DM DEPARTAMENTO DE MATEMÁTICA TÉCNICO LISBOA

Probabilidades e Estatística

LEE, LEIC-A, LEIC-T, LEMat, LERC, MEBiol, MEBiom, MEEC, MEFT, MEMec, MEQ

1º semestre – 2017/2018
11/01/2018 – **11:00**

Duração: 90 minutos **2º teste B**

Justifique convenientemente todas as respostas

Grupo I	10 valores

1. A concentração de um nutriente em determinado produto alimentar é uma variável aleatória com função de densidade de probabilidade dada por

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

onde θ é um parâmetro positivo desconhecido. Seja $(X_1, \ldots, X_n)$ uma amostra aleatória de X.

(a) Mostre que o estimador de máxima verosimilhança do parâmetro θ, com base na amostra aleatória referida acima, é dado por $\frac{-n}{\sum_{i=1}^n \ln(X_i)} - 1$. (2.5)

- **V.a. de interesse**
  X = concentração de um nutriente em determinado produto alimentar
- **F.d.p. de X**
  $$f_X(x) = \begin{cases} (\theta+1)\, x^{\theta}, & 0 < x < 1 \\ 0\,, & \text{caso contrário} \end{cases}$$
- **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

$$\begin{aligned} L(\theta|\underline{x}) &= f_{\underline{X}}(\underline{x}) \\ &\overset{X_i\ indep}{=} \prod_{i=1}^n f_{X_i}(x_i) \\ &\overset{X_i \sim X}{=} \prod_{i=1}^n f_X(x_i) \\ &= \prod_{i=1}^n \left[(\theta+1)\, x_i^{\theta}\right] \\ &= (\theta+1)^n \left(\prod_{i=1}^n x_i\right)^{\theta}, \quad \theta > 0 \end{aligned}$$

  Passo 2 — Função de log-verosimilhança

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

  Passo 3 — Maximização

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

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

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

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

Passo 4 — Estimador de MV de θ

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

(b) Obtenha a estimativa de máxima verosimilhança de $h(\theta) = \frac{\theta+1}{\theta+2}$ baseada na concretização $(x_1, x_2, x_3, x_4, x_5) = (0.32, 0.24, 0.56, 0.67, 0.58)$ para a qual $\sum_{i=1}^{n} \ln(x_i) \simeq -4.09$. (1.5)

- **Estimativa de MV de θ**

$$\begin{aligned} \hat{\theta} &= \frac{-n}{\sum_{i=1}^{n} \ln(x_i)} - 1 \\ &\simeq \frac{-5}{-4.09} - 1 \\ &\simeq 0.222494 \end{aligned}$$

- **Outro parâmetro desconhecido**

  $h(\theta) = \frac{\theta+1}{\theta+2}$

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

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

$$\begin{aligned} \widehat{h(\theta)} &= h(\hat{\theta}) \\ &= \frac{\hat{\theta}+1}{\hat{\theta}+2} \\ &\simeq \frac{0.222494+1}{0.222494+2} \\ &\simeq 0.550055. \end{aligned}$$

(c) Averigúe se $\bar{X}$ é um estimador centrado de $h(\theta)$. (1.0)

- **Parâmetro desconhecido**

  $h(\theta) = \frac{\theta+1}{\theta+2}$

- **Estimador de $h(\theta)$**

  $\bar{X}$

- **Valor esperado de $\bar{X}$**

  Uma vez que $X_i \overset{i.i.d.}{\sim} X$, $i = 1, \ldots, n$, segue-se que:

$$\begin{aligned} E(\bar{X}) &= E(X) \\ &= \int_0^1 (\theta+1)\, x^{\theta+1}\, dx \\ &= \left. \frac{\theta+1}{\theta+2}\, x^{\theta+2} \right|_0^1 \\ &= \frac{\theta+1}{\theta+2} \\ &\equiv h(\theta). \end{aligned}$$

- **Conclusão**

  Uma vez que

  - T se diz um estimador centrado de $h(\theta)$ caso $E(T) = h(\theta), \forall \theta > 0$,

  podemos afirmar que $\bar{X}$ é um estimador centrado de $h(\theta)$.

