Probabilidades e Estatística

LEAN, LEGI, LEGM, LMAC, MEAer, MEAmbi, MEC

1º semestre – 2017/2018
11/01/2018 – **09:00**

Duração: 90 minutos **2º teste A**

Justifique convenientemente todas as respostas

Grupo I	10 valores

1. Admita que o custo de retrabalho (em 10^5 Euro) associado a certo tipo de projeto de construção civil é representado pela variável aleatória X com função de densidade de probabilidade dada por

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

onde λ é um parâmetro positivo desconhecido. Seja $(X_1, X_2, \dots, X_n)$ uma amostra aleatória de X.

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

- **V.a. de interesse**
  X = custo de retrabalho associado ao projeto de construção civil
- **F.d.p. de** X
  $$f_X(x) = \begin{cases} \lambda\, x^{-(\lambda+1)}, & x > 1 \\ 0\,, & \text{caso contrário} \end{cases}$$
- **Parâmetro desconhecido**
  $\lambda, \quad \lambda > 0$
- **Amostra**
  $\underline{x} = (x_1, \dots, x_n)$ amostra de dimensão n proveniente da população X
- **Obtenção do estimador de MV de** θ

  Passo 1 — Função de verosimilhança

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

  Passo 2 — Função de log-verosimilhança

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

  Passo 3 — Maximização

  A estimativa de MV de λ passa a ser representada por $\hat\lambda$ e

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

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

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

Passo 4 — Estimador de MV de λ

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

(b) Obtenha a estimativa de máxima verosimilhança de $\frac{1}{\lambda}$ baseada na concretização $(x_1, x_2, x_3, x_4, x_5) = (1.01, 1.34, 1.19, 2.67, 1.58)$ para a qual $\sum_{i=1}^{5} \ln(x_i) \simeq 1.92$. (1.5)

- **Estimativa de MV de λ**

$$\begin{aligned} \hat{\lambda} &= \frac{n}{\sum_{i=1}^{n} \ln(x_i)} \\ &\simeq \frac{5}{1.92} \\ &\simeq 2.6042 \end{aligned}$$

- **Outro parâmetro desconhecido**
  $h(\lambda) = \frac{1}{\lambda}$

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

$$\begin{aligned} \widehat{h(\lambda)} &= h(\hat{\lambda}) \\ &= \frac{1}{\hat{\lambda}} \\ &\simeq \frac{1}{2.6042} \\ &\simeq 0.3840. \end{aligned}$$

(c) Tendo em conta que a variável aleatória $\ln(X)$ possui distribuição exponencial com parâmetro λ, averigúe se $\frac{1}{n}\sum_{i=1}^{n} \ln(X_i)$ é um estimador centrado de $\frac{1}{\lambda}$. (1.0)

- **Parâmetro desconhecido**
  $\frac{1}{\lambda}$

- **Estimador de $\frac{1}{\lambda}$**
  $\frac{1}{n}\sum_{i=1}^{n} \ln(X_i)$

- **Valor esperado de $\frac{1}{n}\sum_{i=1}^{n} \ln(X_i)$**
  Uma vez que $X_i \overset{i.i.d.}{\sim} X,\ i = 1, \ldots, n$ e $\ln(X) \sim \text{exponencial}(\lambda)$, segue-se:

$$\ln(X_i) \overset{i.i.d.}{\sim} \text{exponencial}(\lambda),\ i = 1, \ldots, n;$$

$$\begin{aligned} E\left[\frac{1}{n}\sum_{i=1}^{n} \ln(X_i)\right] &= \frac{1}{n}\sum_{i=1}^{n} E[\ln(X_i)] \\ &= \frac{1}{n}\sum_{i=1}^{n} \frac{1}{\lambda} \\ &= \frac{1}{\lambda}. \end{aligned}$$

- **Conclusão**
  Uma vez que
  - T se diz um estimador centrado de $\frac{1}{\lambda}$ caso $E(T) = \frac{1}{\lambda}, \forall \lambda > 0$,

  podemos afirmar que $\frac{1}{n}\sum_{i=1}^{n} \ln(X_i)$ é um estimador centrado de $\frac{1}{\lambda}$.

