Métodos Estatísticos
Capítulo 5

Resumos e exemplos

Exercícios

Cap 5 - Estimação

Estatística é qualquer função da amostra aleatória (logo também é uma v.a!) cuja expressão não contém parâmetros desconhecidos

Exemplo: $\overline{X}_n = \dfrac{X_1 + X_2 + \cdots X_n}{n}$

Estimador é qualquer estatística escalar para estimar um parâmetro populacional.

Estimativa é um valor do estimador para uma amostra em concreto.
↳ é um número real!

População ⟶ Amostra aleatória ⟶ Concretização da amostra aleatória

v.a. X → μ parâmetro desconhecido

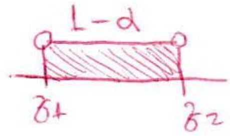
$X_1, X_2 \ldots X_n$ iid → $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ Estimador

$x_1, x_2, \ldots, x_n$ → $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ Estimativa

Método da variável fulcral (para determinar um intervalo aleatório)

- X v.a com um parâmetro desconhecido (habitualmente μ ou σ^2)
- Pretende-se um intervalo de confiança $(1-\alpha)\%$ para esse parâmetro

1) Escolher a variável fulcral (Pag 4 tabelas)
2) Determinar z_1, z_2 tais que $P(Z \in]z_1, z_2[) = 1-\alpha$
3) Determinar L_1, L_2 tais que $P(\underbrace{\Theta}_{\text{parâmetro desconhecido}} \in]L_1, L_2[) = 1-\alpha$



4) Dada uma amostra particular, determinar estimativas para L_1 e L_2.

- $\hat{\Theta}$ é um bom estimador centrado para estimar θ se $E(\hat{\Theta}) = \Theta$. Caso contrário diz-se enviesado.
- As propriedades dos estimadores não são dependentes de m

1)

2) $(x_1, X_2 \dots x_m)$ é amostra de X, $f_\Theta(x) = \begin{cases} \frac{2}{\Theta}\left(1-\frac{x}{\Theta}\right) & , 0<x<\Theta \\ 0 & , cc. \end{cases}$

$E(x) = \frac{\Theta}{3}$

a) $E(\hat{\Theta}_1) = E\left(\sum_{i=1}^{m} x_i\right) = \sum_{i=1}^{m} E(x_i) = \sum_{i=1}^{m} \frac{\Theta}{3} = \frac{m\Theta}{3} \neq \Theta$, logo $\hat{\Theta}_1$ é enviesado

b) $E(\hat{\Theta}_1) = \frac{m\Theta}{3}$ $\times\frac{3}{m}$ $\times\frac{3}{m}$

$\Leftrightarrow \frac{3}{m} E(\hat{\Theta}_1) = \frac{3\cancel{m}\Theta}{3\cancel{m}}$

$\Leftrightarrow E\left(\frac{3}{m}\hat{\Theta}_1\right) = \Theta$

Então $\hat{\Theta}_2 = \frac{3}{m}\hat{\Theta}_1 = \frac{3}{m}\sum_{i=1}^{m} x_i = 3\overline{X}$

c) $X_1, X_2 \dots X_m$, $m=100$, $\overline{x} = 20{,}2$

Uma estimativa cêntrica para Θ será um valor particular de $\hat{\Theta}_2$, ou seja $\hat{\Theta}_2 = 3\overline{x} = 3 \cdot 20{,}2 = 60{,}6$

Uma estimativa cêntrica para $E(x)$ será $\overline{x}$, ou seja, $20{,}2$.

Estimação intervalar

→ Construir ao grau $(1-\alpha)\%$ um intervalo de confiança para θ

Passo 1) Escolher a variável fulcral (Tabelas pag 4).

Passo 2) Enquadrar a variável fulcral num intervalo com probabilidade igual ao grau $(1-\alpha)\%$.

Passo 3) Partindo do enquadramento feito no passo 2, resolver as inequações por forma a enquadrar o parâmetro θ que se quer estimar.

