(W \leq a) = P(W > b) = 0.1$, então

  $a =$ __________________ e $b =$ __________________

- $a \leq W \leq b \Leftrightarrow$ ______________ $\leq \theta \leq 0.1^{-1/n}\,\widehat{\theta}$

- $IC_{80\%}(\theta) \equiv [$______________ , ______________$]$

(e) Considere a amostra $(1.23, 1.38, 1.16, 0.15)$ de observações de uma população com distribuição $U(0, \theta)$.

i. Calcule a estimativa dos momentos de θ assim como o seu erro padrão.

ii. Determine a estimativa de θ por intervalo de 80% de confiança.

Solução: a) $\theta^*/2$, b) –, c) $IC_{80\%}(\theta) \equiv \left[\frac{\sqrt{3n}\,\theta^*}{\sqrt{3n}+1.28}, \frac{\sqrt{3n}\,\theta^*}{\sqrt{3n}-1.28}\right]$ d) –, e) 1.96, 0.565803264, $IC_{80\%}(\theta) = [1.416832333, 2.454025586]$

Teste de Hipóteses

29. De entre os alunos do 2º ano de uma escola de ensino básico, foram escolhidos ao acaso 9 alunos para serem inquiridos sobre o n.º de horas dispendidas/semana a ver televisão, X. As respostas foram:

$$16 \quad 7 \quad 5 \quad 11 \quad 12 \quad 10 \quad 14 \quad 9 \quad 15$$

Por outros estudos, podemos aceitar que X tem distribuição Normal com $\sigma^2 = 36$.

(a) Esta amostra permite confirmar as suspeitas de que o tempo médio dispendido semanalmente a ver televisão, μ, difere de 14 horas, ao nível de significância de 10%?

(b) Responda à alínea anterior calculando o $p - value$.

(c) Admita agora que se desconhece o valor da variância da população, σ^2.

i. Determine a estimativa pontual de σ^2, usando o estimador centrado, S^2.

ii. Repita os pedidos das alíneas a) e b).

(d) Admita agora que é recomendável que o tempo médio dispendido semanalmente a ver televisão, μ, não exceda as 10 horas. Ao nível de 10% de significância, esta amostra corrobora esta recomendação? Responda tendo em conta as situações em que σ é conhecido ou σ é desconhecido. Determine também o $p - value$.

Informação amostral: $\bar{x} = 11$, $s^2 = 13.5$

Solução: a) $z_{obs} = -1.5$, $z_{0.05} = 1.65$, $R_{0.1} =]-\infty, -1.65[\cup]1.65, +\infty[$, Não rejeitar H_0

b) $p - value$ =0.1336 c) i. $s^2 = 13.5$

c) ii. $t_{obs} = -2.449489743$, $t_{8:0.05} = 1.86$, $R_{0.1} =]-\infty, -1.86[\cup]1.86, +\infty[$, Rejeitar H_0

$p - value = 0.039968524$, $0.02 \leq p - value \leq 0.05$

d) σ conhecido: $z_{obs} = 0.5$, $R_{0.1} =]1.28, +\infty[$, Não rejeitar H_0, $p - value$ =0.3085

σ desconhecido: $t_{obs} = 0.816496581$, $R_{0.1} =]1.40, +\infty[$, Não rejeitar H_0, $p - value$ =0.218925811,

$0.2 \leq p - value \leq 0.25$

30. A temperatura média da água que é expelida de uma conduta da torre de refrigeração de uma central termoeléctrica não deve ultrapassar os 100^{o}F. Sabe-se que o desvio padrão da temperatura da água é de 4^{o}F. Foi medida a temperatura da água em 64 dias aleatoriamente seleccionados, e a média correspondente foi de 101^{o}F. A temperatura da água encontra-se fora dos valores aceitáveis, ao nível de 5% de significância? Determine o $p - value$.

Solução: $z_{obs} = 2$, $R_{0.05}(\mu) \approx]1.65, +\infty[$, Rejeitar H_0, $p - value = 0.0228$

31. A quantidade de lixo (toneladas) produzida no concelho do Xeisal, por dia, é uma variável aleatória com distribuição Normal. De forma a avaliar o que se passa no concelho, em relação a esta variável, seleccionaram-se 15 dias ao acaso para os quais se registaram as correspondentes quantidades de lixo produzidas, resultando numa média $\bar{x} = 100.1$ toneladas e num desvio padrão amostral $s = 0.18$ toneladas.