2. O diâmetro (X, em mm) dos cilindros hidráulicos produzidos por determinada fábrica possui distribuição normal com parâmetros desconhecidos μ e σ^2. A concretização de uma amostra aleatória de dimensão 10 conduziu aos seguintes resultados: $\sum_{i=1}^{10} x_i = 846$ e $\sum_{i=1}^{10} x_i^2 = 71\,607$.

(a) Determine um intervalo de confiança a 95% para σ^2. (2.5)

- **V.a. de interesse**
  X = diâmetro de cilindro hidráulico produzido pela fábrica

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

- **Obtenção do IC para σ^2**
  Passo 1 — Selecção da v.a. fulcral para σ^2

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

[uma vez que é suposto determinar um IC para a variância de uma população normal, com valor esperado desconhecido.]

Passo 2 — Obtenção dos quantis de probabilidade
Ao ter-se em consideração que $n = 10$ e $(1-\alpha)\times 100\% = 95\%$, far-se-á uso dos quantis

$$\begin{cases} a_\alpha = F^{-1}_{\chi^2_{(n-1)}}(\alpha/2) = F^{-1}_{\chi^2_{(9)}}(0.025) \stackrel{tabela/calc.}{=} 2.700 \\ b_\alpha = F^{-1}_{\chi^2_{(n-1)}}(1-\alpha/2) = F^{-1}_{\chi^2_{(9)}}(0.975) \stackrel{tabela/calc.}{=} 19.02. \end{cases}$$

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

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

$$P\left[a_\alpha \le \frac{(n-1)S^2}{\sigma^2} \le b_\alpha\right] = 1-\alpha$$

$$P\left[\frac{1}{b_\alpha} \le \frac{\sigma^2}{(n-1)S^2} \le \frac{1}{a_\alpha}\right] = 1-\alpha$$

$$P\left[\frac{(n-1)S^2}{b_\alpha} \le \sigma^2 \le \frac{(n-1)S^2}{a_\alpha}\right] = 1-\alpha$$

Passo 4 — Concretização
Atendendo ao par de quantis acima e ao facto de

$$\begin{aligned} s^2 &= \frac{1}{n-1}\left[\sum_{i=1}^n x_i^2 - n\,(\bar{x})^2\right] \\ &= \frac{1}{10-1}\left[71\,607 - 10\times(846/10)^2\right] \\ &= 3.9(3) \end{aligned}$$

$$IC_{(1-\alpha)\times 100\%}(\sigma^2) = \left[\frac{(n-1)\,s^2}{F^{-1}_{\chi^2_{(n-1)}}(1-\alpha/2)},\ \frac{(n-1)\,s^2}{F^{-1}_{\chi^2_{(n-1)}}(\alpha/2)}\right],$$

segue-se:

$$\begin{aligned} IC_{95\%}(\sigma^2) &= \left[\frac{(10-1)\times 3.9(3)}{19.02},\ \frac{(10-1)\times 3.9(3)}{2.700}\right] \\ &\simeq [1.860928,\ 13.109220]. \end{aligned}$$

(b) Teste $H_0: \sigma^2 = 4$ contra $H_1: \sigma^2 > 4$. Decida com base no valor-p. (2.5)

- **V.a. de interesse e situação**
  Ver alínea (a).

- **Hipóteses**
  $H_0 : \sigma^2 = \sigma_0^2 = 4$
  $H_1 : \sigma^2 > \sigma_0^2$

- **Estatística de teste**

$$T = \frac{(n-1)S^2}{\sigma_0^2} \sim_{H_0} \chi^2_{(n-1)}$$

  [dado que se pretende efectuar um teste sobre a variância de uma população normal, com valor esperado desconhecido.]

- **Região de rejeição de H_0** (para valores da estatística de teste)
  Estamos a lidar com um teste unilateral superior ($H_1 : \sigma^2 > \sigma_0^2$), logo 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 igual a

$$\begin{aligned} t &= \frac{(n-1)s^2}{\sigma_0^2} \\ &= \frac{(10-1)\times 3.9(3)}{4} \\ &\simeq 8.85. \end{aligned}$$

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

$$\begin{aligned} valor-p &= P(T > t \mid H_0) \\ &= 1 - F_{\chi^2_{(n-1)}}(t) \\ &= 1 - F_{\chi^2_{(9)}}(8.85) \\ &\stackrel{calc/tabela}{=} 0.451234. \end{aligned}$$

  Deste modo é suposto:
  - não rejeitar H_0 a qualquer n.s. $\alpha_0 \le 45.1234\%$, pelo que H_0 não é contrariada pelos dados aos n.u.s. $(1\%, 5\%, 10\%)$;
  - rejeitar H_0 a qualquer n.s. $\alpha_0 > 45.1234\%$.

