ESTIMAÇÃO

Bibliografia: BUSSAB, Wilton de O.; MORETTIN, Pedro A. **Estatística básica**. Saraiva, 2010, capítulo 11.

ESTIMAÇÃO

- Inferência estatística: generalização de afirmações para a população a partir de amostras

- Problemas de inferência:
  - estimação de parâmetros
  - testes de hipóteses

ESTIMADOR

- É uma regra para calcular a estimativa de um dado **parâmetro** com base em dados observacionais
  - Regra (estimador), quantidade de interesse (estimando), resultado (estimativa)

- **Estimativa pontual**: fornece o resultado na forma de um valor único

- **Estimativa intervalar**: fornece o resultado na forma de uma faixa de valores plausíveis

ESTIMADOR

- Um estimador pontual, assim como sua estimativa, é uma estatística, isto é, uma função dos dados obtidos a partir da amostra

- Se $(x_1, \dots, x_n)$ é uma amostra, $T = f(x_1, \dots, x_n)$ é uma estatística

- Exemplos:

$$T^1 = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad\qquad T^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

ESTIMADOR

- O estimador de um parâmetro θ é representado por $\hat{\theta}$, de modo que:

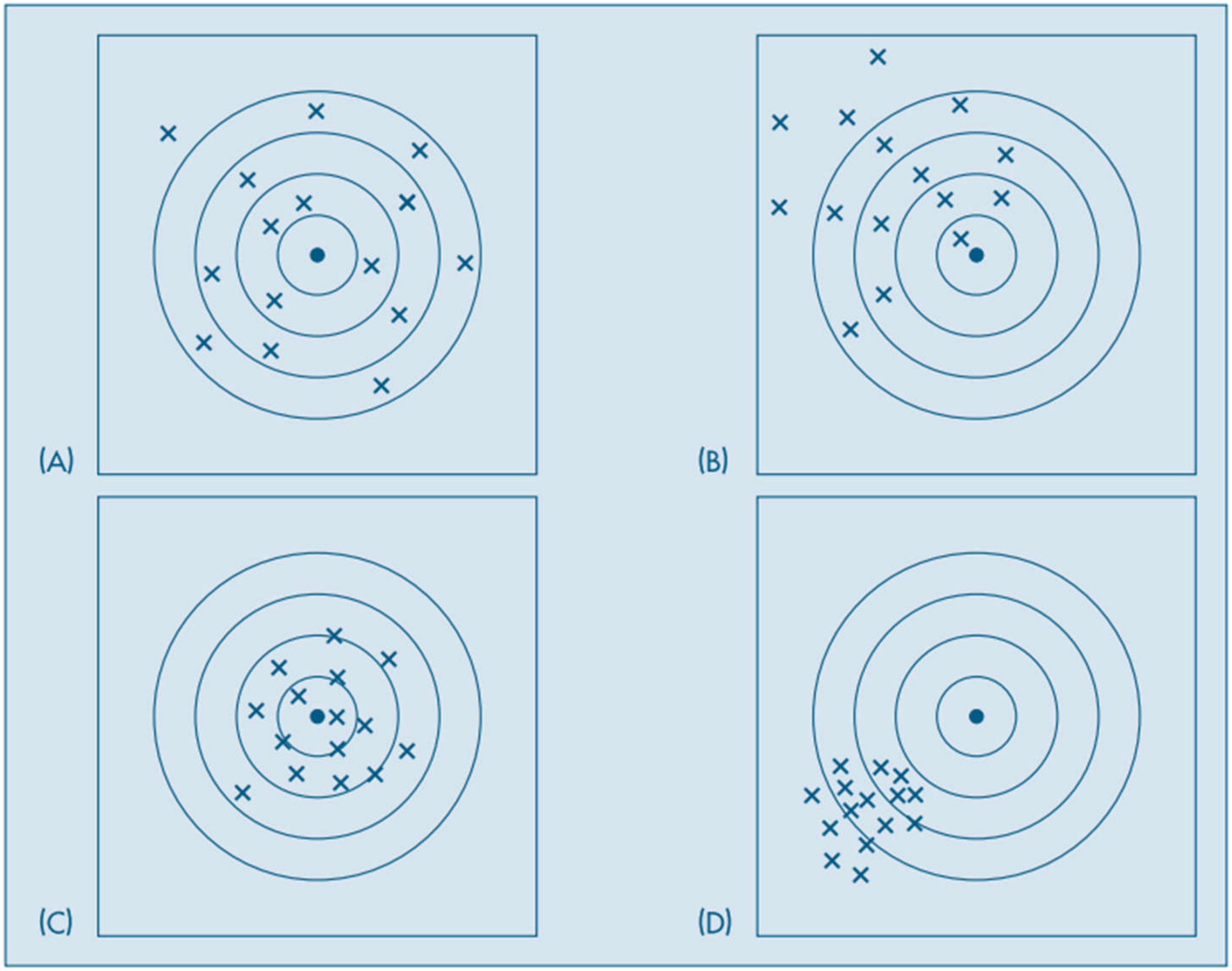
$$\hat{\theta} = T = f(x_1, \ldots, x_n)$$

- O valor de T observado na amostra (t_{obs}) é a estimativa de θ

- Se tivermos mais de um estimador disponível para um dado problema de estimação, como escolher o melhor deles?