(a) Assumindo que $\sigma = 0.2$, teste a hipótese de que a quantidade média de lixo produzido por dia difere de 100 toneladas. Complete o seguinte output estatístico e decida com um nível de significância de 5%.

n:
$\overline{x}$:
z_{obs}:
$p - value$:
α:
Decisão:

(b) É desejável que a quantidade média diária de lixo produzida não ultrapasse as 100 toneladas. Teste a violação deste objectivo, admitindo agora que não se conhece o valor de σ. Preencha o seguinte output estatístico e apresente a sua decisão para um nível de 5% de significância. Confirme, usando uma calculadora, o valor apresentado para o $p - value$.

$H_0 : \mu$ __ __ $H_1 : \mu$ __ __
n:
$\overline{x}$:
s:
t_{obs}:
Graus de liberdade, gl:
α:
$t_{gl:\alpha}$:

Decisão:

$p - value$ 0.024682182

Solução: a) $z_{obs} = 1.936491673$, $R_{0.05} =]-\infty, -1.96[\cup]1.96, +\infty[$, Não rejeitar H_0, $p - value = 0.0524$
b) i) $t_{obs} = 2.151657415$, $R_{0.05} =]1.76, +\infty[$, Rejeitar H_0

32. Uma empresa tenciona importar um grande número de artigos, de custo elevado. O fabricante nacional garante que o peso médio/artigo é superior ou igual a 92 g. Como a qualidade do produto depende essencialmente do seu peso, a administração da empresa decide pela importação dos artigos quando não se assegura a garantia do fabricante nacional. Para tal, solicita ao seu departamento técnico os esclarecimentos necessários sobre o peso médio/instrumento do fabricante nacional, afim de decidir sobre a importação (com custos acrescidos) ou não. Para o efeito, o departamento técnico seleccionou aleatoriamente 49 instrumentos que forneceram os seguintes resultados referentes aos respectivos pesos:

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

Diga qual a resposta que espera do departamento técnico, considerando o seguinte output de uma aplicação estatística (que deverá preencher) e um nível de significância de 2%.

$H_0 : \mu$ __ __ $H_1 : \mu$ __ __
n :
$\bar{x}$:
s^2 :
Valor observado da estatística teste:
$p - value$:

Solução: $n = 49$, $\overline{x} = 90$, $s^2 = 22.5$, $w_{obs} = -2.951459149$, $R_{0.02} =]-\infty, -2.05[$, $p - value = 0.0016$, Rejeitar H_0

33. O departamento de segurança de uma fábrica quer assegurar um tempo esperado superior a 30 minutos, para os tempos que o empregado nocturno da segurança dispende nas vistorias á fabrica. Admita que o tempo (em minutos) que o empregado dispende numa volta à fábrica tem distribuição Normal.
Em 16 voltas, a média dos tempos observados foi de 28.8 minutos, com um desvio padrão $s = 1.5$ minutos. Complete o output estatístico abaixo exibido e diga se, ao nível de 1% de significância, é de admitir a hipótese considerada.

$H_0 : \mu$ __ __ $H_1 : \mu$ __ __
n:
$\bar{x}$:
s:
Valor observado estatística teste:
α:
Graus de liberdade (gl):
$t_{gl:\alpha}$:

$P(T > 3.2)$: 0.002981924
$p - value$:

Decisão:

Solução: $t_{obs} = -3.2$, $gl = 15$, $t_{15:0.01} = 2.60$, $R_{0.01} =]-\infty, -2.60[$, $p - value = 0.002981924$, Rejeitar H_0

34. Um comerciante retalhista recebe carregamentos de roupa de homem em que habitualmente $p_0 = 10\%$ das peças estão defeituosas. A fim de se certificar de que a qualidade do produto não se alterou, resolve inspeccionar 100 peças, determinar a percentagem de peças defeituosas e conduzir um teste de hipóteses com nível de significância igual a 4%.
   (a) Especifique a hipótese nula que está em causa assim como a hipótese alternativa.
   (b) Indique a estatística de teste e determine a região de rejeição.
   (c) Suponha que, nas peças verificadas, foram encontradas 12 defeituosas. O comerciante deve ou não rejeitar o carregamento?

