Probabilidades e Estatística

— TODOS OS CURSOS —

1º semestre – 2019/2020
29/01/2020 – **15:00**

Duração: 90 minutos **2º Teste C**

Justifique convenientemente todas as respostas

Grupo I	10 valores

1. Suponha que o controlo de qualidade de determinado equipamento produzido por uma unidade fabril passa pelo registo do número X de equipamentos inspecionados de modo independente até que se observe o primeiro equipamento que sofra uma falha.

(a) Deduza o estimador de máxima verosimilhança da probabilidade de falha do equipamento em cada inspeção, com base numa amostra aleatória de dimensão n proveniente da população X. (3.0)

- **V.a. de interesse**
  $X =$ no. total de inspecções até à ocorrência de uma falha
  $X \sim \text{geométrica}(p)$
- **F.p. de X**
  $P(X = x) = (1-p)^x\, p, \quad x = 1, 2, \ldots$
- **Parâmetro desconhecido**
  $p, \quad p \in (0,1)$
- **Amostra aleatória**
  $\underline{X} = (X_1, \ldots, X_n)$
  $X_i \overset{i.i.d}{\sim} X, \quad i = 1, \ldots, n$
- **Amostra**
  $\underline{x} = (x_1, \ldots, x_n)$ amostra de dimensão n proveniente da população X
- **Obtenção da estimativa de MV de p**

  Passo 1 — Função de verosimilhança

$$
\begin{aligned}
L(p|\underline{x}) &= P(\underline{X} = \underline{x}) \\
&\overset{X_i \overset{i.i.d}{\sim} X}{=} \prod_{i=1}^{n} P(X = x_i) \\
&= \prod_{i=1}^{n} \left[(1-p)^{x_i - 1}\, p\right] \\
&= (1-p)^{\sum_{i=1}^n x_i - n}\, p^n, \quad 0 < p < 1
\end{aligned}
$$

  Passo 2 — Função de log-verosimilhança

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

  Passo 3 — Maximização
  A estimativa de MV de p é doravante representada por $\hat{p}$ e

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

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

$$\begin{cases} -\hat{p}\sum_{i=1}^n x_i + n\hat{p} + n - n\hat{p} = 0 & (\hat{p} \neq 0,1) \\ \text{prop. verdadeira já que} \quad \sum_{i=1}^n x_i \geq n, \quad n \in \mathbb{N} \end{cases}$$

$$\begin{cases} \hat{p} = \frac{n}{\sum_{i=1}^n x_i} \\ --- \end{cases}$$

- **Passo 4 — Estimador de MV de p**

$$EMV(p) = \frac{n}{\sum_{i=1}^n X_i} \qquad [\equiv 1/\bar{X}]$$

(b) Determine a estimativa de máxima verosimilhança da variância de X sabendo que dispomos dos seguintes sete registos do número de inspeções até que o primeiro equipamento sofra uma falha: $4, 5, 3, 6, 4, 3, 3$. (2.0)

- **Estimativa de MV de p**

$$\begin{aligned} \hat{p} &= \frac{n}{\sum_{i=1}^n x_i} \qquad [= 1/\bar{x}] \\ &= \frac{7}{4+5+3+6+4+3+3} \\ &= \frac{7}{28} \\ &= 0.25 \end{aligned}$$

- **Outro parâmetro desconhecido**

$$\begin{aligned} h(p) &= V(X) \\ &\overset{form.}{=} \frac{1-p}{p^2} \end{aligned}$$

- **Estimativa de MV de $h(p)$**
Pela propriedade de invariância dos estimadores de MV, concluímos que a estimativa de MV de $h(p)$ é dada por

$$\begin{aligned} \widehat{h(p)} &= h(\hat{p}) \\ &= \frac{1-\hat{p}}{\hat{p}^2} \qquad \left[\equiv \frac{1-\frac{1}{\bar{x}}}{\left(\frac{1}{\bar{x}}\right)^2} = \bar{x}(\bar{x}-1)\right] \\ &= \frac{1-0.25}{0.25^2} \qquad [\equiv 28/7 \times (28/7-1)] \\ &\simeq 12. \end{aligned}$$

2. Numa empresa que produz e comercializa granulado de borracha, enchem-se dois tipos de caixas, umas contendo granulado de borracha reciclada (caixas do tipo 1) e outras com granulado de borracha não reciclada (caixas do tipo 2). Admita que as massas (em kg) das caixas dos tipos 1 e 2 são, respectivamente, variáveis aleatórias X_1 e X_2 normalmente distribuídas com variâncias iguais. Ao selecionarem-se, casualmente, 12 caixas do tipo 1 e 17 caixas do tipo 2 obtiveram-se os seguintes resultados: $\bar{x}_1 = 1.2433$ kg, $s_1 = 0.6654$ kg, $\bar{x}_2 = 1.2247$ kg, $s_2 = 0.6271$ kg.