2. Por forma a estudar o desempenho de dois modelos de baterias de carros eléctricos, considerou-se a variável aleatória X_1 (respetivamente X_2) que representa a distância percorrida (em km) com uma bateria escolhida ao acaso do modelo 1 (respetivamente 2). Ao selecionarem-se casualmente 40 baterias do modelo 1 e 35 baterias do modelo 2, obtiveram-se os seguintes resultados: $\bar{x}_1 = 242.5$, $s_1^2 = 56.34$, $\bar{x}_2 = 238.6$, $s_2^2 = 35.85$.

(a) Determine um intervalo aproximado de confiança a 95% para a diferença dos valores esperados das distâncias percorridas com as baterias dos modelos 1 e 2. (2.5)

- **V.a. de interesse**
  $X_i =$ distância percorrida com uma bateria do modelo i, $i = 1,2$
- **Situação**
  X_i v.a. com dist. arbitrária [possivelmente normal], valor esperado μ_i e variância σ_i^2, $i = 1,2$
  $X_1 \perp\!\!\!\perp X_2$
  $(\mu_1 - \mu_2)$ DESCONHECIDO
  σ_1^2 e σ_2^2 desconhecidas [não necessariamente iguais]
  $n_1 = 40 > 30$ e $n_2 = 35 > 30$ [i.e., ambas as amostras são suficientemente grandes].
- **Obtenção do IC para $\mu_1 - \mu_2$**

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

  $$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \overset{a}{\sim} \text{normal}(0,1)$$

  [dado que se pretende determinar um IC para a diferença de valores esperados de duas populações com distribuições arbitrárias independentes e com variâncias desconhecidas, dispondo de duas amostras suficientemente grandes.]

  Passo 2 — Obtenção dos quantis de probabilidade

  Ao ter-se em consideração que $(1-\alpha) \times 100\% = 95\%$, far-se-á uso dos quantis

  $$\begin{cases} a_\alpha = \Phi^{-1}(\alpha/2) = \Phi^{-1}(0.025) = -\Phi^{-1}(1-0.025) \overset{tabela/calc.}{=} -1.96 \\ b_\alpha = \Phi^{-1}(1-\alpha/2) = \Phi^{-1}(0.975) \overset{tabela/calc.}{=} 1.96. \end{cases}$$

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

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

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

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

  Passo 4 — Concretização

  Tendo em conta que

  $$IC_{(1-\alpha)\times 100\%}(\mu_1 - \mu_2) = \left[(\bar{x}_1 - \bar{x}_2) \pm \Phi^{-1}(1-\alpha/2) \times \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}\right],$$

  e aos valores dos quantis acima e de n_i, $\bar{x}_i$, s_i^2 $(i = 1,2)$, temos

  $$\begin{aligned} IC_{95\%}(\mu_1 - \mu_2) &= \left[(242.5 - 238.6) \pm 1.96 \times \sqrt{\frac{56.34}{40} + \frac{35.85}{35}}\right] \\ &\simeq [3.9 \pm 1.96 \times 1.559739] \\ &\simeq [0.842912,\ 6.957088]. \end{aligned}$$

(b) Teste a hipótese de igualdade dos valores esperados das distâncias percorridas com as baterias dos modelos 1 e 2. Decida com base no valor-p. (2.5)

- **V.a. de interesse e situação**
  Ver alínea (a).
- **Hipóteses**
  $H_0 : \mu_1 - \mu_2 = \mu_0 = 0$
  $H_1 : \mu_1 - \mu_2 \neq \mu_0$
- **Estatística de teste**

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - \mu_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \overset{a}{\sim}_{H_0} \text{normal}(0,1)$$