Solução: a) $H_0 : p = 0.1$ vs $H_1 : p \neq 0.1$ b) $\frac{10}{0.3}\left(\hat{P} - 0.1\right) \underset{p=0.1}{\overset{a}{\sim}} N(0,1)$, $R_{0.04} =]-\infty, -2.05[\cup]2.05, +\infty[$
c) $\hat{p} = 0.12$, $w_{obs} = 0.6(6)$, Não rejeitar H_0

35. Um laboratório lançou no mercado um novo medicamento para o tratamento de uma alergia, afirmando que a sua eficácia, num período de 8 horas, é de 90% ou mais. A sua aplicação a 200 indivíduos sofrendo de tal alergia revelou-se eficaz em 168 casos.
Será que a afirmação do laboratório é consistente com os dados obtidos, ao nível de 3% de significância?

---

n:
$\hat{p}$:
Hipóteses: H_0 : vs H_1 :
Valor observado estatística teste:
$p - value$:
Decisão ao nível de 3% de significância:

Solução: $n = 200$, $\hat{p} = 0.84$, $w_{obs} = -2.828427125$, $p - value$=0.0023, Não

36. Considere a informação e a amostra tratada no exercício 4, relativa ao número de veículos que chegam ao cruzamento (por minuto). Admita que se decide construir uma rotunda caso se comprove estatisticamente, que o número médio de veículos que chegam ao cruzamento (em cada minuto), é *superior* a 2.

   (a) As hipóteses a testar são: H_0 : __ __ __ vs H_1 : __ __ __

   (b) Justifique porque a estatística de teste apropriada é $W = \sqrt{n}\frac{\overline{X}-2}{\sqrt{2}} \underset{\lambda=2}{\overset{a}{\sim}} N(0,1)$.

   (c) Determine o valor observado da estatística W e o valor aproximado do $p - value$ associado ao teste.

   (d) Qual a decisão a tomar para um nível de 10% de significância?

Informação amostral: $n = 100$, $\overline{x} = 2.16$
Solução: a) $H_0 : \lambda \leq 2$ vs $H_1 : \lambda > 2$, – , c) $w_{obs} = 1.13137085$, $p - value$=0.1292 d) Não rejeitar H_0

37. A cadeia de supermercados referida no exercício 25, abrirá uma nova loja se a proporção de pessoas do bairro que poderão vir a ser clientes habituais, se cifrar em mais de 75%.

   (a) Especifique as hipóteses que devem ser testadas.

   (b) Para um nível de 10% de significância, apresente a correspondente região de rejeição.

   (c) Face aos dados obtidos na sondagem, qual a atitude a tomar?

   (d) Determine o $p - value$.

Solução: a) $H_0 : p \leq 0.75$ vs $H_1 : p > 0.75$ b) $R_{0.1} =]1.28, +\infty[$ c) $\hat{p} = 0.775$, $w_{obs} \approx 1.1547$, Não rejeitar H_0
d) $p - value$=0.1251

38. No registo do peso (em kg) de 25 sacos de adubo, verificou-se que $\sum_{i=1}^{25} x_i = 1255$ e que $\sum_{i=1}^{25} x_i^2 = 63253$.

   (a) Apresente uma estimativa pontual e centrada da variância do peso dos sacos de adubo.

   (b) Admitindo que o peso/saco de adubo tem distribuição normal, teste a hipótese do peso ter um desvio padrão diferente de 3.5 kg, com um nível de 10% de significância.

Solução: a) $s^2 = 10.5$, b) $R_{0.10} = [0, 13.8[\cup]36.4, +\infty[$, $x^2_{obs} = 20.57142857$, Não rejeitar H_0

39. Considere a informação prestada no exercício 19.

   (a) Teste a hipótese da variância do tempo de duração dos fatos da marca A ser superior a 20000 $horas^2$, para um nível de significância de 2.5%. Sabendo que $P(X^2 > 8.2375) = 0.143623161$, indique o p-value associado ao teste.

   (b) Ao nível de 5% de significância, poderá afirmar que o desvio padrão do tempo de duração dos fatos de marca B é inferior a 100 horas?
   Recorrendo aos valores tabelados para a distribuição da estatística de teste, apresente o intervalo de valores possíveis para o p-value.

