

Probabilidades e Estatística

LEAN, LEE, LEGI, LERC, LMAC, MEAer, MEAmbi, MEBiol,
MEEC, MEMec

2º semestre – 2011/2012

2º Teste B

08/06/2012 – 11:00

Duração: 1 hora e 30 minutos

Justifique convenientemente **todas as respostas!**

Grupo I

3.5 + 2.5 + 3.0 + 1.0 = 10 valores

1. Para inferir sobre a probabilidade p de cancelamento de reservas de voos por parte dos seus clientes, uma agência de viagens decidiu analisar os processos relativos a $n = 230$ reservas de voos passados, escolhidos ao acaso da sua vasta carteira, registando-se para cada reserva se a mesma tinha sido cancelada ou não.

(a) Deduza o estimador de máxima verosimilhança de p . Diga se o estimador é centrado. (3.5)

• **V.a. de interesse**

X v.a. indicadora de cancelamento de reserva

• **Distribuição**

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

• **Parâmetro DESCONHECIDO**

$p = P(X = 1) = P(\text{cancelamento da reserva}), 0 \leq p \leq 1$

• **F.p.**

$P(X = x) \stackrel{\text{form}}{=} p^x (1 - p)^{1-x}, x = 0, 1$

• **Amostra**

$\underline{x} = (x_1, \dots, x_n)$ amostra de dimensão n proveniente da população X

• **Obtenção do estimador de MV de p**

Passo 1 — Função de verosimilhança

$$\begin{aligned} L(p|\underline{x}) &\stackrel{X_i \text{ indep}}{=} \prod_{i=1}^n P(X_i = x_i) \\ &\stackrel{X_i \sim X}{=} \prod_{i=1}^n [p^{x_i} (1 - p)^{1-x_i}] \\ &= p^{\sum_{i=1}^n x_i} (1 - p)^{n - \sum_{i=1}^n x_i}, 0 < p < 1 \end{aligned}$$

Passo 2 — Função de log-verosimilhança

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

Passo 3 — Maximização¹

A estimativa de MV de p é aqui representada por \hat{p} e

$$\hat{p} : \begin{cases} \frac{d \ln L(p|\underline{x})}{dp} \Big|_{p=\hat{p}} = 0 & (\text{ponto de estacionaridade}) \\ \frac{d^2 \ln L(p|\underline{x})}{dp^2} \Big|_{p=\hat{p}} < 0 & (\text{ponto de máximo}) \\ \frac{\sum_{i=1}^n x_i}{\hat{p}} - \frac{n - \sum_{i=1}^n x_i}{1 - \hat{p}} = 0 \\ -\frac{\sum_{i=1}^n x_i}{\hat{p}^2} - \frac{n - \sum_{i=1}^n x_i}{(1 - \hat{p})^2} < 0 \\ \hat{p} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}, \text{ média da amostra} \\ \text{Proposição verdadeira já que } n \in \mathbb{N} \text{ e } 0 \leq \sum_{i=1}^n x_i \leq n \end{cases}$$

¹Este procedimento só deve ser aplicado se $\sum_{i=1}^n x_i \neq 0, n$, sendo o resultado obtido, $\hat{p} = \bar{x}$, também válido se $\sum_{i=1}^n x_i \in \{0, n\}$.

Passo 4 — Estimador de MV de p

Será representado pela v.a. $EMV(p) = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}$, média da amostra aleatória.

- **Estimador de MV de p é centrado?**

$EMV(p) = \bar{X}$ é um estimador centrado de p sse $E(\bar{X}) = p, \forall p \in [0, 1]$. Ora,

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

$\forall p \in [0, 1]$. Assim, conclui-se que \bar{X} é um estimador centrado de p .

- (b) Sabendo que exactamente 11 reservas (das 230 analisadas) tinham sido canceladas, construa um intervalo, com nível de confiança de aproximadamente 95%, para o parâmetro p . (2.5)

- **V.a.**

X_i = indicador de cancelamento da i - ésima reserva, $i = 1, \dots, n$

$X_i \stackrel{i.i.d.}{\sim} X$

- **Situação**

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

p DESCONHECIDO

$n = 230 \gg 30$ (suficientemente grande)

- **Obtenção de IC para p**

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

Utilizaremos a v.a. fulcral para p

$$Z = \frac{\bar{X} - E(\bar{X})}{\sqrt{EMV[V(\bar{X})]}} = \frac{\bar{X} - p}{\sqrt{EMV[p(1-p)/n]}} = \frac{\bar{X} - p}{\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}} \stackrel{a}{\sim} \text{normal}(0, 1),$$

uma vez que nos foi solicitada a determinação de um IC aproximado para uma probabilidade de sucesso e a dimensão da amostra justifica o recurso a uma aproximação distribucional.

