Probabilidades e Estatística

TODOS OS CURSOS

Exame Época Especial
2016/2017
24/07/2017 – **09:00**

Duração: 3 horas

Justifique convenientemente todas as respostas

Grupo I	5 valores

1. Uma companhia de seguros divide os seus clientes em três classes A, B e C. De acordo com os registos desta companhia: 20%, 50% e 30% dos clientes pertencem às classes A, B e C (respetivamente); 5%, 15% e 30% dos clientes das classes A, B e C (respetivamente) estiveram envolvidos em acidentes no último ano.

Admitindo que se selecionou ao acaso um cliente desta companhia, calcule:

(a) A probabilidade de o cliente selecionado ter estado envolvido em acidentes no último ano. (1.5)

- **Quadro de acontecimentos e probabilidades**

Evento	**Probabilidade**
$A = \{$cliente pertence à classe $A\}$	$P(A) = 0.2$
$B = \{$cliente pertence à classe $B\}$	$P(B) = 0.5$
$C = \{$cliente pertence à classe $C\}$	$P(C) = 0.3$
$E = \{$cliente envolvido em acidentes no último ano$\}$	$P(E) =?$
	$P(E \mid A) = 0.05$
	$P(E \mid B) = 0.15$
	$P(E \mid C) = 0.30$

- **Probabilidade pedida**
Ao aplicar-se a lei da probabilidade total, obtemos

$$\begin{aligned} P(E) &= P(E \mid A) \times P(A) + P(E \mid B) \times P(B) + P(E \mid C) \times P(C) \\ &= 0.05 \times 0.2 + 0.15 \times 0.5 + 0.3 \times 0.3 \\ &= 0.175. \end{aligned}$$

(b) A probabilidade de o cliente selecionado pertencer à classe A sabendo que não esteve envolvido em acidentes no último ano. (1.0)

- **Prob. pedida**
Tirando partido do teorema de Bayes, segue-se

$$\begin{aligned} P(A \mid \overline{E}) &= \frac{P(\overline{E} \mid A) \times P(A)}{P(\overline{E})} \\ &= \frac{[1 - P(E \mid A)] \times P(A)}{1 - P(E)} \\ &\overset{(a)}{=} \frac{(1 - 0.05) \times 0.2}{1 - 0.175} \\ &\simeq 0.230303. \end{aligned}$$

2. Considere uma urna contendo 3 bolas brancas e 27 bolas pretas. Admita que um jogador retira duas bolas dessa urna, ao acaso e sem reposição, recebendo um prémio caso retire duas bolas brancas.

(a) Qual é a probabilidade de o jogador receber o prémio? (1.5)

- **Variável aleatória de interesse**
X = número de bolas brancas em extracção de 2 bolas, ao acaso e SEM reposição, de uma urna contendo 3 bolas brancas e 27 bolas pretas

- **Distribuição de** X
  $X \sim \text{Hipergeométrica}(N, M, n)$
  com:
  $N = 3 + 27 = 30$ (bolas na urna);
  $M = 3$ (bolas brancas);
  $n = 2$ (bolas extraídas ao acaso e SEM reposição).

- **F.p. de** X

$$P(X = x) \stackrel{form.}{=} \begin{cases} \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}, & x = \max\{0, n-(N-M)\}, \ldots, \min\{n, M\} \\ 0, & \text{caso contrário} \end{cases}$$
$$= \begin{cases} \frac{\binom{3}{x}\binom{30-3}{2-x}}{\binom{30}{2}}, & x = 0, 1, 2 \\ 0, & \text{caso contrário} \end{cases}$$

- **Prob. pedida**
  O jogador receberá o prémio com probabilidade

$$\begin{aligned} P(X = 2) &= \frac{\binom{3}{2}\binom{30-3}{2-2}}{\binom{30}{2}} \\ &= \frac{\frac{3!}{2!\,(3-2)!} \times \frac{27!}{0!\,(27-0)!}}{\frac{30!}{2!\,(30-2)!}} \\ &= \frac{3 \times 1}{\frac{30 \times 29}{2}} \\ &= \frac{2}{290} \\ &= \frac{1}{145} \\ &\simeq 0.006897. \end{aligned}$$

(b) Considere que o jogador efetua uma sequência de 100 extrações independentes de duas bolas ao acaso e sem reposição. Determine a probabilidade de o jogador receber 2 ou mais prémios. (1.0)

Nota: Caso não tenha resolvido a alínea (a), considere que a probabilidade de o jogador receber o prémio é igual a 0.006897.