Informação amostral: $n_A = 6$, $\bar{a} = 1645$, $s^2_A = 32950$ $n_B = 7$, $\bar{b} = 1685$, $s^2_B = 5454$
Solução: a) $R_{0.025} = [12.8, +\infty[$, $x^2_{obs} = 8.2375$, Não rejeitar H_0, $p - value = 0.143623161$
Solução: b) $R_{0.05} = [0, 1.64[$, $x^2_{obs} = 3.2724$, Não rejeitar H_0, $0.2 \leq p - value \leq 0.3$

40. Um fabricante garante um mínimo de 5000 horas para a vida útil média de determinado componente. Ultimamente as reclamações das empresas clientes apontam no sentido de que a referida duração média não atinge as 5000 horas. A fim de se certificar se as reclamações são de atender, o fabricante escolheu ao acaso 49 componentes e registou as respectivas horas de vida útil. Obteve então a seguinte informação: uma média amostral da vida útil de 4940 horas e um desvio padrão amostral de 200 horas.

   (a) Para $\alpha = 2.5\%$, que pode concluir o fabricante? Determine o valor aproximado do $p-value$ associado a este teste.

   (b) Admita agora que é conhecida a variância, σ^2, da vida útil/componente, e que $\sigma^2 = 150^2$.
   Para $\alpha = 0.025$, que pode concluir o fabricante? Determine o valor aproximado do $p-value$ associado a este teste.

   Solução: a) $w_{obs} = -2.1$, $R_{0.025} \approx]-\infty, -1.96[$, Rejeitar H_0, $p - value \approx 0.0179$
   Solução: b) $w_{obs} = -2.8$, $R_{0.025} \approx]-\infty, -1.96[$, Rejeitar H_0, $p - value \approx 0.0026$

41. Seja p a probabilidade de obter cara no lançamento de uma moeda.
Para testar as hipóteses $H_0: \ p = 1/2 \ vs \ H_1: \ p = 3/4$, serão feitos 4 lançamentos independentes desta moeda e, no final, registar-se-á X - n.º total de caras observadas.
Considere X a estatística de teste.

   (a) Analise a veracidade das seguintes afirmações:

$$(X \,|H_0 \text{ verdadeira}) \sim B\,(4, 1/2) \equiv X \underset{p=1/2}{\sim} B\,(4, 1/2)$$

   e

$$(X \,|H_0 \text{ falsa}) \sim B\,(4, 3/4) \equiv X \underset{p=3/4}{\sim} B\,(4, 3/4)\,.$$

   (b) Para seguintes regiões de rejeição:

$$R_1 = \{2, 3, 4\} \qquad R_2 = \{3, 4\} \qquad R_3 = \{4\}$$

   calcule, com base nos valores da tabela seguinte, as probabilidades dos erros de tipo I e II associados a cada uma das regiões de rejeição.

X-n.º caras observadas	$H_0: \ p = 1/2$	$H_1: \ p = 3/4$
0	0.0625	0.0039
1	0.2500	0.0469
2	0.3750	0.2109
3	0.2500	0.4219
4	0.0625	0.3164

   (c) Escolha, justificando, a região de rejeição mais apropriada para definir o teste.

   Solução: b) R_1: $\alpha = 0.6875$, $\beta = 0.0508$; R_2: $\alpha = 0.3125$, $\beta = 0.2617$; R_3: $\alpha = 0.0625$, $\beta = 0.6836$ c) R_2

Testes de Hipóteses para a diferença de valores médios populacionais e para a comparação de variâncias populacionais

42. Durante um ano e em duas cidades (A e B), numa amostra de freguesias, o número de licenças concedidas para alojamento local foram:

Cidade	N.º licenças							
A	46	51	66	43	60	68	41	65
B	56	32	61	43	53	22	41	

Admita que, para as duas cidades, o n.º de licenças tem distribuição Normal.

Nas perguntas seguintes, arredonde todos os valores intermédios e finais a duas casas decimais.

(a) Complete a seguinte tabela referente aos seguintes valores amostrais relevantes:

Cidade	Média	Variância
A		
B		

(b) Use o nível de significância de 0.01 para testar se é diferente o n.º médio de licenças atribuídas por cidade. Indique o pressuposto relevante para a realização deste teste.

(c) Teste ao nível 0.10 de significância se existe algum motivo para duvidar da suposição de que as variâncias das duas populações são iguais.

