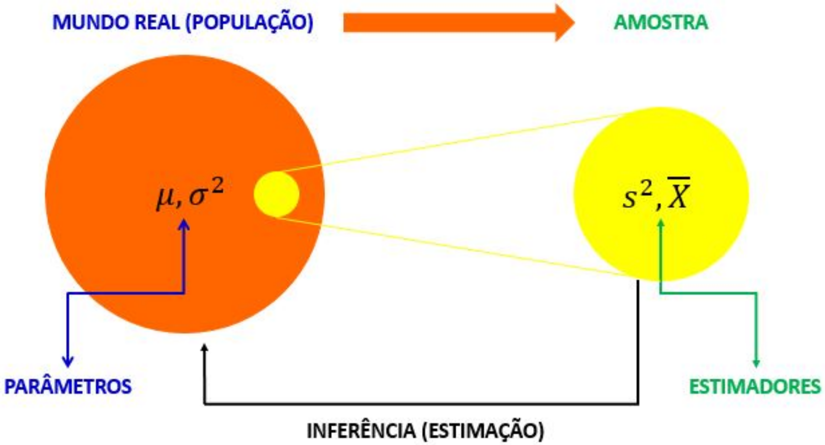
# Estatística

Professora: Patrícia Ferreira Paranaíba

## Introdução à Inferência Estatística

- Inferência estatística: é o processo pelo qual pode-se tirar conclusões acerca de um conjunto maior (a população) usando informação de um conjunto menor (a amostra). O objetivo principal da Inferência Estatística, é estimar os parâmetros populacionais (média, variância, etc), deduzidos a partir da estatística amostral correspondente.


MUNDO REAL (POPULAÇÃO)
AMOSTRA
$\mu, \sigma^2$
$s^2, \overline{X}$
PARÂMETROS
ESTIMADORES
INFERÊNCIA (ESTIMAÇÃO)

## Tipos de inferência

- Estimação pontual: o objetivo é encontrar os valores do parâmetro desconhecido;
- Estimação por intervalos: o objetivo é encontrar um intervalo que contenha o parâmetro de interesse com uma probabilidade especificada;
- Testes de hipóteses: o objetivo é criar conjecturas sobre os valores possíveis do parâmetro e verificar se, estas conjecturas, são muito ou pouco prováveis (isto é, testar as hipóteses).

## Estimação Intervalar

- Uma outra maneira de se calcular um estimativa de um parâmetro desconhecido, é construir um intervalo de confiança $[a, b]$ para esse parâmetro com uma probabilidade de $1 - \alpha$ (nível de confiança) de que o intervalo contenha o verdadeiro parâmetro ($\theta$). Podemos obter expressões do tipo:

$$P(a \leq \theta \leq b) = 1 - \alpha$$

- Dessa maneira $\alpha$ será o nível de significância, isto é, o erro que se estará cometendo ao afirmar que o parâmetro está entre o limite inferior e o superior calculado.

## Intervalo de Confiança para média $\mu$ com variância $\sigma^2$ conhecida

- Um intervalo de $(1-\alpha)$ de confiança para $\mu$ será dado por:

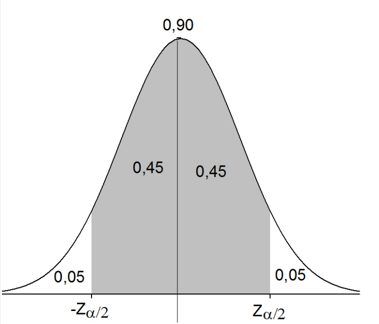
$$P\left(\overline{X} - Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

- Exemplo: Um pesquisador obteve a partir de uma amostra uma média $\overline{X} = 180cm$ para altura de uma determinado grupo de pessoas utilizando uma amostra n=40, sabe-se que a variância populacional da altura é de $\sigma^2 = 100cm^2$. Qual o intervalo de confiança a $90\%$ para a média populacional.

- Primeiramente temos que obter o valor tabelado de Z. Como queremos o intervalo de confiança a 90%, temos que:

$$1 - \alpha = 0,90 \Rightarrow \alpha = 1 - 0,90 = 0,10 \Rightarrow \frac{\alpha}{2} = \frac{0,10}{2} = 0,05$$

- Assim, temos que procurar na tabela qual o valor de Z que deixa $0,05$ de probabilidade acima dele.



- Olhando na tabela o valor em que $P(0 < Z < z) = 0,45$, temos que $z = 1,65$, logo o valor $Z_{\frac{\alpha}{2}} = 1,65$

$$P\left(\overline{X} - Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$
$$P\left(180 - 1,65\frac{\sqrt{200}}{\sqrt{40}} \leq \mu \leq 180 + 1,65\frac{\sqrt{200}}{\sqrt{40}}\right) = 0,90$$
$$P\left(176,31 \leq \mu \leq 183,69\right) = 0,90$$

Logo, o intervalo de confiança a $90\%$ para a média é:

$$IC_{90\%}(\mu) = [176,31;\ 183,69]$$

## Intervalo de Confiança para média $\mu$ com variância $\sigma^2$ desconhecida

- Um intervalo de $(1-\alpha)$ de confiança para $\mu$ será dado por:

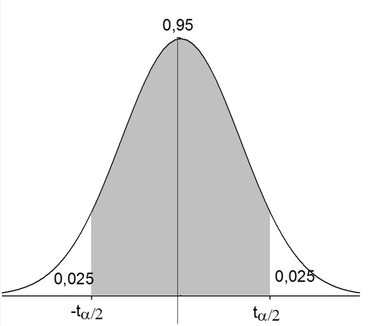
$$P\left(\overline{X} - t_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}} \leq \mu \leq \overline{X} + t_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}}\right) = 1-\alpha$$

- Exemplo: Em uma determinada indústria para verificar a qualidade dos rolamentos esféricos produzidos foi retirada uma amostra ao acaso de 15 peças, fornecendo um diâmetro médio de 240cm com desvio padrão de 15cm . Encontre um intervalo de confiança de 95% para o diâmetro.

- Primeiramente temos que obter o valor tabelado de t, como queremos o intervalo de confiança a 95%, temos que:

$$1 - \alpha = 0,95 \Rightarrow \alpha = 1 - 0,95 = 0,05 \Rightarrow \frac{\alpha}{2} = \frac{0,05}{2} = 0,025$$

- Olhando na tabela o valor que deixa 0,025 de área acima com $\nu = 15 - 1 = 14$ gl, temos $t_{\frac{\alpha}{2}} = 2,145$



$$P\left(\overline{X} - t_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}} \leq \mu \leq \overline{X} + t_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}}\right) = 1 - \alpha$$

$$P\left(240 - 2,145\frac{15}{\sqrt{15}} \leq \mu \leq 240 + 2,145\frac{15}{\sqrt{15}}\right) = 0,95$$

$$P\left(231,69 \leq \mu \leq 248,31\right) = 0,95$$

Assim,

$$IC_{95\%}(\mu) = [231,69;\ 248,31]$$

## Intervalo de Confiança para proporção $p$

- Um intervalo de $(1-\alpha)$ de confiança para $p$ será dado por:

$$P\left(\hat{p} - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\hat{q}}{n}} \leq p \leq \hat{p} + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\hat{q}}{n}}\right) = 1-\alpha$$

- Exemplo: Uma empresa de pesquisa de mercado faz contato com 30 pessoas para saber a satisfação de uma determinada marca de refrigerante, 12 delas respondem que gostam da referida marca. Obtenha um intervalo de confiança de $95\%$ para a proporção de pessoas que gostam da marca.

$$\hat{p} = \frac{x}{n} = \frac{12}{30} = 0,40$$

- Como $\hat{p} = 0,40$, temos que $\hat{q} = 1 - \hat{p} = 1 - 0,40 = 0,60$
- Como queremos o intervalo de confiança a $95\%$, temos que:

$$1 - \alpha = 0,95 \Rightarrow \alpha = 1 - 0,95 = 0,05 \Rightarrow \frac{\alpha}{2} = \frac{0,05}{2} = 0,025$$

- Assim, temos que o valor tabelado de $z_{\frac{\alpha}{2}} = 1,96$

$$P\left(\hat{p} - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\hat{q}}{n}} \leq p \leq \hat{p} + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\hat{q}}{n}}\right) = 0,95$$

$$P\left(0,40 - 1,96\sqrt{\frac{0,40 \times 0,60}{30}} \leq p \leq 0,40 + 1,96\sqrt{\frac{0,40 \times 0,60}{30}}\right)$$

$$P\left(0,40 - 0,17 \leq p \leq 0,40 + 0,17\right) = 0,95$$

$$P\left(0,23 \leq p \leq 0,57\right) = 0,95$$

- Assim,

$$IC_{95\%}(\mu) = [0,23;\ 0,57]$$

## Intervalo de Confiança para variância $\sigma^2$

- Um intervalo de confiança para a variância $\sigma^2$ é dado por:

$$P\left(\frac{(n-1)S^2}{\chi_{\frac{\alpha}{2}}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi_{\left(1-\frac{\alpha}{2}\right)}}\right) = 1-\alpha$$

- Exemplo: No exemplo das 15 peças de rolamentos esféricos, obter um intervalo de confiança de 95% para a variância dos rolamentos.

- Temos que: $\frac{0,05}{2} = 0,025$. Nesse caso precisamos obter na tabela Qui-Quadrado os valores $\chi_{0,025}$ e $\chi_{1-0,025} = \chi_{0,975}$, com $\nu = 14$ graus de liberdade, então:

$$\chi_{0,025} = 26,119 \quad \chi_{0,975} = 5,629$$

- Nesse exemplo foi fornecido a variância amostral é $S^2 = 225$.

$$P\left(\frac{(n-1)S^2}{\chi_{\frac{\alpha}{2}}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi_{\left(1-\frac{\alpha}{2}\right)}}\right) = 0,95$$

$$P\left(\frac{14 \times 225}{26,119} \leq \sigma^2 \leq \frac{14 \times 225}{5,629}\right) = 0,95$$

$$P\left(120,6 \leq \sigma^2 \leq 559,6\right) = 0,95$$

- Assim,

$$IC_{95\%}(\sigma^2) = [120,6;\ 559,6]$$