- **Variável aleatória de interesse**
  $Y =$ número de prémios recebidos em 100 extracções independentes

- **Distribuição de** Y
  $Y \sim \text{Binomial}(n, p)$, com $n = 100$ e $p = 0.006897$

- **F.p. de** Y
  $P(Y = y) = \binom{100}{y} 0.006897^y (1 - 0.006897)^{100-y}, \quad y = 0, 1, 2, \ldots, 100$

- **Prob. pedida**

$$\begin{aligned} P(Y \geq 2) &= 1 - P(Y \leq 1) \\ &= 1 - \sum_{y=0}^{1} \binom{100}{y} 0.006897^y (1 - 0.006897)^{100-y} \\ &= 1 - (1 - 0.006897)^{100} - 100 \times 0.006897 \times (1 - 0.006897)^{99} \\ &\simeq 0.151858. \end{aligned}$$

Grupo II	5 valores

1. O atraso (em minutos) de um voo entre Lisboa e Porto em determinada companhia aérea é uma variável aleatória X com função de distribuição

$$F_X(x) = \begin{cases} 0, & x < -6 \\ \frac{1}{2} + \frac{1}{288}\left(36x - \frac{x^3}{3}\right), & -6 \le x < 6 \\ 1, & x \ge 6, \end{cases}$$

onde os valores negativos de X indicam adiantamento do voo.

(a) Calcule a mediana de X. (1.0)

- **Variável aleatória de interesse**
  X = atraso (em minutos) de um voo entre Lisboa e Porto em determinada companhia aérea
- **Obtenção da mediana de X**
  Tratando-se de uma v.a. contínua, tem-se

$$\begin{aligned} me(X) = x \in (-6,6) \quad &: \quad F_X(x) = 0.5 \\ & \quad \frac{1}{2} + \frac{1}{288}\left(36x - \frac{x^3}{3}\right) = \frac{1}{2} \\ & \quad \frac{108x - x^3}{3} = 0 \\ & \quad x \times (108 - x^2) = 0 \\ & \quad x = 0 \quad \text{ou} \quad x = \pm\sqrt{108} \simeq \pm 10.392305 \\ & \quad x = 0 \end{aligned}$$

  dado que $\pm\sqrt{108} \not\in (-6,6)$.

(b) Qual a probabilidade de um voo entre Lisboa e Porto se atrasar pelo menos 1 minuto? E de se adiantar pelos menos 3 minutos? (1.0)

- **Prob. pedidas**
  A probabilidade de um voo se atrasar pelo menos 1 minuto é

$$\begin{aligned} P(X \ge 1) &= 1 - P(X < 1) \\ &= 1 - F_X(1) \\ &= 1 - \left[\frac{1}{2} + \frac{1}{288}\left(36 - \frac{1}{3}\right)\right] \\ &= \frac{325}{864} \\ &\simeq 0.376157, \end{aligned}$$

  ao passo que a de um voo se adiantar pelos menos 3 minutos é dada por

$$\begin{aligned} P(X \le -3) &= F_X(-3) \\ &= \frac{1}{2} + \frac{1}{288}\left[36 \times (-3) - \frac{(-3)^3}{3}\right] \\ &= \frac{5}{32} \\ &\simeq 0.15625, \end{aligned}$$

  respetivamente.

(c) Deduza a função de densidade de probabilidade de X. (0.5)

- **F.d.p. de X**

$$\begin{aligned} f_X(x) &= \frac{d\,F_X(x)}{dx} \\ &= \begin{cases} \frac{1}{288}(36 - x^2), & -6 < x < 6 \\ 0, & \text{caso contrário.} \end{cases} \end{aligned}$$

2. Seja (X, Y) um par aleatório, em que X e Y representam o número de telemóveis das marcas A e B (respetivamente) vendidos diariamente numa pequena loja. Admita que a função de probabilidade conjunta de (X, Y) é dada por

	Y		
X	0	1	2
0	a	0.10	0.06
1	0.12	0.03	0.01
2	0.03	0.07	0.04

(a) Determine o valor da constante a. (0.5)