  [Em alternativa, poderíamos recorrer às tabelas de quantis da distribuição do qui-quadrado com 9 graus de liberdade e adiantar um intervalo para o *p-value*:

$$\begin{aligned} F^{-1}_{\chi^2_{(9)}}(0.50) = 8.343 &< t = 8.85 < 9.414 = F^{-1}_{\chi^2_{(9)}}(0.60) \\ 0.50 &< F_{\chi^2_{(9)}}(8.85) < 0.60 \\ 1-0.60 &< 1 - F_{\chi^2_{(9)}}(8.85) < 1-0.50 \\ 0.40 &< valor-p < 0.50. \end{aligned}$$

  Assim, é suposto:
  - não rejeitar H_0 a qualquer n.s. $\alpha_0 \le 40\%$, pelo que H_0 não é contrariada pelos dados aos n.u.s. $(1\%, 5\%, 10\%)$;
  - rejeitar H_0 a qualquer n.s. $\alpha_0 \ge 50\%$.

Grupo II — 10 valores

1. Ao recorrer a um pequeno programa destinado a gerar 250 números pseudo-aleatórios no intervalo [0, 10], obtiveram-se os seguintes dados:

Classe	$[0,2]$	$]2,4]$	$]4,6]$	$]6,8]$	$]8,10]$
Frequência absoluta observada	38	55	54	41	62

Uma engenheira informática defende a hipótese H_0 de que o programa gera números pseudo-aleatórios que seguem uma distribuição uniforme contínua no intervalo [0, 10].

(a) Calcule os valores das frequências absolutas esperadas sob H_0 de cada uma das classes. (1.0)

- **V.a. de interesse**
  X = número pseudo-aleatório gerado pelo programa

- **Distribuição, f.d.p. e f.d. conjecturadas**
  $X \sim$ uniforme contínua$(0,10)$

  $$f_X(x) = \begin{cases} \frac{1}{10}, & 0 \le x \le 10 \\ 0, & \text{c.c.} \end{cases}$$

  $$F_X(x) = P(X \le x) = \int_{-\infty}^{x} f_X(t)\,dt = \begin{cases} \int_{-\infty}^{x} 0\,dt = 0, & x < 0 \\ \int_{-\infty}^{0} 0\,dt + \int_{0}^{x} \frac{1}{10}\,dt = \frac{x}{10}, & 0 \le x \le 10 \\ \int_{-\infty}^{0} 0\,dt + \int_{0}^{10} \frac{1}{10}\,dt + \int_{10}^{x} 0\,dt = 1, & x > 10. \end{cases}$$

- **Frequências absolutas esperadas**
  Atendendo à dimensão da amostra $n = 250$ e à f.d. conjecturada, segue-se, para $i = 1, \ldots, 5$:

  $$\begin{aligned} E_i &= n \times [F(2i) - F(2i-2)] \\ &= 250 \times \left(\frac{2i}{10} - \frac{2i-2}{10}\right) \\ &= 250 \times \frac{1}{5} \\ &= 50. \end{aligned}$$

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

- **Hipóteses**
  $H_0 : X \sim$ uniforme contínua$(0,10)$
  $H_1 : X \not\sim$ uniforme contínua$(0,10)$

- **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 = Frequência absoluta esperada, sob H_0, da classe i

  β = No. de parâmetros a estimar = 0 [dado que em H_0 se conjectura uma distribuição específica.]

- **Frequências absolutas esperadas sob H_0**
  De acordo com (a), os valores das freq. absolutas esperadas sob H_0 são: $E_i = 50$, $i = 1, \ldots, 5$.

