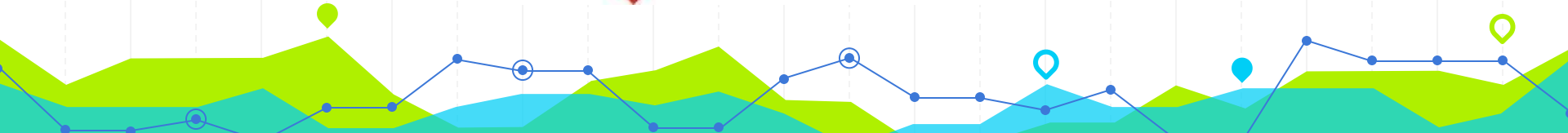




Lisbon School
of Economics
& Management
Universidade de Lisboa



Estatística II

Licenciatura em Gestão do Desporto
2.º Ano/2.º Semestre
2025/2026

Aulas Teórico-Práticas N.ºs 10 e 11 (Semana 6)

Docente: Elisabete Fernandes

E-mail: efernandes@iseg.ulisboa.pt



<https://doity.com.br/estatistica-aplicada-a-nutricao>



<https://basiccode.com.br/produto/informatica-basica/>

Conteúdos Programáticos

Aulas Teórico-Práticas (Semanas 1 a 3)

- **Capítulo 1:** Revisões e Distribuições de Amostragem

Aulas Teórico-Práticas (Semanas 4 a 7)

- **Capítulo 2:** Estimação

Aulas Teórico-Práticas (Semanas 7 a 9)

- **Capítulo 3:** Testes de Hipóteses

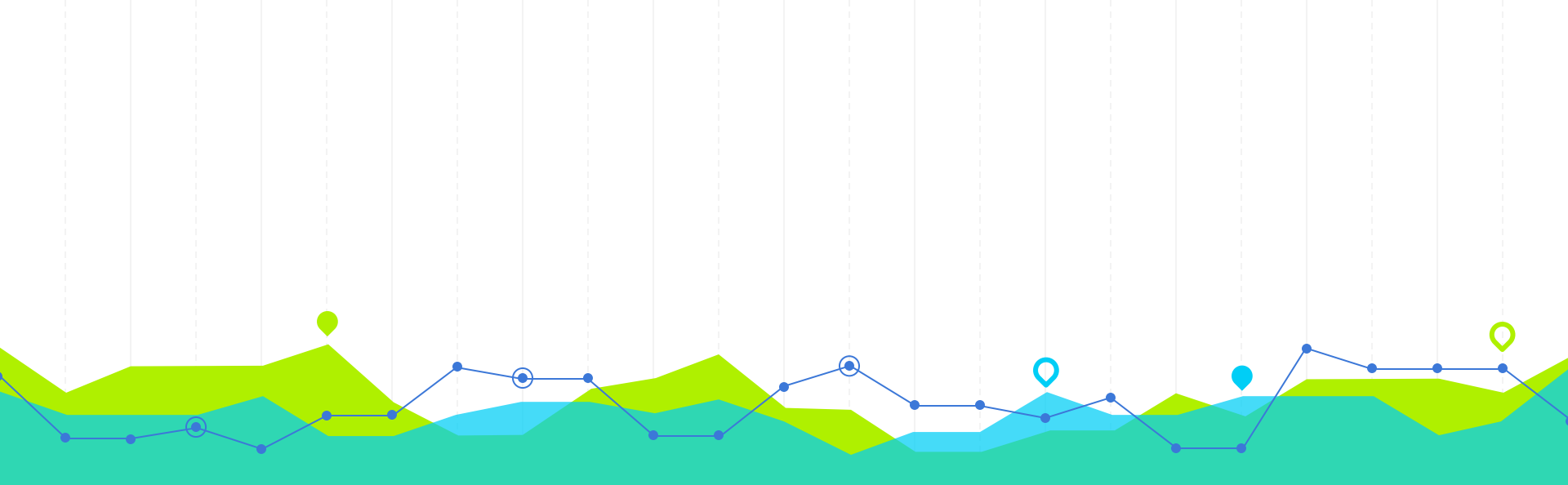
Aulas Teórico-Práticas (Semanas 10 a 13)

- **Capítulo 4:** Modelo de Regressão Linear Múltipla

Material didático: Exercícios do Livro Murteira et al (2015), Formulário e Tabelas Estatísticas

Bibliografia: B. Murteira, C. Silva Ribeiro, J. Andrade e Silva, C. Pimenta e F. Pimenta; *Introdução à Estatística*, 2ª ed., Escolar Editora, 2015.

<https://cas.iseg.ulisboa.pt>



Propriedades dos Estimadores: Exercícios

Estimador Centrado, Eficiente e Convergente

1

1. Suponha que uma variável aleatória X representa o número de avarias de um dispositivo durante um período de tempo e que obedece a uma lei de Poisson de parâmetro λ desconhecido. Para este parâmetro foram sugeridos dois estimadores:

$$\hat{\lambda} = \frac{X_1 + \dots + X_n}{n} \text{ e } \tilde{\lambda} = \frac{X_1 + X_n}{2}.$$

- a) Compare-os quanto ao enviesamento.
- b) Deduza a variância para cada um deles.
- c) Qual dos dois estimadores é mais eficiente? Justifique a sua escolha.
- d) Estude os dois estimadores quanto à consistência.



Exercício 1 a): Estimadores Centrados

Se $X \sim P(\lambda)$, então $E(X) = \text{Var}(X) = \lambda$.

Se $X_1, X_2, \dots, X_n$ é uma a. a. duma dada população $X \sim P(\lambda)$, então $X_i \sim P(\lambda), i = 1, \dots, n$.

$$\begin{aligned} \text{a) } E(\hat{\lambda}) &= E\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right) = \frac{1}{n}E(X_1 + X_2 + \dots + X_n) = \frac{1}{n}(E(X_1) + E(X_2) + \dots + E(X_n)) \\ &= \frac{1}{n}(\lambda + \lambda + \dots + \lambda) = \frac{1}{n}n\lambda = \lambda. \end{aligned}$$

$$E(\tilde{\lambda}) = E\left(\frac{X_1 + X_n}{2}\right) = \frac{1}{2}(E(X_1) + E(X_n)) = \frac{1}{2}(\lambda + \lambda) = \lambda.$$

Portanto, ambos os estimadores são centrados.

[ProbabilidadesEstatistica_2019 \(uevora.pt\)](#)

VALOR ESPERADO, MOMENTOS E PARÂMETROS

$$\text{Var}(X) = E(X - \mu)^2 = E(X^2) - \mu^2;$$

$$\text{Cov}(X, Y) = E\{(X - \mu_X)(Y - \mu_Y)\} = E(XY) - E(X)E(Y); \quad \rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

$$E(aX + bY) = aE(X) + bE(Y) \text{ e } \text{Var}(aX + bY) = a^2 \text{Var}(X) + b^2 \text{Var}(Y) + 2ab \text{Cov}(X, Y) \text{ com } a, b \text{ constantes}$$

Formulário

Exercício 1 b): Variâncias dos Estimadores

$$\begin{aligned} b) \operatorname{Var}(\hat{\lambda}) &= \operatorname{Var}\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) \\ &= \frac{1}{n} \operatorname{Var}(X_1 + X_2 + \cdots + X_n) \stackrel{\substack{\text{indep.} \\ \bar{X}_i}}{=} \frac{1}{n^2} (\operatorname{Var}[X_1] + \operatorname{Var}[X_2] + \cdots + \operatorname{Var}[X_n]) \\ &= \frac{1}{n^2} (\lambda + \lambda + \cdots + \lambda) = \frac{1}{n^2} n\lambda = \frac{\lambda}{n}. \end{aligned}$$

$$\operatorname{Var}(\tilde{\lambda}) = \operatorname{Var}\left(\frac{X_1 + X_n}{2}\right) \stackrel{\substack{\text{indep.} \\ \bar{X}_i}}{=} \frac{1}{2^2} (\operatorname{Var}(X_1) + \operatorname{Var}(X_n)) = \frac{1}{4} (\lambda + \lambda) = \frac{\lambda}{2}.$$

Exercício 1 c): Eficiência

- c) Se $n = 1$, $Var(\hat{\lambda}) > Var(\tilde{\lambda}) \rightarrow \tilde{\lambda}$ é mais eficiente do que $\hat{\lambda}$;
Se $n = 2$, $Var(\hat{\lambda}) = Var(\tilde{\lambda}) \rightarrow \hat{\lambda}$ é tão eficiente como $\tilde{\lambda}$;
Se $n > 2$, $Var(\hat{\lambda}) < Var(\tilde{\lambda}) \rightarrow \hat{\lambda}$ é mais eficiente do que $\tilde{\lambda}$

Exercício 1 d): Consistência

- d) $\hat{\lambda}$ é um estimador consistente, pois $E(\hat{\lambda}) = \lambda$ e $Var(\hat{\lambda}) = \frac{\lambda}{n} \rightarrow 0$, quando $n \rightarrow \infty$.
 $\tilde{\lambda}$ não é um estimador consistente: $E(\tilde{\lambda}) = \lambda$ mas $Var(\tilde{\lambda}) = \frac{\lambda}{2}$, seja qual for o valor de n .

Exercício 6.4 [Estimadores, erro quadrático médio, eficiência relativa.]

Se (X_1, X_2, X_3) constitui uma amostra aleatória de dimensão 3 extraída de uma população normal com valor esperado μ e variância σ^2 , qual a eficiência de $\frac{1}{4}(X_1 + 2X_2 + X_3)$ relativamente a $\bar{X}$?



Exercício 6.4: Eficiência

- **V.a. de interesse**

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

- **Amostra aleatória**

$\underline{X} = (X_1, X_2, X_3)$ é a.a. de dimensão 3 proveniente da população X , donde $X_i \sim_{i.i.d.} X, i = 1, 2, 3$.

- **Parâmetro desconhecido a estimar**

$$\mu = E(X)$$

- **Estimador de $\mu = E(X)$**

$$\bar{X} = \frac{1}{3} \sum_{i=1}^3 X_i \quad [\text{média da a.a. } \underline{X}]$$

- **Erro quadrático médio de $\bar{X}$**

$$\begin{aligned} EQM_{\mu}(\bar{X}) &= V(\bar{X}) + [bias_{\mu}(\bar{X})]^2 \\ &= V(\bar{X}) + [E(\bar{X}) - \mu]^2 \\ &\stackrel{X_i \sim_{i.i.d.} X}{=} \frac{V(X)}{n} + [E(X) - E(X)]^2 \\ &= \frac{V(X)}{n} \\ &= \frac{\sigma^2}{3} \end{aligned}$$

Exercício 6.4: Eficiência

- Outro estimador de $\mu = E(X)$

$$T = \frac{1}{4} (X_1 + 2X_2 + X_3)$$

- Erro quadrático médio de T

$$\begin{aligned} EQM_{\mu}(T) &= V(T) + [bias_{\mu}(T)]^2 \\ &= V(T) + [E(T) - \mu]^2 \\ &= V\left[\frac{1}{4}(X_1 + 2X_2 + X_3)\right] + \left\{E\left[\frac{1}{4}(X_1 + 2X_2 + X_3)\right] - E(X)\right\}^2 \\ &\stackrel{X_i \text{ indep}}{=} \frac{1}{4^2} [V(X_1) + 4V(X_2) + V(X_3)] + \left\{\frac{1}{4} [E(X_1) + 2E(X_2) + E(X_3)] - E(X)\right\}^2 \\ &\stackrel{X_i \sim X}{=} \frac{1}{4^2} (1 + 4 + 1) V(X) + \left[\frac{1}{4} (1 + 2 + 1) E(X) - E(X)\right]^2 \\ &= \frac{3}{8} V(X) \\ &= \frac{3\sigma^2}{8} \end{aligned}$$

Exercício 6.4: Eficiência

- **Eficiência do estimador** $T = \frac{1}{4}(X_1 + 2X_2 + X_3)$ **relativamente a** $\tilde{X}$

$$\begin{aligned}e_{\mu}(T, \tilde{X}) &= \frac{EQM_{\mu}(\tilde{X})}{EQM_{\mu}(T)} \\&= \frac{\frac{\sigma^2}{3}}{\frac{3\sigma^2}{8}} \\&= \frac{8}{9}\end{aligned}$$

- **Comentário**

Tendo em conta que

$$e_{\mu}(T, \tilde{X}) = \frac{8}{9} < 1,$$

i.e.,

$$EQM_{\mu}(\tilde{X}) < EQM_{\mu}(T),$$

pode afirmar-se que $T = \frac{1}{4}(X_1 + 2X_2 + X_3)$ é um estimador menos eficiente que $\tilde{X}$ no que respeita à estimação de $\mu = E(X)$.