- **Par aleatório** (X, Y)
  X = número de telemóveis vendidos da marca A
  Y = número de telemóveis vendidos da marca B
- **F.p. conjunta**
  $P(X = x, Y = y)$ é dada pela tabela do enunciado.
- **Obtenção da constante** a

$$\begin{aligned} a \quad : \quad & \sum_{x=0}^{2}\sum_{y=0}^{2} P(X = x, Y = y) = 1 \\ & a = 1 - (0.10 + 0.06 + 0.12 + 0.03 + 0.01 + 0.03 + 0.07 + 0.04) \\ & a = 0.54. \end{aligned}$$

(b) Determine a função de probabilidade de $Y \mid X = 2$. (1.0)

- **V.a.**
  $Y \mid X = 2$
- **F.p. de** $Y \mid X = 2$
  Atendendo a que

$$\begin{aligned} P(X = 2) &= \sum_{y=0}^{2} P(X = 2, Y = y) \\ &= 0.03 + 0.07 + 0.04 \\ &= 0.14, \end{aligned}$$

temos

$$\begin{aligned} P(Y = y \mid X = 2) &= \frac{P(X = 2, Y = y)}{P(X = 2)} \\ &= \begin{cases} \frac{0.03}{0.14} = \frac{3}{14} \simeq 0.214286, & y = 0 \\ \frac{0.07}{0.14} = \frac{7}{14} = 0.5, & y = 1 \\ \frac{0.04}{0.14} = \frac{4}{14} \simeq 0.285714, & y = 2 \\ 0, & \text{restantes valores de } y \end{cases} \end{aligned}$$

(c) Calcule o valor esperado do número de telemóveis vendidos diariamente na loja. (1.0)

- **V.a. de interesse**
  $X + Y$ = número de telemóveis vendidos diariamente na loja
- **Valor esperado pedido**
  $E(X + Y) = E(X) + E(Y)$
- **F.p. marginais**
  $P(X = x) = \sum_{y=0}^{2} P(X = x, Y = y)$ e $P(Y = y) = \sum_{x=0}^{2} P(X = x, Y = y)$ encontram-se sumariadas na tabela seguinte:

X	Y: 0	1	2	$P(X=x)$
0	0.54	0.10	0.06	0.70
1	0.12	0.03	0.01	0.16
2	0.03	0.07	0.04	0.14
$P(Y=y)$	0.69	0.20	0.11	1

- **Valor esperados de X e Y**

$$\begin{aligned} E(X) &= \sum_{x=0}^{2} x \times P(X=x) \\ &= 0 \times 0.70 + 1 \times 0.16 + 2 \times 0.14 \\ &= 0.44 \end{aligned}$$

$$\begin{aligned} E(Y) &= \sum_{y=0}^{2} y \times P(Y=y) \\ &= 0 \times 0.69 + 1 \times 0.20 + 2 \times 0.11 \\ &= 0.42 \end{aligned}$$

- **Valor esperado pedido** (cont.)

$$\begin{aligned} E(X+Y) &= E(X)+E(Y) \\ &= 0.44+0.42 \\ &= 0.86 \end{aligned}$$

Grupo III	5 valores

1. Admita que o tamanho de um ficheiro transferido usando o protocolo TCP é representado pela variável aleatória X com função de densidade de probabilidade

$$f_X(x) = \begin{cases} \theta\,(1+x)^{-(\theta+1)}, & x>0 \\ 0, & \text{caso contrário,} \end{cases}$$

onde θ é um parâmetro positivo desconhecido.

(a) Mostre que o estimador de máxima verosimilhança de θ, com base numa amostra aleatória $(X_1,\ldots,X_n)$ proveniente da população X, é dado por $\frac{n}{\sum_{i=1}^{n}\ln(1+X_i)}$. (1.5)

- **V.a. de interesse**
  X = tamanho de um ficheiro transferido usando o protocolo TCP
- **F.d.p. de X**

$$f_X(x) = \begin{cases} \theta\,(1+x)^{-(\theta+1)}, & x>0 \\ 0, & \text{caso contrário} \end{cases}$$

- **Parâmetro desconhecido**
  $\theta, \quad \theta>0$
- **Amostra**
  $\underline{x}=(x_1,\ldots,x_n)$ amostra de dimensão n proveniente da população X
- **Obtenção do estimador de MV de θ**

