Departamento de Matemática

Probabilidades e Estatística

LEE, LEGI, LEIC-T, LERC, LMAC, MEBiom, MEFT

2º Teste — 1º semestre – 2010/2011

Duração: 1 hora e 30 minutos — 8/01/2011 – **11** horas

Grupo III — **1.5 + 1.0 + 1.5 + 1.0 = 5.0** valores

Exercício 1

(a)

- **V.a. de interesse**

  X = tensão eléctrica aos terminais (em Volt)

- **A.a.**

  $\underline{X} = (X_1, \ldots, X_n)$ a.a. de dimensão n proveniente da população X

  $X_i \overset{i.i.d.}{\sim} X$, $i = 1, \ldots, n$

- **Parâmetros** desconhecidos

  $E(X_i) = E(X) = \mu$

  $V(X_i) = V(X) = \sigma^2$

- **Estimador de μ**

  $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ (média da a.a.)

- **Estimador centrado de μ ?**

  Trata-se de um estimador centrado de μ se $E(\bar{X}) = \mu$. Efectivamente

$$\begin{aligned} E(\bar{X}) &= E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\ &= \frac{1}{n}\sum_{i=1}^{n} E(X_i) \\ &\overset{X_i \sim X}{=} \frac{1}{n}\sum_{i=1}^{n} E(X) \\ &= \frac{1}{n} \times n\mu \\ &= \mu \end{aligned}$$

- **Estimador de σ^2**

  $S^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} X_i^2 - n\,(\bar{X})^2\right]$ (variância corrigida da a.a.)

- **Estimador centrado de σ^2 ?**

  Ao tirar-se partido desta fórmula alternativa de S^2 e ao notar-se que $E(Z^2) = V(Z) + E^2(Z)$, $E(\bar{X}) = \mu$ e

$$\begin{aligned} V(\bar{X}) &\overset{X_i \, indep}{=} \frac{1}{n^2}\sum_{i=1}^{n} V(X_i) \\ &\overset{X_i \sim X}{=} \frac{1}{n^2}\sum_{i=1}^{n} V(X) \\ &= \frac{1}{n^2} n V(X) \\ &= \sigma^2/n \end{aligned}$$

segue-se

$$\begin{aligned}
E(S^2) &= E\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2\right] \\
&= \frac{1}{n-1}E\left[\left(\sum_{i=1}^{n}X_i^2\right)-n(\bar{X})^2\right] \\
&= \frac{1}{n-1}\left[\sum_{i=1}^{n}E(X_i^2)-nE[(\bar{X})^2]\right] \\
&= \frac{1}{n-1}\left\{\sum_{i=1}^{n}[V(X_i)+E^2(X_i)]-n\times[V(\bar{X})+E^2(\bar{X})]\right\} \\
&= \frac{1}{n-1}\left[\sum_{i=1}^{n}(\sigma^2+\mu^2)-n\times(\sigma^2/n+\mu^2)\right] \\
&= \frac{1}{n-1}\left(n\sigma^2+n\mu^2-\sigma^2-n\mu^2\right) \\
&= \frac{1}{n-1}\left(n-1\right)\sigma^2 \\
&= \sigma^2.
\end{aligned}$$

Assim, conclui-se que S^2 é de facto um estimador centrado de σ^2.

(b)
- **Outro parâmetro desconhecido**

  $\mu^2+\sigma^2$, valor esperado da potência eléctrica dissipada pelo circuito ($\frac{X^2}{R}$) quando $R=1$

- **Obtenção de k**

  Recordando que $E(Z^2)=V(Z)+E^2(Z)$ e tendo em conta os seguintes resultados da alínea anterior

$$\begin{aligned}
E(\bar{X}) &= \mu \\
V(\bar{X}) &= \frac{\sigma^2}{n} \\
E(S^2) &= \sigma^2,
\end{aligned}$$

  segue-se

$$\begin{aligned}
k \;:\; & E(\bar{X}^2+k\times S^2)=\mu^2+\sigma^2 \\
& E(\bar{X}^2)+k\times E(S^2)=\mu^2+\sigma^2 \\
& V(\bar{X})+E^2(\bar{X})+k\times E(S^2)=\mu^2+\sigma^2 \\
& \frac{\sigma^2}{n}+\mu^2+k\times\sigma^2=\mu^2+\sigma^2 \\
& \left(\frac{1}{n}+k\right)\times\sigma^2=\sigma^2 \\
& \frac{1}{n}+k=1 \\
& k=1-\frac{1}{n} \\
& k=\frac{n-1}{n}.
\end{aligned}$$