  [uma vez que se pretende efectuar um teste de igualdade dos valores esperados de duas populações com distribuições arbitrárias independentes e com variâncias desconhecidas, dispondo de duas amostras suficientemente grandes.]
- **Região de rejeição de H_0** (para valores da estatística de teste)
  Estamos a lidar com um teste bilateral ($H_1 : \mu_1 - \mu_2 \neq \mu_0$), logo a região de rejeição de H_0 (para valores da estatística de teste) é do tipo $W = (-\infty, -c) \cup (c, +\infty)$.
- **Decisão (com base no valor-p)**
  O valor observado da estatística de teste é igual a

$$\begin{aligned} t &= \frac{(\bar{x}_1 - \bar{x}_2) - \mu_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \\ &= \frac{(242.5 - 238.6) - 0}{\sqrt{\frac{56.34}{40} + \frac{35.85}{35}}} \\ &\simeq 2.50. \end{aligned}$$

  Dado que a região de rejeição deste teste é a reunião de dois intervalos simétricos, temos:

$$\begin{aligned} valor - p &= 2 \times P(T > |t| \mid H_0) \\ &\simeq 2 \times [1 - \Phi(|t|)] \\ &\simeq 2 \times [1 - \Phi(2.50)] \\ &\overset{calc/tabela}{=} 2 \times (1 - 0.9938) \\ &= 0.0124. \end{aligned}$$

  Logo é suposto:
  - não rejeitar H_0 a qualquer n.s. $\alpha_0 \leq 1.24\%$, pelo que H_0 não é contrariada pelos dados ao n.u.s. de 1%;
  - rejeitar H_0 a qualquer n.s. $\alpha_0 > 1.24\%$, nomeadamente aos n.u.s. de 5% e 10%.

  [É curioso notar que $\mu_0 = 0 \not\in IC_{95\%}(\mu_1 - \mu_2)$, donde (invocando a analogia entre testes de hipóteses bilaterais e IC) devemos rejeitar $H_0 : \mu_1 - \mu_2 = \mu_0 = 0$ ao n.s. de 5%. Ora esta decisão coaduna-se com a decisão tomada com base no valor-p do teste.]

Grupo II	10 valores

1. Um perito de uma companhia de seguros defende a hipótese H_0 de que o número semestral de pedidos de reembolso de despesas de saúde por segurado possui uma distribuição de Poisson com valor esperado igual a 2. Uma amostra casual de 200 segurados num dado semestre conduziu ao seguinte quadro de frequências:

Nº de pedidos	0	1	2	3	4 ou mais
Frequência absoluta observada	22	53	58	39	28
Frequência absoluta esperada sob H_0	27.06	54.14	54.14	E_4	E_5

(a) Calcule os valores das frequências absolutas esperadas E_4 e E_5 (aproximando-os às centésimas). (1.0)

- **V.a. de interesse**
  X = número semestral de pedidos de reembolso de despesas de saúde por segurado
- **Distribuição e f.p. conjecturadas**
  $X \sim \text{Poisson}(2)$
  $P(X=x) = \frac{e^{-2}\,2^x}{x!}$, $x = 0, 1, 2, \ldots$
- **Frequências absolutas esperadas omissas**
  Atendendo à dimensão da amostra $n = 200$ e à f.p. conjecturada, segue-se:

$$\begin{aligned}
E_4 &= P(X=3) \\
&= 200 \times \frac{e^{-2}\,2^3}{3!} \\
&\simeq 36.09; \\
E_5 &= n \times P_0(X \geq 4) \\
&= n - \sum_{i=1}^{4} E_i \\
&\simeq 200 - (27.06 + 54.14 + 54.14 + 36.09) \\
&= 28.57.
\end{aligned}$$

(b) Teste H_0, ao nível de significância de 1%. (3.0)

- **Hipóteses**
  $H_0 : X \sim \text{Poisson}(2)$
  $H_1 : X \not\sim \text{Poisson}(2)$
- **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

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

- **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: $E_1 \simeq 27.06$; $E_2 \simeq 54.14$; $E_3 \simeq 54.14$; $E_4 \simeq 36.08$; $E_5 \simeq 28.58$.
  [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)
  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

$$\begin{aligned}
c &= F^{-1}_{\chi^2_{(k-\beta-1)}}(1-\alpha_0) \\
&= F^{-1}_{\chi^2_{(5-0-1)}}(1-0.01) \\
&= F^{-1}_{\chi^2_{(4)}}(0.99) \\
&\overset{tabela/calc.}{=} 13.28.
\end{aligned}$$