43. Num programa de formação profissional, alguns formandos são instruídos pelo método 1 e outros formandos são instruídos pelo método 2. Ambos os métodos de instrução envolvem lições administradas por computador, mas o método 1 também envolve a atenção pessoal de um formador. A atribuição de formandos a cada método de intrução é aleatória. Amostras aleatórias de dimensão 10 são retiradas de grandes grupos de formandos instruídos pelos dois métodos. Os valores a seguir são as classificações resultantes de um teste de desempenho.

Método	Classificação										Média	s^2
1	81	71	79	83	76	75	84	90	83	78	80	29.11
2	59	65	62	59	57	64	60	56	66	62	61	11.33

Sejam X_1 e X_2 as classificações dos formandos intruídos pelos métodos 1 e 2, respectivamente. Admita-se que $X_1 \sim N\left(\mu_1, \sigma_1^2\right)$, que $X_2 \sim N\left(\mu_2, \sigma_2^2\right)$ e que $\sigma_1^2 = \sigma_2^2$.

(a) Teste a hipótese $H_1 : \mu_1 - \mu_2 < 10$, com um nível de significância de 5%.

(b) Realize o teste ao pressuposto $\sigma_1^2 = \sigma_2^2$, ao nível de 0.02 de significância.

44. Considere o tema descrito no exercício 19 e o resumo amostral referente a outras amostras:

Marca	n	Média	s^2
A	32	1645	32950
B	45	1685	5454

Teste a hipótese da duração média dos fatos da marca A ser inferior à duração média dos fatos da marca B, para um nível de significância de 2%.
Determine o valor aproximado do p-value associado a este teste.

Regressão Linear

45. Determinada empresa está interessada em estudar o n.º de horas de funcionamento/dia do ar condicionado, em função da temperatura exterior diária (em ºC), e durante o verão. Assim, seleccionaram-se 14 dias de verão, ao acaso, para os quais se mediram as temperaturas (x) e se registaram o número de horas de funcionamento do ar condicionado (Y):

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14
x_i	29	28	29	34	26	25	32	30	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	13.4	14.2	12.3	8.7

(a) Disponha os dados em gráfico.

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

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

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

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

Solução: b) $\widehat{Y}(x) = -24.90254237 + 1.205084746\, x$, c) $R^2 = 0.871921291$, d) $R_{0.05} =]-\infty, -2.18[\cup]2.18, +\infty[$, $t_{obs} = 9.038384089$, e) $\widehat{Y}(30) = 11.25$, –

46. Pretende-se utilizar um modelo de regressão linear simples para se explicar a quantidade (Kg) de vidro, Y, depositado num ecoponto, usando como variável independente, x, o número de dias sem despejar o mesmo. Para tal, registaram-se os seguintes dados:

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

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

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

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

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

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

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

Solução: a) $\widehat{Y}(x) = 10.632184 + 56.130268\, x$, $R^2 = 0.985981$, b) $R_{0.05} =]-\infty, -2.45[\cup]2.45, +\infty[$, $t_{obs} = 20.54265454$, c) $IC_{95\%}(\beta_0) = [-78.051801, 99.31616882]$, d) $\widehat{E}(Y(10)) = 571.9348659$, – , e) $IC_{90\%}(\sigma^2) = [1855.804902, 14258.01327]$, f) $IC_{95\%}(Y(14)) = [632.5475305, 960.3643469]$

47. Considere a amostra apresentada no exercício 14 sobre a duração, tipo e custo das chamadas telefónicas. Se admitir um modelo de regressão linear para x-duração e Y-custo, de cada chamada,

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

   (b) Determine o coeficiente de determinação.

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

   Solução: a) $\widehat{Y}(x) = -0.038343194 + 0.003872547\,x$, b) $R^2 = 0.599382109$, c) $R_{0.05} =]-\infty, -2.04[\cup]2.04, +\infty[$, $t_{obs} = 6.699576412$

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

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

   (a) Esboce o gráfico dos dados amostrais.

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

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

   (d) Interprete as estimativas dos parâmetros do modelo de regressão.