- **Conclusão**

  $\bar{X}^2+\frac{n-1}{n}\times S^2=\bar{X}^2+\frac{1}{n}\left[\sum_{i=1}^{n}X_i^2-n\left(\bar{X}\right)^2\right]=\sum_{i=1}^{n}X_i^2$ é um estimador centrado de $E(X^2)=\mu^2+\sigma^2$.

Exercício 2

(a)

- **V.a. de interesse**

  X = duração do díodo fornecido pelo primeiro fabricante (em centenas de horas)

  Y = duração do díodo fornecido pelo segundo fabricante (em centenas de horas)

- **Situação**

  X e Y, com distribuições arbitrárias (eventualmente não normais) e tais que:

  $E(X) = \mu_X, V(X) = \sigma_X^2$

  $E(Y) = \mu_X, V(X) = \sigma_X^2$

  $(\mu_X - \mu_Y)$ DESCONHECIDO

  σ_X^2 e σ_Y^2 desconhecidos

  n_X e n_Y suficientemente grandes (maiores que 30).

  As a.a. associadas a X e Y são independentes, pelo que $\bar{X}$ e $\bar{Y}$ são v.a. independentes.

- **Passo 1 — Selecção da v.a. fulcral para** $(\mu_1 - \mu_2)$

$$Z = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{\sqrt{\frac{S_X^2}{n_X} + \frac{S_Y^2}{n_Y}}} \overset{a}{\sim} \text{normal}(0, 1)$$

  uma vez que pretendemos um intervalo aproximado para a diferença de valores esperados de duas populações com distribuições arbitrárias variâncias desconhecidas e dispomos de amostras independentes com dimensões suficientemente grandes.

- **Passo 2 — Obtenção dos quantis de probabilidade**

  Como o nível aproximado de confiança é de $(1 - \alpha) \times 100\% = 99\%$ (i.e., $\alpha = 0.01$), recorre-se aos quantis

$$\begin{cases} a_\alpha = -\Phi^{-1}(1 - \alpha/2) = -\Phi^{-1}(0.995) \overset{tabela}{=} -2.5758 \\ b_\alpha = \Phi^{-1}(1 - \alpha/2) = \Phi^{-1}(0.995) \overset{tabela}{=} 2.5758, \end{cases}$$

  sendo que estes quantis enquadram a v.a. fulcral Z com probabilidade aproximadamente igual a $(1 - \alpha) = 0.99$.

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

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

$$P\left[a_\alpha \leq \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{\sqrt{\frac{S_X^2}{n_X} + \frac{S_Y^2}{n_Y}}} \leq b_\alpha\right] \simeq 1 - \alpha$$

$$P\left[(\bar{X} - \bar{Y}) - b_\alpha \times \sqrt{\frac{S_X^2}{n_X} + \frac{S_Y^2}{n_Y}} \leq \mu_X - \mu_Y \leq (\bar{X} - \bar{Y}) - a_\alpha \times \sqrt{\frac{S_X^2}{n_X} + \frac{S_Y^2}{n_Y}}\right] \simeq 1 - \alpha$$

$$P\left[(\bar{X} - \bar{Y}) - \Phi^{-1}(1 - \alpha/2) \times \sqrt{\frac{S_X^2}{n_X} + \frac{S_Y^2}{n_Y}} \leq \mu_X - \mu_Y \right.$$
$$\left. \leq (\bar{X} - \bar{Y}) + \Phi^{-1}(1 - \alpha/2) \times \sqrt{\frac{S_X^2}{n_X} + \frac{S_Y^2}{n_Y}}\right] \simeq 1 - \alpha.$$

- **Passo 4 — Concretização**

  Neste caso o IC aproximado a $(1 - \alpha) \times 100\%$ é dado por

$$IC_{(1-\alpha)\times 100\%}(\mu_X - \mu_Y) = \left[(\bar{x} - \bar{y}) - \Phi^{-1}(1 - \alpha/2) \times \sqrt{\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}}, \right.$$
$$\left. (\bar{x} - \bar{y}) + \Phi^{-1}(1 - \alpha/2) \times \sqrt{\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}}\right].$$