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

Dado que $(1 - \alpha) \times 100\% = 95\% \Leftrightarrow \alpha = 0.05$, os quantis a utilizar são

$$\begin{cases} a_\alpha = -\Phi^{-1}(1 - \alpha/2) = -\Phi^{-1}(0.975) \stackrel{tabela}{=} -1.9600 \\ b_\alpha = \Phi^{-1}(1 - \alpha/2) = \Phi^{-1}(0.975) = 1.9600. \end{cases}$$

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

- **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 consideração que

$$* n = 230$$

$$* \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{11}{230} = \text{proporção observada de reservas canceladas}$$

$$* \Phi^{-1}(1 - \alpha/2) = 1.9600,$$

conclui-se que o IC aproximado a 95% para p é dado por

$$\begin{aligned} IC(p) &= \left[\bar{x} \pm \Phi^{-1}(1 - \alpha/2) \times \sqrt{\frac{\bar{x}(1 - \bar{x})}{n}}, \right] \\ &= \left[\frac{11}{230} \pm 1.9600 \times \sqrt{\frac{\frac{11}{230} \times (1 - \frac{11}{230})}{230}} \right] \\ &= \left[0.047826 \pm 1.9600 \times \sqrt{\frac{0.047826 \times (1 - 0.047826)}{230}} \right] \\ &\simeq [0.020247, 0.075405]. \end{aligned}$$

2. Um fabricante admite que o conteúdo de nicotina dos cigarros produzidos possui distribuição normal, com valor esperado (μ) e desvio padrão (σ) desconhecidos. Uma amostra de 61 cigarros, seleccionados ao acaso da produção diária, conduziu a uma média e um desvio padrão iguais a $\bar{x} = 1.05$ mg e $s = 0.1$ mg, respectivamente.

- (a) O fabricante afirma que $\mu = 1$ mg (teor máximo imposto legalmente em 2003), ao passo que a representante de uma agência de vigilância sanitária afirma que μ é superior a esse valor. Confronte estas duas hipóteses ao nível de significância de 5%. (3.0)

• **V.a. de interesse**

X = conteúdo de nicotina de cigarros da produção (em mg)

• **Situação**

$$X \sim \text{normal}(\mu, \sigma^2)$$

μ DESCONHECIDO

σ^2 desconhecido

• **Hipóteses**

$$H_0 : \mu = \mu_0 = 1$$

$$H_1 : \mu > \mu_0 = 1$$

• **Nível de significância**

$$\alpha_0 = 0.05$$

• **Estatística de teste**

$$T = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}} \sim_{H_0} t_{(n-1)}$$

pois pretendemos efectuar um teste sobre o valor esperado de uma população normal com variância desconhecida.

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

Tratando-se de um teste unilateral superior ($H_1 : \mu > \mu_0$), a região de rejeição de H_0 é um intervalo à direita do tipo $W = (c, +\infty)$, onde $c : P(\text{Rejeitar } H_0 | \mu = \mu_0) = \alpha_0$, i.e.,

$$\begin{aligned} c &= F_{t_{(n-1)}}^{-1}(1 - \alpha_0) \\ &= F_{t_{(61-1)}}^{-1}(1 - 0.05) \\ &= F_{t_{(60)}}^{-1}(0.95) \\ &\stackrel{\text{tabela}}{=} 1.671. \end{aligned}$$

- **Decisão**

Uma vez que $n = 61$, $\bar{x} = 1.05$ e $s = 0.10$, o valor observado da estatística é igual a

$$\begin{aligned} t &= \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} \\ &= \frac{1.05 - 1}{\frac{0.10}{\sqrt{61}}} \\ &= 3.905. \end{aligned}$$

Ora $t \simeq 3.905 \in W = (1.671, +\infty)$, pelo que devemos rejeitar H_0 (hipótese do fabricante) a qualquer n.s. maior ou igual a 5%.