(a) Determine um intervalo de confiança a 90% para $E(X_1) - E(X_2) = \mu_1 - \mu_2$. (2.0)

- **V.a. de interesse**
$X_i =$ massa (em kg) de caixa do tipo i, $i = 1,2$

- **Situação**
$X_1 \overset{indep}{\sim} \text{normal}(\mu_i, \sigma_i^2)$, $i = 1,2$
$(\mu_1 - \mu_2)$ DESCONHECIDO
σ_1^2 e σ_2^2 desconhecidos, no entanto, assume-se que são IGUAIS: $\sigma_1^2 = \sigma_2^2 = \sigma^2$
$n_1 = 12$, $n_2 = 17$

- **Obtenção do IC para $\mu_1 - \mu_2$**

  Passo 1 — V.a. fulcral para $\mu_2 - \mu_2$

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{(n_1-1)\,S_1^2 + (n_2-1)\,S_2^2}{n_1+n_2-2} \times \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t_{(n_1+n_2-2)}$$

  Passo 2 — Quantis de probabilidade

  Já que $(1-\alpha) \times 100\% = 90\%$, temos $\alpha = 0.10$ e lidaremos com os quantis

$$\begin{cases} a_\alpha = F^{-1}_{t_{(n_1+n_2-2)}}(\alpha/2) = -F^{-1}_{t_{(12+17-2)}}(1-0.10/2) = -F^{-1}_{t_{(27)}}(0.95) \overset{tabela/calc.}{=} -1.703 \\ b_\alpha = F^{-1}_{t_{(n_1+n_2-2)}}(1-\alpha/2) = F^{-1}_{t_{(12+17-2)}}(1-0.10/2) = F^{-1}_{t_{(27)}}(0.95) \overset{tabela/calc.}{=} 1.703. \end{cases}$$

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

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

$$P\left[a_\alpha \le \frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{(n_1-1)\,S_1^2+(n_2-1)\,S_2^2}{n_1+n_2-2} \times \left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \le b_\alpha\right] = 1 - \alpha$$

$$P\left[(\bar{X}_1 - \bar{X}_2) - b_\alpha \times \sqrt{\frac{(n_1-1)\,S_1^2+(n_2-1)\,S_2^2}{n_1+n_2-2}} \times \left(\frac{1}{n_1}+\frac{1}{n_2}\right) \le \mu_1 - \mu_2 \right.$$
$$\left. \le (\bar{X}_1 - \bar{X}_2) - a_\alpha \times \sqrt{\frac{(n_1-1)\,S_1^2+(n_2-1)\,S_2^2}{n_1+n_2-2} \times \left(\frac{1}{n_1}+\frac{1}{n_2}\right)}\right] = 1 - \alpha$$

- **Passo 4 — Concretização**

  A expressão geral do IC pedido é

$$\begin{aligned} IC_{(1-\alpha)\times 100\%}(\mu_1 - \mu_2) \quad &= \quad \left[\, (\bar{x}_1 - \bar{x}_2) \pm F^{-1}_{t_{(n_1+n_2-2)}}(1-\alpha/2)\,\right] \\ &\qquad \times \sqrt{\frac{(n_1-1)\,s_1^2 + (n_2-1)\,s_2^2}{n_1+n_2-2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}\;\Bigg], \end{aligned}$$

  temos

$$\begin{aligned} IC_{95\%}(\mu_1 - \mu_2) \quad &\simeq \quad [(1.2433 - 1.2247) \pm 1.703 \\ &\qquad \times \sqrt{\frac{(12-1)\times 0.6654^2 + (17-1)\times 0.6271^2}{12+17-2}\left(\frac{1}{12}+\frac{1}{17}\right)}\,] \\ &\simeq \quad [0.0186 \pm 1.703 \times 0.242427] \\ &\simeq \quad [-0.394253, 0.431453]. \end{aligned}$$

(b) Teste a hipótese $H_0 : \mu_1 = \mu_2$ contra a alternativa $H_1 : \mu_1 > \mu_2$. Decida com base num intervalo para o valor-p. (3.0)

- **Hipóteses**

  $H_0 : \mu_1 - \mu_2 = \mu_0 = 0$
  $H_1 : \mu_1 - \mu_2 > \mu_0$

- **Estatística de teste**

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - \mu_0}{\sqrt{\frac{(n_1-1)\,S_1^2 + (n_2-1)\,S_2^2}{n_1+n_2-2} \times \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim_{H_0} t_{(n_1+n_2-2)}$$

  [dado que se pretende efectuar um teste sobre a diferença de valores esperados de duas populações normais independentes, com variâncias desconhecidas mas que se assume serem iguais.]