Passo 1 — Função de verosimilhança

$$\begin{aligned} L(\theta|\underline{x}) &= f_{\underline{X}}(\underline{x}) \\ &\overset{X_i\,indep}{=} \prod_{i=1}^{n} f_{X_i}(x_i) \\ &\overset{X_i\sim X}{=} \prod_{i=1}^{n} f_X(x_i) \end{aligned}$$

$$
\begin{aligned}
L(\theta|\underline{x}) &= \prod_{i=1}^{n} \left[\theta\,(1+x_i)^{-(\theta+1)}\right] \\
&= \theta^n \left[\prod_{i=1}^{n}(1+x_i)\right]^{-(\theta+1)}, \quad \theta > 0
\end{aligned}
$$

Passo 2 — Função de log-verosimilhança

$$\ln L(\theta|\underline{x}) = n\ln(\theta) - (\theta+1)\sum_{i=1}^{n}\ln(1+x_i)$$

Passo 3 — Maximização

A estimativa de MV de θ passa a ser representada por $\hat{\theta}$ e

$$\hat{\theta} \;:\; \begin{cases} \left.\frac{d\ln L(\theta|\underline{x})}{d\theta}\right|_{\theta=\hat{\theta}} = 0 & \text{(ponto de estacionaridade)} \\ \left.\frac{d^2\ln L(\theta|\underline{x})}{d\theta^2}\right|_{\theta=\hat{\theta}} < 0 & \text{(ponto de máximo)} \end{cases}$$

$$\begin{cases} \frac{n}{\hat{\theta}} - \sum_{i=1}^{n}\ln(1+x_i) = 0 \\ -\frac{n}{\hat{\theta}^2} < 0 \end{cases}$$

$$\begin{cases} \hat{\theta} = \frac{n}{\sum_{i=1}^{n}\ln(1+x_i)} \\ -\frac{\left[\sum_{i=1}^{n}\ln(1+x_i)\right]^2}{n} < 0 & \text{(proposição verdadeira porque } \sum_{i=1}^{n}\ln(1+x_i) > 0\text{).} \end{cases}$$

Passo 4 — Estimador de MV de θ

$$EMV(\theta) = \frac{n}{\sum_{i=1}^{n}\ln(1+X_i)}.$$

(b) Determine a estimativa de máxima verosimilhança de $P(X<1) = 1-\frac{1}{2^\theta}$ baseada na amostra $(x_1, x_2, \ldots, x_7) = (1.42, 1.31, 1.53, 1.05, 1.63, 2.65, 2.22)$ para a qual $\sum_{i=1}^{7}\ln(1+x_i) \simeq 6.80$. (0.5)

- **Estimativa de MV de θ**

$$
\begin{aligned}
\hat{\theta} &= \frac{n}{\sum_{i=1}^{n}\ln(1+x_i)} \\
&\simeq \frac{7}{6.80} \\
&\simeq 1.029412
\end{aligned}
$$

- **Outro parâmetro desconhecido**

$$
\begin{aligned}
h(\theta) &= P(X<1) \\
&= 1-\frac{1}{2^\theta}
\end{aligned}
$$

- **Estimativa de MV de $h(\theta)$**

Invocando a propriedade de invariância dos estimadores de máxima verosimilhança, pode concluir-se que a estimativa de MV de $h(\theta)$ é dada por

$$
\begin{aligned}
\widehat{h(\theta)} &= h(\hat{\theta}) \\
&= 1-\frac{1}{2^{\hat{\theta}}} \\
&\simeq 1-\frac{1}{2^{1.029412}} \\
&\simeq 0.510090.
\end{aligned}
$$

2. O consumo diário individual de calorias (X, em kcal/g) de uma certa espécie de roedor possui distribuição normal com valor esperado μ e desvio padrão σ desconhecidos. Sabendo que uma amostra de dimensão 24 proveniente da população X conduziu à média e variância amostrais $\bar{x} = 0.388$ e $s^2 = 0.057181$:

(a) Calcule um intervalo de confiança a 95% para μ. (1.5)

- **V.a. de interesse**
  X = consumo diário individual de calorias de certa espécie de roedor

- **Situação**
  $X \sim \text{Normal}(\mu, \sigma^2)$
  μ DESCONHECIDO
  σ^2 desconhecido

- **Obtenção do IC para μ**

  Passo 1 — Selecção da v.a. fulcral para μ

$$Z = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{(n-1)}$$

  [dado que é suposto determinar um IC para o valor esperado de uma população normal, com variância desconhecida.]