1ª) Sendo $X_1, \dots, X_n$ é uma amostra aleatória de uma $U(0, \theta)$.

a) Encontre o estimador pelo método dos momentos de θ e o seu EQM.



Exercício 1 a): Método dos Momentos e EQM

- **UNIFORME (CONTÍNUA)** $X \sim U(\alpha, \beta)$, $(\alpha < \beta)$

$$f(x|\alpha, \beta) = \frac{1}{\beta - \alpha} \quad \alpha < x < \beta \quad ; \quad E(X) = \frac{\alpha + \beta}{2} \quad ; \quad \text{Var}(X) = \frac{(\beta - \alpha)^2}{12} ;$$

Formulário

Dado que $\mathbb{E}(X) = \frac{\theta}{2}$ e o primeiro momento amostral é $m_1 = \bar{X}$, temos que pelo método dos momentos,

$$\mathbb{E}(X) = m_1 \Rightarrow \frac{\hat{\theta}_{mm}}{2} = \bar{X} \Rightarrow \hat{\theta}_{mm} = 2\bar{X}.$$

Sabendo que o EQM de um estimador $\hat{\theta}$ de um parâmetro θ é definido como

$$EQM(\hat{\theta}) = \text{Var}(\hat{\theta}) + (\mathbb{E}(\hat{\theta}) - \theta)^2,$$

e dado que $\text{Var}(X) = \frac{\theta^2}{12}$, temos que para $\hat{\theta}_{mm} = 2\bar{X}$ o EQM é dado por,

$$\mathbb{E}(2\bar{X}) = \frac{2 \sum_{i=1}^n \mathbb{E}(X_i)}{n} = \frac{2n \frac{\theta}{2}}{n} = \theta,$$

$$\text{Viés}(2\bar{X}) = \mathbb{E}(2\bar{X}) - \theta = \theta - \theta = 0,$$

$$\text{Var}(2\bar{X}) = \frac{4 \sum_{i=1}^n \text{Var}(X_i)}{n^2} = \frac{4n \frac{\theta^2}{12}}{n^2} = \frac{\theta^2}{3n},$$

$$EQM(2\bar{X}) = \text{Var}(2\bar{X}) + (\text{Viés}(2\bar{X}))^2 = \text{Var}(2\bar{X}) = \frac{\theta^2}{3n}.$$

2. Pretende-se ter uma ideia da densidade do tráfego em certo local, entre as 8 e as 9 horas. Sabe-se que o número de veículos que aí passam durante esse período segue um processo de Poisson. Fizeram-se as seguintes 9 observações casuais:

(95, 100, 80, 70, 110, 98, 97, 90, 70).

- a) Obtenha o estimador e a estimativa da máxima verosimilhança para o número médio de veículos que passam naquele local.
- b) Mostre que se trata de um estimador consistente e que é o mais eficiente.
- c) Obtenha a estimativa de máxima verosimilhança para a probabilidade de decorrerem pelo menos dois minutos sem passar qualquer veículo.
- d) Mostre que $\sum_{i=1}^n X_i$ é estatística suficiente para a média da população.



Exercício 2

$X \equiv$ nº de veículos entre as 8h e as 9h (intervalo de 1 hora)

$$X \sim \text{Poi}(\lambda) \quad (\lambda > 0) \quad n = 9$$

$$f_x(x|\lambda) = \frac{e^{-\lambda} \lambda^x}{x!} \quad (x=0,1,2,\dots) \quad E(X) = \lambda$$

$$\text{Var}(X) = \lambda$$

amostra ($n=9$):

$$(x_1, x_2, \dots, x_9) = (95, 100, 80, 70, 110, 98, 97, 90, 70)$$

Exercício 2 a)

a) $\lambda = ?$

$$\begin{aligned} L(\lambda) &= \prod_{i=1}^m f(x_i | \lambda) = \prod_{i=1}^m \frac{e^{-\lambda} \lambda^{x_i}}{x_i!} = \\ &= \frac{e^{-m\lambda} \lambda^{\sum_{i=1}^m x_i}}{\prod_{i=1}^m x_i!} \quad (\lambda > 0) \end{aligned}$$

$$\begin{aligned} l(\lambda) &= \ln L(\lambda) = \ln \left(\frac{e^{-m\lambda} \lambda^{\sum_{i=1}^m x_i}}{\prod_{i=1}^m x_i!} \right) = \\ &= -m\lambda + \sum_{i=1}^m x_i \ln(\lambda) - \sum_{i=1}^m \ln(x_i!) \end{aligned}$$

$$\frac{d}{d\lambda} l(\lambda) = 0 \quad (\Leftrightarrow) \quad -m + \frac{\sum_{i=1}^m x_i}{\lambda} = 0 \quad (\Leftrightarrow)$$

$$(\Leftrightarrow) \quad \frac{\sum_{i=1}^m x_i}{\lambda} = m \quad (\Leftrightarrow) \quad \lambda = \frac{\sum_{i=1}^m x_i}{m} = \bar{x}$$

Exercício 2 a)

$$\begin{aligned}\frac{d^2}{d\lambda^2} \ell(\hat{\lambda}) &= \frac{d}{d\lambda} \left(-n + \frac{\sum_{i=1}^n x_i}{\lambda} \right) \Big|_{\lambda=\bar{x}} = - \frac{\sum_{i=1}^n x_i}{\lambda^2} \Big|_{\lambda=\bar{x}} \\ &= - \frac{n\bar{x}}{\bar{x}^2} = - \frac{n}{\bar{x}} < 0 \text{ porque } n > 0 \text{ e } \bar{x} > 0.\end{aligned}$$

Conclusão:

- $\bar{x}$ é a estimada da M.V.
- $\bar{x} = \frac{95+100+80+70+110+98+97+90+70}{9} = \frac{810}{9} = 90$ é a estimativa da M.V.

Exercício 2 b)

b)

Um estimador $\hat{\lambda}$ é consistente para λ se:

$$\begin{cases} \lim_{n \rightarrow +\infty} E(\hat{\lambda}) = \lambda \\ \lim_{n \rightarrow +\infty} \text{Var}(\hat{\lambda}) = 0 \end{cases}$$

Neste caso temos $\hat{\lambda} = \bar{X}$

$$\begin{aligned} \lim_{n \rightarrow +\infty} E(\hat{\lambda}) &= \lim_{n \rightarrow +\infty} E(\bar{X}) = \lim_{n \rightarrow +\infty} E(X) = \\ &= \lim_{n \rightarrow +\infty} \lambda = \lambda \end{aligned}$$

Exercício 2 b)

$$\begin{aligned}\lim_{n \rightarrow +\infty} \text{Var}(\hat{\lambda}) &= \lim_{n \rightarrow +\infty} \text{Var}(\bar{X}) = \lim_{n \rightarrow +\infty} \left(\frac{\text{Var}(X)}{n} \right) = \\ &= \lim_{n \rightarrow +\infty} \left(\frac{\lambda}{n} \right) = 0\end{aligned}$$

Conclusão: $\hat{\lambda}$ é estimador consistente de λ .

Exercício 2 c)

c) $Y \equiv$ nº de veículos num intervalo de 2 minutos

$$\frac{60 \text{ min} - 1}{2 \text{ min} - \lambda_Y} \quad \lambda_Y = \frac{2\lambda}{60} = \frac{\lambda}{30} \quad Y \sim \text{Poi}\left(\frac{\lambda}{30}\right)$$

$$f_Y(0) = \frac{e^{-\frac{\lambda}{30}} \left(\frac{\lambda}{30}\right)^0}{0!} = e^{-\frac{\lambda}{30}} \rightarrow \text{função de } \lambda \text{ (estritamente decrescente)}$$

Como $e^{-\frac{\lambda}{30}}$ é função biunívoca de λ podemos usar a propriedade de invariância do estimado da M.V.:

$$\hat{f}_Y(0) = \widehat{e^{-\frac{\lambda}{30}}} = e^{-\frac{\hat{\lambda}}{30}} = e^{-\frac{\bar{x}}{30}} = e^{-\frac{90}{30}} = e^{-3}$$

Exercício 2 c)

Resolução alternativa usando a distribuição exponencial:

$$X \sim \text{Poi}(\lambda)$$

nº de veículos por hora

$$Y \sim \text{ex}(\lambda)$$

tempo (em horas) até à chegada do 1º veículo (ou entre dois veículos consecutivos)

$$f_Y(y|\lambda) = \lambda e^{-\lambda y} \quad (y > 0)$$

$$2 \text{ min} = \frac{2}{60} \text{ horas} = \frac{1}{30} \text{ horas}$$

Porque Y é V.A.
contínua

$$\begin{aligned} P(Y \geq \frac{1}{30}) &= 1 - P(Y < \frac{1}{30}) = 1 - F_Y(\frac{1}{30}) = \\ &= 1 - \int_0^{\frac{1}{30}} \lambda e^{-\lambda y} dy = 1 - [-e^{-\lambda y}]_{y=0}^{y=\frac{1}{30}} = \\ &= 1 - (-e^{-\frac{\lambda}{30}} + e^0) = 1 - (1 - e^{-\frac{\lambda}{30}}) = \\ &= e^{-\lambda/30} \rightarrow \text{Função estritamente decrescente de } \lambda \end{aligned}$$

Exercício 2 c)

Como $e^{-\lambda/30}$ é função binomial de λ podemos usar a propriedade de invariância do estimador da M.V.:

$$\begin{aligned} P(Y \geq \frac{1}{30}) &= e^{-\lambda/30} = e^{-\hat{\lambda}/30} = e^{-\frac{90}{30}} = \\ &= e^{-3} \approx 0.0498 \end{aligned}$$

Nota: Sabendo que $F_Y(y|\lambda) = 1 - e^{-\lambda y}$ podia-se fazer:

$$F_Y\left(\frac{1}{30}\right) = 1 - e^{-\frac{\lambda}{30}}$$

11. Admite-se que o tempo de funcionamento, em milhares de horas, de determinada componente electrónica, X , tem distribuição $G(3;1/\theta)$, ou seja, a sua função densidade é dada por

$$f(x|\theta) = \frac{x^2 e^{-x/\theta}}{2\theta^3} \quad (x > 0), \quad \theta > 0.$$

Com base numa amostra casual de dimensão n , definiram-se os seguintes estimadores para o parâmetro θ :

$$T_1 = \frac{\bar{X}}{3} \text{ e } T_2 = \frac{n\bar{X} + 1}{3n}.$$

- a) Estude o enviesamento e a consistência destes estimadores.
- b) Qual dos estimadores tem menor erro quadrático médio?
- c) Observada uma amostra casual de 10 componentes, registou-se um tempo médio de funcionamento, por componente, de 1600 horas. Calcule o estimador e a estimativa da máxima verosimilhança para θ .