- **Região de rejeição de H_0** (para valores da estatística de teste)
  Estamos a lidar com um teste unilateral superior ($H_1 : \mu_1 - \mu_2 > \mu_0$), 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)**
  Uma vez que
  $$n_1 = 12, \quad \bar{x}_1 = 1.2433, \quad s_1 = 0.6654$$
  $$n_2 = 17 \quad \bar{x}_2 = 1.2247, \quad s_2 = 0.6271$$
  o valor observado da estatística de teste é igual a
  $$\begin{aligned} t &= \frac{(\bar{x}_1 - \bar{x}_2) - \mu_0}{\sqrt{\frac{(n_1-1)\,s_1^2+(n_2-1)\,s_2^2}{n_1+n_2-2} \times \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \\ &= \frac{(1.2433 - 1.2247) - 0}{\sqrt{\frac{(12-1)\times 0.6654^2+(17-1)\times 0.6271^2}{12+17-2}\left(\frac{1}{12} + \frac{1}{17}\right)}} \\ &\overset{(a)}{=} \frac{0.0186}{0.242427} \\ &\simeq 0.076724. \end{aligned}$$
  Uma vez que a região de rejeição deste teste é um intervalo à direita, temos:
  $$\begin{aligned} valor - p &= P(T > t \mid H_0) \\ &= 1 - F_{t_{(n_1+n_2-2)}}(t) \\ &= 1 - F_{t_{(27)}}(0.076724). \end{aligned}$$
  Ao recorrer às tabelas de quantis da distribuição t-student, podemos adiantar um intervalo para o valor-p:
  $$\begin{aligned} F^{-1}_{t_{(27)}}(0.5) = 0 &< t = 0.076724 < 0.256 = F^{-1}_{t_{(27)}}(0.6) \\ 0.5 &< F_{t_{(27)}}(0.076724) < 0.6 \\ 1 - 0.6 &< F_{t_{(27)}}(0.076724) < 1 - 0.5 \\ 0.4 &< valor - p < 0.5. \end{aligned}$$
  Assim, é suposto:
  - não rejeitar H_0 a qualquer n.s. $\alpha_0 \leq 40\%$, nomeadamente a qualquer dos n.u.s. $(1\%, 5\%, 10\%)$;
  - rejeitar H_0 a qualquer n.s. $\alpha_0 \geq 50\%$.

  [Alternativamente, poderíamos recorrer a uma calculadora e obter
  $$\begin{aligned} valor - p &= P(T > t \mid H_0) \\ &= 1 - F_{t_{(n_1+n_2-2)}}(t) \\ &= 1 - F_{t_{(27)}}(0.076724) \\ &\simeq 0.469704. \end{aligned}$$
  Deste modo é suposto:
  - não rejeitar H_0 a qualquer n.s. $\alpha_0 \leq 46.9704\%$, nomeadamente a qualquer dos n.u.s. $(1\%, 5\%, 10\%)$;
  - rejeitar H_0 a qualquer n.s. $\alpha_0 > 46.9704\%$.]

Grupo II	10 valores

1. Considere testes *on-line* para alunos/as de engenharia com cinco questões cada. Pretende-se testar a hipótese H_0 segundo a qual o número de questões corretas em cada teste deste tipo segue uma distribuição binomial$(5, 0.5)$. Os resultados de uma amostra de 200 testes, selecionados casualmente, estão resumidos na seguinte tabela de frequências:

Número de questões corretas por teste	0	1	2	3	4	5
Frequência absoluta observada	2	15	56	67	49	11
Frequência absoluta esperada sob H_0	E_1	E_2	62.5	62.5	31.25	6.25

(a) Complete a tabela dada, considerando os valores omissos com duas casas decimais. (1.0)

- **V.a. de interesse**
  $X =$ no. de questões correctas por teste

- **F.p. conjecturada**
  $\binom{5}{x} 0.5^x (1-0.5)^{5-x}$, $x = 0, 1, \ldots, 5$

- **Frequências absolutas esperadas omissas**
  Tendo em conta a dimensão da amostra $n = 200$ e a f.p. conjecturada, temos:

$$\begin{aligned} E_1 &= 200 \times \binom{5}{0} 0.5^0 (1-0.5)^{5-0} \\ &= 200 \times 0.03125 \\ &= 6.25; \\ E_2 &= n - E_1 - \sum_{i=3}^{5} E_i \\ &= 200 - (6.25 + 62.5 + 62.5 + 31.25 + 6.25) \\ &= 31.25. \end{aligned}$$