  [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)
  Tratando-se de 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

  $$\begin{aligned} c &= F^{-1}_{\chi^2_{(k-\beta-1)}}(1-\alpha_0) \\ &= F^{-1}_{\chi^2_{(5-0-1)}}(1-0.10) \\ &= F^{-1}_{\chi^2_{(4)}}(0.90) \\ &\overset{tabela/calc.}{=} 7.779. \end{aligned}$$

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

i	Classe i	Freq. abs. obs. o_i	Freq. abs. esp. sob H_0 E_i	Parcelas valor obs. estat. teste $\frac{(o_i-E_i)^2}{e_i}$
1	$[0,2]$	38	50	$\frac{(38-50)^2}{50} = 2.88$
2	$]2,4]$	55	50	0.50
3	$]4,6]$	54	50	0.32
4	$]6,8]$	41	50	1.62
5	$]8,10]$	62	50	2.88
		$\sum_{i=1}^{k} o_i = n$ $= 250$	$\sum_{i=1}^{k} e_i = n$ $= 250$	$t = \sum_{i=1}^{k} \frac{(o_i-e_i)^2}{e_i}$ $\simeq 8.20$

  Como $t \simeq 8.20 \in W = (7.779, +\infty)$, devemos rejeitar H_0 ao n.s. de $\alpha_0 = 10\%$ [ou a qualquer outro n.s. superior a 10%].

2. A densidade relativa de certa madeira é descrita pela variável x e a força máxima necessária para o seu esmagamento em compressão paralela ao grão é representada pela variável aleatória Y (em psi). Uma amostra de dimensão 10 conduziu aos seguintes valores:

$$\sum_{i=1}^{10} x_i = 4.31, \quad \sum_{i=1}^{10} x_i^2 = 1.8629, \quad \sum_{i=1}^{10} y_i = 24\,520, \quad \sum_{i=1}^{10} y_i^2 = 61\,761\,600, \quad \sum_{i=1}^{10} x_i\, y_i = 10\,632.9,$$

onde $\left[\min_{i=1,\ldots,10} x_i, \max_{i=1,\ldots,10} x_i\right] = [0.39\,, 0.47]$.

(a) Após ter enunciado as hipóteses de trabalho que entender convenientes, calcule as estimativas de máxima verosimilhança dos parâmetros da reta de regressão linear simples de Y em x, bem como a estimativa de máxima verosimilhança de $E(Y \mid x = 0.431)$. (2.5)

- **Hipóteses de trabalho**
  $\epsilon_i \overset{i.i.d.}{\sim} \text{Normal}(0, \sigma^2)$, $i = 1, \ldots, n$

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

  $n = 10$

  $\sum_{i=1}^{n} x_i = 4.31$
  $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{4.31}{10} = 0.431$
  $\sum_{i=1}^{n} x_i^2 = 1.8629$
  $\sum_{i=1}^{n} x_i^2 - n\,(\bar{x})^2 = 1.8629 - 10 \times 0.431^2 = 0.00529$

  $\sum_{i=1}^{n} y_i = 24\,520$
  $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{24\,520}{10} = 2\,452$
  $\sum_{i=1}^{n} y_i^2 = 61\,761\,600$
  $\sum_{i=1}^{n} y_i^2 - n\,(\bar{y})^2 = 61\,761\,600 - 10 \times 2\,452^2 = 1\,638\,560$

  $\sum_{i=1}^{n} x_i y_i = 10\,632.9$
  $\sum_{i=1}^{n} x_i y_i - n\,\bar{x}\,\bar{y} = 10\,632.9 - 10 \times 0.431 \times 2\,452 = 64.78,$

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

$$\begin{aligned}\hat{\beta}_1 &= \frac{\sum_{i=1}^{n} x_i y_i - n\,\bar{x}\,\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\,(\bar{x})^2} \\ &= \frac{64.78}{0.00529} \\ &\simeq 12\,245.7\end{aligned}$$

$$
\begin{aligned}
\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \times \bar{x} \\
&\simeq 2\,452 - 12\,245.7 \times 0.431 \\
&\simeq -2\,825.9.
\end{aligned}
$$

- **Estimativa de MV para** $E(Y \mid x = x_0) = \beta_0 + \beta_1 x_0$, com $x_0 = 0.431$

$$
\begin{aligned}
\hat{E}(Y \mid x = x_0) &= \hat{\beta}_0 + \hat{\beta}_1 x_0 \\
&= -2\,825.9 + 12\,245.7 \times 0.431 \\
&\simeq 2452.
\end{aligned}
$$