Exercício 11 a)

$X \equiv$ Tempo de funcionamento (em milhares de horas) da componente electrónica

$$X \sim G(3, \frac{1}{\theta}) \quad E(X) = \frac{3}{1/\theta} = 3\theta$$

$$\text{Var}(X) = \frac{3}{(1/\theta)^2} = 3\theta^2$$

$$f_X(x|\theta) = \frac{x^2 e^{-x/\theta}}{2\theta^3} \quad (x > 0), \theta > 0.$$

$$T_1 = \frac{\bar{X}}{3}$$

$$T_2 = \frac{n \bar{X} + 1}{3n}$$

Exercício 11 a)

$$a) E(T_1) = E\left(\frac{\bar{X}}{3}\right) = \frac{1}{3} E(\bar{X}) = \frac{1}{3} E(X) = \frac{1}{3} 3\theta = \theta$$

T_1 é centrado.

$$\begin{aligned} \text{Var}(T_1) &= \text{Var}\left(\frac{\bar{X}}{3}\right) = \frac{1}{9} \text{Var}(\bar{X}) = \frac{1}{9m} \text{Var}(X) = \\ &= \frac{1}{9m} 3\theta^2 = \frac{1}{3m} \theta^2 \end{aligned}$$

$$\lim_{m \rightarrow +\infty} E(T_1) = \lim_{m \rightarrow +\infty} \theta = \theta$$

$$\lim_{m \rightarrow +\infty} \text{Var}(T_1) = \lim_{m \rightarrow +\infty} \frac{1}{3m} \theta^2 = 0$$

$\left\{ \begin{array}{l} T_1 \text{ é} \\ \text{consistente} \end{array} \right.$

Exercício 11 a)

$$\begin{aligned} E(T_2) &= E\left(\frac{n\bar{X} + 1}{3n}\right) = \frac{n E(\bar{X}) + 1}{3n} = \\ &= \frac{n E(X) + 1}{3n} = \frac{3n\theta + 1}{3n} = \theta + \frac{1}{3n} \end{aligned}$$

T_2 é enviesado

$$\text{Env}(T_2) = E(T_2) - \theta = \frac{1}{3n}$$

$$\begin{aligned} \text{Var}(T_2) &= \text{Var}\left(\frac{n\bar{X} + 1}{3n}\right) = \frac{n^2 \text{Var}(\bar{X})}{9n^2} = \\ &= \frac{\text{Var}(\bar{X})}{9} = \frac{\text{Var}(X)}{9n} = \frac{3}{9n} \theta^2 = \frac{1}{3n} \theta^2 \end{aligned}$$

$$\lim_{n \rightarrow +\infty} E(T_2) = \lim_{n \rightarrow +\infty} \left(\theta + \frac{1}{3n}\right) = \theta$$

$$\lim_{n \rightarrow +\infty} \text{Var}(T_2) = \lim_{n \rightarrow +\infty} \left(\frac{1}{3n} \theta^2\right) = 0$$

T_2 é

consistente

Exercício 11 b)

b)

$$EQM(T_1) = Var(T_1) + E_{err}(T_1)^2 = Var(T_1) = \frac{\theta^2}{3n}$$

Porque T_1 é centrado

$$\begin{aligned} EQM(T_2) &= Var(T_2) + E_{err}(T_2)^2 = \\ &= \frac{\theta^2}{3n} + \left(\frac{1}{3n}\right)^2 > EQM(T_1) = \frac{\theta^2}{3n} \end{aligned}$$

Conclusão: T_1 tem menor EQM.

Exercício 11 c)

c) $n = 10$ componentes $\bar{x} = 1.6$ milhões de horas.

$$L(\theta) = \prod_{i=1}^{10} \frac{x_i^2 e^{-x_i/\theta}}{2\theta^3} = \frac{1}{2^{10}\theta^{30}} \left(\prod_{i=1}^{10} x_i^2 \right) e^{-\frac{1}{\theta} \sum_{i=1}^{10} x_i} \quad (\theta > 0)$$

$$\begin{aligned} l(\theta) &= \ln L(\theta) = \ln \left(\frac{1}{2^{10}\theta^{30}} \right) + \sum_{i=1}^{10} \ln(x_i^2) - \frac{1}{\theta} \sum_{i=1}^{10} x_i \quad (\theta > 0) \\ &= -10 \ln(2) - 30 \ln(\theta) + \sum_{i=1}^{10} 2 \ln(x_i) - \frac{1}{\theta} \sum_{i=1}^{10} x_i \quad (\theta > 0) \end{aligned}$$

$$\frac{d}{d\theta} l(\theta) = -\frac{30}{\theta} - \left(-\frac{1}{\theta^2} \sum_{i=1}^{10} x_i \right) = 0 \quad (=)$$

$$(=) \frac{30}{\theta} = \frac{1}{\theta^2} \sum_{i=1}^{10} x_i \Rightarrow \hat{\theta} = \frac{\sum_{i=1}^{10} x_i}{30} = \frac{\bar{x}}{3}$$

Exercício 11 c)

$$\begin{aligned}\frac{d^2}{d\theta^2} l(\theta) &= \frac{d}{d\theta} \left(-\frac{30}{\theta} + \frac{1}{\theta^2} \sum_{i=1}^{10} x_i \right) = \\ &= \frac{30}{\theta^2} - \frac{2\theta}{\theta^4} \sum_{i=1}^{10} x_i = \frac{30}{\theta^2} - \frac{2}{\theta^3} \sum_{i=1}^{10} x_i = \\ &= \frac{30}{\theta^2} - \frac{20 \bar{x}}{\theta^3}\end{aligned}$$

$$\frac{d^2}{d\theta^2} l(\hat{\theta}) = \frac{30}{\left(\frac{\bar{x}}{3}\right)^2} - \frac{20 \bar{x}}{\left(\frac{\bar{x}}{3}\right)^3} = \frac{270}{\bar{x}^2} - \frac{540}{\bar{x}^2} < 0$$

Conclusão: $\hat{\theta} = \frac{\bar{X}}{3}$ é θ estimado da M.V.

$$\hat{\theta} = \frac{\bar{x}}{3} = \frac{1.6}{3} \text{ é uma estimativa da M.V.}$$

14. Seja X uma variável aleatória com função densidade

$$f(x|\theta) = \frac{3\sqrt{x}}{2\theta} \exp\left\{-\frac{x^{3/2}}{\theta}\right\} \quad (x > 0), \text{ para } \theta > 0.$$

a) Mostre que

$$\hat{\theta} = \frac{\sum_{i=1}^n X_i^{3/2}}{n},$$

é o estimador da máxima verosimilhança para θ .

b) Sabendo que $X^{3/2}$ tem distribuição exponencial de média θ , estude o enviesamento e a consistência do estimador da alínea a).



Exercício 14 a)

$$f(x|\theta) = \frac{3\sqrt{x}}{2\theta} \exp\left\{-\frac{x^{\frac{3}{2}}}{\theta}\right\} \quad (x>0), \quad \text{para } \theta>0$$

a)

$$\begin{aligned} L(\theta) &= \prod_{i=1}^n f(x_i|\theta) = \prod_{i=1}^n \frac{3\sqrt{x_i}}{2\theta} \exp\left\{-\frac{x_i^{\frac{3}{2}}}{\theta}\right\} = \\ &= \frac{3^m \left(\prod_{i=1}^n x_i\right)^{\frac{1}{2}}}{2^m \theta^m} \exp\left\{-\frac{1}{\theta} \sum_{i=1}^n x_i^{\frac{3}{2}}\right\} \end{aligned}$$

$$\begin{aligned} l(\theta) &= \log\{L(\theta)\} = \\ &= m \log(3) + \frac{1}{2} \sum_{i=1}^n \log(x_i) - m \log(2) - m \log(\theta) - \frac{1}{\theta} \sum_{i=1}^n x_i^{\frac{3}{2}} \end{aligned}$$

Exercício 14 a)

$$\frac{d}{d\theta} \{l(\theta)\} = -\frac{n}{\theta} + \frac{\sum_{i=1}^n x_i^{\frac{3}{2}}}{\theta^2} = 0 \quad (=)$$

$$(=) \quad \frac{n}{\theta} = \frac{\sum_{i=1}^n x_i^{\frac{3}{2}}}{\theta^2} \quad (=)$$

$$\Rightarrow \hat{\theta} = \frac{\sum_{i=1}^n x_i^{\frac{3}{2}}}{n} \rightarrow \text{Candidato a EMV}$$

$$\begin{aligned} \frac{d^2}{d\theta^2} l(\theta) &= \frac{d}{d\theta} \left(-\frac{n}{\theta} + \frac{\sum_{i=1}^n x_i^{\frac{3}{2}}}{\theta^2} \right) \\ &= \frac{n}{\theta^2} - \frac{\sum_{i=1}^n x_i^{\frac{3}{2}}}{\theta^4} \cdot 2\theta = \\ &= \frac{n}{\theta^2} - 2 \frac{\sum_{i=1}^n x_i^{\frac{3}{2}}}{\theta^3} \end{aligned}$$

Exercício 14 a)

$$\begin{aligned}\frac{d^2}{d\theta^2} l(\hat{\theta}) &= \frac{n}{\left(\frac{\sum_{i=1}^n x_i^{3/2}}{n}\right)^2} - 2 \frac{\sum_{i=1}^n x_i^{3/2}}{\left(\frac{\sum_{i=1}^n x_i^{3/2}}{n}\right)^3} = \\ &= \frac{n^3}{\left(\sum_{i=1}^n x_i^{3/2}\right)^2} - \frac{2n^3}{\left(\sum_{i=1}^n x_i^{3/2}\right)^2} = -\frac{n^3}{\left(\sum_{i=1}^n x_i^{3/2}\right)^2} < 0 \text{ porque } n > 0\end{aligned}$$

Conclusão: $\hat{\theta} = \frac{\sum_{i=1}^n x_i^{\frac{3}{2}}}{n}$ é $\theta \in MV$, $Q \in D$

Exercício 14 b)

b) $Y = X^{\frac{3}{2}} \sim \text{LX} \left(\frac{1}{\theta} \right) \quad \begin{aligned} E(Y) &= \theta \\ \text{Var}(Y) &= \theta^2 \end{aligned}$

Um estimador de θ é consistente se: $\lim_{n \rightarrow +\infty} E(\hat{\theta}) = \theta$
 $\lim_{n \rightarrow +\infty} \text{Var}(\hat{\theta}) = 0$

$$\begin{aligned} E(\hat{\theta}) &= E\left(\frac{\sum_{i=1}^n X_i^{3/2}}{n}\right) = \frac{1}{n} E\left(\sum_{i=1}^n X_i^{3/2}\right) = \\ &= \frac{1}{n} \sum_{i=1}^n E(X_i^{3/2}) = \frac{1}{n} \sum_{i=1}^n E(Y) = \\ &= \frac{1}{n} \cdot n \cdot E(Y) = E(Y) = \theta \end{aligned}$$

$\hookrightarrow$ porque $X_i^{3/2}$ são i.i.d. a $Y (i=1, 2, \dots, n)$

Conclusão: $\hat{\theta}$ é um estimador centrado em θ
(não enviesado)

Exercício 14 b)

$$\lim_{n \rightarrow +\infty} E(\hat{\theta}) = \lim_{n \rightarrow +\infty} (\theta) = \theta$$

$$\begin{aligned} \text{Var}(\hat{\theta}) &= \text{Var}\left(\frac{\sum_{i=1}^n X_i^{3/2}}{n}\right) = \frac{1}{n^2} \text{Var}\left(\sum_{i=1}^n X_i^{3/2}\right) = \\ &= \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i^{3/2}) = \frac{1}{n^2} n \text{Var}(Y) = \frac{\theta^2}{n} \end{aligned}$$

$$\lim_{n \rightarrow +\infty} \text{Var}(\hat{\theta}) = \lim_{n \rightarrow +\infty} \left(\frac{\theta^2}{n}\right) = 0$$

Como: $\lim_{n \rightarrow +\infty} E(\hat{\theta}) = \theta$ e $\lim_{n \rightarrow +\infty} \text{Var}(\hat{\theta}) = 0$ podemos

concluir que $\hat{\theta}$ é um estimador consistente.