Passo 4) Concretizar para a amostra recolhida

Exe 3) X = "diâmetro de um parafuso" ; $X \sim N(\mu, \sigma)$; media = 150mm

Sabemos que: $n = 20$; $\sum_{i=1}^{20} x_i = 2900$; $\sum_{i=1}^{20} x_i^2 = 432500$; Confiança $= 0,95$

a) Como podem para $\sigma = 25$mm, $n = 20 < 30$ e distribuição normal:

Passo 1) $Z = \sqrt{n}\,\dfrac{\overline{X}-\mu}{\sigma} \sim N(0,1)$

Passo 2)

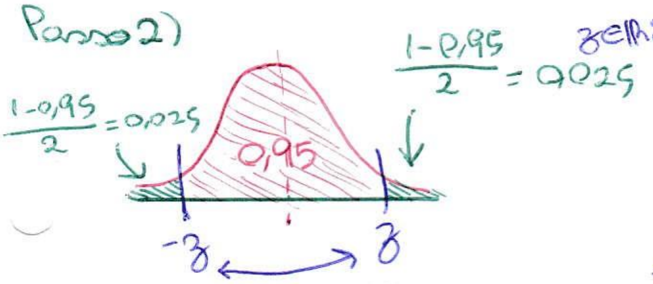


gráfico simétrico em zero

$|-z| = z$

$z \in \mathbb{R}$: $P(-z < Z < z) = 0,95$

$\Leftrightarrow P(Z < z) = 0,95 - (-z)$

$\Leftrightarrow P(Z < z) = 0,95 + 0,025$

$\Leftrightarrow P(Z < z) = 0,975$

$\Leftrightarrow z = F_N^{-1}(0,975)$

$\Leftrightarrow z = 1,96$

z	0,06
1,9	0,975

Tabela Distribuição Normal Centrada e reduzida

Passo 3) $-1,96 < Z < 1,96$

$$\Leftrightarrow -1,96 < \sqrt{n}\,\frac{\overline{X}-\mu}{\sigma} < 1,96$$

$$\Leftrightarrow -\frac{1,96}{\sqrt{n}} < \frac{\overline{X}-\mu}{\sigma} < \frac{1,96}{\sqrt{n}}$$

$$\Leftrightarrow -\frac{1,96\sigma}{\sqrt{n}} < \overline{X}-\mu \leq \frac{1,96\sigma}{\sqrt{n}}$$

$$\Leftrightarrow -\overline{X}-\frac{1,96\sigma}{\sqrt{n}} < -\mu < -\overline{X}+\frac{1,96\sigma}{\sqrt{n}}$$

$$\Leftrightarrow \overline{X}-\frac{1,96\sigma}{\sqrt{n}} < \mu < \overline{X}+\frac{1,96\sigma}{\sqrt{n}}$$

O intervalo para μ com grau de confiança a 0,95 é:

$$\left]\overline{X}-\frac{1,96\sigma}{\sqrt{n}}, \overline{X}+\frac{1,96\sigma}{\sqrt{n}}\right[$$

Passo 4) $\sum_{i=1}^{20} x_i \rightarrow \overline{x} = \frac{1}{20}\sum_{i=1}^{20} x_i = \frac{2900}{20} = 145$; $n=20$; $\sigma=25$

$$\left]\overline{x}-\frac{1,96\sigma}{\sqrt{n}}, \overline{x}+\frac{1,96\sigma}{\sqrt{n}}\right[$$

$$= \left]145-\frac{1,96\cdot 25}{\sqrt{20}}, 145+\frac{1,96\cdot 25}{\sqrt{20}}\right[$$

$$= \left]134,0433 ; 155,9567\right[$$

R: Como 150 pertence ao intervalo com grau 0,95, podemos concluir que, com 95% de confiança, não é possível admitir irregularidade de produção na máquina.

b) Aqui podem para σ desconhecido

Continuamos com:

$$n=20;\ \sum_{i=1}^{20} x_i = 2900;\ \sum_{i=1}^{20} x_i^2 = 432500;\ \text{confiança} = 0,95$$

Passo 1) Como $\sigma = ?$ e $n=20<30$ e distribuição normal, então:

$$Z = \sqrt{n}\,\frac{\overline{X}-\mu}{S} \sim \underline{t_{n-1} = t_{19}}$$ ← Distribuição t-student!