  Passo 2 — Obtenção dos quantis de probabilidade
  Uma que vez que $(1-\alpha) \times 100\% = 95\%$, far-se-á uso dos quantis

$$(a_\alpha, b_\alpha) \quad : \quad \begin{cases} P(a_\alpha \leq Z \leq b_\alpha) = 1 - \alpha \\ P(Z < a_\alpha) = P(Z > b_\alpha) = \alpha/2. \end{cases}$$

$$\begin{cases} a_\alpha = F^{-1}_{t_{(n-1)}}(\alpha/2) = F^{-1}_{t_{(24-1)}}(0.05) = -F^{-1}_{t_{(9)}}(1-0.025) \stackrel{tabela/calc.}{=} -2.069 \\ b_\alpha = F^{-1}_{t_{(n-1)}}(1-\alpha/2) = F^{-1}_{t_{(24-1)}}(0.975) \stackrel{tabela/calc.}{=} 2.069. \end{cases}$$

  Passo 3 — Inversão da desigualdade $a_\alpha \leq Z \leq b_\alpha$

$$P(a_\alpha \leq Z \leq b_\alpha) = 1 - \alpha$$

$$P\left[a_\alpha \leq \frac{\bar{X}-\mu}{S/\sqrt{n}} \leq b_\alpha\right] = 1 - \alpha$$

$$P\left[\bar{X} - b_\alpha \times \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} - a_\alpha \times \frac{S}{\sqrt{n}}\right] = 1 - \alpha$$

  Passo 4 — Concretização
  Atendendo aos quantis acima, às concretizações de $\bar{X}$ e S^2,

$$\begin{aligned} \bar{x} &= 0.388 \\ s^2 &= 0.057181, \end{aligned}$$

  e ao facto de

$$IC_{(1-\alpha)\times 100\%}(\mu) = \left[\bar{x} - F^{-1}_{t_{(n-1)}}(1-\alpha/2) \times \frac{s}{\sqrt{n}}, \quad \bar{x} + F^{-1}_{t_{(n-1)}}(1-\alpha/2) \times \frac{s}{\sqrt{n}}\right],$$

  temos

$$\begin{aligned} IC_{95\%}(\mu) &= \left[0.388 - 2.069 \times \frac{\sqrt{0.057181}}{\sqrt{24}}, \quad 0.388 + 2.069 \times \frac{\sqrt{0.057181}}{\sqrt{24}}\right] \\ &= [0.287009,\ 0.488991]. \end{aligned}$$

(b) Confronte as hipóteses $H_0 : \mu = 0.35$ e $H_1 : \mu > 0.35$, calculando para o efeito o valor-p. (1.5)

- **Hipóteses**
  $H_0 : \mu = \mu_0 = 0.35$
  $H_1 : \mu > \mu_0$

- **Estatística de teste**

  $T = \frac{\bar{X}-\mu_0}{\frac{S}{\sqrt{n}}} \sim_{H_0} t_{(n-1)}$

  [pois pretendemos efectuar um teste sobre o valor esperado de uma população normal, com variância desconhecida.]

- **Região de rejeição de H_0** (para valores de T)
  Tratando-se de um teste unilateral superior ($H_1 : \mu > \mu_0$), a região de rejeição de H_0 (para valores da estatística de teste) é do tipo $W = (c, +\infty)$.

- **Decisão (com base no valor-p)**
  O valor observado da estatística de teste é dado por

$$
\begin{aligned}
t &= \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} \\
&= \frac{0.388 - 0.35}{\sqrt{\frac{0.057181}{24}}} \\
&\simeq 0.778508.
\end{aligned}
$$

  Dado que a região de rejeição deste teste é um intervalo à direita, temos:

$$
\begin{aligned}
valor - p &= P(T > t \mid H_0) \\
&= 1 - F_{t_{(n-1)}}(t) \\
&= 1 - F_{t_{(23)}}(0.778508) \\
&\overset{calc.}{=} 0.222103.
\end{aligned}
$$

  Consequentemente, é suposto:
  - não rejeitar H_0 a qualquer n.s. $\alpha_0 \leq 22.2103\%$, pelo que $H_0 : \mu = \mu_0 = 0.35$ é consistente a qualquer dos n.u.s. $(1\%, 5\%, 10\%)$;
  - rejeitar H_0 a qualquer n.s. $\alpha_0 > 22.2103\%$.

