Probabilidades e Estatística

LEGM, LEIC-A, LEIC-T, MA, MEMec

1º semestre – 2018/2019
12/06/2019 – **11:00**

Duração: 90 minutos **2º Teste B**

Justifique convenientemente todas as respostas

Grupo I	10 valores

1. A variável aleatória X representa o erro associado a medições do teor de sílica em rochas magmáticas. Admita que X possui distribuição normal com valor esperado μ e variância σ^2, ambos desconhecidos. Seja $(X_1, X_2, \ldots, X_n)$ uma amostra aleatória de X.

(a) Sabendo que a variável aleatória $Z = \frac{1}{\sigma^2}\sum_{i=1}^n (X_i - \bar{X})^2$ é tal que $E(Z) = n-1$ e $V(Z) = 2(n-1)$, averigúe se $T = \frac{n-1}{n+1} S^2$ é um estimador centrado de σ^2 e calcule a respetiva variância. (2.0)

- **V.a. de interesse**
  $X \sim \text{normal}(\mu, \sigma^2)$
- **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**
  $E(X_i) = E(X) = \mu$
  $V(X_i) = V(X) = \sigma^2$
- **V.a. e resultados auxiliares**
  $Z = \frac{1}{\sigma^2}\sum_{i=1}^n (X_i - \bar{X})^2$
  $E(Z) = n-1$
  $V(Z) = 2(n-1)$
- **Estimador de σ^2**

$$\begin{aligned} T &= \frac{n-1}{n+1} S^2 \\ &= \frac{1}{n+1}\sum_{i=1}^n (X_i - \bar{X})^2 \\ &= \frac{\sigma^2}{n+1} Z \end{aligned}$$

- **Valor esperado de T**
  Ao tirar-se partido desta fórmula alternativa de T e do primeiro dos resultados auxiliares, tem-se:

$$\begin{aligned} E(T) &= E\left(\frac{\sigma^2}{n+1} Z\right) \\ &= \frac{\sigma^2}{n+1} \times E(Z) \\ &= \frac{\sigma^2}{n+1} \times (n-1) \\ &= \frac{n-1}{n+1}\sigma^2. \end{aligned}$$

- **Comentário**
  Uma vez que
  - T se diz um estimador centrado de σ^2 caso $E(T) = \sigma^2,\ \forall \sigma^2 > 0$,
  - $E(T) = \frac{n-1}{n+1}\sigma^2 \neq \sigma^2,\ \forall \sigma^2 > 0$,

  podemos concluir que T não é um estimador centrado de σ^2.

- **Variância de T**

$$\begin{aligned} V(T) &= V\left(\frac{\sigma^2}{n+1} Z\right) \\ &= \left(\frac{\sigma^2}{n+1}\right)^2 \times V(Z) \\ &= \frac{\sigma^4}{(n+1)^2} \times 2(n-1). \end{aligned}$$

(b) Determine a eficiência relativa do estimador S^2 com respeito ao estimador T. (2.5)

- **Outro estimador de σ^2**

$$\begin{aligned} S^2 &= \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 \\ &= \frac{\sigma^2}{n-1} Z \end{aligned}$$

- **Erro quadrático médio de S^2**

$$\begin{aligned} EQM_{\sigma^2}(S^2) &= V(S^2) + \left[bias_{\sigma^2}(S^2)\right]^2 \\ &= V(S^2) + \left[E(S^2) - \sigma^2\right]^2 \\ &= V\left(\frac{\sigma^2}{n-1} Z\right) + \left(\frac{n-1}{n-1}\sigma^2 - \sigma^2\right)^2 \\ &= \left(\frac{\sigma^2}{n-1}\right)^2 \times V(Z) \\ &= \frac{\sigma^4}{(n-1)^2} \times 2(n-1) \\ &= \frac{2\sigma^4}{n-1} \end{aligned}$$