Ora,

$$
\begin{aligned}
&n_X = 80\\
&\bar{x} = \tfrac{1}{n_X}\sum_{i=1}^{n_X} x_i = \tfrac{1}{80}\times 140 = 1.75\\
&s_X^2 = \tfrac{1}{n_X-1}\left[\sum_{i=1}^{n_X} x_i^2 - n_X\,(\bar{x})^2\right] = \tfrac{1}{80-1}(280 - 80\times 1.75^2) = 0.4430\\
&n_Y = 72\\
&\bar{y} = \tfrac{1}{n_Y}\sum_{i=1}^{n_Y} y_i = \tfrac{1}{72}\times 145 = 2.0139\\
&s_Y^2 = \tfrac{1}{n_Y-1}\left[\sum_{i=1}^{n_Y} y_i^2 - n_Y\,(\bar{y})^2\right] = \tfrac{1}{72-1}(310 - 72\times 2.0139^2) = 0.2533\\
&\alpha = 0.01\\
&\Phi^{-1}(1-\alpha/2) = 2.5858,
\end{aligned}
$$

logo

$$
\begin{aligned}
IC_{99\%}(\mu_X-\mu_Y) &= \left[(1.75-2.0139) \pm 2.5858\times\sqrt{\frac{0.4430}{80}+\frac{0.2533}{72}}\right]\\
&= [-0.5090,\ -0.0188].
\end{aligned}
$$

- **Comentário**

  Invocando a relação entre intervalos de confiança e testes de hipóteses, devemos rejeitar a hipótese $H_0 : \mu_X - \mu_Y = 0$ [1] ao n.s. de $\alpha \times 100\% = 1\%$ (ou a qualquer outro n.s. maior que 1%) dado que $0 \notin IC_{99\%}(\mu_X - \mu_Y) = [-0.5090,\ -0.0188]$.

  Esta analogia é válida porque:

  – a v.a. fulcral para $(\mu_X - \mu_Y)$ usada na construção do IC aproximado é também utilizada para definir a estatística de teste sobre $(\mu_X - \mu_Y)$;

  – o nível de significância do teste, $\alpha\times 100\% = 1\%$, é igual a (100%−nível de confiança aproximado do IC).

(b) • **Hipóteses de trabalho**

$V(X) = V(Y) = 1$

$n_X = n_Y = n \geq 30$ logo $\frac{(\bar{X}-\bar{Y})-(\mu_X-\mu_Y)}{\sqrt{\frac{V(X)}{n_X}+\frac{V(Y)}{n_Y}}} = \frac{(\bar{X}-\bar{Y})-(\mu_X-\mu_Y)}{\sqrt{\frac{2}{n}}} \overset{a}{\sim} \text{Normal}(0,1)$.

- **Obtenção de n**

$$
\begin{aligned}
n \ :\ & P\left[|(\bar{X}-\bar{Y})-(\mu_X-\mu_Y)| < 0.3\right] \simeq 0.99\\
& P\left[-\frac{0.3}{\sqrt{\frac{1}{n}+\frac{1}{n}}} \leq \frac{(\bar{X}-\bar{Y})-(\mu_X-\mu_Y)}{\sqrt{\frac{1}{n}+\frac{1}{n}}} \leq \frac{0.3}{\sqrt{\frac{1}{n}+\frac{1}{n}}}\right] \simeq 0.99\\
& \Phi\left(\frac{0.3}{\sqrt{\frac{2}{n}}}\right) - \Phi\left(-\frac{0.3}{\sqrt{\frac{2}{n}}}\right) \simeq 0.99\\
& \Phi\left(\frac{0.3}{\sqrt{2}}\sqrt{n}\right) - 1 + \Phi\left(\frac{0.3}{\sqrt{2}}\sqrt{n}\right) \simeq 0.99\\
& \Phi\left(\frac{0.3}{\sqrt{2}}\sqrt{n}\right) \simeq \frac{1+0.99}{2}\\
& \frac{0.3}{\sqrt{2}}\sqrt{n} \simeq \Phi^{-1}(0.995) = 2.5758\\
& n \simeq \left(\frac{2.5758\times\sqrt{2}}{0.3}\right)^2\\
& n \simeq 147.4388
\end{aligned}
$$

- **Conclusão**

  O valor mínimo de n é de 148 para que haja uma confiança aproximada de pelo menos 99% de que o erro de estimação da diferença entre os valores esperados de X e Y seja inferior a 0.3.