(b) *Confronte as hipóteses referidas na alínea anterior usando o valor-p.* (1.0)

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

Uma vez que este teste é unilateral superior temos:

$$\begin{aligned} \text{valor} - p &= P(T > t \mid H_0) \\ &= P(T > 3.905 \mid H_0) \\ &= 1 - F_{t(60)}(3.905). \end{aligned}$$

Recorrendo às tabelas de quantis da distribuição de t-student podemos adiantar um intervalo para o *valor-p* deste teste. Com efeito, ao enquadrarmos convenientemente $t = 3.905$, obtemos sucessivamente

$$\begin{aligned} F_{t(60)}^{-1}(0.9995) = 3.460 &< 3.905 < +\infty \\ 0.9995 &< F_{t(60)}(3.905) < 1 \\ 0 = 1 - 1 &< \text{valor} - p < 1 - 0.9995 = 0.0005. \end{aligned}$$

Logo devemos rejeitar H_0 a qualquer n.s. $\alpha_0 \geq 0.05\%$, por exemplo a qualquer dos níveis usuais de significância de 1%, 5% e 10%.

- **Alternativa — Decisão (com base em valor-p determinado usando máquina de calcular)**

Uma vez que este teste é unilateral superior

$$\begin{aligned} \text{valor} - p &= P(T > t \mid H_0) \\ &= P(T > 3.905 \mid H_0) \\ &= 1 - F_{t(60)}(3.905) \\ &= 0.000121, \end{aligned}$$

pelo que devemos rejeitar H_0 a qualquer n.s. $\alpha_0 > 0.0121\%$, por exemplo a qualquer dos níveis usuais de significância de 1%, 5% e 10%.

Grupo II	3.0 + 2.0 + 3.0 + 3.0 = 10 valores
-----------------	---

1. *O responsável pela segurança de uma fábrica de automóveis recolheu uma amostra casual de registos de acidentes de trabalho que foi classificada de acordo com o período do dia de trabalho em que estes acidentes ocorreram, tendo-se obtido a seguinte tabela de frequências:*

Período	8:00-9:59	10:00-11:59	13:00-14:59	15:00-16:59
Número de acidentes	12	28	15	25

Teste, ao nível de significância de 5%, se os acidentes de trabalho se distribuem de modo uniforme pelos 4 períodos considerados. (3.0)

- **V.a. de interesse**

X = indicador do período de duas horas em que ocorreu o acidente
= 1, 2, 3 e 4 se o acidente ocorreu nos períodos 8:00-9:59, 10:00-11:59, 13:00-14:59 e 15:00-16:59, respectivamente

- **Hipóteses**

Seja $p_i = P(X = i)$, $i = 1, 2, 3, 4$. Se os acidentes se distribuem de modo uniforme ao longo do dia de trabalho, com 4 períodos de duas horas, $p_i = p_i^0 = \frac{1}{4}$, $i = 1, 2, 3, 4$. Assim sendo, confrontar-se-ão as seguintes hipóteses:

$$H_0 : p_i = p_i^0 = \frac{1}{4}, i = 1, 2, 3, 4$$

$$H_1 : \exists i : p_i \neq p_i^0$$

- **Nível de significância**

$$\alpha_0 = 0.05$$

- **Estatística de Teste**

$$T = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} \stackrel{a}{\sim}_{H_0} \chi^2_{(k-\beta-1)},$$

onde:

k = No. de classes = 4;

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.

- **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 $c : P(\text{Rejeitar } H_0 | H_0) \simeq \alpha_0$, em particular,

$$\begin{aligned} c &= F_{\chi^2_{(k-\beta-1)}}^{-1}(1 - \alpha_0) \\ &= F_{\chi^2_{(4-0-1)}}^{-1}(1 - 0.05) \\ &= F_{\chi^2_{(3)}}^{-1}(0.95) \\ &\stackrel{\text{tabela}}{=} 7.815. \end{aligned}$$

- **Frequências absolutas esperadas sob H_0**

Atendendo a que, sob H_0 , $p_i = p_i^0 = \frac{1}{4}$, $i = 1, 2, 3, 4$, tem-se $E_i = n \times p_i^0 = 80 \times \frac{1}{4} = 20$, $i = 1, 2, 3, 4$.

Note-se que não é necessário qualquer agrupamento de classes uma vez que em pelo menos 80% das classes se verifica $E_i \geq 5$ e em todas elas $E_i \geq 1$.

- **Decisão**

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

i	Período 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	8:00–9:59	12	20	$\frac{(12-20)^2}{20} = 3.2$
2	10:00–11:59	28	20	3.2
3	13:00–14:59	15	20	1.25
4	15:00–16:59	25	20	1.25
		$\sum_{i=1}^k o_i = n = 80$	$\sum_{i=1}^k E_i = n = 80$	$t = \sum_{i=1}^k \frac{(o_i - E_i)^2}{E_i} = 8.9$

Dado que $t = 8.9 \in W = (7.815, +\infty)$, devemos rejeitar H_0 a qualquer n.s. maior ou igual a 5%.