Passo 2) Como gráfico simétrico em zero, $|-z| = z$. Então:

$$z \in \mathbb{R}: P(-z \leq Z < z) = 0,95 \iff P(Z<z) = 0,95 + 0,025$$

$$\iff P(Z<z) = 0,975$$

$$\iff z = F_T^{-1}(0,975)$$

$$\iff z = 2,093$$

n\P	0,975
19	2,093

Tabela distribuição t-student

Passo 3) $$-2,093 < \sqrt{n}\,\frac{\overline{X}-\mu}{S} < 2,093$$

$$\iff \overline{X} - \frac{2,093 \cdot S}{\sqrt{n}} < \mu < \overline{X} + \frac{2,093 \cdot S}{\sqrt{n}}$$

O intervalo para μ com grau 0,95 e μ desconhecido é:

$$\left]\overline{X} - \frac{2,093 \cdot S}{\sqrt{n}},\ \overline{X} + \frac{2,093 \cdot S}{\sqrt{n}}\right[$$

Passo 4: $$\overline{x} = \frac{1}{20}\sum_{i=1}^{20} x_i = \frac{1}{20} \cdot 2900 = 145\ ;\ n=20;$$

$$S^2 = \frac{1}{19}\sum_{i=1}^{20} x_i^2 - \frac{20}{19}\overline{X}^2 = \frac{1}{19}\cdot 432500 - \frac{20}{19}145^2$$

$$= 631,5789$$

$$S = \sqrt{631,5789} = 25,1312$$

$$\left]145 - \frac{2,093 \cdot 25,1312}{\sqrt{20}},\ 145 + \frac{2,093 \cdot 25,1312}{\sqrt{20}}\right[= \underline{]133,2384;\ 156,7616[}$$

R: Como 150 pertence ao intervalo com grau 0,95, podemos concluir que, com 95% de confiança, não é possível admitir irregularidade da produção na máquina.

Exe 4) $m = 10 \{7{,}4;\ 7{,}8;\ 7{,}1;\ 6{,}9;\ 7{,}3;\ 7{,}6;\ 7{,}3;\ 7{,}4;\ 7{,}7;\ 7{,}3\}$

X = "Tensão de ruptura", $X \sim N(\mu, \sqrt{0{,}08})$

a) Um bom estimador para μ é $\bar{X}$.

$$\hat{\mu} = \bar{x} = \frac{1}{10}\sum_{i=10}^{10} x_i = \frac{1}{10}(7{,}4+7{,}8+7{,}1+6{,}9+7{,}3+7{,}6+7{,}3+7{,}4+7{,}7+7{,}3) = 7{,}38$$

b) intervalo para 95% de confiança.

Passo 1: $\sigma = \sqrt{0{,}08}$ e $m = 10 < 30 \Rightarrow Z = \sqrt{m}\,\frac{\bar{X}-\mu}{\sigma} \sim N(0,1)$

Passo 2: $P(-z < Z \leq z) = 0{,}95 \Leftrightarrow P(Z < z) = 0{,}95 + \frac{1-0{,}95}{2}$

$\Leftrightarrow P(Z < z) = 0{,}975$

$\Leftrightarrow z = 1{,}96$

z	0,06
1,9	0,975

Passo 3: $-1{,}96 < \sqrt{m}\,\frac{\bar{X}-\mu}{\sigma} < 1{,}96$

$$\Leftrightarrow \bar{X} - \frac{1{,}96\sigma}{\sqrt{m}} < \mu < \bar{X} + \frac{1{,}96\sigma}{\sqrt{m}}$$

$$\left]\bar{X} - \frac{1{,}96\sigma}{\sqrt{m}} < \mu \leq \bar{X} + \frac{1{,}96\sigma}{\sqrt{m}}\right[$$

Passo 4: $\left]7{,}38 - \frac{1{,}96 \cdot \sqrt{0{,}08}}{\sqrt{10}},\ 7{,}38 + \frac{1{,}96 \cdot \sqrt{0{,}08}}{\sqrt{10}}\right[$

$=]7{,}2047;\ 7{,}5553[$

c) Erro da aproximação de μ pelo intervalo Aleatório de confiança é:

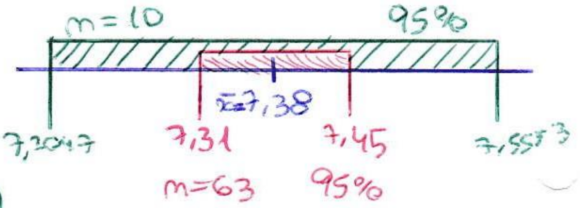
!!! $\varepsilon = z\frac{\sigma}{\sqrt{m}}$ Como queremos que o erro não ultrapasse 0,07, então:

$\varepsilon \leq 0{,}07 \Leftrightarrow z\frac{\sigma}{\sqrt{m}} \leq 0{,}07$

$\Leftrightarrow 1{,}96 \cdot \frac{\sqrt{0{,}08}}{\sqrt{m}} \leq 0{,}07$

$\Leftrightarrow m \geq \left(\frac{1{,}96 \cdot \sqrt{0{,}08}}{0{,}07}\right)^2$