[1] A favor da hipótese $H_1 : \mu_X - \mu_Y \neq 0$.

- **Alternativa**

  Tirando partido do quantil utilizado na alínea anterior e do facto de $\frac{(\bar{X}-\bar{Y})-(\mu_X-\mu_Y)}{\sqrt{\frac{2}{n}}} \overset{a}{\sim} \text{Normal}(0,1)$, pode concluir-se que

$$P\left[-2.5758 \leq \frac{(\bar{X}-\bar{Y})-(\mu_X-\mu_Y)}{\sqrt{\frac{2}{n}}} \leq 2.5758\right] \simeq 0.99$$
$$P\left[(\bar{X}-\bar{Y}) - 2.5758 \times \sqrt{\frac{2}{n}} \leq \mu_X - \mu_Y \leq (\bar{X}-\bar{Y}) - 2.5758 \times \sqrt{\frac{2}{n}}\right] \simeq 0.99,$$

  pelo que o intervalo com extremos aleatórios

$$\left[(\bar{X}-\bar{Y}) \pm 2.5758 \times \sqrt{\frac{2}{n}}\right]$$

  está associado a um erro de estimação (ou margem de erro) igual a $\Delta(n) = 2.5758 \times \sqrt{\frac{2}{n}}$. Ora, pretende-se

$$\begin{aligned} n \quad : \quad & \Delta(n) = 2.5758 \times \sqrt{\frac{2}{n}} < 0.3 \\ & n > \left(\frac{2.5758 \times \sqrt{2}}{0.3}\right)^2 \\ & n > 147.4388. \end{aligned}$$

  Assim, o valor mínimo de n é efectivamente igual a 148.

Grupo IV **0.5 + 1.5 + 2.0 + 1.0 = 5.0** valores

Exercício 1

(a)
- **V.a. de interesse**

  $X =$ duração de aparelho electrónico

- **Hipótese nula**

$$H_0 : F_X(x) = F_0(x),\ -\infty < x < +\infty, \text{ onde } F_0(x) = \begin{cases} 0, & x < 0 \\ \frac{1-e^{-0.1x}}{1-e^{-0.1\times 70}}, & 0 \le x \le 70 \\ 1, & x > 70 \end{cases}$$

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

  As probabilidades de pertença às classe $(16, 24]$ e $(24, 70]$ são, sob H_0, iguais a:

$$\begin{aligned} p_3^0 &= P[X \in (16,24] \mid H_0] \\ &= F_0(24) - F_0(16) \\ &= \frac{1-e^{-0.1\times 24}}{1-e^{-0.1\times 70}} - \frac{1-e^{-0.1\times 16}}{1-e^{-0.1\times 70}} \\ &\simeq 0.111 \\ p_4^0 &= P[X \in (24,70] \mid H_0] \\ &= F_0(70) - F_0(24) \\ &= \frac{1-e^{-0.1\times 70}}{1-e^{-0.1\times 70}} - \frac{1-e^{-0.1\times 24}}{1-e^{-0.1\times 70}} \\ &\simeq 0.090. \end{aligned}$$

  Logo as frequências absolutas esperadas sob H_0, E_3 e E_4, são dadas por:

$$\begin{aligned} E_3 &= n \times p_3^0 \\ &\simeq 100 \times 0.111 \\ &= 11.1 \\ E_4 &= n \times p_4^0 \\ &\simeq 100 \times 0.090 \\ &= 9.0. \end{aligned}$$

(b)
- **Hipóteses**

$$H_0 : F_X(x) = F_0(x),\ -\infty < x < +\infty, \text{ onde } F_0(x) = \begin{cases} 0, & x < 0 \\ \frac{1-e^{-0.1x}}{1-e^{-0.1\times 70}}, & 0 \le x \le 70 \\ 1, & x > 70 \end{cases}$$

  $H_1 : F_X(x) \neq F_0(x)$, para algum x.