2. Uma engenheira informática decidiu estudar a forma como o número de horas despendidas na internet na última semana (Y) se relaciona com o número de anos de escolaridade completos do utilizador (x). A amostra de 200 adultos que recolheu ao acaso conduziu aos seguintes resultados:

$$\sum_{i=1}^{200} x_i = 2\,207, \quad \sum_{i=1}^{200} x_i^2 = 25\,131, \quad \sum_{i=1}^{200} y_i = 1\,333, \quad \sum_{i=1}^{200} y_i^2 = 13\,295, \quad \sum_{i=1}^{200} x_i y_i = 15\,322$$

- (a) Após ter considerado o modelo de regressão linear simples $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ ($i = 1, \dots, 200$), (2.0)
obtenha as estimativas de mínimos quadrados de β_0 e β_1 e interprete a estimativa de β_1 .

• **[Modelo de RLS**

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

Y_i = número de horas despendidas na internet pelo i – éximo indivíduo

x_i = número de anos de escolaridade completados pelo i – éximo indivíduo

ϵ_i = erro aleatório associado à medição da resposta i – éximo indivíduo]

• **Estimativas de β_0 e β_1**

[Uma vez que $n = 200$ e

$$\sum_{i=1}^n x_i = 2\,207,$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{2\,207}{200} = 11.035$$

$$\sum_{i=1}^n x_i^2 = 25\,131$$

$$\sum_{i=1}^n x_i^2 - n(\bar{x})^2 = 25\,131 - 200 \times 11.035^2 = 776.76$$

$$\sum_{i=1}^n y_i = 1\,333,$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i = \frac{1\,333}{200} = 6.665$$

$$\sum_{i=1}^n y_i^2 = 13\,295$$

$$\sum_{i=1}^n y_i^2 - n(\bar{y})^2 = 13\,295 - 200 \times 6.665^2 = 4410.6$$

$$\sum_{i=1}^n x_i y_i = 15\,322$$

$$\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y} = 15\,322 - 200 \times 11.035 \times 6.665 = 612.35,$$

as estimativas dos mínimos quadrados de β_1 e β_0 são, para este modelo, iguais a:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n(\bar{x})^2}$$

$$= \frac{612.35}{776.76}$$

$$= 0.788337$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \times \bar{x}$$

$$= 6.665 - 0.788337 \times 11.035$$

$$= -2.0343.$$

• **Interpretação da estimativa de mínimos quadrados de β_1**

$$\hat{\beta}_1 \simeq 0.788337$$

Por cada ano adicional de escolaridade completado, estima-se que o valor esperado do número de horas semanais despendidas na internet aumente aproximadamente 0.788337 horas (pouco mais de 47 minutos).

- (b) Será que a amostra recolhida permite concluir que existe relação de tipo linear entre o valor (3.0)
esperado da variável resposta (Y) e a variável explicativa (x), ao nível de significância de 5%?

• **[Hipóteses de trabalho**

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

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

• **Hipóteses**

$$H_0 : \beta_1 = \beta_{1,0} = 0$$

$$H_1 : \beta_1 \neq \beta_{1,0} = 0 \text{ (existe relação linear entre } E(Y) \text{ e } x)$$

- **Nível de significância**

$$\alpha_0 = 5\%$$

- **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 bilateral ($H_1 : \beta_1 \neq \beta_{1,0}$), pelo que a região de rejeição de H_0 é $W = (-\infty, -c) \cup (c, +\infty)$, onde

$$c : P(\text{Rejeitar } H_0 | H_0) = \alpha_0$$

$$c = F_{t_{(n-2)}}^{-1}(1 - \alpha_0/2)$$

$$c = F_{t_{(200-2)}}^{-1}(0.975)$$

$$c \simeq \Phi^{-1}(0.975)$$

$$c = 1.96$$

- **Decisão**

Tendo em conta que $\hat{\beta}_1 = 0.788337$, $\sum_{i=1}^n x_i^2 - n \bar{x}^2 = 776.76$ e que a estimativa σ^2 é igual a

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{n-2} \left[\left(\sum_{i=1}^n y_i^2 - n \bar{y}^2 \right) - (\hat{\beta}_1)^2 \left(\sum_{i=1}^n x_i^2 - n \bar{x}^2 \right) \right] \\ &= \frac{1}{200-2} (4410.6 - 0.788337^2 \times 776.76) \\ &\simeq 19.838, \end{aligned}$$

o valor observado da estatística de teste é dado por

$$\begin{aligned} t &= \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2 - n \bar{x}^2}}} \\ &= \frac{0.788337 - 0}{\sqrt{\frac{19.838}{776.76}}} \\ &= 4.9329. \end{aligned}$$

Como $t = 4.9329 \in W = (-\infty, -1.96) \cup (1.96, +\infty)$, devemos rejeitar $H_1 : \beta_1 = \beta_{1,0} = 0$ a favor de $H_1 : \beta_1 \neq \beta_{1,0}$ a qualquer n.s. $\alpha_0 \geq 5\%$.