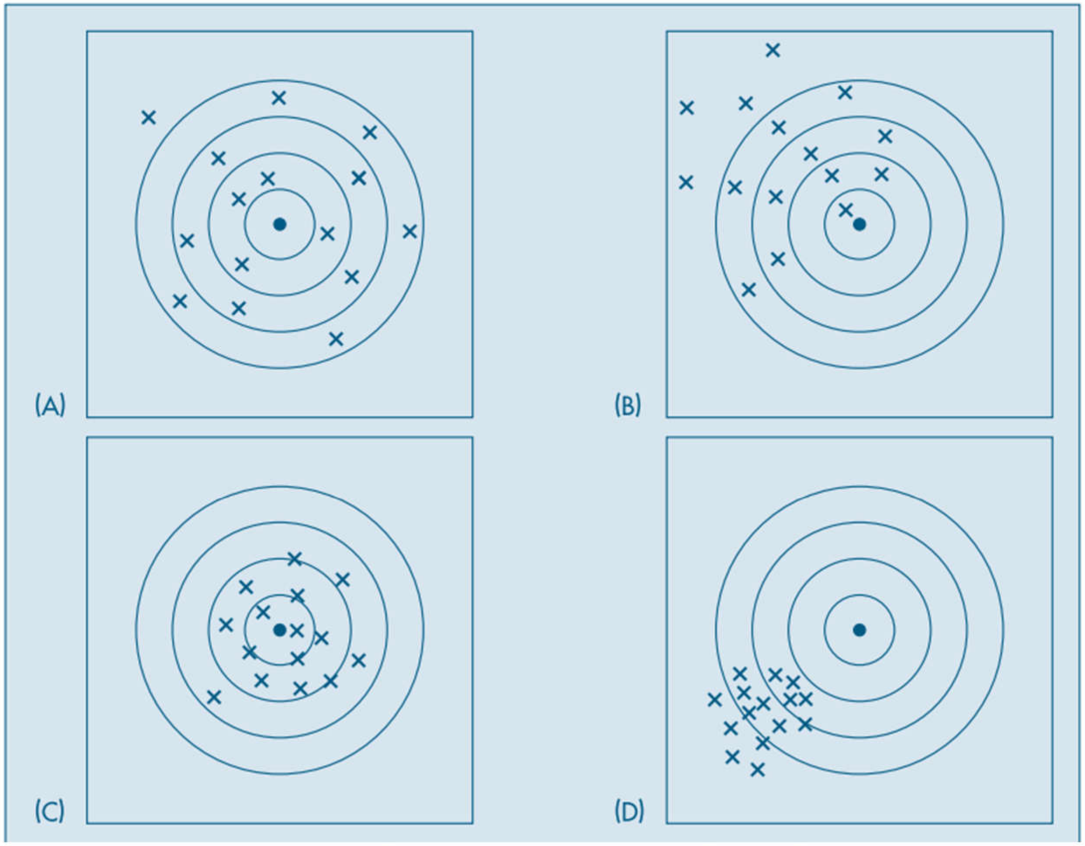
ESTIMAÇÃO

Figura 11.1: Resultados de 15 tiros dados por 4 rifles.



ESTIMAÇÃO

Figura 11.1: Resultados de 15 tiros dados por 4 rifles.



Arma A: não-viesada, pouco acurada e baixa precisão.
Arma B: viesada, pouco acurada e baixa precisão.
Arma C: não-viesada, muito acurada e boa precisão.
Arma D: viesada, pouco acurada e alta precisão.