  [Decisão (com base em intervalo para o valor-p)
  Recorrendo às tabelas de quantis da distribuição de t-student obtemos um intervalo para o valor-p:

$$
\begin{aligned}
F^{-1}_{t_{(23)}}(0.75) = 0.685 \quad &< \quad 0.778508 \ < \ 0.858 = F^{-1}_{t_{(23)}}(0.80) \\
0.20 = 1 - 0.8 \quad &< \quad valor - p = 1 - F_{t_{(23)}}(0.778508)] \ < \ 1 - 0.75 = 0.25.
\end{aligned}
$$

  Logo:
  - não devemos rejeitar H_0 a qualquer n.s. $\alpha_0 \leq 20\%$, nomeadamente aos n.u.s. $(1\%, 5\%, 10\%)$;
  - devemos rejeitar H_0 a qualquer n.s. $\alpha_0 \geq 25\%$.]

Grupo IV 5 valores

1. O tempo que decorre entre o aproveitamento malicioso de uma vulnerabilidade de *software* não conhecida e a altura em que a maior parte dos sistemas vulneráveis já aplicaram as devidas correções de segurança é conhecido por *janela de vulnerabilidade* (X, em dias). Foram recolhidos os seguintes dados relativos a cem vulnerabilidades surgidas recentemente: (2.0)

Janela de vulnerabilidade	$]0,20]$	$]20,40]$	$]40,60]$	$]60,+\infty[$
Frequência absoluta observada	40	25	12	23

Avalie a hipótese de X possuir distribuição exponencial com valor esperado igual a 28, ao nível de significância de 1%.

- **V.a. de interesse**
  X = amplitude da janela de vulnerabilidade

- **Hipóteses**
  $H_0 : X \sim \text{Exponencial}(1/28)$
  $H_1 : X \not\sim \text{Exponencial}(1/28)$

- **Nível de significância**
  $\alpha_0 = 1\%$

- **Estatística de Teste**

$$T = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \overset{a}{\sim}_{H_0} \chi^2_{(k-\beta-1)}$$

onde:

k = No. de classes = 4

O_i = Frequência absoluta observável da classe i

E_i = Frequência absoluta esperada, sob H_0, da classe i

β = No. de parâmetros a estimar = 0 [dado que H_0 é uma hipótese simples.]

- **Cálculo das frequências absolutas esperadas sob H_0**

Tirando partido do facto de

$$F_{X|H_0}(x) = P[X \leq x \mid X \sim \text{Exponencial}(1/28)] = 1 - e^{-\frac{x}{28}}, \quad x > 0,$$

as frequências absolutas esperadas sob H_0 são dadas por

$$\begin{aligned}
E_i &= n \times p_i^0 \\
&= n \times P(X \in \text{ Classe } i \mid H_0) \\
&= 100 \times \begin{cases} P(0 < X \leq 20 \mid H_0) = F_{X|H_0}(20) - F_{X|H_0}(0) = 1 - e^{-\frac{20}{28}}, & i = 1 \\ P(20 < X \leq 40 \mid H_0) = F_{X|H_0}(40) - F_{X|H_0}(20) = e^{-\frac{20}{28}} - e^{-\frac{40}{28}}, & i = 2 \\ P(40 < X \leq 60 \mid H_0) = F_{X|H_0}(60) - F_{X|H_0}(40) = e^{-\frac{40}{28}} - e^{-\frac{60}{28}}, & i = 3 \\ P(X > 60 \mid H_0) = 1 - F_{X|H_0}(60) = e^{-\frac{60}{28}}, & i = 4 \end{cases} \\
&= \begin{cases} 100 \times 0.5105 = 51.05, & i = 1 \\ 100 \times 0.2499 = 24.99, & i = 2 \\ 100 \times 0.1223 = 12.23, & i = 3 \\ 100 \times 0.1173 = 11.73, & i = 5. \end{cases}
\end{aligned}$$

[Constata-se que não é necessário fazer qualquer agrupamento de classes uma vez que em pelo menos 80% das classes se verifica $E_i \geq 5$ e que $E_i \geq 1$ para todo o i. Caso fosse preciso efectuar agrupamento de classes, os valores de k e $c = F^{-1}_{\chi^2_{(k-\beta-1)}}(1-\alpha_0)$ teriam que ser recalculados...]