- **Erro quadrático médio de T**

$$\begin{aligned} EQM_{\sigma^2}(T) &= E\left[\left(T - \sigma^2\right)^2\right] \\ &= V(T) + \left[bias_{\mu}(T)\right]^2 \\ &= V(T) + \left[E(T) - \sigma^2\right]^2 \\ &= V(T) + \left(\frac{n-1}{n+1}\sigma^2 - \sigma^2\right)^2 \\ &= V(T) + \left(\frac{-2}{n+1}\right)^2 \sigma^4 \\ &\overset{(a)}{=} \frac{\sigma^4}{(n+1)^2} \times 2(n-1) + \frac{4\sigma^4}{(n+1)^2} \\ &= \frac{2\sigma^4}{n+1}. \end{aligned}$$

- **Eficiência do estimador S^2 relativamente ao estimador T**

$$\begin{aligned} e_{\sigma^2}(S^2, T) &= \frac{EQM_{\sigma^2}(T)}{EQM_{\sigma^2}(S^2)} \\ &= \frac{\frac{2\sigma^4}{n+1}}{\frac{2\sigma^4}{(n-1)}} \\ &= \frac{n-1}{n+1} \end{aligned}$$

- **Comentário**

Como $e_{\sigma^2}(S^2, T) = \frac{n-1}{n+1} < 1, \forall n \in \mathbb{N}$ concluímos que $EQM_{\sigma^2}(T) < EQM_{\sigma^2}(S^2)$, pelo que podemos afirmar que S^2 é um estimador menos eficiente que T no que respeita à estimação de σ^2.

2. Recolheu-se uma amostra casual de 200 dias de operação de uma rede informática, tendo sido detetados acessos não autorizados em 15 desses 200 dias.

(a) Determine um intervalo de confiança a aproximadamente 95% para a verdadeira fração, p, de dias de operação da rede em que são detetados acessos não autorizados. (2.5)

- **V.a. de interesse**

  $X = \begin{cases} 1, & \text{dia de operação da rede com deteção de acessos não autorizados} \\ 0, & \text{c.c.} \end{cases}$

- **Situação**

  $X \sim \text{Bernoulli}(p)$

  $p = P(\text{dia op. com deteção acessos não autoriz})$ DESCONHECIDA

  $n = 200 > 30$ (suficientemente grande).

- **Obtenção de IC aproximado para p**

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

  [Uma vez que nos foi solicitada a determinação de um IC aproximado para uma probabilidade e a dimensão da amostra é suficientemente grande para justificar o recurso à seguinte v.a. fulcral para p com distribuição aproximada]

$$Z = \frac{\bar{X} - p}{\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}} \overset{a}{\sim} \text{normal}(0,1)$$

  Passo 2 — Obtenção dos quantis de probabilidade

  Os quantis a utilizar são

$$\begin{cases} a_\alpha = \Phi^{-1}(\alpha/2) = -\Phi^{-1}(1-\alpha/2) = -\Phi^{-1}(0.975) \overset{tabela/calc.}{=} -1.96 \\ b_\alpha = \Phi^{-1}(1-\alpha/2) = \Phi^{-1}(0.975) = 1.96. \end{cases}$$

  [Estes enquadram a v.a. fulcral para p com probabilidade aproximadamente igual a $(1-\alpha) = 0.95$.]

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

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

$$P\left[a_\alpha \leq \frac{\bar{X}-p}{\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}} \leq b_\alpha\right] \simeq 1-\alpha$$

$$P\left[\bar{X} - b_\alpha \times \sqrt{\frac{\bar{X}(1-\bar{X})}{n}} \leq p \leq \bar{X} - a_\alpha \times \sqrt{\frac{\bar{X}(1-\bar{X})}{n}}\right] \simeq 1-\alpha$$

$$P\left[\bar{X} - \Phi^{-1}(1-\alpha/2) \times \sqrt{\frac{\bar{X}(1-\bar{X})}{n}} \leq p \leq \bar{X} + \Phi^{-1}(1-\alpha/2) \times \sqrt{\frac{\bar{X}(1-\bar{X})}{n}}\right] \simeq 1-\alpha.$$