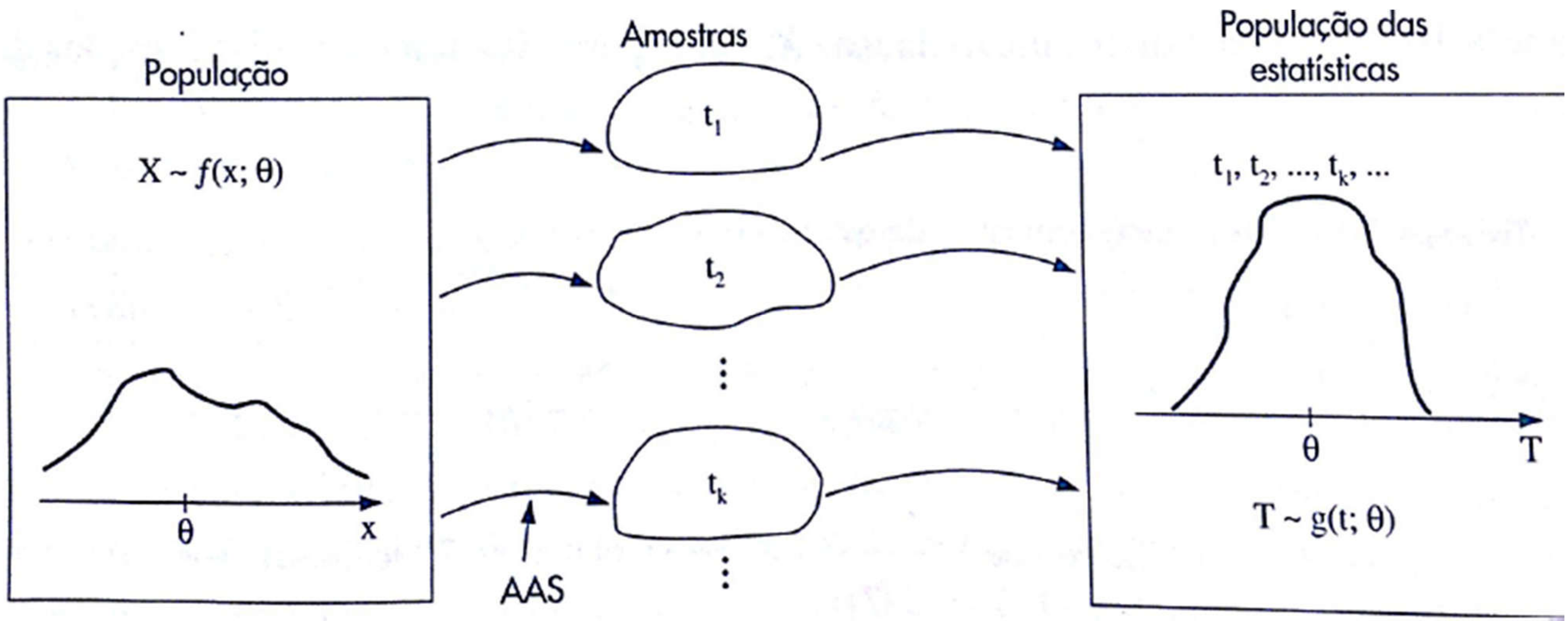
Seleção de estimadores

- Critérios de atratividade de estimadores
  - Ausência de viés
  - Eficiência
  - Consistência

ESTIMADOR

- Note que como a obtenção de uma amostra da população é um experimento, a estatística T é uma variável aleatória, uma vez que o experimento pode produzir diferentes resultados se repetido diversas vezes

INFERÊNCIA ESTATÍSTICA



ESTIMADOR

- A distribuição de probabilidade de um estimador é chamada **distribuição de amostragem**

- Consequentemente, é possível caracterizar a distribuição de amostragem do estimador usando medidas descritivas:
  - Tendência central: $E(T)$
  - Variabilidade: $Var(T)$

SELEÇÃO DE ESTIMADORES

- Critérios de atratividade de estimadores
  - Ausência de viés
  - Eficiência
  - Consistência

Propriedades dos Estimadores

Viés

- Distância entre a média do conjunto de estimativas e o verdadeiro valor do parâmetro sendo estimado

$$V(T) = E(T) - \theta$$

- Um estimador é dito não enviesado, se e somente se, $V(T) = 0$, ou seja:

$$E(T) = \theta$$

Não confundir erro na estimativa com viés do estimador

PROPRIEDADES DOS ESTIMADORES

Eficiência

- Se T e T' são dois estimadores não enviesados de um mesmo parâmetro θ e:

$$Var(T) < Var(T')$$

- Então diz-se que T é mais eficiente que T'

PROPRIEDADES DOS ESTIMADORES

Eficiência

- Entre os possíveis estimadores não enviesados de um parâmetro, frequentemente há um que apresenta a **menor variância**, chamado de:
  - Melhor estimador não enviesado ou
  - Estimador não enviesado de mínima variância