- **Região de rejeição de H_0** (para valores de T)

Ao lidar-se com um teste de ajustamento, a região de rejeição de H_0 escrita para valores de T é o intervalo à direita $W = (c, +\infty)$, onde

$$\begin{aligned}
c &= F^{-1}_{\chi^2_{(k-\beta-1)}}(1-\alpha_0) \\
&= F^{-1}_{\chi^2_{(4-0-1)}}(1-0.01) \\
&= F^{-1}_{\chi^2_{(3)}}(0.99) \\
&\overset{tabela/calc.}{=} 11.34.
\end{aligned}$$

- **Decisão**

No cálculo do valor obs. da estat. de teste convém recorrer à seguinte tabela auxiliar:

i	Classe i	Freq. abs. obs. o_i	Estim. freq. abs. esp. sob H_0 $e_i = n \times \hat{p}_i^0$	Parcelas valor obs. estat. teste $\frac{(o_i-e_i)^2}{e_i}$
1	$]-\infty, 20]$	40	51.05	$\frac{(40-51.05)^2}{51.05} \simeq 2.392$
2	$]20, 40]$	25	24.99	0.000
3	$]40, 60]$	12	12.23	0.004
4	$]60, +\infty]$	23	11.73	10.828
		$\sum_{i=1}^k o_i = n$ $= 100$	$\sum_{i=1}^k e_i = n$ $= 100$	$t = \sum_{i=1}^k \frac{(o_i-e_i)^2}{e_i}$ $\simeq 13.224$

Como $t \simeq 13.224 \in W = (11.34, +\infty)$, devemos rejeitar H_0 ao n.s. de $\alpha_0 = 1\%$ [ou a qualquer outro n.s. superior a α_0].

2. É geralmente aceite que a frequência cardíaca (Y, em *batimentos por minuto*) é influenciada pela temperatura corporal dos seres humanos (x, em ^oC). Um conjunto de 130 medições independentes conduziu aos seguintes resultados:

$$\sum_{i=1}^{130} x_i = 4\,784.7\,,\ \sum_{i=1}^{130} x_i^2 = 176\,121.67\,,\ \sum_{i=1}^{130} y_i = 9\,589,\ \sum_{i=1}^{130} y_i^2 = 713\,733\,,\ \sum_{i=1}^{130} x_i y_i = 353\,018.5\,.$$

Admitindo que os erros aleatórios associados ao modelo de regressão linear simples de Y em x satisfazem $\epsilon_i \stackrel{i.i.d.}{\sim} \text{Normal}(0, \sigma^2)$, $i = 1, \ldots, 130$:

(a) Calcule as estimativas de máxima verosimilhança dos parâmetros da reta de regressão linear simples de Y em x. (1.0)

- **Estimativas de MV de β_0 e β_1**
  Dado que

$$n = 130$$

$$\sum_{i=1}^n x_i = 4\,784.7$$
$$\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i = \frac{4\,784.7}{130} = 36.80538$$
$$\sum_{i=1}^n x_i^2 = 176\,121.67$$
$$\sum_{i=1}^n x_i^2 - n\,(\bar{x})^2 = 176\,121.67 - 130 \times 36.80538^2 = 18.946231$$

$$\sum_{i=1}^n y_i = 9\,589$$
$$\bar{y} = \frac{1}{n}\sum_{i=1}^n y_i = \frac{9589}{130} = 73.761538$$
$$\sum_{i=1}^n y_i^2 = 713\,733$$
$$\sum_{i=1}^n y_i^2 - n\,(\bar{y})^2 = 713\,733 - 130 \times 73.761538^2 = 6\,433.616544$$

$$\sum_{i=1}^n x_i y_i = 353\,018.5$$
$$\sum_{i=1}^n x_i y_i - n\,\bar{x}\,\bar{y} = 353\,018.5 - 130 \times 36.80538 \times 73.761538 = 91.669131,$$

as estimativas de MV de β_1 e β_0 são, para este modelo de RLS, iguais a:

$$\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_{i=1}^n x_i y_i - n\,\bar{x}\bar{y}}{\sum_{i=1}^n x_i^2 - n\,(\bar{x})^2} \\
&= \frac{91.669131}{18.946231} \\
&\simeq 4.838383 \\
\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \times \bar{x} \\
&\simeq 73.761538 - 4.838383 \times 36.80538 \\
&\simeq -104.317009.
\end{aligned}$$