[Não cometemos qualquer erro de extrapolação ao estimar pontualmente $E(Y \mid x = x_0) = \beta_0 + \beta_1 \times x_0$ uma vez que $x_0 \in [\min_{i=1,\dots,n} x_i, \max_{i=1,\dots,n} x_i]$. Note-se que, neste caso $x_0 = \bar{x}$, logo $\hat{E}(Y \mid x = x_0) = \hat{\beta}_0 + \hat{\beta}_1\,\bar{x} = (\bar{y} - \hat{\beta}_1 \times \bar{x}) + \hat{\beta}_1\,\bar{x} \equiv \bar{y}$, como se pôde constatar acima.]

(b) Teste a hipótese $E(Y \mid x = 0.431) = 2\,500$, ao nível de significância de 5%. (2.5)

- **Hipóteses** (com $x_0 = 0.431$)
  $H_0 : E(Y \mid x = x_0) = E_0(Y \mid x_0) = 2\,500$ vs.
  $H_1 : E(Y \mid x = x_0) \neq E_0(Y \mid x_0)$

- **Nível de significância**
  $\alpha_0 = 5\%$

- **Estatística de teste**

$$
T = \frac{(\hat{\beta}_0 + \hat{\beta}_1 x_0) - E_0(Y|x_0)}{\sqrt{\hat{\sigma}^2 \times \left[\frac{1}{n} + \frac{(x_0-\bar{x})^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}\right]}} \sim_{H_0} t_{(n-2)}
$$

- **Região de rejeição de** H_0 (para valores de T)
  Estamos a lidar com um teste bilateral ($H_1 : E(Y \mid x_0) \neq E_0(Y \mid x_0)$), pelo que a região de rejeição de H_0 [escrita para valores observados da estatística de teste] é $W = (-\infty, -c) \cup (c, +\infty)$, onde

$$
\begin{aligned}
c \;:\;\; & P(\text{Rejeitar } H_0 \mid H_0) = \alpha_0 \\
& c = F^{-1}_{t_{(n-2)}}(1 - \alpha_0/2) \\
& c = F^{-1}_{t_{(8)}}(0.975) \\
& c \overset{tabela/calc}{=} 2.306
\end{aligned}
$$

- **Decisão**
  Tendo em conta que

$$
\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{n-2}\left[\left(\sum_{i=1}^n y_i^2 - n\,\bar{y}^2\right) - (\hat{\beta}_1)^2 \left(\sum_{i=1}^n x_i^2 - n\,\bar{x}^2\right)\right] \\
&\simeq \frac{1}{10-2}\left(1\,638\,560 - 12\,245.7^2 \times 0.00529\right) \\
&\simeq 105660.1,
\end{aligned}
$$

o valor observado da estatística de teste é igual a

$$
\begin{aligned}
t &= \frac{(\hat{\beta}_0 + \hat{\beta}_1 x_0) - E_0(Y|x_0)}{\sqrt{\hat{\sigma}^2 \times \left[\frac{1}{n} + \frac{(x_0-\bar{x})^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}\right]}} \\
&= \frac{2\,452 - 2500}{\sqrt{105660.1 \times \left[\frac{1}{10} + \frac{(0.431-0.431)^2}{0.00529}\right]}} \\
&\simeq -0.466967.
\end{aligned}
$$

Como $t \simeq -0.466967 \not\in W = (-\infty, -2.306) \cup (2.306, +\infty)$, não devemos rejeitar H_0, ao nível de significância $\alpha_0 = 5\%$ [ou a qualquer outro n.s. inferior a 5%].

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

- **Cálculo do coeficiente de determinação**

$$
\begin{aligned}
r^2 &= \frac{\left(\sum_{i=1}^n x_i y_i - n\,\bar{x}\,\bar{y}\right)^2}{\left(\sum_{i=1}^n x_i^2 - n\,\bar{x}^2\right) \times \left(\sum_{i=1}^n y_i^2 - n\,\bar{y}^2\right)} \\
&= \frac{64.78^2}{0.00529 \times 1638560} \\
&\simeq 0.484132.
\end{aligned}
$$

- **Interpretação coeficiente de determinação**
Cerca de 48.4% 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 não parece ajustar-se bem ao conjunto de dados.