PROPRIEDADES DOS ESTIMADORES

Consistência

- Uma sequência $\{T_n\}$ de estimadores é consistente se:

$$\lim_{n \to \infty} E(T_n) = \theta$$

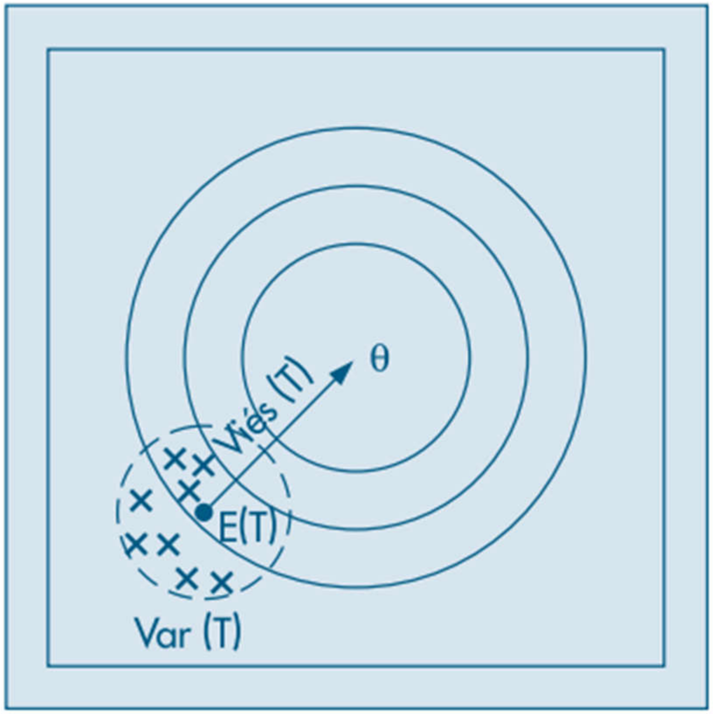
$$\lim_{n \to \infty} Var(T_n) = 0$$

Erro quadrático médio do estimador T

$$EQM(T;\theta) = Var(T) + V^2$$

Exercícios 1 a 5 do capítulo 11

Figura 11.2: Representação gráfica para o EQM.



APLICAÇÕES

Um professor deseja estimar a competência adquirida por cada aluno na disciplina que ele ministra.

- Suponha que, ao final da disciplina, a competência adquirida por um aluno de siga uma distribuição de probabilidade desconhecida, com parâmetros μ e σ^2
- O professor vai usar uma estimativa de μ para decidir se aprova ou não o aluno

APLICAÇÕES

- As opções de estimadores que o professor está considerando são:

$$T^1 = \sum_{i=1}^{3} \frac{X_i}{3} \qquad T^2 = \frac{X_1 + 2X_2 + 3X_3}{3} \qquad T^3 = \frac{2X_1 + 3X_2 + 4X_3}{10}$$

- Avalie se os estimadores são livres de viés (caso algum não seja, calcule o viés)
- Determine a variância e o erro quadrático médio dos estimadores

UM ESTIMADOR PARA σ^2

- Considere uma população com N elementos e a variância populacional σ^2
- Um possível estimador para σ^2, baseado numa AAS de tamanho n extraída dessa população é:

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

- Entretanto, é possível demonstrar que

$$E(\hat{\sigma}^2) = \left(\frac{n-1}{n}\right)\sigma^2$$

- E que

$$V(\hat{\sigma}^2) = \frac{-\sigma^2}{n}$$

UM ESTIMADOR PARA σ^2

- Podemos obter um estimador não enviesado para σ^2 tomando

$$T = \left(\frac{n}{n-1}\right)\hat{\sigma}^2 = \left(\frac{n}{n-1}\right)\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

- De modo que

$$E(T) = E\left[\left(\frac{n}{n-1}\right)\hat{\sigma}^2\right] = \left(\frac{n}{n-1}\right)\left(\frac{n-1}{n}\right)\sigma^2 = \sigma^2$$

- Esse estimador T é denominado S^2

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

ESTIMAÇÃO POR INTERVALO DE CONFIANÇA

- Estimação por ponto não permite avaliar a magnitude do erro cometido
- Estimadores de intervalos de confiança são baseados na distribuição amostral do estimador pontual
- Se quisermos estimar a média $\boldsymbol{\mu}$ de uma população qualquer, e para isso vamos usar o **estimador $\boldsymbol{\bar{X}}$**, calculado para uma amostra com **tamanho $\boldsymbol{n}$**
- Pelo Teorema do Limite Central, se n for suficientemente grande:

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$$

ESTIMAÇÃO POR INTERVALO DE CONFIANÇA

- Consequentemente, o erro na estimativa, definido como:

$$e = (\bar{X} - \mu) \sim N(0, \frac{\sigma^2}{n})$$

- Assim, podemos calcular a probabilidade de cometermos erros de determinadas magnitudes. Por exemplo:

$$P\left(|e| < 1{,}96\frac{\sigma}{\sqrt{n}}\right) = 0{,}95$$

ESTIMAÇÃO POR INTERVALO DE CONFIANÇA

$$P\left(|e| < 1{,}96\frac{\sigma}{\sqrt{n}}\right) = 0{,}95$$

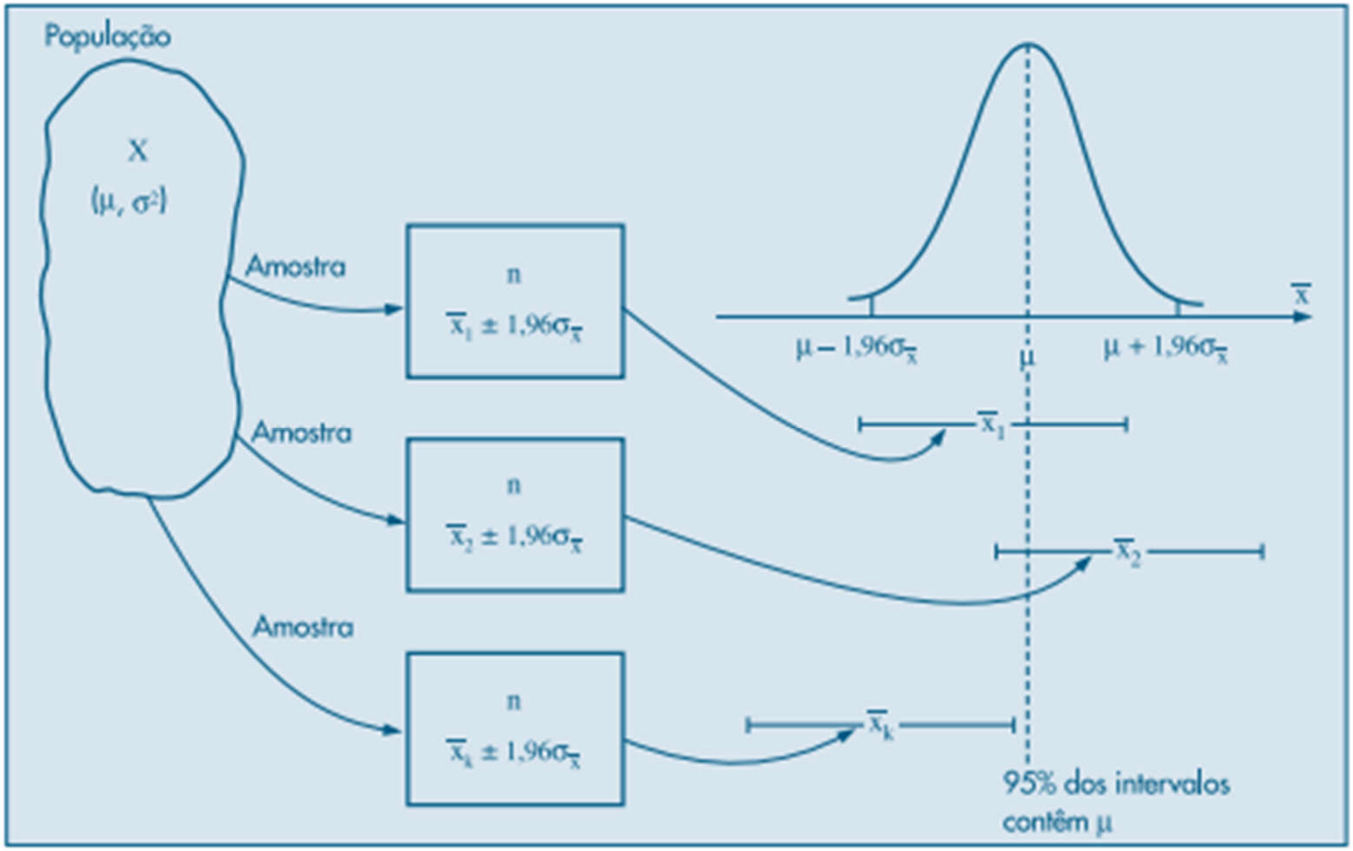
$$P\left(|\bar{X} - \mu| < 1{,}96\frac{\sigma}{\sqrt{n}}\right) = 0{,}95$$

$$P\left(-1{,}96\frac{\sigma}{\sqrt{n}} < \bar{X} - \mu < 1{,}96\frac{\sigma}{\sqrt{n}}\right) = 0{,}95$$

$$P\left(\bar{X} - 1{,}96\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 1{,}96\frac{\sigma}{\sqrt{n}}\right) = 0{,}95$$

- Intervalo de confiança, a 95%, para a média populacional:

$$IC_{95}: \mu = \bar{X} \pm 1{,}96\frac{\sigma}{\sqrt{n}}$$



Estimação por Intervalo de Confiança

Caso em que $\boldsymbol{X \sim N}$ com μ desconhecido e $\boldsymbol{\sigma^2}$ **conhecido**. Então, para **qualquer tamanho de amostra, $\boldsymbol{n}$**:

- Intervalo de confiança, a 95%, para a média populacional:

$$IC_{95}(\mu) = \bar{X} \pm 1{,}96 \frac{\sigma}{\sqrt{n}}$$

- Intervalo de confiança para a média populacional, para um coeficiente de confiança qualquer γ:

$$IC(\mu; \gamma) = \bar{X} \pm z(\gamma) \frac{\sigma}{\sqrt{n}}$$

ESTIMAÇÃO POR INTERVALO DE CONFIANÇA

Caso em que $\boldsymbol{X \sim N}$ com μ desconhecido e $\boldsymbol{\sigma^2}$ **desconhecido**. Então:

- **Se $\boldsymbol{n > 120}$**:
  - Intervalo de confiança, a 95%, para a média populacional:

$$IC_{95}(\mu) = \bar{X} \pm 1{,}96\frac{S}{\sqrt{n}}$$

  - Intervalo de confiança para a média populacional, para um coeficiente de confiança qualquer γ:

$$IC(\mu;\gamma) = \bar{X} \pm z(\gamma)\frac{S}{\sqrt{n}}$$

Estimação por intervalo de confiança

Caso em que $\boldsymbol{X \sim N}$ com μ desconhecido e $\boldsymbol{\sigma^2}$ **desconhecido**. Então:

- **Se $\boldsymbol{n < 120}$**:
  - Intervalo de confiança para a média populacional, para um coeficiente de confiança qualquer γ:

$$IC(\mu;\gamma) = \bar{X} \pm t(\gamma)\frac{S}{\sqrt{n}}$$

ESTIMAÇÃO POR INTERVALO DE CONFIANÇA

Caso em que $\boldsymbol{X \sim f}$ com μ desconhecido e σ^2 desconhecido. Então:

- **Se $\boldsymbol{30 < n < 120}$**:
  - Intervalo de confiança para a média populacional, para um coeficiente de confiança qualquer γ:

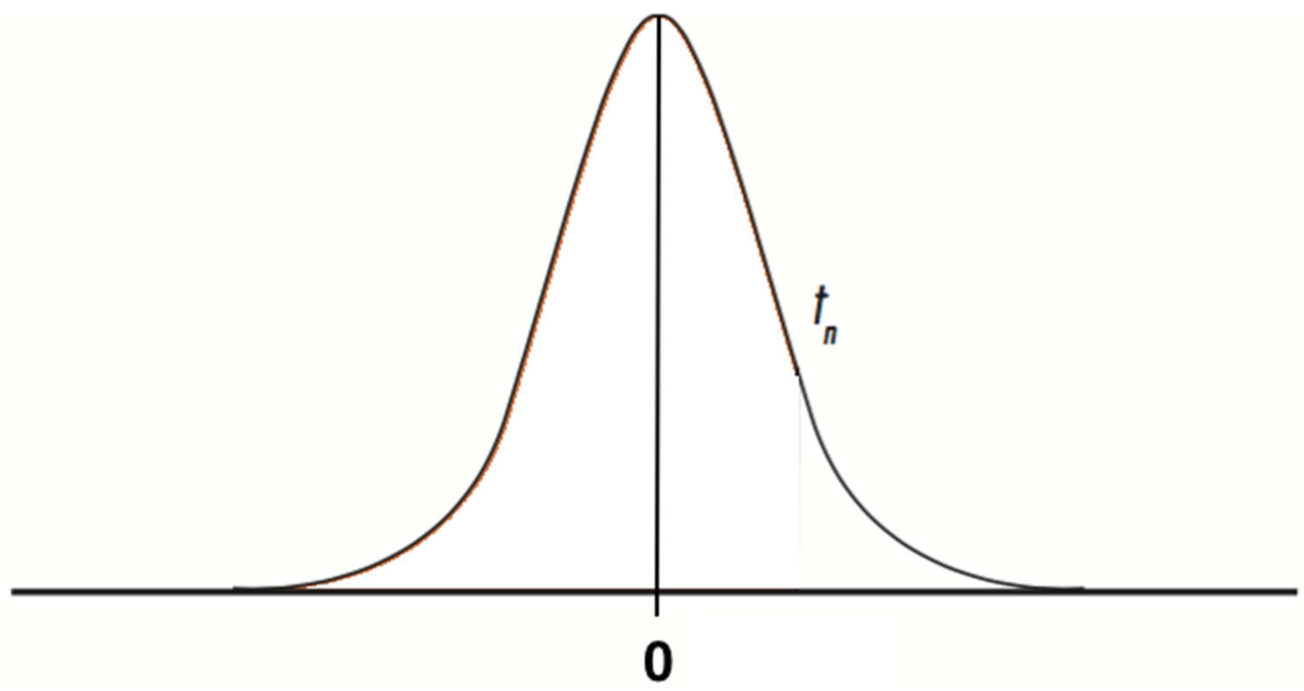
$$IC(\mu;\gamma) = \bar{X} \pm t(\gamma)\frac{S}{\sqrt{n}}$$