(b) Deduza um intervalo de confiança a 90% para o declive da reta de regressão linear simples de Y em x. (1.5)

Nota: Pode vir a necessitar do seguinte quantil $F^{-1}_{t_{(128)}}(0.95) \simeq 1.65685$.

- **Obtenção de IC para β_1**

  Passo 1 — Selecção da v.a. fulcral para β_1

$$Z = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}}} \sim t_{(n-2)}$$

  Passo 2 — Obtenção dos quantis de probabilidade
  Neste caso $n = 130$ e $(1-\alpha) \times 100\% = 90\%$, logo usaremos os quantis de probabilidade

$$(a_\alpha, b_\alpha) \quad : \quad \begin{cases} P(a_\alpha \le Z \le b_\alpha) = 1-\alpha \\ P(Z < a_\alpha) = P(Z > b_\alpha) = \alpha/2. \end{cases}$$

$$\begin{cases} a_\alpha = -F^{-1}_{t_{(n-2)}}(1-\alpha/2) = -F^{-1}_{t_{(128)}}(0.95) \overset{tabela/calc.}{=} -1.65685 \\ b_\alpha = F^{-1}_{t_{(n-2)}}(1-\alpha/2) = F^{-1}_{t_{(128)}}(0.95) \overset{tabela/calc.}{=} 1.65685. \end{cases}$$

Passo 3 — Inversão da desigualdade $a_\alpha \le Z \le b_\alpha$

$P\,(a_\alpha \le Z \le b_\alpha) = 1-\alpha$

$\ldots$

$$P\left[\hat{\beta}_1 - F^{-1}_{t_{(n-2)}}(1-\alpha/2) \times \sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}} \le \beta_1 \le \hat{\beta}_1 + F^{-1}_{t_{(n-2)}}(1-\alpha/2) \times \sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}}\right] = 1-\alpha.$$

Passo 4 — Concretização

Atente-se que

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{n-2}\left[\left(\sum_{i=1}^n y_i^2 - n\,\bar{y}^2\right) - \left(\hat{\beta}_1\right)^2 \left(\sum_{i=1}^n x_i^2 - n\,\bar{x}^2\right)\right] \\ &= \frac{1}{130-2}\left[6433.616544 - 4.838383^2 \times 18.946231\right] \\ &= 46.797549 \end{aligned}$$

$$IC_{(1-\alpha)\times 100\%}(\beta_1) = \left[\hat{\beta}_1 \pm F^{-1}_{t_{(n-2)}}(1-\alpha/2) \times \sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}}\right].$$

Logo

$$\begin{aligned} IC_{90\%}(\beta_1) &\simeq \left[4.83838 \pm 1.65685 \times \sqrt{\frac{46.797549}{18.946231}}\right] \\ &\simeq [4.83838 \pm 2.603954]] \\ &= [2.234429, 7.442337]. \end{aligned}$$

(c) Calcule o valor do coeficiente de determinação e comente a utilidade do modelo ajustado. (0.5)

- **Cálculo do coeficiente de determinação**

$$\begin{aligned} r^2 &= \frac{\left(\sum_{i=1}^n x_i y_i - n\,\bar{x}\,\bar{y}\right)^2}{\left(\sum_{i=1}^n x_i^2 - n\,\bar{x}^2\right) \times \left(\sum_{i=1}^n y_i^2 - n\,\bar{y}^2\right)} \\ &= \frac{91.669131^2}{6433.616544 \times 18.946231} \\ &\simeq 0.068940. \end{aligned}$$

- **Interpretação coeficiente de determinação**

  Cerca de 6.8940% da variação total da frequência cardíaca é explicada pela temperatura corporal, através do modelo de regressão linear simples considerado. Assim sendo, podemos afirmar que a recta estimada parece ajustar-se muito mal ao conjunto de dados.