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

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

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

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

   Solução: b) $\widehat{Y}(x) = 110.301964 + 1.577741408\,x$, c) $\hat{\sigma}^2 = 29.57682079$, e) $R^2 = 0.927828$, f) $R_{0.01} =]-\infty, -3.36[\cup]3.36, +\infty[$, $t_{obs} = 10.1413479$, g) $\hat{E}(Y(50)) = 189.1890344$, $IC_{95\%}(E(Y(50))) = [184.9642345, 189.1890344]$, h) $IC_{95\%}(Y(50)) = [175.9348348, 202.443234]$

Teste de Ajustamento do Qui-Quadrado

49. O fluxo de tráfego num cruzamento, definido pelo número de veículos X que chegam ao cruzamento por minuto. Suspeita-se que X tem distribuição de Poisson com parâmetro $\lambda = 2$, $X \sim P(2)$.
Efectuou-se um conjunto de 1000 observações independentes de X, com a distribuição de frequências absolutas que se apresenta na seguinte tabela:

Nº veículos	0	1	2	3	4	5	6	≥ 7
Frequência	149	294	257	165	81	44	8	2

(a) Averigue, recorrendo ao teste de ajustamento do Qui-Quadrado, se existe evidência estatística para rejeitarmos a hipótese $H_0: X \sim P(2)$, ao nível de significância de 20%.

(b) O que concluiria, para um nível de significância de 10%?

Classes	O_i	p_{0i}	E_i	$\frac{(O_i - E_i)^2}{E_i}$
0	149	0.1353	135.3	1.3872
1	294	0.2707		
2	257			0.6934
3		0.1804	180.4	1.3146
4	81	0.0902	90.2	
5	44	0.0361	36.1	1.7288
6	8	0.012	12.0	1.3333
≥ 7				
Total		1		

Solução: $x^2_{obs} = 10.8708$, Rejeitar H_0, Não rejeitar H_0

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

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

Sabendo que a média amostral é $\bar{x} = 70$, teste a hipótese $H_1: X \nsim N(\mu, 2^2)$, com um nível de 10% de significância.

Classes	O_i	p_{0i}	E_i	$\frac{(O_i - E_i)^2}{E_i}$
$]-\infty, 65]$	15			27.5841
$]65, 67]$	26	0.0606		
$]67, 69]$	151			1.0853
$]69, 71]$	293	0.383	260.44	4.0706
$]71, 73]$	167	0.2417	164.356	0.0425
$]73, 75]$	17	0.0606	41.208	14.2212
$]75, +\infty[$	11		4.216	
Total	680	1		

Solução: $x^2_{obs} = 63.5325$, Rejeitar H_0

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

Valores observados em 30 horas

41	31	28	41	28	26	28	41	30	34
40	36	30	20	43	36	36	20	42	43
42	40	32	26	28	41	34	24	42	40

Considere $\alpha = 5\%$ e os dados agrupados nas seguintes classes: $]-\infty, 24]$, $]24, 28]$, $]28, 32]$, $]32, 36]$, $]36, 40]$ e $]40, \infty]$. Outras informações amostrais: $\overline{x} = 34.1$ e $s \approx 7.14$

Classes	O_i	p_{0i}	E_i	$\frac{(O_i - E_i)^2}{E_i}$
$]-\infty, 24]$				
$]24, 28]$		0.1184		
$]28, 32]$				0.4799
$]32, 36]$		02205		0.3943
$]36, 40]$		0.1903	5.709	1.2855
$]40, +\infty[$			6.099	
Total	680	1		

Solução: $x^2_{obs} = 5.3888$, $g.l. = 6 - 2 - 1 = 3$; $\chi^2_{3:0.05} = 7.81$, Não rejeitar a hipótese $H_0 : X \sim N(\mu, \sigma^2)$

52. Num centro de inspecção automóvel, para $n = 40$ veículos inspeccionados registaram-se os correspondentes valores acerca do n.º de visitas até ficarem aprovados:

1	1	2	1	2	3	1	2	2	1	1	1	1	2	1	1	2	1	1	1
1	1	1	2	2	1	1	1	1	3	1	1	4	2	1	1	1	1	2	1

Teste se, X-n.º de visitas necessárias até um veículo ficar aprovado, tem distribuição Geométrica de parâmetro $p = 0.8$. Para tal complete o seguinte output:

- $H_0 : X \sim G(0.8)$ vs $H_1 : X \nsim G(0.8)$

- Tabela de frequências observadas e esperadas

i	A_i	O_i	p_{0i}	E_i	$(O_i - E_i)^2 / E_i$
1	1				
2	2		0.16	6.4	2.025
3	≥ 3				
Total	–				

- $x^2_{obs} =$
- $g.l. =$ $R_{0.1} =]$ $, +\infty[$
- Decisão: ________________ H_0, ao nível de 10% de significância.

Solução: $x^2_{obs} = 4.0313$