24. Classifique as afirmações apresentadas abaixo de verdadeiras ou falsas, justificando sucintamente.

- a) Com base numa amostra casual de dimensão $n=3$, o estimador $T = 0.5 X_1 + 0.4 X_2 + 0.1 X_3$, sendo centrado é também o estimador mais eficiente para μ .
- b) Comparar os Erros Quadráticos Médios de dois estimadores centrados (referentes ao mesmo parâmetro) equivale a comparar as respectivas variâncias.
- c) Um estimador consistente é sempre centrado.



Exercício 24 a)

a) **Falso.** Nada garante que não há estimadores mais eficientes. Por exemplo:

$$\begin{aligned} \text{Var}(T) &= 0.5^2 \text{Var}(X_1) + 0.4^2 \text{Var}(X_2) + 0.1^2 \text{Var}(X_3) = \\ &= 0.5^2 \sigma_x^2 + 0.4^2 \sigma_x^2 + 0.1^2 \sigma_x^2 = \\ &= \sigma_x^2 (0.5^2 + 0.4^2 + 0.1^2) \simeq 0.42 \sigma_x^2 \end{aligned}$$

$E(\bar{X}) = E(X) = \mu$ também é estimador centrado.

$$\text{Var}(\bar{X}) = \frac{\sigma_x^2}{3} \simeq 0.33 \sigma_x^2 < \text{Var}(T) \simeq 0.42 \sigma_x^2$$

Logo $\bar{X}$ é mais eficiente que T .

Exercício 24 b)

b) Verdade

$$EQM(T) = Var(T) + bias(T)^2$$

Se o estimado for centrado temos $bias(T) = 0$

logo $EQM(T) = Var(T)$.

Exercício 24 c)

c) Falso. Existem estimadores consistentes que são enviesados.

Exemplo: S^2

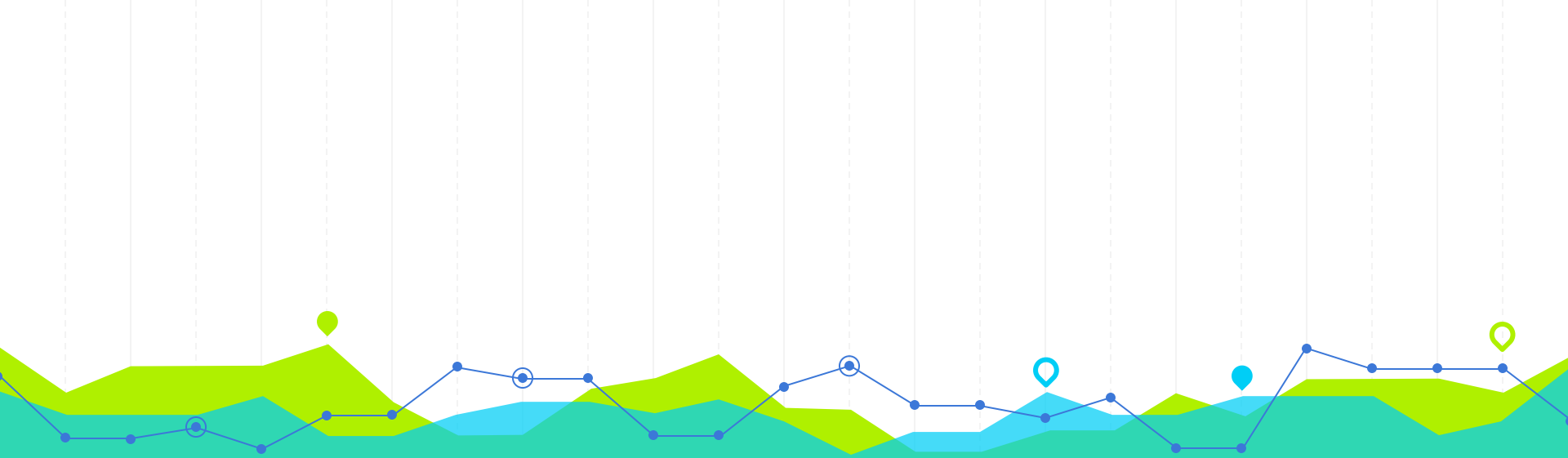
Demonstração no livro (página 365)

$$E(S^2) = \frac{n-1}{n} \sigma^2 \rightarrow S^2 \text{ é estimador enviesado de } \sigma^2$$

$$\lim_{n \rightarrow +\infty} E(S^2) = \lim_{n \rightarrow +\infty} \frac{n-1}{n} \sigma^2 = \sigma^2$$

$$\begin{aligned} \lim_{n \rightarrow +\infty} \text{Var}(S^2) &= \lim_{n \rightarrow +\infty} \left(\frac{\mu_4 - \mu_2^2}{n} - 2 \frac{(\mu_4 - 2\mu_2^2)}{n^2} + \frac{\mu_4 - 3\mu_2^2}{n^3} \right) = \\ &= 0 \end{aligned}$$

Conclusão: S^2 é enviesado e consistente.



Estimação Intervalar

Intervalos de Confiança (ICs)

2

Estimação Pontual vs Estimação Intervalar

Existem dois processos de estimação paramétrica:

- ✓ **Estimação Pontual:** produção de um valor (estimativa), que se pretende que seja o melhor, para um determinado parâmetro da população, com base na informação amostral;
- ✓ **Estimação Intervalar:** construção de um intervalo que, com certo grau de certeza previamente estipulado, se pretende que contenha o verdadeiro valor do parâmetro da população.

Um **estimador** dum parâmetro da população é uma variável aleatória (v. a.) ou função que depende da informação amostral e cujas realizações fornecem aproximações para o parâmetro desconhecido. A um valor específico assumido por este estimador para uma amostra em concreto chama-se **estimativa**.

Afonso & Nunes, 2019

Estimação Pontual vs Estimação Intervalar

Falámos até aqui da estimação pontual e métodos de determinar estimadores de um parâmetro desconhecido θ .

Iremos agora tratar a questão da **estimação intervalar**.

Os intervalos são preferíveis quando, em vez de se propôr uma estimativa isolada, $\hat{\theta}$, podemos associar-lhe uma medida de erro $\hat{\theta} \pm \epsilon$, para significar que provavelmente o verdadeiro valor do parâmetro estará em $\hat{\theta} - \epsilon, \hat{\theta} + \epsilon$.

Intervalo de Confiança (IC)

Definição

Considere-se uma amostra aleatória $(X_1, X_2, \dots, X_n)$ de uma população com função de distribuição $F(x|\theta)$. Sejam $\Theta_1^*(X_1, \dots, X_n)$ e $\Theta_2^*(X_1, \dots, X_n)$ duas estatísticas, tais que

$$P(\Theta_1^* < \theta < \Theta_2^*) = 1 - \alpha, \quad 0 < \alpha < 1,$$

onde α é uma constante, não dependente do parâmetro θ .

Diz-se que (Θ_1^*, Θ_2^*) é um intervalo aleatório, que contém θ com probabilidade $1 - \alpha$.

Intervalo de Confiança

Com a utilização de um intervalo de confiança para estimarmos um parâmetro ficamos a ganhar?

Efectivamente, pensemos por exemplo, no estimador $\bar{X}$.
Tem-se $P[\bar{X} = \mu] = 0$, mas já temos uma probabilidade positiva se considerarmos

$$P\{\mu \in]\bar{X} - a, \bar{X} + a[\} \quad \text{com } a > 0$$

ou seja, há uma probabilidade positiva de o intervalo aleatório conter o parâmetro desconhecido.

Intervalo de Confiança

Definição

A qualquer intervalo (θ_1^*, θ_2^*) , com $\theta_1^* < \theta_2^*$, números reais, que resulta da concretização do intervalo aleatório chama-se **intervalo de confiança** a $(1 - \alpha)100\%$ para θ .

Observações: Chama-se **precisão da estimativa** à semi-amplitude do intervalo de confiança e **confiança** ou **grau de confiança** a $(1 - \alpha) \times 100\%$. Quanto maior for o intervalo, maior é o grau de confiança, mas menor a precisão da estimativa.

A **margem de erro** ou **precisão da estimativa** é metade da largura do intervalo de confiança.

[modulo1_aula5.pdf](#)

ICs vs Distribuições por Amostragem

A determinação de intervalos de confiança para os parâmetros necessita do conhecimento da distribuição dos estimadores envolvidos **distribuições por amostragem**, isto é, são distribuições de funções da amostra aleatória $(X_1, X_2, \dots, X_n)$, que vamos usar para obter **Intervalos de Confiança**

[modulo1_aula5.pdf](#)

ICs vs Distribuições por Amostragem: Exemplos

AMOSTRAGEM. DISTRIBUIÇÕES POR AMOSTRAGEM

Formulário

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} ; \quad S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n} = \frac{\sum_{i=1}^n X_i^2}{n} - \bar{X}^2 ; \quad (n-1)S'^2 = n S^2$$

$$E(\bar{X}) = \mu \quad ; \quad \text{Var}(\bar{X}) = \frac{\sigma^2}{n} \quad ; \quad E(S^2) = \frac{n-1}{n} \sigma^2 ; \quad E(S'^2) = \sigma^2$$

IC vs Distribuições de Amostragem: Exemplos

Relembrando

- Se $X \sim N(\mu, \sigma)$ e σ conhecido $\rightarrow \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$

Para obter o I.C. para μ com σ desconhecido

Variável usada	Condições	Distribuição
$\frac{\bar{X} - \mu}{S/\sqrt{n}}$ $S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$	$X_i \sim N(\mu, \sigma)$ $i = 1, 2, \dots, n$	$t_{(n-1)}$