- **Se $\boldsymbol{n > 120}$**:
  - Intervalo de confiança para a média populacional, para um coeficiente de confiança qualquer γ:

$$IC(\mu;\gamma) = \bar{X} \pm z(\gamma)\frac{S}{\sqrt{n}}$$

DISTRIBUIÇÃO t DE STUDENT

- A distribuição t de Student

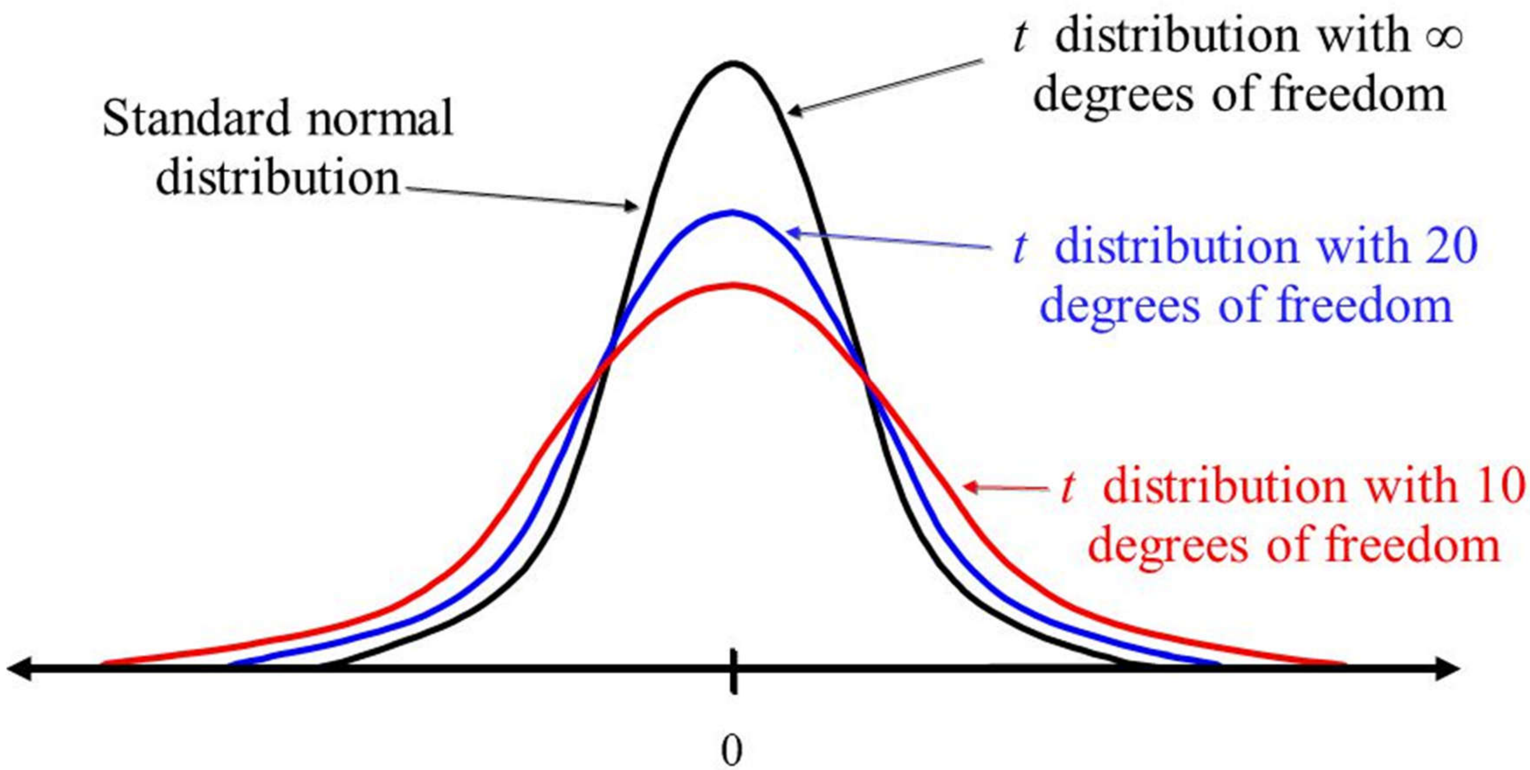


DISTRIBUIÇÃO t DE STUDENT

- Membro de uma família de distribuições contínuas que surge quando **estimamos a média de uma população normalmente distribuída**, em situações nas quais a **amostra é pequena** e o **desvio padrão populacional é desconhecido**
- A distribuição t:
  - Descreve as amostras extraídas dessa população
  - É diferente para cada tamanho de amostra
  - Quanto maior o tamanho da amostra mais ela se aproxima da Normal