(b) Teste a hipótese considerada, ao nível de significância de 5%. (3.0)

- **Hipóteses**
  $H_0 : P(X = x) = \binom{5}{x} 0.5^x (1-0.5)^{5-x}$, $x = 0, 1, \ldots, 5$
  $H_1 : \neg H_0$

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

- **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 $= 6$

  $O_i =$ Frequência absoluta observável da classe i

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

  $\beta =$ No. de parâmetros a estimar $= 0$ [dado que a distribuição conjecturada em H_0 está completamente especificada, i.e., H_0 é uma hipótese simples.]

- **Frequências absolutas esperadas sob H_0**
  De acordo com a tabela facultada e a alínea (a), os valores das frequências absolutas esperadas sob H_0 aproximados às décimas são: $E_1 = 6.25$; $E_2 = 31.25$; $E_3 = 62.5$; $E_4 = 62.5$; $E_5 = 31.25$; $E_6 = 6.25$;

  [De notar 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 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

$$c = F^{-1}_{\chi^2_{(k-\beta-1)}}(1-\alpha_0) = F^{-1}_{\chi^2_{(6-0-1)}}(1-0.05) = F^{-1}_{\chi^2_{(5)}}(0.95) \overset{tabela/calc}{=} 11.07.$$

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

i	Classe i	Freq. abs. obs. o_i	Freq. abs. esper. sob H_0 E_i	Parcelas valor obs. estat. teste $\frac{(o_i-E_i)^2}{E_i}$
1	{0}	2	6.25	$\frac{(2-6.25)^2}{6.25} \simeq 2.890$
2	{1}	15	31.25	8.450
3	{2}	56	62.5	0.676
4	{3}	67	62.5	0.324
5	{4}	49	31.25	10.082
6	{5}	11	6.25	3.610
		$\sum_{i=1}^k o_i = n = 200$	$\sum_{i=1}^k E_i = n = 200$	$t = \sum_{i=1}^k \frac{(o_i-E_i)^2}{E_i} \simeq 26.032$

  Uma vez que $t \simeq 26.032 \in W = (11.07, +\infty)$, devemos rejeitar H_0 ao n.s. de $\alpha_0 = 5\%$ [ou qualquer outro n.s. superior a α_0].

2. Com o objetivo de estudar a relação entre o grau de concentração de cloreto de sódio (Y, em miligrama por litro) nas correntes de superfície na ilha de Rodes e a correspondente percentagem de área da bacia hidrográfica total da ilha (x, em percentagem), foram recolhidas 18 observações independentes, tendo-se obtido:

$$\sum_{i=1}^{18} x_i = 14.51, \quad \sum_{i=1}^{18} x_i^2 = 14.7073, \quad \sum_{i=1}^{18} y_i = 306.9, \quad \sum_{i=1}^{18} y_i^2 = 6727.13, \quad \sum_{i=1}^{18} x_i y_i = 309.316,$$

onde $\left[\min_{i=1,\dots,18} x_i, \max_{i=1,\dots,18} x_i\right] = [0.15, 1.74]$.

(a) Considere o modelo de regressão linear simples de Y em x e determine a estimativa de mínimos quadrados do valor esperado do grau de concentração de cloreto de sódio para $x = 0.5$. (2.0)

- **[Hipóteses de trabalho**
  $E(\epsilon_i) = 0, \; V(\epsilon_i) = \sigma^2, \; i = 1, \dots, n$
  $cov(\epsilon_i, \epsilon_j) = 0, \; i \neq j, \; i, j = 1, \dots, n]$

- **Estimativa de MQ de** $E(Y \mid x) = \beta_0 + \beta_1 x$ **com** $x = 0.5$
  Dado que

  $n = 18$

  $\sum_{i=1}^n x_i = 14.51$
  $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i \simeq \frac{14.51}{18} \simeq 0.806111$
  $\sum_{i=1}^n x_i^2 = 14.7073$
  $\sum_{i=1}^n x_i^2 - n(\bar{x})^2 \simeq 14.7073 - 18 \times 0.806111^2 \simeq 3.010628$

  $\sum_{i=1}^n y_i = 306.9$
  $\bar{y} = \frac{1}{n}\sum_{i=1}^n y_i = \frac{306.9}{18} = 17.05$
  $\sum_{i=1}^n y_i^2 = 6727.13$
  $\sum_{i=1}^n y_i^2 - n(\bar{y})^2 = 6727.13 - 18 \times 17.05^2 \simeq 1494.4850$