Variância corrigida

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

Definição da distribuição t – Student

Se $Z \sim N(0, 1)$ e $X \sim \chi^2_{(n)}$ são v.a. independentes

$\Downarrow$

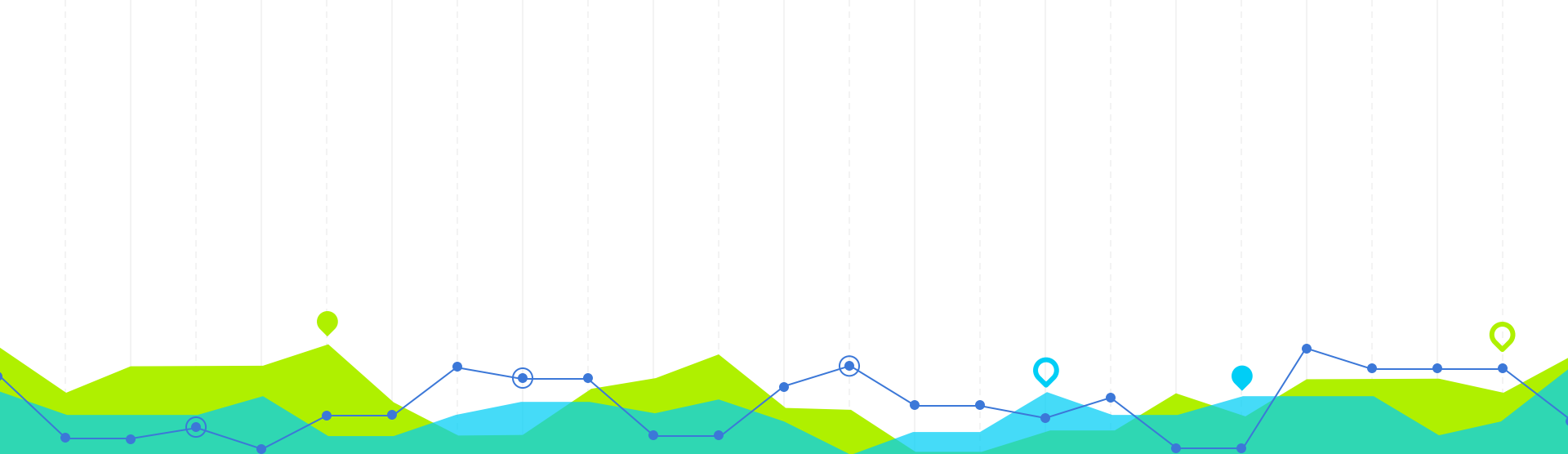
$$T = \frac{Z}{\sqrt{X/n}} \sim t_{(n)}$$

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

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

$$S' = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{(n-1)}$$

modulol_aula5.pdf



Intervalo de Confiança para o Valor Médio μ

3

Intervalo de Confiança para μ

Intervalo de confiança a $(1 - \alpha) \times 100\%$ para μ quando $X \sim N(\mu, \sigma)$

- Se σ conhecido $\left[\bar{X} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}; \bar{X} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right]$

$$\bar{X} - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Variância corrigida

- Se σ desconhecido $\left[\bar{X} - t_{n-1, 1-\frac{\alpha}{2}} \frac{S'}{\sqrt{n}}; \bar{X} + t_{n-1, 1-\frac{\alpha}{2}} \frac{S'}{\sqrt{n}} \right]$

$$S'^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

$$\bar{X} - t_{1-\alpha/2}(n-1) \frac{S'}{\sqrt{n}} < \mu < \bar{X} + t_{1-\alpha/2}(n-1) \frac{S'}{\sqrt{n}}$$

[modulo1_aula5.pdf](#)

Intervalo de Confiança para μ

Amplitude do IC para μ : $2 \times z_{1-\alpha/2} (\sigma / n^{1/2})$

Amplitude do IC para μ : $2 \times t_{1-\alpha/2; n-1} (s / n^{1/2})$

$$\left(\bar{x} - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}; \bar{x} + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \right)$$

$$\left(\bar{x} - t_{1-\alpha/2} \frac{s'}{\sqrt{n}}; \bar{x} + t_{1-\alpha/2} \frac{s'}{\sqrt{n}} \right)$$

A **margem de erro** é metade da largura do intervalo de confiança.

O **grau ou nível de confiança** ($1-\alpha$) de um intervalo de confiança é a probabilidade de este vir a incluir o verdadeiro valor do parâmetro populacional.

Os **níveis de significância** (α) são as probabilidades complementares dos níveis de confiança e são usados para testar a hipótese nula (H_0) num teste de hipóteses.

- 1%, 5%, 10% = alfa => **Níveis de significância.**
- 99%, 95%, 90% = (1-alfa) => **Níveis de confiança**

O **erro padrão da média** é uma medida de variação de uma média amostral em relação à média da população. Sendo assim, é uma medida que ajuda a verificar a confiabilidade da média amostral calculada.

Quanto menor o erro padrão mais precisas são as estimativas!

O **desvio padrão (s)** é uma medida que indica a dispersão dos dados dentro de uma amostra com relação à média.

Este tipo de ICs só são válidos se a variável em estudo tem distribuição normal ou se a amostra em análise é de grande dimensão (i.e., $n \geq 30$) (ver Teorema do Limite Central).

Intervalo de Confiança para μ : Exemplo

Exemplo de construção de um I.C. no , para o valor médio de uma normal com variância conhecida (exemplo académico!)

Exemplo Dada a amostra referente a 10 alturas, admita-se que os erros de medição são normais de média 0 e desvio padrão 1.5.

```
> x<-c(175,176, 173, 175, 174, 173, 173, 176, 173, 179)
> int.conf.z<-function(x,sigma,conf.level=0.95)
  n <-length(x);xbar<-mean(x)
  alpha <- 1 - conf.level
  zstar <- qnorm(1-alpha/2)
  SE <- sigma/sqrt(n)
  xbar + c(-zstar*SE,zstar*SE)  ## definimos uma função
> int.conf.z(x,1.5)           # basta fazer isto
```

Obteve-se o I.C a 95% para μ]173.7703; 175.6297[

Intervalo de Confiança para μ e $n \geq 30$: X tem qualquer Distribuição

Intervalo de confiança a $(1 - \alpha) \times 100\%$ para μ

Se X tem **dist. qualquer não normal**

É necessário dispor de uma **amostra de dimensão elevada**, i.e., n grande $\rightarrow$ aplicação do Teorema Limite Central

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1) \quad \text{se } \sigma \text{ conhecido}$$

Ou, que é o caso mais frequente,

$$\frac{\bar{X} - \mu}{s/\sqrt{n}} \sim \mathcal{N}(0, 1) \quad \text{se } \sigma \text{ desconhecido}$$

Intervalo a $(1 - \alpha) \times 100\%$ de confiança para μ

$$\bar{X} - z_{1-\alpha/2} \frac{s'}{\sqrt{n}} < \mu < \bar{X} + z_{1-\alpha/2} \frac{s'}{\sqrt{n}}$$

$$\left[\bar{X} - z_{1-\frac{\alpha}{2}} \frac{s'}{\sqrt{n}}; \bar{X} + z_{1-\frac{\alpha}{2}} \frac{s'}{\sqrt{n}} \right]$$

IC para μ : Formulário

Variância corrigida

$$S'^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

• POPULAÇÕES NORMAIS

Média	$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$ Variância Conhecida	$\frac{\bar{X} - \mu}{S'/\sqrt{n}} \sim t(n-1)$ Variância Desconhecida
Diferença de médias	$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \sim N(0,1)$	$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1'^2}{m} + \frac{S_2'^2}{n}}} \sim t(v)$
	$T = \frac{\frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{1}{m} + \frac{1}{n}}}}{\sqrt{\frac{(m-1)S_1'^2 + (n-1)S_2'^2}{m+n-2}}} \sim t(m+n-2)$	onde v é o maior inteiro contido em r , $r = \frac{\left(\frac{S_1'^2}{m} + \frac{S_2'^2}{n}\right)^2}{\frac{1}{m-1} \left(\frac{S_1'^2}{m}\right)^2 + \frac{1}{n-1} \left(\frac{S_2'^2}{n}\right)^2}$
Variância	$\frac{nS^2}{\sigma^2} = \frac{(n-1)S'^2}{\sigma^2} \sim \chi^2(n-1)$	
Relação de variâncias	$\frac{S_1'^2}{S_2'^2} \frac{\sigma_2^2}{\sigma_1^2} \sim F(m-1, n-1)$	

IC para μ : Formulário

Variância corrigida

$$S'^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

- GRANDES AMOSTRAS

Caso geral

Variância Conhecida

Variância Desconhecida

Média	$\frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \stackrel{a}{\sim} N(0,1)$	$\frac{\bar{X} - \mu}{S' / \sqrt{n}} \stackrel{a}{\sim} N(0,1)$
Diferença de médias	$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \stackrel{a}{\sim} N(0,1)$	$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1'^2}{m} + \frac{S_2'^2}{n}}} \stackrel{a}{\sim} N(0,1)$

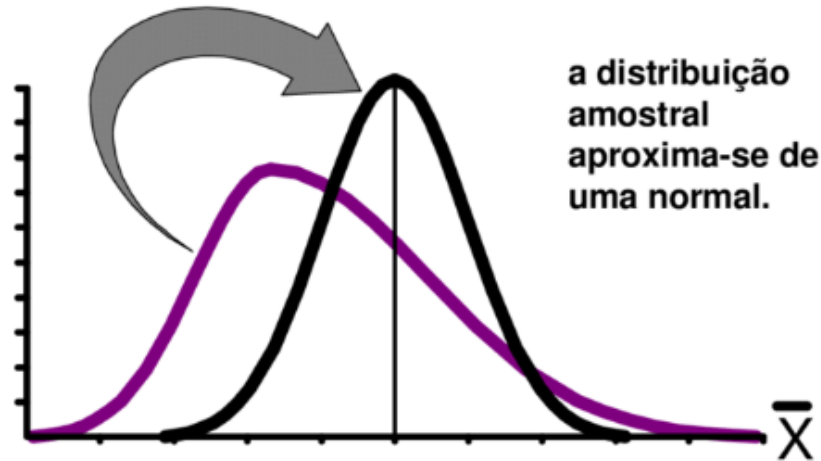
Distribuição Normal: Teorema do Limite Central

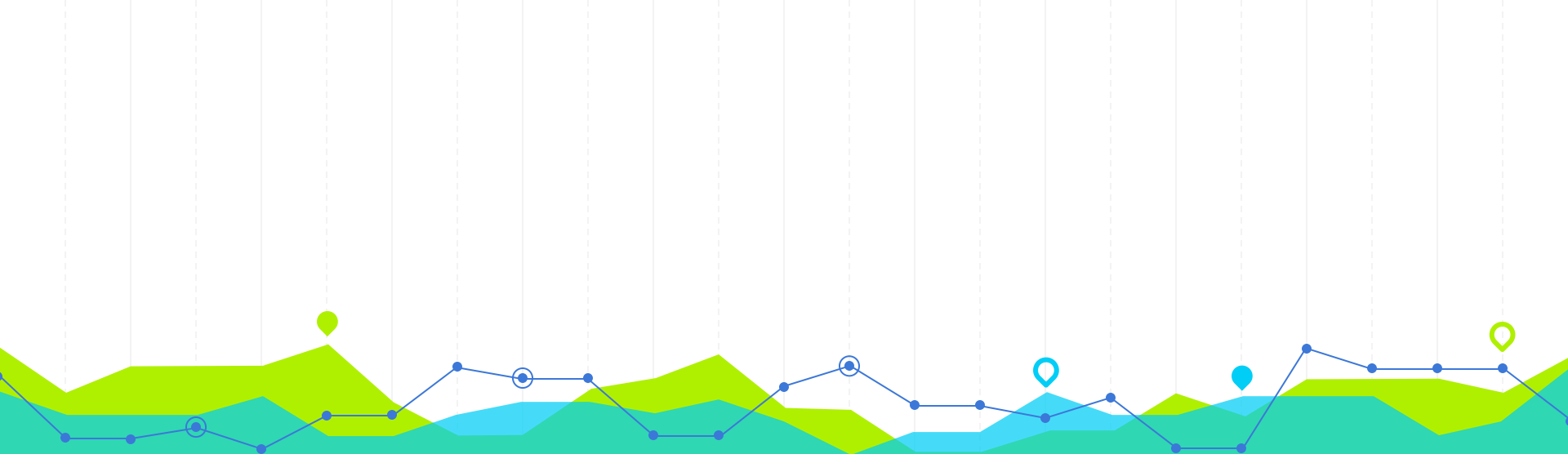


- ✓ The **central limit theorem** (CLT) states that the distribution of sample means approximates a normal distribution as the sample size gets larger.
- ✓ Sample sizes equal to or greater than 30 are considered sufficient for the CLT to hold.
- ✓ A key aspect of CLT is that the average of the sample means and standard deviations will equal the population mean and standard deviation.
- ✓ A sufficiently large sample size can predict the characteristics of a population accurately.