DISTRIBUIÇÃO t DE STUDENT

- Quanto maior o tamanho da amostra mais ela se aproxima da Normal

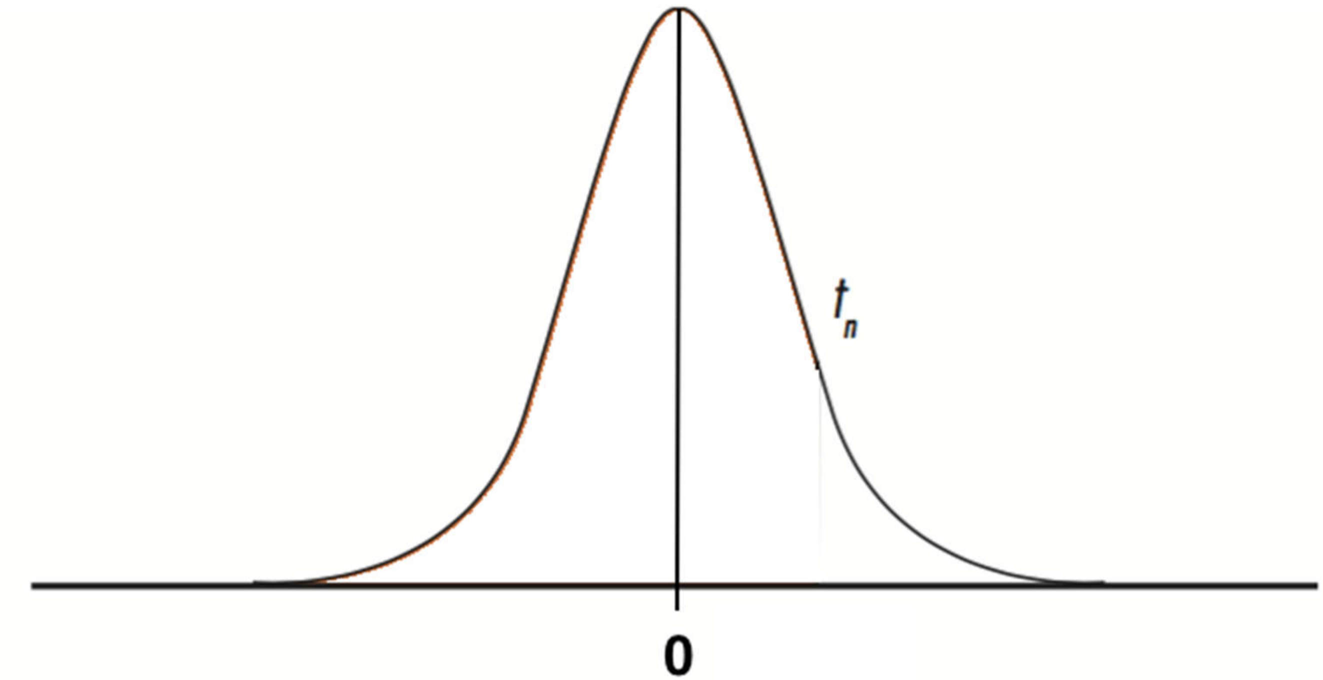


DISTRIBUIÇÃO t DE STUDENT

- A distribuição t é usada em várias análises estatísticas, incluindo o teste de médias, a construção de intervalos de confiança e na análise de regressão
- Se tomarmos uma amostra de n observações de uma distribuição normal, então a distribuição t com $\nu = n - 1$ graus de liberdade pode ser definida como a distribuição da localização da média amostral, relativa à média populacional, dividida pelo desvio padrão, e depois multiplicada pelo termo normalizador $\sqrt{n}$

DISTRIBUIÇÃO t DE STUDENT

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$



INTERVALO DE CONFIANÇA PARA A PROPORÇÃO AMOSTRAL

- IC conservador:

$$\hat{p} - \frac{z(\gamma)}{\sqrt{4n}} \leq p \leq \hat{p} + \frac{z(\gamma)}{\sqrt{4n}}$$

- Se $n > 100$, pode-se usar uma fórmula alternativa:

$$\hat{p} - z(\gamma)\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z(\gamma)\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

EXERCÍCIOS

Exercícios capítulo 11:

- 14 a 25; 27 a 30; 32; 48.