  $\sum_{i=1}^n x_i y_i = 309.316$
  $\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y} \simeq 309.316 - 18 \times 0.806111 \times 17.05 = 61.9205,$

  as estimativas de MQ de β_1, β_0 e $\beta_0 + \beta_1 x$ são, para este modelo de RLS, iguais a:

$$\begin{aligned}\hat{\beta}_1 &= \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^n x_i^2 - n(\bar{x})^2} \\ &\simeq \frac{61.9205}{3.010628} \\ &\simeq 20.567305\end{aligned}$$

$$
\begin{aligned}
\hat\beta_0 &= \bar y - \hat\beta_1 \times \bar x \\
&\simeq 17.05 - 20.567305 \times 0.806111 \\
&\simeq 0.470467 \\
\hat\beta_0 + \hat\beta_1 x &\simeq 0.470467 + 20.567305 \times 0.5 \\
&\simeq 10.754120.
\end{aligned}
$$

(b) Após ter enunciado as hipóteses de trabalho que entender convenientes, construa um intervalo de confiança a 95% para o declive da reta de regressão linear. (3.0)

- **Hipóteses de trabalho**
  $\epsilon_i \overset{i.i.d.}{\sim} \text{normal}(0, \sigma^2)$, $i = 1, \dots, n$
- **Obtenção do IC para β_1**
  Passo 1 — 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 — Quantis de probabilidade
Já que $(1-\alpha) \times 100\% = 95\%$, temos $\alpha = 0.05$ e lidaremos com os quantis

$$
\begin{cases}
a_\alpha = F^{-1}_{t_{(n-2)}}(\alpha/2) = -F^{-1}_{t_{(18-2)}}(1-0.05/2) = -F^{-1}_{t_{(16)}}(0.975) \overset{tabela/calc.}{=} -2.120 \\
b_\alpha = F^{-1}_{t_{(18-2)}}(1-0.05/2) = F^{-1}_{t_{(16)}}(0.975) \overset{tabela/calc.}{=} 2.120.
\end{cases}
$$

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

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

$$
P\left[a_\alpha \le \frac{\hat\beta_1 - \beta_1}{\sqrt{\frac{\hat\sigma^2}{\sum_{i=1}^n x_i^2 - n\bar x^2}}} \le b_\alpha\right] = 1 - \alpha
$$

$$
P\left[\hat\beta_1 - b_\alpha \times \sqrt{\frac{\hat\sigma^2}{\sum_{i=1}^n x_i^2 - n\bar x^2}} \le \beta_1 \le \hat\beta_1 - b_\alpha \times \sqrt{\frac{\hat\sigma^2}{\sum_{i=1}^n x_i^2 - n\bar x^2}}\right] = 1 - \alpha
$$

- **Passo 4 — Concretização**
  Dado que a estimativa de σ^2 é igual a

$$
\begin{aligned}
\hat\sigma^2 &= \frac{1}{n-2}\left[\left(\sum_{i=1}^n y_i^2 - n\bar y^2\right) - (\hat\beta_1)^2 \left(\sum_{i=1}^n x_i^2 - n\bar x^2\right)\right] \\
&\simeq \frac{1}{18-2}\left(1\,494.4850 - 20.567305^2 \times 3.010628\right) \\
&\simeq 13.809194
\end{aligned}
$$

e a expressão geral do IC pretendido é

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

temos

$$
\begin{aligned}
IC_{95\%}(\beta_1) &\simeq \left[20.567305 \pm 2.120 \times \sqrt{\frac{13.809194}{3.010628}}\right] \\
&\simeq [20.567305 \pm 2.120 \times 2.141686] \\
&\simeq [16.026931, 25.107679].
\end{aligned}
$$

(c) Calcule o coeficiente de determinação e comente o valor obtido. (1.0)

- **Cálculo do coeficiente de determinação**

$$\begin{aligned} r^2 &= \frac{\left(\sum_{i=1}^n x_i y_i - n\,\bar{x}\,\bar{y}\right)^2}{\left(\sum_{i=1}^n x_i^2 - n\,\bar{x}^2\right) \times \left(\sum_{i=1}^n y_i^2 - n\,\bar{y}^2\right)} \\ &\simeq \frac{61.9205^2}{3.010628 \times 1494.4850} \\ &\simeq 0.852158. \end{aligned}$$

- **Interpretação coeficiente de determinação**
  Cerca de 85.2% da variação total da variável resposta Y é explicada pela variável x, através do modelo de regressão linear simples ajustado, logo podemos afirmar que a recta estimada parece ajustar-se bem ao conjunto de dados.