Distribuição Normal: Teorema do Limite Central

À medida que o tamanho da amostra se torna suficientemente grande (≥ 30) ...





Intervalo de Confiança para o Valor Médio

μ : Exercícios

4

7.1 Medições do comprimento de 25 peças produzidas por uma máquina conduziram a uma média $\bar{x} = 140$ mm. Admita que cada peça tem comprimento aleatório com distribuição normal de valor esperado μ e desvio padrão $\sigma = 10$ mm, e que o comprimento de cada peça é independente das restantes. Construa um intervalo de confiança a 95% para o valor esperado da população.



Exercício 7.1: IC para μ

Dados:

$X = \text{comprimento de uma peça} \sim \mathcal{N}(\mu, 10^2)$

θ (parâmetro desconhecido)

$n = 25 \quad (x_1, x_2, \dots, x_{25}) \quad \bar{x} = \frac{1}{25} \sum_{i=1}^{25} x_i = 140$

Pretende-se:

Intervalo de confiança a 95% para μ :

I.C. 0.95 (μ)

Exercício 7.1: IC para μ

Passo 0: ("Entender" a situação)

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

$$I.C._{0.95}(\mu)$$

[pretende-se um I.C. para o valor esperado de uma população normal de variância conhecida]

Passo 1: (escolha de variável funcional)

Definição: é uma v.a. que depende do parâmetro desconhecido mas cuja distribuição é conhecida.

Exercício 7.1: IC para μ

Média	$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$	$\frac{\bar{X} - \mu}{S'/\sqrt{n}} \sim t(n-1)$
-------	---	---

① variável fidede que vamos aplicar para construir um
população normal (velocidade, de uma
população normal) ($\mu = E[X]$; $\sigma^2 = \text{var}(X)$)

v. fidede : $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \text{cn}(0,1)$

Exercício 7.1: IC para μ

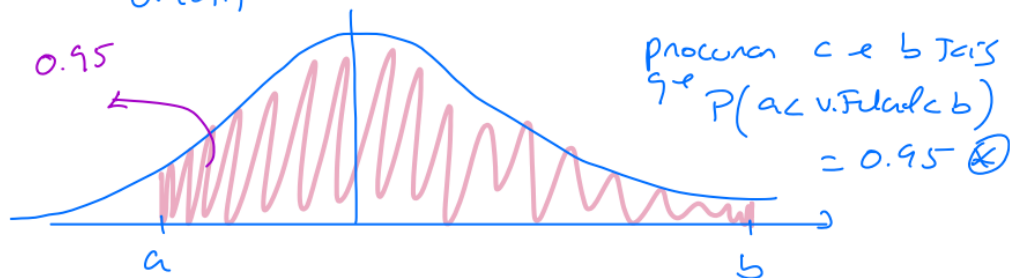
Passo 2 (encontrar os quantis)

diss. de
v. f. n. c.

(Neste caso é uma normal)
 $N(0,1)$

confiança

(Neste caso é 95%)



problema: existe uma infinidade de pontos
a e b com esta propriedade

Exercício 7.1: IC para μ

Solução: encontrar o par (a, b) tal que
 $b - a$ seja o menor possível (mas
 sempre de tal forma que $\otimes$ é
 válido)



Slides Professora Cláudia Nunes

2 formas de
 escrever a
 mesma
 quantidade:

$$Z_{0,975} = Z^*_{0,025} = 1,96$$

Tabela da Normal

ε	.0005	.0010	.0050	.0100	.0200	.0250	.0500	.1000	.2000	.3000	.4000
z_ε	3.290	3.090	2.576	2.326	2.054	1.960	1.645	1.282	.842	.524	.253
$z_{\varepsilon/2}$	3.481	3.290	2.807	2.576	2.326	2.241	1.960	1.645	1.282	1.036	.842

$$z_\varepsilon : P(Z > z_\varepsilon) = \varepsilon; \quad z_{\varepsilon/2} : P(|Z| > z_{\varepsilon/2}) = \varepsilon.$$

Cálculo do Quantil da Distribuição Normal Padrão de Probabilidade α

Alternativamente

O nível de confiança é $1-\alpha = 0,95 \Leftrightarrow \alpha = 0,05$,
então tem-se $1-\alpha/2 = 0,975$

Logo, pretende-se calcular o quantil da distribuição
normal padrão de probabilidade 0,975
 $Z_{1-\alpha/2} = Z_{0,975} = 1,96$ (ver tabela)

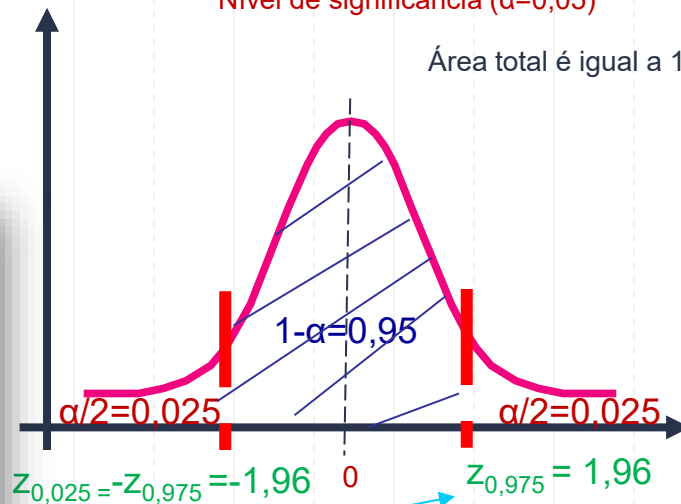
Tabela da Normal

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767

Nível de confiança ($1-\alpha=0,95$)

Nível de significância ($\alpha=0,05$)

Área total é igual a 1



Exercício 7.1: IC para μ

Passo 3 (constituição do intervalo)
 $\cong$ *mostrar como chegar lá*

v. final $\swarrow$ $\searrow$ quantis

$$P\left(-1.96 < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < 1.96\right) = 0.95$$

$\Leftrightarrow$

$$P\left(-1.96 \frac{\sigma}{\sqrt{n}} < \bar{X} - \mu < 1.96 \frac{\sigma}{\sqrt{n}}\right) = 0.95$$

$\Leftrightarrow$

$$P\left(-\bar{X} - 1.96 \frac{\sigma}{\sqrt{n}} < -\mu < -\bar{X} + 1.96 \frac{\sigma}{\sqrt{n}}\right) = 0.95$$

Exercício 7.1: IC para μ

$$P\left(\bar{X} - 1.96 \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 1.96 \frac{\sigma}{\sqrt{n}}\right) = 0.95$$

Conclusão:

$$\text{I.C.A.}_{0.95}(\mu) = \left] \bar{X} - 1.96 \frac{\sigma}{\sqrt{n}} ; \bar{X} + 1.96 \frac{\sigma}{\sqrt{n}} \right[$$

$$\bar{x} = 140 ; \sigma = 10 ; n = 25$$

$$\text{I.C.}_{0.95}(\mu) = \left] 140 - 1.96 \times \frac{10}{5} ; 140 + 1.96 \frac{10}{5} \right[$$

Slides Professora Cláudia Nunes

$$IC_{95\%}(\mu) = (136.08; 143.92)$$

ICs: Método da Variável Fulcral

Em geral, os intervalos de confiança são "derivados" através do **Método da Variável Fulcral**, que é seguinte:

Passo 0: identificar a distribuição de v.c. X de interesse

- (A) {
- identificar qual o parâmetro θ o qual queremos derivar o i.c.
 - quais os parâmetros conhecidos.

Exemplos

- $X \sim \text{env}(\mu, \sigma^2)$
 - (i) I.C. p/ μ , com σ conhecido
 - (ii) I.C. p/ μ , com σ desconhecido
 - (iii) I.C. p/ σ^2 , com μ desconh.
- $X \sim \text{Bernoulli}(p) \rightarrow$ I.C. para p . (iv)

ICs: Método da Variável Fulcral

Passo 1 : identificar a v. fulcral
(usando (A))

I.C. p/ μ , σ
conhecido,
 $X \sim N(\mu, \sigma^2)$

I.C. p/ μ , σ
desconhecido,
 $X \sim N(\mu, \sigma^2)$

idem, n grande

variância
amostral
conhecida

$$\frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0, 1)$$

$$\frac{\bar{X} - \mu}{S / \sqrt{n}} \sim t_{(n-1)}$$

$$\frac{\bar{X} - \mu}{S / \sqrt{n}} \overset{a}{\sim} N(0, 1)$$

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

$$= \left(\frac{1}{n-1} \sum_{i=1}^n X_i \right)^2 - \left(\frac{1}{n-1} \sum_{i=1}^n X_i^2 \right)$$

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

$$\sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_{(k-m-1)}$$

I.C. p/ σ , com
 μ desconhecido
 $X \sim N(\mu, \sigma^2)$

iii

ICs: Método da Variável Fulcral

Passo 2 : encontrar os q-antis
(distribuição de v. fulcral)

$U(0,1)$ $t_{(n-1)}$ $\chi^2_{(n-1)}$
(t de Student) (qui-quadrado)

?

ICs: Método da Variável Fulcral

Passo 3 : inverter as desigualdades
(v. fulcral e dos quantis)

$$I.C.A._{1-\alpha}(\theta)$$

(int. de confiança aleatório, p/ θ , com
confiança $1-\alpha$)

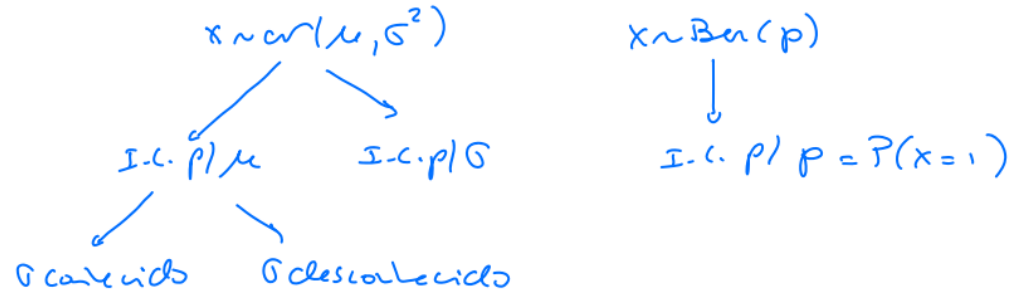
$$I.C._{1-\alpha}(\theta)$$

(int. de confiança, que resulta de concetivação do I.C.A. para uma amostra)

ICs: Método da Variável Fulcral - Resumo...

Regras:

Passo 0: identificar a situação



Passo 1: identificar a v. fulcral

Passo 2: encontrar os quantis tendo em conta a distribuição de v. fulcral e de confiança pretendida.