  Passo 4 — Concretização

  Ao ter-se em conta que

  - $n = 200$
  - $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i = \frac{15}{200} = 0.075$ [≡ proporção observada de dias com acessos não autorizados]
  - $\Phi^{-1}(1-\alpha/2) = 1.96$,

  conclui-se que o intervalo de confiança a aproximadamente 95% para p é dado por

$$\left[\bar{x} - \Phi^{-1}(1-\alpha/2) \times \sqrt{\frac{\bar{x}(1-\bar{x})}{n}}, \quad \bar{x} + \Phi^{-1}(1-\alpha/2) \times \sqrt{\frac{\bar{x}(1-\bar{x})}{n}}\right]$$

$$= \left[0.075 - 1.96 \times \sqrt{\frac{0.075 \times (1-0.075)}{200}}, \quad 0.075 + 1.96 \times \sqrt{\frac{0.075 \times (1-0.075)}{200}}\right]$$

$$= [0.038496, 0.111504].$$

(b) Com base na amostra referida, confronte as hipóteses $H_0 : p = 0.06$ e $H_1 : p > 0.06$. Decida com base no valor-p. (3.0)

- **Hipóteses**
  $H_0 : p = p_0 = 0.06$
  $H_1 : p > p_0$

- **Estatística de teste**
  [Sabe-se que o estimador de MV de p é $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, onde $X_i \sim_{i.i.d.} X$. Para além disso, $E(\bar{X}) = E(X) = p$ e $V(\bar{X}) = \frac{1}{n}V(X) = \frac{p(1-p)}{n} < +\infty$. Então pelo TLC pode afirmar-se que $\frac{\bar{X}-E(\bar{X})}{\sqrt{V(\bar{X})}} = \frac{\bar{X}-p}{\sqrt{\frac{p(1-p)}{n}}} \overset{a}{\sim}$ normal$(0,1)$, pelo que a estatística de teste é]

$$T = \frac{\bar{X}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \overset{a}{\sim}_{H_0} \text{normal}(0,1).$$

- **Região de rejeição de H_0** (para valores de T)
  Tratando-se de um teste unilateral superior ($H_1 : p > p_0$), a região de rejeição de H_0, escrita 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

$$\begin{aligned} t &= \frac{\bar{x}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \\ &= \frac{0.075-0.06}{\sqrt{\frac{0.06\,(1-0.06)}{200}}} \\ &\simeq 0.89. \end{aligned}$$

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

$$\begin{aligned} valor-p &= P(T > t \mid H_0) \\ &= 1 - P(T \le t \mid H_0)] \\ &\simeq 1 - \Phi(t) \\ &\simeq 1 - \Phi(0.89) \\ &\overset{tabela/calc.}{=} 0.1867. \end{aligned}$$

  Logo é suposto:
  - não rejeitar H_0 a qualquer n.s. $\alpha_0 \le 18.67\%$, por exemplo, a qualquer dos n.u.s. (1%, 5% e 10%);
  - rejeitar H_0 a qualquer n.s. $\alpha_0 > 18.67\%$.

Grupo II	10 valores

1. Uma engenheira defende a hipótese H_0 de que, em determinada região do globo, o tempo (em ano, X) entre dois eventos sísmicos consecutivos possui função de distribuição $F(x) = 1 - e^{-\lambda x}$, $x > 0$, onde λ é um parâmetro positivo desconhecido.

Uma amostra casual de 100 registos de tais tempos conduziu à seguinte tabela de frequências:

Classe	$]0,1]$	$]1,2]$	$]2,3]$	$]3,4]$	$]4,+\infty[$
Frequência absoluta observada	24	11	8	11	46
Estimativa da frequência absoluta esperada sob H_0	e_1	14.84	12.15	9.95	e_5

(a) Sabendo que a estimativa de máxima verosimilhança de λ é $\hat{\lambda} \simeq 0.2$, obtenha os valores das estimativas e_1 e e_5 (aproximando-as às centésimas). (1.0)

- **V.a. de interesse**
  X = tempo entre dois eventos sísmicos consecutivos em determinada região do globo

- **F.d. conjecturada**
  $F(x) = P(X \le x \mid \lambda) = 1 - e^{-\lambda x}$, $x > 0$, onde λ é um parâmetro positivo desconhecido.