- **Decisão**
No cálculo do valor observado da estatística 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. este $\frac{(o_i-E_i)^2}{E_i}$
1	$\{0\}$	22	27.06	$\frac{(22-27.06)^2}{27.06} \simeq 0.946$
2	$\{1\}$	53	54.14	0.024
3	$\{2\}$	58	54.14	0.275
4	$\{3\}$	39	36.09	0.235
5	$\{4,5,\ldots\}$	28	28.57	0.011
		$\sum_{i=1}^k o_i = n$ $= 200$	$\sum_{i=1}^k E_i = n$ $= 200$	$t = \sum_{i=1}^k \frac{(o_i-E_i)^2}{E_i}$ $\simeq 1.491$

Uma vez que $t \simeq 1.491 \not\in W = (13.28, +\infty)$, não devemos rejeitar H_0 ao n.s. de $\alpha_0 = 1\%$ [nem a qualquer outro n.s. inferior a α_0].

2. Suponha que o rendimento semanal de agregados familiares de uma dada população é descrito pela variável x e que a despesa semanal em bens e serviços culturais dos mesmos agregados é representada pela variável aleatória Y, com valores em Euro. Um estudo envolvendo 10 agregados familiares conduziu aos seguintes valores:

$$\sum_{i=1}^{10} x_i = 2\,910, \quad \sum_{i=1}^{10} x_i^2 = 985\,100, \quad \sum_{i=1}^{10} y_i = 379, \quad \sum_{i=1}^{10} y_i^2 = 15\,673, \quad \sum_{i=1}^{10} x_i\, y_i = 116\,700,$$

onde $\left[\min_{i=1,\ldots,10} x_i,\, \max_{i=1,\ldots,10} x_i\right] = [125,\ 308]$.

(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 $E(Y \mid x = 291)$. (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**
Atendendo a que

$$n = 10$$

$$\sum_{i=1}^n x_i = 2\,910$$
$$\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i = \frac{2\,910}{10} = 291$$
$$\sum_{i=1}^n x_i^2 = 985\,100$$
$$\sum_{i=1}^n x_i^2 - n\,(\bar{x})^2 = 985\,100 - 10 \times 291^2 = 138\,290$$

$$\sum_{i=1}^n y_i = 379$$
$$\bar{y} = \frac{1}{n}\sum_{i=1}^n y_i = \frac{379}{10} = 37.9$$
$$\sum_{i=1}^n y_i^2 = 15\,673$$
$$\sum_{i=1}^n y_i^2 - n\,(\bar{y})^2 = 15\,673 - 10 \times 37.9^2 = 1\,308.9$$

$$\sum_{i=1}^n x_i y_i = 116\,700$$
$$\sum_{i=1}^n x_i y_i - n\,\bar{x}\,\bar{y} = 116\,700 - 10 \times 291 \times 37.9 = 6\,411,$$

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

$$\begin{aligned}\hat{\beta}_1 &= \frac{\sum_{i=1}^n x_i y_i - n\,\bar{x}\,\bar{y}}{\sum_{i=1}^n x_i^2 - n\,(\bar{x})^2} \\ &= \frac{6411}{138\,290} \\ &\simeq 0.046359\end{aligned}$$

$$\begin{aligned}\hat\beta_0 &= \bar y - \hat\beta_1 \times \bar x\\ &\simeq 37.9 - 0.046359 \times 291\\ &\simeq 24.409531.\end{aligned}$$

- **Estimativa de MV para** $E(Y \mid x = x_0) = \beta_0 + \beta_1 x_0$, com $x_0 = 291$

$$\begin{aligned}\hat E(Y \mid x = x_0) &= \hat\beta_0 + \hat\beta_1 x_0\\ &\simeq 24.409531 + 0.046359 \times 291\\ &\simeq 37.9\\ [&= \bar y].\end{aligned}$$

[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 $\hat E(Y \mid x = x_0) = \hat\beta_0 + \hat\beta_1 \bar x = (\bar y - \hat\beta_1 \times \bar x) + \hat\beta_1 \bar x = \bar y$, como se pôde constatar acima.]