Passo 3: inverter as desigualdades $[a < v.F < b]$ de forma a colocar em evidência o parâmetro p o qual vamos construir o I.C.

7.2 Admita que a densidade de construção, X , num projecto de urbanização tem distribuição normal. Uma amostra aleatória de 50 lotes desse projecto conduziu a

$$\sum_{i=1}^{50} x_i = 227.2; \sum_{i=1}^{50} x_i^2 = 2242.6$$

~~Assumindo que o desvio padrão de X é igual a 4~~, construa um intervalo de confiança a 95% para a densidade média de construção. Que dimensão deveria ter a amostra para que a amplitude desse intervalo fosse reduzida a metade?

Vamos supor que a variância populacional é desconhecida.



Exercício 7.2: IC para μ

Supondo que a variância populacional é desconhecida, tem-se:

Nota: Saber a variância da amostra não significa que se saiba a variância da população!!

$$s^2 = \frac{1}{50-1} 2242.6 - \frac{50}{50-1} \left(\frac{227.2}{50} \right)^2 = 24.698$$

$$\bar{x} = \frac{227.2}{50} = 4.544$$

Tabela da t-Student

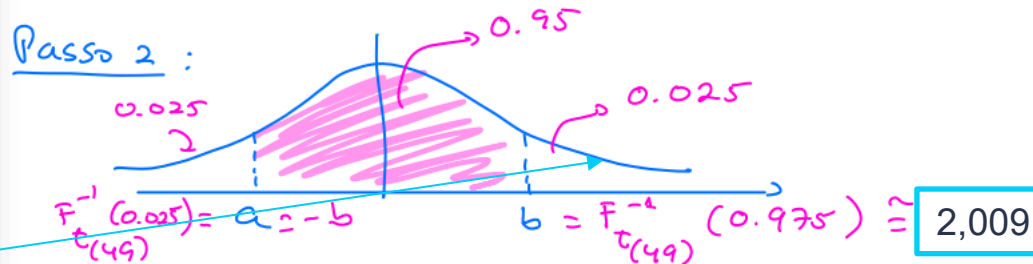
$$t_{n,\varepsilon} : P(X > t_{n,\varepsilon}) = \varepsilon$$

X tem distribuição normal
Supondo-se que a variância populacional é desconhecida, tem-se:

Exercício 7.2: IC para μ

Passo 0: $X \sim \mathcal{N}(\mu, \sigma^2)$ $\mu, \sigma^2 = ?$
 $I.C._{0.95}(\mu) = 2$

Passo 1: $\frac{\bar{X} - \mu}{S/\sqrt{50}} \sim t_{(49)}$



ε	.400	.250	.100	.050	.025	.010	.005	.001
n								
1	.325	1.000	3.078	6.314	12.706	31.821	63.656	318.289
2	.289	.816	1.886	2.920	4.303	6.965	9.925	22.328
3	.277	.765	1.638	2.353	3.182	4.541	5.841	10.214
4	.271	.741	1.533	2.132	2.776	3.747	4.604	7.173
5	.267	.727	1.476	2.015	2.571	3.365	4.032	5.894
6	.265	.718	1.440	1.943	2.447	3.143	3.707	5.208
7	.263	.711	1.415	1.895	2.365	2.998	3.499	4.785
8	.262	.706	1.397	1.860	2.306	2.896	3.355	4.501
9	.261	.703	1.383	1.833	2.262	2.821	3.250	4.297
10	.260	.700	1.372	1.812	2.228	2.764	3.169	4.144
11	.260	.697	1.363	1.796	2.201	2.718	3.106	4.025
12	.259	.695	1.356	1.782	2.179	2.681	3.055	3.930
13	.259	.694	1.350	1.771	2.160	2.650	3.012	3.852
14	.258	.692	1.345	1.761	2.145	2.624	2.977	3.787
15	.258	.691	1.341	1.753	2.131	2.602	2.947	3.733
16	.258	.690	1.337	1.746	2.120	2.583	2.921	3.686
17	.257	.689	1.333	1.740	2.110	2.567	2.898	3.646
18	.257	.688	1.330	1.734	2.101	2.552	2.878	3.610
19	.257	.688	1.328	1.729	2.093	2.539	2.861	3.579
20	.257	.687	1.325	1.725	2.086	2.528	2.845	3.552
21	.257	.686	1.323	1.721	2.080	2.518	2.831	3.527
22	.256	.686	1.321	1.717	2.074	2.508	2.819	3.505
23	.256	.685	1.319	1.714	2.069	2.500	2.807	3.485
24	.256	.685	1.318	1.711	2.064	2.492	2.797	3.467
25	.256	.684	1.316	1.708	2.060	2.485	2.787	3.450
26	.256	.684	1.315	1.706	2.056	2.479	2.779	3.435
27	.256	.684	1.314	1.703	2.052	2.473	2.771	3.421
28	.256	.683	1.313	1.701	2.048	2.467	2.763	3.408
29	.256	.683	1.311	1.699	2.045	2.462	2.756	3.396
30	.256	.683	1.310	1.697	2.042	2.457	2.750	3.385
40	.255	.681	1.303	1.684	2.021	2.423	2.704	3.307
50	.255	.679	1.299	1.676	2.009	2.403	2.678	3.261

$$t_{0.975;49} = t_{0.025;49}^* = 2,009$$

Exercício 7.2: IC para μ

$$\text{Pelo 3} \quad P\left(-2,009 < \frac{\bar{X} - \mu}{s/\sqrt{50}} < 2,009\right) = 0,95$$

$$\bar{X} - 2,009 \frac{s}{\sqrt{50}} < \mu < \bar{X} + 2,009 \frac{s}{\sqrt{50}}$$

$$I.C.A._{0,95}(\mu) = \left] \bar{X} - 2,009 \frac{s}{\sqrt{50}} ; \bar{X} + 2,009 \frac{s}{\sqrt{50}} \right[$$

Concretizando p/ a amostra em causa:

$$\bar{x} = 4,544 \quad s^2 = 24,698$$

$$I.C._{0,95}(\mu) = \left] 4,544 - 2,009 \sqrt{\frac{24,698}{50}} ; \right.$$

$$4,544 + 2,009 \sqrt{\frac{24,698}{50}} \left[, \right.$$

Supondo que a variância populacional é desconhecida, tem-se:

$$IC_{95\%}(\mu): (3,132; 5,956)$$

Exercício 7.2: Amplitude Amostral

Supondo variância populacional desconhecida, tem-se:

$IC_{95\%}(\mu): (3,132; 5,956)$

Amplitude do IC para $\mu = 5,956 - 3,132 = 2,824$

Amplitude do IC para $\mu: 2 \times t_{1-\alpha/2; n-1} (s' / n^{1/2})$

n?

$$2 \times t_{1-\alpha/2; n-1} (s' / n^{1/2}) = 2 \times 2,009 \times 4,970 / n^{1/2} \leq 2,824/2 \Leftrightarrow n \geq 200$$

Pelo menos $n = 200$

30. Com base numa amostra casual com 16 observações, retirada de uma população normal, construiu-se, segundo o processo habitual, o seguinte intervalo de confiança para o valor esperado: (7.398, 12.602).
- a) Sabendo que, com a informação da amostra, se obteve $s = 3.872$, qual o grau de confiança que pode atribuir ao intervalo atrás referido?
 - b) Com base na mesma amostra, construa um intervalo de confiança a 95% para a variância da população.
 - c) Suponha que a verdadeira variância é 36. Se pretender construir um intervalo de confiança a 95% para a média, de modo que a respectiva amplitude não exceda 6.5, qual é a dimensão mínima da amostra a recolher?



Exercício 30 a)

$$X \sim N(\mu, \sigma^2) \quad m = 16 \text{ amostra casual}$$

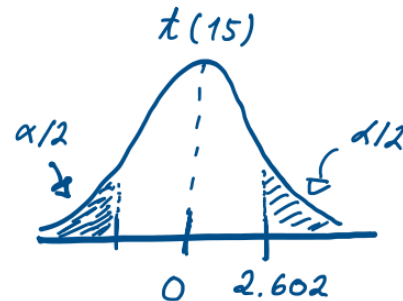
$$I.C._{1-\alpha}(\mu) = (7.398; 12.602)$$

$$a) \hat{s} = 3.872 \quad \hat{s}'^2 = \frac{m}{m-1} \hat{s}^2$$

$$\hat{s}' = \sqrt{\frac{m}{m-1}} \hat{s} = \sqrt{\frac{16}{15}} \times 3.872 \approx 4$$

$$\frac{\bar{X} - \mu}{\hat{s}'/\sqrt{m}} \sim t(m-1) = t(15)$$

$$I.C._{1-\alpha}(\mu) = \bar{x} \pm t_{m-1, \alpha/2} \frac{\hat{s}'}{\sqrt{m}}$$



Exercício 30 a)

$$\text{Amplitude do I.C.} = 2 t_{m-1, \alpha/2} \frac{s'}{\sqrt{m}} = 2 t_{15, \alpha/2} \times \frac{4}{\sqrt{16}}$$

$$= 2 t_{15, \alpha/2} = 12.602 - 7.398 \quad (=)$$

$$(\Rightarrow) t_{15, \alpha/2} = 2.602 \quad \Rightarrow \frac{\alpha}{2} = 0.01 \quad (=)$$

Tabela $\Rightarrow$ $(\Rightarrow) \alpha = 0.02$

$$\text{Grau de confiança} = (1 - 0.02) 100\% = 98\%$$

Exercício 30 c)

$$c) \sigma^2 = 36$$

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

$$IC_{95\%}(\mu) = \bar{x} \pm z_{0.025} \frac{\sigma}{\sqrt{n}} = \bar{x} \pm 1.96 \times \frac{6}{\sqrt{n}}$$

$$= \bar{x} \pm \frac{11.76}{\sqrt{n}}$$

$$\text{Amplitude} = 2 \times \frac{11.76}{\sqrt{n}} = \frac{23.52}{\sqrt{n}} \leq 6.5 \quad (=)$$

$$(=) \sqrt{n} \geq \frac{23.52}{6.5} \quad (=) n \geq \left(\frac{23.52}{6.5} \right)^2 \simeq 13.09$$

Resposta: $n = 14$ é a dimensão mínima.

31. Considere uma população com distribuição normal de parâmetros desconhecidos. Dessa população foi retirada uma amostra casual de dimensão 25. Suponha que a amostra forneceu os seguintes resultados:

$$\sum_{i=1}^{25} x_i = 75 \text{ e } \sum_{i=1}^{25} x_i^2 = 321.$$

- a) Construa um intervalo de confiança a 95% para a média.
- b) Construa um intervalo de confiança a 95% para o desvio padrão.