- **Estimativas das frequências absolutas esperadas omissas**
  Atendendo à dimensão da amostra $n = 100$, à f.d. conjecturada, à estimativa de MV facultada $\hat{\lambda} \simeq 0.2$ e à propriedade de invariância dos EMV, temos:

$$
\begin{aligned}
e_1 &= n \times \widehat{P(X \le 1)} \\
&= n \times P(X \le 1 \mid \lambda = \hat{\lambda}) \\
&= 100 \times \left(1 - e^{-\hat{\lambda}}\right) \\
&\simeq 18.13; \\
e_5 &= n \times \widehat{P(X > 4)} \\
&= n \times P(X > 4 \mid \lambda = \hat{\lambda}) \\
&= n - \sum_{i=1}^{4} e_i \\
&\simeq 100 - (18.13 + 14.84 + 12.15 + 9.95) \\
&= 44.93.
\end{aligned}
$$

(b) Teste H_0, ao nível de significância de 10%. (3.0)

- **Hipóteses**
  $H_0 : X \sim \text{Exponencial}(\lambda)$ $\qquad$ (λ desconhecido)
  $H_1 : X \not\sim \text{Exponencial}(\lambda)$

- **Nível de significância**
  $\alpha_0 = 10\%$

- **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 = ESTIMADOR da frequência absoluta esperada, sob H_0, da classe i

  β = No. de parâmetros a estimar = 1 [dado que λ é um parâmetro desconhecido.]

- **Estimativas das frequências absolutas esperadas sob H_0**
  De acordo com a tabela facultada e a alínea (a), as estimativas [de MV] das frequências absolutas esperadas sob H_0 aproximadas às centésimas são: $e_1 \simeq 18.13$; $e_2 \simeq 14.84$; $e_3 \simeq 12.15$; $e_4 \simeq 9.95$; $e_5 \simeq 44.93$.
  [Não é necessário fazer qualquer agrupamento de classes uma vez que em pelo menos 80% das classes se verifica $e_i \ge 5$ e que $e_i \ge 1$ 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)
  Lidamos com um teste de ajustamento, logo a região de rejeição de H_0 é o intervalo à direita $W = (c, +\infty)$, onde

$$
\begin{aligned}
c &= F^{-1}_{\chi^2_{(k-\beta-1)}}(1-\alpha_0) \\
&= F^{-1}_{\chi^2_{(5-1-1)}}(1-0.1) \\
&\overset{tabela/calc.}{=} 6.251.
\end{aligned}
$$

- **Decisão**

i	Classe i	Freq. abs. obs. o_i	Estim. freq. abs. esp. sob H_0 e_i	Parcelas valor obs. estat. teste $\frac{(o_i-e_i)^2}{e_i}$
1	$]0,1]$	24	18.13	$\frac{(24-18.13)^2}{18.13} \simeq 1.901$
2	$]1,2]$	11	14.84	$\frac{(11-14.84)^2}{14.84} \simeq 0.994$
3	$]2,3]$	8	12.15	1.417
4	$]3,4]$	11	9.95	0.111
5	$]4,+\infty[$	46	44.93	0.025
		$\sum_{i=1}^k o_i = n = 100$	$\sum_{i=1}^k e_i = n = 100$	$t = \sum_{i=1}^k \frac{(o_i-e_i)^2}{e_i} \simeq 4.448$

Uma vez que $t \simeq 4.448 \not\in W = (6.251, +\infty)$, não devemos rejeitar H_0 ao n.s. de $\alpha_0 = 10\%$ [nem a qualquer outro n.s. inferior a α_0].