(c) *Deduz a um intervalo de confiança a 90% para o valor esperado do número de horas semanais despendidas na internet por um adulto com 12 anos de escolaridade completos.* (2.0)

- **Outro parâmetro desconhecido**

$E(Y|x_0 = 12) = \beta_0 + \beta_1 \times 12$, valor esperado do número de horas semanais despendidas na internet por um adulto com 12 anos de escolaridade completos.

- **IC a 95% para $E(Y|x_0)$**

– **Passo 1 — V.a. fulcral para $E(Y|x_0) = \beta_0 + \beta_1 x_0$**

$$Z = \frac{(\hat{\beta}_0 + \hat{\beta}_1 x_0) - (\beta_0 + \beta_1 x_0)}{\sqrt{\hat{\sigma}^2 \times \left[\frac{1}{n} + \frac{(\bar{x} - x_0)^2}{\sum_{i=1}^n x_i^2 - n \bar{x}^2} \right]}} \sim t_{(n-2)}$$

– **Passo 2 — Quantis de probabilidade**

Já que $(1 - \alpha) \times 100\% = 90\%$ temos $\alpha = 0.10$ e lidaremos com dois quantis simétricos $a_\alpha = -b_\alpha$ iguais a:

$$\pm F_{t_{(n-2)}}^{-1}(1 - \alpha/2) = \pm F_{t_{(200-2)}}^{-1}(1 - 0.10/2) \simeq \begin{cases} \pm \Phi^{-1}(0.95) \stackrel{\text{tabela}}{=} \pm 1.6449 \\ \text{ou} \\ \pm F_{t_{(\infty)}}^{-1}(0.95) \stackrel{\text{tabela}}{=} \pm 1.65. \end{cases}$$

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

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

$$P \left[-F_{t_{(n-2)}}^{-1}(1 - \alpha/2) \leq \frac{(\hat{\beta}_0 + \hat{\beta}_1 x_0) - (\beta_0 + \beta_1 x_0)}{\sqrt{\hat{\sigma}^2 \times \left[\frac{1}{n} + \frac{(\bar{x} - x_0)^2}{\sum_{i=1}^n x_i^2 - n \bar{x}^2} \right]}} \leq F_{t_{(n-2)}}^{-1}(1 - \alpha/2) \right] = 1 - \alpha$$

$$P \left\{ (\hat{\beta}_0 + \hat{\beta}_1 x_0) - F_{t_{(n-2)}}^{-1}(1 - \alpha/2) \times \sqrt{\hat{\sigma}^2 \times \left[\frac{1}{n} + \frac{(\bar{x} - x_0)^2}{\sum_{i=1}^n x_i^2 - n \bar{x}^2} \right]} \leq \beta_0 + \beta_1 x_0 \leq (\hat{\beta}_0 + \hat{\beta}_1 x_0) + F_{t_{(n-2)}}^{-1}(1 - \alpha/2) \times \sqrt{\hat{\sigma}^2 \times \left[\frac{1}{n} + \frac{(\bar{x} - x_0)^2}{\sum_{i=1}^n x_i^2 - n \bar{x}^2} \right]} \right\} = 1 - \alpha.$$

– **Passo 4 — Concretização**

$$\begin{aligned} IC_{(1-\alpha) \times 100\%}(\beta_0 + \beta_1 x_0) &= \left[(\hat{\beta}_0 + \hat{\beta}_1 x_0) \pm F_{t_{(n-2)}}^{-1}(1 - \alpha/2) \times \sqrt{\hat{\sigma}^2 \times \left[\frac{1}{n} + \frac{(\bar{x} - x_0)^2}{\sum_{i=1}^n x_i^2 - n \bar{x}^2} \right]} \right] \\ &= [(-2.0343 + 0.788337 \times 12) \\ &\quad \pm 1.6449 \times \sqrt{19.838 \times \left[\frac{1}{200} + \frac{(11.035 - 12)^2}{776.76} \right]}] \\ &= [7.42574 \pm 1.6449 \times \sqrt{0.122973}] \\ &= [6.8489, 8.0026]. \end{aligned}$$