32. Seja X uma população normal com desvio padrão 1.5. Com base numa amostra casual de dimensão 25 seleccionada dessa população, construiu-se, pelo processo habitual, o seguinte intervalo de confiança para a média:

$$(\bar{x} - 0.6162, \bar{x} + 0.6162).$$

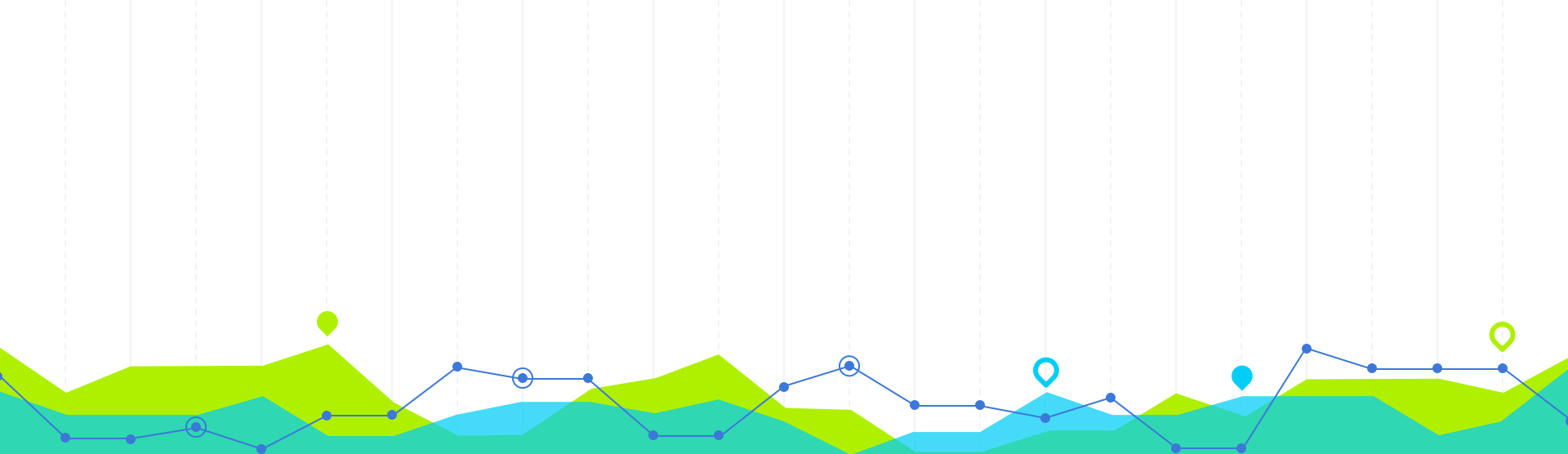
a) Qual o grau de confiança deste intervalo?

b) Qual deve ser a dimensão da amostra para diminuir para metade a amplitude do intervalo, para o mesmo grau de confiança?



33. Pode considerar-se que o grau de acidez do azeite de determinada marca é uma variável aleatória com distribuição normal.
- a) Analisada uma amostra de 16 garrafas, obteve-se para o grau de acidez uma média de 1.3 graus e uma variância corrigida de 0.16. Sabendo que essa marca garante, nos rótulos das embalagens do seu produto, um grau de acidez médio não superior a 1, diga, com base num intervalo de confiança a 95%, se pode acusar a empresa que comercializa a marca de publicidade enganosa?
 - b) Supondo que o desvio padrão da população é igual a 0.4 graus, qual deve ser a dimensão da amostra a recolher de forma a garantir, com confiança de 99%, um desvio máximo de ± 0.3 graus entre a média observada e o seu verdadeiro valor?





Intervalo de Confiança para a Diferença de Valores Médios $\mu_1 - \mu_2$

5

Intervalo de Confiança para $\mu_1 - \mu_2$: Variâncias Conhecidas

Portanto, quando as populações são Normais com variâncias conhecidas, o I. C. para $\mu_1 - \mu_2$ com $100(1 - \alpha)\%$ de confiança é dado por:

$$\left[(\bar{X}_1 - \bar{X}_2) - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}; (\bar{X}_1 - \bar{X}_2) + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right]$$

ProbabilidadesEstatistica_2019 (uevora.pt)

Intervalo de Confiança para $\mu_1 - \mu_2$: Variâncias Desconhecidas e Iguais

Portanto, quando as populações são Normais com variâncias desconhecidas, mas iguais, o I. C. para $\mu_1 - \mu_2$ com $100(1 - \alpha)\%$ de confiança é dado por:

$$\left[(\bar{X}_1 - \bar{X}_2) - t_{n_1+n_2-2; 1-\frac{\alpha}{2}} S^*; (\bar{X}_1 - \bar{X}_2) + t_{n_1+n_2-2; 1-\frac{\alpha}{2}} S^* \right].$$

Probabilidades Estatística 2019 (uevora.pt)

$$S^* = \sqrt{\frac{(n_1 - 1) S_1'^2 + (n_2 - 1) S_2'^2}{n_1 + n_2 - 2}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Variâncias corrigidas

$$S'^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

Intervalo de Confiança para $\mu_1 - \mu_2$: Variâncias Desconhecidas e Diferentes

Portanto, quando **as populações são Normais com variâncias desconhecidas,** , o I. C. para $\mu_1 - \mu_2$ com $100(1 - \alpha)\%$ de confiança é dado por:

$$\left[(\bar{X}_1 - \bar{X}_2) - t_{v; 1-\frac{\alpha}{2}} \sqrt{\frac{S_1'^2}{n_1} + \frac{S_2'^2}{n_2}}, (\bar{X}_1 - \bar{X}_2) + t_{v; 1-\frac{\alpha}{2}} \sqrt{\frac{S_1'^2}{n_1} + \frac{S_2'^2}{n_2}} \right]$$

ProbabilidadesEstatistica 2019 (uevora.pt)

Variâncias corrigidas

$$v = \left[\frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} \right)^2}{\frac{1}{n_1 - 1} \left(\frac{S_1^2}{n_1} \right)^2 + \frac{1}{n_2 - 1} \left(\frac{S_2^2}{n_2} \right)^2} \right]$$

Variâncias corrigidas

$$S'^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

Intervalo de Confiança para $\mu_1 - \mu_2$ e $n \geq 30$: Variâncias Conhecidas e Desconhecidas

- **Parâmetro:** $\mu_1 - \mu_2$

(populações quaisquer independentes com variâncias finitas)

V. F.

Variâncias Conhecidas

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \stackrel{a}{\sim} N(0,1)$$

ou

Variâncias Desconhecidas

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1'^2}{m} + \frac{s_2'^2}{n}}} \stackrel{a}{\sim} N(0,1).$$

I. C.

$$(\bar{x}_1 - \bar{x}_2 - z_{\alpha/2} \sigma^*, \bar{x}_1 - \bar{x}_2 + z_{\alpha/2} \sigma^*)$$

ou

$$(\bar{x}_1 - \bar{x}_2 - z_{\alpha/2} s^*, \bar{x}_1 - \bar{x}_2 + z_{\alpha/2} s^*),$$

com $\sigma^* = \sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}$

$$s^* = \sqrt{\frac{s_1'^2}{m} + \frac{s_2'^2}{n}}$$

e $z_{\alpha/2}$ a verificar $\Phi(z_{\alpha/2}) = 1 - \alpha/2$.

Murteira et al (2015)

IC para $\mu_1 - \mu_2$: Formulário

Variação corrigida

$$S'^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

• POPULAÇÕES NORMAIS

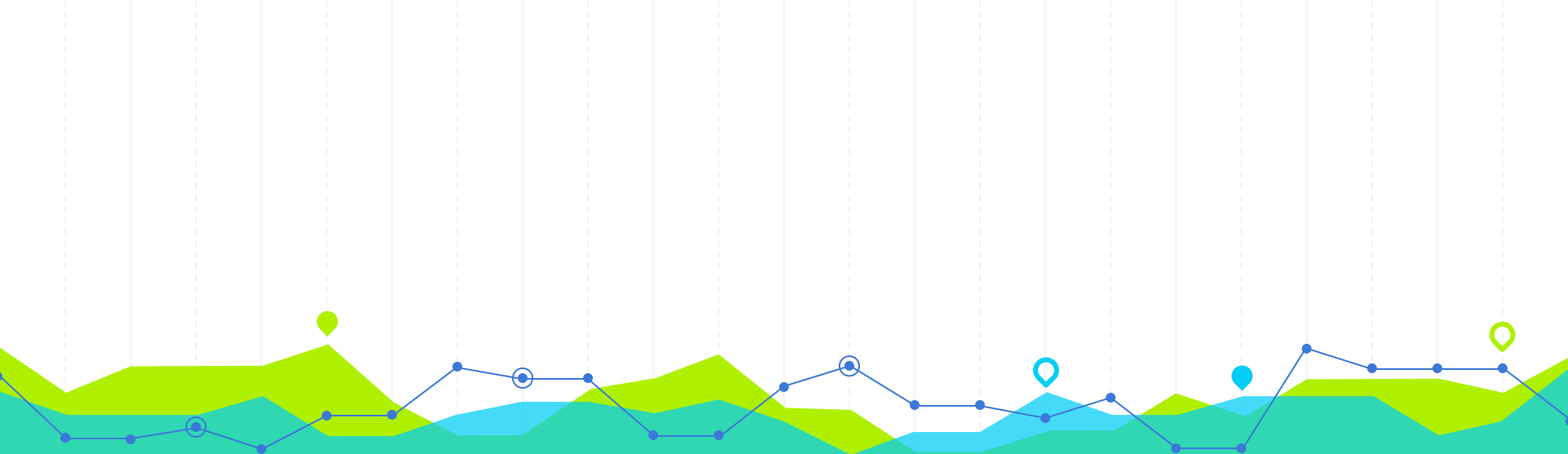
Média	$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$	$\frac{\bar{X} - \mu}{S'/\sqrt{n}} \sim t(n-1)$
Diferença de médias	$\frac{(X_1 - X_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \sim N(0,1)$ Variações Conhecidas	$Z = \frac{(X_1 - X_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1'^2}{m} + \frac{S_2'^2}{n}}} \sim t(v)$ onde v é o maior inteiro contido em r, $r = \frac{\left(\frac{S_1'^2}{m} + \frac{S_2'^2}{n}\right)^2}{\frac{1}{m-1} \left(\frac{S_1'^2}{m}\right)^2 + \frac{1}{n-1} \left(\frac{S_2'^2}{n}\right)^2}$
	Variações Desconhecidas e Iguais $T = \frac{\frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{1}{m} + \frac{1}{n}}}}{\sqrt{\frac{(m-1)S_1'^2 + (n-1)S_2'^2}{m+n-2}}} \sim t(m+n-2)$	
Variação	$\frac{nS^2}{\sigma^2} = \frac{(n-1)S'^2}{\sigma^2} \sim \chi^2(n-1)$	Variações Desconhecidas e Diferentes
Relação de variações	$\frac{S_1'^2}{S_2'^2} \frac{\sigma_2^2}{\sigma_1^2} \sim F(m-1, n-1)$	

IC para $\mu_1 - \mu_2$: Formulário

- GRANDES AMOSTRAS

Caso geral

Média	$\frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \stackrel{a}{\sim} N(0,1)$	$\frac{\bar{X} - \mu}{S' / \sqrt{n}} \stackrel{a}{\sim} N(0,1)$
Diferença de médias	$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \stackrel{a}{\sim} N(0,1)$	$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1'^2}{m} + \frac{S_2'^2}{n}}} \stackrel{a}{\sim} N(0,1)$
Variâncias Conhecidas		Variâncias Desconhecidas



Intervalo de Confiança para a Diferença de Valores Médios $\mu_1 - \mu_2$: Exercícios

6

Havendo indícios de que o esquema de avaliação e as classificações finais atribuídas diferem fortemente entre duas escolas, decidiu-se comprovar estatisticamente esta hipótese. Os desvios-padrão são conhecidos sendo 2,1 valores na escola A e 1,8 valores na escola B. Assim, retirou-se uma amostra de testes de alunos em cada uma das escolas que levaram aos seguintes resultados:

Escola	n_i	$\bar{x}_i$
A	41	12,9
B	31	14,7

Recorrendo a um intervalo de confiança a 90%, diga se há diferenças entre as classificações médias das escolas A e B. Justifique.



Obrigada!

Questões?