- **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 $= 5$

  $O_i =$ Frequência absoluta observável da classe i

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

  $\beta =$ No. de parâmetros a estimar $= 0$.[2]

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

  A região de rejeição de H_0 escrita para valores de T é um intervalo à direita $W = (c, +\infty)$, onde

[2] Dado que a distribuição conjecturada em H_0 está completamente especificada, i.e., H_0 é uma hipótese simples.

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

- **Frequências absolutas esperadas sob H_0**

  Foram adiantadas (ver enunciado) ou calculadas na alínea anterior: $E_1 = 55.1$, $E_2 = 24.8$, $E_3 = 11.1$ e $E_4 = 9.0$.[3]

- **Decisão**

  No cálculo do valor observado da estatística de teste convém adiantar a seguinte tabela auxiliar:

i	Classe i	Freq. abs. obs. o_i	Freq. abs. esper. sob H_0 $E_i = n \times p_i^0$	Parcelas valor obs. estat. teste $\frac{(o_i-E_i)^2}{E_i}$
1	$[0, 8]$	52	55.1	$\frac{(52-55.1)^2}{55.1} = 0.174410$
2	$(8, 16]$	23	24.8	0.130345
3	$(16, 24]$	12	11.1	0.072973
4	$(24, 70]$	13	9.0	1.777782
		$\sum_{i=1}^k o_i = n = 100$	$\sum_{i=1}^k E_i = n = 100$	$t = \sum_{i=1}^k \frac{(o_i-E_i)^2}{E_i} = 2.15581$

  Assim, temos

$$\begin{aligned} t &= \sum_{i=1}^{k} \frac{(o_i - E_i)^2}{E_i} \\ &= 2.15581 \\ &\notin W = (11.34, +\infty) \end{aligned}$$

  pelo que não devemos rejeitar a hipótese de os dados serem provenientes de uma população com f.d. $F_0(x)$, a qualquer nível de significância menor ou igual que 1%.

Exercício 2

(a)
- [**Modelo de RLS**

  $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$

  $Y_i =$ peso animal i

  $x_i =$ quantidade de produto de engorda na ração do animal i

  $\epsilon_i =$ erro aleatório associado à medição do peso do animal i

- **Hipóteses de trabalho**

  $\epsilon_i \sim_{i.i.d.} \text{Normal}(0, \sigma^2)$, $i = 1, \ldots, n$ (hipótese de trabalho)

  $\beta_0, \beta_1, \sigma^2$ DESCONHECIDOS]

- **Estimativas de MV de β_0 e β_1**

  [Dado que

  $n = 10$

  $\sum_{i=1}^n x_i = 668$

  $\bar{x} = \frac{668}{10} = 66.8$

  $\sum_{i=1}^n x_i^2 = 44794$

  $\sum_{i=1}^n x_i^2 - n\,(\bar{x})^2 = 44794 - 10 \times 66.8^2 = 171.6$

  $\sum_{i=1}^n y_i = 670.1$

  $\bar{y} = \frac{670.1}{10} = 67.01$

  $\sum_{i=1}^n y_i^2 = 44941.93$

  $\sum_{i=1}^n y_i^2 - n\,(\bar{y})^2 = 44941.93 - 10 \times 67.01^2 = 38.529$

[3] É interessante notar que não é necessário qualquer agrupamento de classes uma vez que em pelo menos 80% das classes se verifica $E_i \geq 5$.

$\sum_{i=1}^{n} x_i y_i = 44842.4$

$\sum_{i=1}^{n} x_i y_i - n\,\bar{x}\,\bar{y} = 44842.4 - 10 \times 66.8 \times 67.01 = 79.72.]$

As estimativas de MV para este modelo são iguais a:

$$\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_{i=1}^{n} x_i y_i - n\,\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\,(\bar{x})^2} \\
&= \frac{79.72}{171.6} \\
&= 0.464569 \\
\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \times \bar{x} \\
&= 67.01 - 0.464569 \times 66.8 \\
&= 35.9768 \text{ (não era necessário calcular esta estimativa)}
\end{aligned}$$