$\Leftrightarrow m > 62{,}72 \rightarrow m \geq 63$



5) X = "Classificação de determinado curso", $X \sim N(\mu, \sigma)$

$(n=)$ 42 observações, tais que:

$$\sum_{i=1}^{42} x_i = 588 = x_1 + x_2 + \dots + x_{42} \; ; \quad \sum_{i=1}^{42} x_i^2 = 8400 = x_1^2 + x_2^2 + \dots + x_{42}^2$$

a) Estimativa para $\mu \Rightarrow \bar{x} = \frac{1}{42} \sum_{i=1}^{42} x_i = \frac{588}{42} = 14$

Estimativa para Variância $= \sigma^2 \Rightarrow s^2 = \frac{1}{41} \sum_{i=1}^{42} x_i^2 - \frac{42}{41} \bar{x}^2 = \frac{8400}{41} - \frac{42}{41} \cdot 14^2$

$\simeq 4,0956$

b) ? $(1-\alpha)\%$ que permite afirmar que $\mu \in \,]a, b[$ de amplitude $a - b = \underline{\underline{1,224}}$

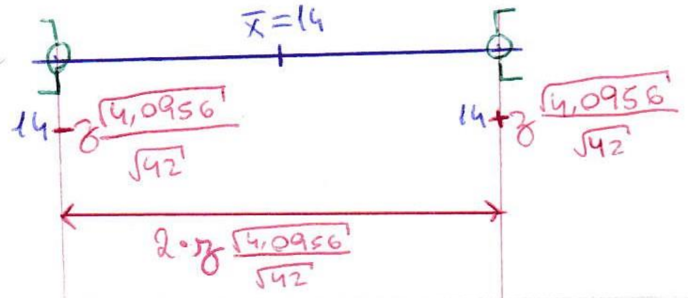
① σ desconhecido $\rightarrow Z = \sqrt{n} \cdot \frac{\bar{X} - \mu}{S} \sim t_{n-1} = t_{41}$

② $z \in \mathbb{R}: P(-z < Z < z) = 1 - \alpha$ (Não se sabe $1-\alpha$, saltar para ③)

③ $-z < \sqrt{n} \frac{\bar{X} - \mu}{S} < z \Leftrightarrow \bar{X} - z \frac{S}{\sqrt{n}} < \mu < \bar{X} + z \frac{S}{\sqrt{n}}$

$IAC_\mu(1-\alpha\%) = \,]\bar{X} - z \frac{S}{\sqrt{n}}, \bar{X} + z \frac{S}{\sqrt{n}}[$

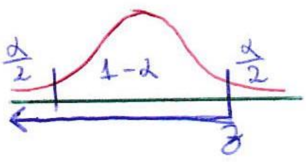
④ $IC_\mu(1-\alpha\%) = \,]14 - z \frac{\sqrt{4,0956}}{\sqrt{42}}, 14 + z \frac{\sqrt{4,0956}}{\sqrt{42}}[$



$2 \cdot z \cdot \frac{\sqrt{4,0956}}{\sqrt{42}} = 1,224$

$\Leftrightarrow z = \frac{1,224 \cdot \sqrt{42}}{2 \sqrt{4,0956}} \Leftrightarrow z = \underline{\underline{1,9593}}$

Como $n = 42$ não existe nas Tabelas, utilizar valor mais próximo e z mais próximo



$P(-z < Z < z) = 1 - \alpha$

$\Leftrightarrow P(Z < z) = 1 - \alpha + \frac{\alpha}{2}$

$n \backslash P$	0,975
40	2,045 $\simeq z$

$42 \simeq$

$1 - \alpha + \frac{\alpha}{2} = 0,975 \Leftrightarrow 1 - \frac{\alpha}{2} = 0,975$

$\Leftrightarrow \alpha = 0,05$

$P(Z < 1,9593) = 1 - 0,05 = 0,95 = 95\%$ de confiança

6) X = "Medida de uma dada grandeza", $X \sim N(\mu, \sigma)$

10 observações $\}$ 125,3   124,8   124,8   125,1   125,0   125,1   124,7   124,4   125,2   125,0 $\{$

$\uparrow n$

a) Estimativa para $\mu \Rightarrow \bar{x} = \frac{1}{10}\sum_{i=1}^{10} x_i = \frac{125{,}3 + \dots + 125{,}0}{10} = \frac{1250{,}4}{10} = \underline{125{,}04}$