(b) Deduza um intervalo de confiança a 95% para $E(Y \mid x = 291)$. (2.5)

- **Obtenção do IC para** $E(Y \mid x = x_0) = \beta_0 + \beta_1 x_0$, com $x_0 = 291$

Passo 1 — V.a. fulcral para $E(Y \mid x = x_0) = \beta_0 + \beta_1 x_0$

$$Z = \frac{(\hat\beta_0 + \hat\beta_1 x_0) - (\beta_0 + \beta_1 x_0)}{\sqrt{\hat\sigma^2 \times \left[\frac{1}{n} + \frac{(x_0 - \bar x)^2}{\sum_{i=1}^n x_i^2 - n\bar x^2}\right]}} \sim t_{(n-2)}$$

Passo 2 — Quantis de probabilidade

Já que $(1-\alpha) \times 100\% = 95\%$, temos $\alpha = 0.05$ e lidaremos com os quantis

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

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

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

$$P\left[a_\alpha \le \frac{(\hat\beta_0 + \hat\beta_1 x_0) - (\beta_0 + \beta_1 x_0)}{\sqrt{\hat\sigma^2 \times \left[\frac{1}{n} + \frac{(x_0 - \bar x)^2}{\sum_{i=1}^n x_i^2 - n\bar x^2}\right]}} \le b_\alpha\right] = 1 - \alpha$$

$$P\left[(\hat\beta_0 + \hat\beta_1 x_0) - b_\alpha \times \sqrt{\hat\sigma^2 \times \left[\frac{1}{n} + \frac{(x_0 - \bar x)^2}{\sum_{i=1}^n x_i^2 - n\bar x^2}\right]} \le \beta_0 + \beta_1 x_0 \right.$$
$$\left. \le (\hat\beta_0 + \hat\beta_1 x_0) - a_\alpha \times \sqrt{\hat\sigma^2 \times \left[\frac{1}{n} + \frac{(x_0 - \bar x)^2}{\sum_{i=1}^n x_i^2 - n\bar x^2}\right]}\right] = 1 - \alpha$$

- **Passo 4 — Concretização**

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

$$\begin{aligned}\hat\sigma^2 &= \frac{1}{n-2}\left[\left(\sum_{i=1}^n y_i^2 - n\bar y^2\right) - (\hat\beta_1)^2\left(\sum_{i=1}^n x_i^2 - n\bar x^2\right)\right]\\ &= \frac{1}{10-2}\left(1308.9 - 0.046359^2 \times 138290\right)\\ &\simeq 126.461637\end{aligned}$$

e a expressão geral do IC pretendido é

$$IC_{(1-\alpha)\times 100\%}(\beta_0 + \beta_1 x_0)$$
$$= \left[(\hat\beta_0 + \hat\beta_1 x_0) \pm F^{-1}_{t_{(n-2)}}(1 - \alpha/2) \times \sqrt{\hat\sigma^2 \times \left[\frac{1}{n} + \frac{(x_0 - \bar x)^2}{\sum_{i=1}^n x_i^2 - n\bar x^2}\right]}\right],$$

temos

$$
\begin{aligned}
& IC_{95\%}(\beta_0+\beta_1\times 74)\\
&= \left[(24.409531+0.046359\times 291)\pm 2.306\times\sqrt{126.461637\times\left[\tfrac{1}{10}+\tfrac{(291-291)^2}{138290}\right]}\right]\\
&= [37.9\pm 2.306\times 3.556144]\\
&= [37.9\pm 8.200469]\\
&= [29.699531,\ 46.100469].
\end{aligned}
$$

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

- **Cálculo do coeficiente de determinação**

$$
\begin{aligned}
r^2 &= \frac{\left(\sum_{i=1}^n x_i y_i - n\,\bar{x}\,\bar{y}\right)^2}{\left(\sum_{i=1}^n x_i^2 - n\,\bar{x}^2\right)\times\left(\sum_{i=1}^n y_i^2 - n\,\bar{y}^2\right)}\\
&= \frac{6411^2}{138290\times 1308.9}\\
&\simeq 0.227067.
\end{aligned}
$$

- **Interpretação coeficiente de determinação**
  Cerca de 22.71% 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.