2. Um conjunto de 10 medições independentes conduziu aos seguintes resultados que dizem respeito ao quadrado do número de *frames* do GOP (*group of pictures*) de um vídeo (x) e ao seu tempo de transcodificação (Y, em segundo):

$$\sum_{i=1}^{10} x_i = 48\,193\,, \quad \sum_{i=1}^{10} x_i^2 = 318\,634\,213\,, \quad \sum_{i=1}^{10} y_i = 73.8\,, \quad \sum_{i=1}^{10} y_i^2 = 731.30\,, \quad \sum_{i=1}^{10} x_i y_i = 475\,295.1\,.$$

(a) Considerando o modelo de regressão linear simples de Y em x, determine a estimativa de mínimos quadrados dos coeficientes β_0 e β_1. (1.5)

- **Estimativas de MQ de β_0 e β_1**

Dado que

$n = 10$

$\sum_{i=1}^n x_i = 48\,193$

$\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i = \frac{48\,193}{10} = 4\,819.3$

$\sum_{i=1}^n x_i^2 = 318\,634\,213$

$\sum_{i=1}^n x_i^2 - n\,(\bar{x})^2 = 318\,634\,213 - 10 \times 4\,819.3^2 = 86\,377\,688.1$

$\sum_{i=1}^n y_i = 73.8$

$\bar{y} = \frac{1}{n}\sum_{i=1}^n y_i = \frac{73.8}{10} = 7.38$

$\sum_{i=1}^n y_i^2 = 731.30$

$\sum_{i=1}^n y_i^2 - n\,(\bar{y})^2 = 731.30 - 10 \times 7.38^2 = 186.6560$

$\sum_{i=1}^n x_i y_i = 475\,295.1$

$\sum_{i=1}^n x_i y_i - n\,\bar{x}\,\bar{y} = 475\,295.1 - 10 \times 4\,819.3 \times 7.38 = 119\,630.760,$

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

$$\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_{i=1}^n x_i y_i - n\,\bar{x}\,\bar{y}}{\sum_{i=1}^n x_i^2 - n\,(\bar{x})^2} \\
&= \frac{119\,630.760}{86\,377\,688.1} \\
&\simeq 0.00138497 \\
\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \times \bar{x} \\
&\simeq 7.38 - 0.00138497 \times 4\,819.3 \\
&\simeq 0.705414.
\end{aligned}$$

(b) Admitindo a validade das hipóteses de trabalho habituais para inferências neste modelo de regressão, obtenha um intervalo de confiança a 95% para β_0. (3.5)

Confronte $H_0 : \beta_0 = 0$ e $H_1 : \beta_0 \neq 0$, ao nível de significância de 5%, tirando partido do intervalo que obteve.

- **[Hipóteses de trabalho**

  $\epsilon_i \stackrel{i.i.d.}{\sim} \text{Normal}(0,\sigma^2),\ i = 1,\dots,n$]

- **Obtenção do IC para β_0**

  Passo 1 — V.a. fulcral para β_0

$$Z = \frac{\hat\beta_0 - \beta_0}{\sqrt{\hat\sigma^2 \times \left(\frac{1}{n} + \frac{\bar x^2}{\sum_{i=1}^n x_i^2 - n\,\bar x^2}\right)}} \sim t_{(n-2)}$$

Passo 2 — Quantis de probabilidade

Como $n = 10$ e $(1-\alpha)\times 100\% = 95\%$, usaremos os quantis de probabilidade simétricos $a_\alpha = -b_\alpha$ dados por:

$$\begin{cases} a_\alpha = F^{-1}_{t_{(n-2)}}(\alpha/2) = -F^{-1}_{t_{(10-2)}}(1-0.05/2) = -F^{-1}_{t_{(8)}}(0.975) \stackrel{tabela/calc.}{=} -2.306 \\ b_\alpha = F^{-1}_{t_{(10-2)}}(1-0.05/2) = F^{-1}_{t_{(8)}}(0.975) \stackrel{tabela/calc.}{=} 2.306. \end{cases}$$