Estimativa para variância $\Rightarrow s^2 = \frac{1}{9}\sum_{i=1}^{10} x_i^2 - \frac{10}{9}\bar{x}^2$

$= \frac{125{,}3^2 + \dots + 125{,}0^2}{9} - \frac{10}{9} \cdot 14^2 \cong 0{,}0516$

$s = \sigma = \sqrt{0{,}0516} = \underline{0{,}227}$

b) Intervalo de confiança para a média μ a 98%

1) σ desconhecido e $X \sim N \longrightarrow Z = \sqrt{n} \cdot \frac{\bar{X} - \mu}{S} \sim t_{n-1} = t_9$

2) $z \in \mathbb{R}: P(-z < Z < z) = 0{,}98$

$\Leftrightarrow P(Z < z) = 0{,}98 + \frac{1 - 0{,}98}{2}$

$\Leftrightarrow P(Z < z) = 0{,}99 \rightarrow z = 2{,}821$

$n \backslash p$	0,99
9	2,821

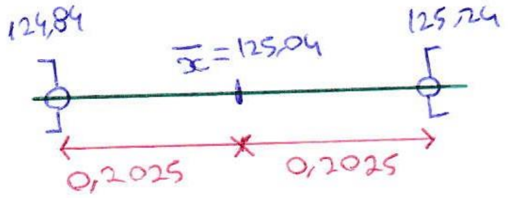
3) $-2{,}821 < \sqrt{n} \cdot \frac{\bar{X} - \mu}{S} < 2{,}821$

$\Leftrightarrow \bar{X} - 2{,}821 \frac{S}{\sqrt{n}} < \mu < \bar{X} + 2{,}821 \frac{S}{\sqrt{n}}$

$IAC_\mu(98\%) =]\bar{X} - 2{,}821 \frac{S}{\sqrt{n}} < \mu < \bar{X} + 2{,}821 \frac{S}{\sqrt{n}}$

4) $IC_\mu(98\%) =]125{,}04 - 2{,}821 \frac{0{,}227}{\sqrt{10}} \ ; \ 125{,}04 + 2{,}821 \frac{0{,}227}{\sqrt{10}}[$

$=]124{,}84 \ ; \ 125{,}24[$



• Erro máximo da estimativa $\bar{x} = 125{,}04$

$= 0{,}2025$

• Amplitude do intervalo

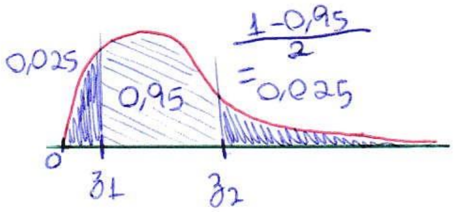
$= 2 \cdot 2{,}821 \frac{0{,}227}{\sqrt{10}} = 0{,}4050$

c) Intervalo de confiança para o desvio padrão de 95%

$(\sigma = \sqrt{V(X)})$

① μ desconhecido e $X \sim N(\mu, \sigma) \rightarrow Z = \frac{n-1}{\sigma^2} \cdot S^2 \sim \chi^2_{n-1} = \chi^2_9$

② $z_1, z_2 \in \mathbb{R}: P(z_1 < Z < z_2) = 0{,}95$



Não é simétrica!

$z_1 \neq z_2$

Distribuição Qui-Quadrado χ^2_{n-1}

$P(Z < z_1) = 0{,}025 \rightarrow z_1 = 2{,}7$

$P(Z < z_2) = 0{,}975 \rightarrow z_2 = 19{,}02$

n\P	0,025	0,975
9	2,700	19,02

③ $2{,}7 < \frac{n-1}{\sigma^2} S^2 < 19{,}02$

$\Leftrightarrow \frac{(n-1) \cdot S^2}{19{,}02} < \sigma^2 < \frac{(n-1) \cdot S^2}{2{,}7}$

$IAC_{\sigma^2}(95\%) = \left] \frac{(n-1) S^2}{19{,}02} , \frac{(n-1) S^2}{2{,}7} \right[$

a)

④ $IC_{\sigma^2}(95\%) = \left] \frac{9 \cdot 0{,}0516}{19{,}02} , \frac{9 \cdot 0{,}0516}{2{,}7} \right[$

variância

$=]0{,}02438 \, ; \, 0{,}17176[$

$IC_{\sigma}(95\%) =]\sqrt{0{,}02438} ; \sqrt{0{,}17176}[$

$=]0{,}1561 \, ; \, 0{,}4144[$