- **Estimativa de σ^2**

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{n-2}\left[\left(\sum_{i=1}^{n} y_i^2 - n\,\bar{y}^2\right) - (\hat{\beta}_1)^2\left(\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2\right)\right] \\
&= \frac{1}{10-2}\left(38.529 - 0.464569^2 \times 171.6\right) \\
&= 0.186697
\end{aligned}$$

- **Hipóteses**

  $H_0 : \beta_1 = \beta_{1,0} = 0.4$ vs. $H_1 : \beta_1 > \beta_{1,0}$

- **Estatística de teste**

$$T = \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2}}} \sim_{H_0} t_{(n-2)}$$

- **Região de rejeição de H_0** (para valores da estatística de teste)

  Estamos a lidar com um teste unilateral superior ($H_1 : \beta_1 > \beta_{1,0}$), pelo que a região de rejeição de H_0 é um intervalo à direita.

- **Decisão (com base em intervalo para o p-value)**

  Tendo em conta que

  $\hat{\beta}_1 = 0.464569$

  $\hat{\sigma}^2 = 0.186697$

  $\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2 = 171.6,$

  o valor observado da estatística de teste é dado por

$$\begin{aligned}
t &= \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2}}} \\
&= \frac{0.464569 - 0.4}{\sqrt{\frac{0.186697}{171.6}}} \\
&= 1.95755.
\end{aligned}$$

  Uma vez que este teste é unilateral superior temos:

$$\begin{aligned}
p - value &= P(T > t \mid H_0) \\
&= P(T > 1.95755 \mid H_0) \\
&= 1 - F_{t_{(8)}}(1.95755).
\end{aligned}$$

  Recorrendo às tabelas de quantis da distribuição de t-student podemos adiantar um intervalo para o *p-value* deste teste. Com efeito, ao enquadrarmos convenientemente $t = 1.95755$, obtemos sucessivamente

$$\begin{aligned}
F^{-1}_{t_{(8)}}(0.95) = 1.860 &< 1.95755 < 2.306 = F^{-1}_{t_{(8)}}(0.975) \\
0.95 &< F_{t_{(8)}}(1.95755) < 0.975 \\
0.025 = 1 - 0.975 &< p - value < 1 - 0.95 = 0.05.
\end{aligned}$$

Logo

- não devemos rejeitar H_0 a qualquer n.s. $\alpha_0 \leq 2.5\%$, nomeadamente ao nível usual de significância de 1%;
- devemos rejeitar H_0 a qualquer n.s. $\alpha_0 \geq 2.5\%$, por exemplo aos níveis usuais de significância de 5% e 10%.

- **Decisão (com base em p-value determinado usando máquina de calcular)**

O valor observado da estatística de teste é igual a

$$\begin{aligned} t &= \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}}} \\ &= \frac{0.464569 - 0.4}{\sqrt{\frac{0.186697}{171.6}}} \\ &= 1.95755. \end{aligned}$$

Uma vez que este teste é unilateral superior

$$\begin{aligned} p-value &= P(T > t \mid H_0) \\ &= P(T > 1.95755 \mid H_0) \\ &= 1 - F_{t_{(8)}}(1.95755) \\ &= 0.0429918, \end{aligned}$$

pelo que

- não devemos rejeitar H_0 a qualquer n.s. $\alpha_0 \leq 4.29918\%$, nomeadamente aos nível usual de significância de 1%;
- rejeitar H_0 a qualquer n.s. $\alpha_0 > 4.29918\%$, como é o caso dos níveis usuais de significância de 5% e 10%.

(b) • **Coeficiente de determinação**

Tirando partido dos valores obtidos anteriormente, segue-se

$$\begin{aligned} r^2 &= \frac{\left(\sum_{i=1}^n x_i y_i - n\,\bar{x}\,\bar{y}\right)^2}{\left(\sum_{i=1}^n x_i^2 - n\,\bar{x}^2\right) \times \left(\sum_{i=1}^n y_i^2 - n\,\bar{y}^2\right)} \\ &= \frac{79.72^2}{171.6 \times 38.529} \\ &= 0.961235, \end{aligned}$$

i.e., cerca de 96.12% da variação total da variável resposta Y é explicada pelo modelo de regressão, o que indicia um muito bom ajustamento da recta estimada ao nosso conjunto de dados.