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

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

$$P\left[-F^{-1}_{t_{(n-2)}}(1-\alpha/2) \leq \frac{\hat\beta_0-\beta_0}{\sqrt{\hat\sigma^2\times\left(\frac{1}{n}+\frac{\bar x^2}{\sum_{i=1}^n x_i^2-n\,\bar x^2}\right)}} \leq F^{-1}_{t_{(n-2)}}(1-\alpha/2)\right] = 1-\alpha$$

$$P\left[\hat\beta_0 - F^{-1}_{t_{(n-2)}}(1-\alpha/2)\times\sqrt{\hat\sigma^2\times\left(\frac{1}{n}+\frac{\bar x^2}{\sum_{i=1}^n x_i^2-n\,\bar x^2}\right)} \leq \beta_0 \leq \hat\beta_0 + F^{-1}_{t_{(n-2)}}(1-\alpha/2)\times\sqrt{\hat\sigma^2\times\left(\frac{1}{n}+\frac{\bar x^2}{\sum_{i=1}^n x_i^2-n\,\bar x^2}\right)}\right] = 1-\alpha$$

Passo 4 — Concretização

Uma vez que a estimativa de σ^2 é igual a

$$\begin{aligned} \hat\sigma^2 &= \frac{1}{n-2}\left[\left(\sum_{i=1}^n y_i^2 - n\,\bar y^2\right) - (\hat\beta_1)^2\left(\sum_{i=1}^n x_i^2 - n\,\bar x^2\right)\right] \\ &\simeq \frac{1}{10-2}\left(186.6560 - 0.00138497^2\times 86\,377\,688.1\right) \\ &\simeq 2.621417 \end{aligned}$$

e a expressão geral do IC pretendido é

$$IC_{(1-\alpha)\times 100\%}(\beta_0) = \left[\hat\beta_0 \pm F^{-1}_{t_{(n-2)}}(1-\alpha/2)\times\sqrt{\hat\sigma^2\times\left(\frac{1}{n}+\frac{\bar x^2}{\sum_{i=1}^n x_i^2-n\,\bar x^2}\right)}\right],$$

temos

$$\begin{aligned} IC_{95\%}(\beta_0) &\simeq \left[0.705414 \pm 2.306\times\sqrt{2.621417\times\left(\frac{1}{10}+\frac{4\,819.3^2}{86\,377\,688.1}\right)}\right] \\ &\simeq [0.705414 \pm 2.306\times 0.983362] \\ &\simeq [-1.562219, 2.973047]. \end{aligned}$$

- **Hipóteses**
  $H_0 : \beta_0 = \beta_{0,0} = 0$
  $H_1 : \beta_0 \neq \beta_{0,0}$

- **N.s.**
  $\alpha_0 = 0.05$

- **Decisão**
  Invocando a relação entre intervalos de confiança e testes de hipóteses (bilaterais), não devemos rejeitar a hipótese $H_0 : \beta_0 = \beta_{0,0} = 0$ (a favor da hipótese $H_1 : \beta_0 \neq \beta_{0,0}$), ao n.s. de $\alpha \times 100\% = 100\% - 95\% = 5\%$ [ou a qualquer n.s. inferior a 5%] já que

  $$\beta_{0,0} = 0 \in IC_{95\%}(\beta_0) = [-1.562219, 2.973047].$$

(c) Determine e interprete o valor do coeficiente de determinação do modelo ajustado. (1.0)

- **Cálculo do coeficiente de determinação**

  $$\begin{aligned} r^2 &= \frac{\left(\sum_{i=1}^n x_i y_i - n\,\bar{x}\,\bar{y}\right)^2}{\left(\sum_{i=1}^n x_i^2 - n\,\bar{x}^2\right) \times \left(\sum_{i=1}^n y_i^2 - n\,\bar{y}^2\right)} \\ &\stackrel{(a)}{=} \frac{119\,630.760^2}{86\,377\,688.1 \times 186.6560} \\ &\simeq 0.887651. \end{aligned}$$

- **Interpretação coeficiente de determinação**
  Cerca de 88.8% da variação total da variável resposta Y é explicada pela variável x, através do modelo de regressão linear simples ajustado, donde possamos afirmar que a recta estimada parece ajustar-se bem ao conjunto de dados.