## INTERVALOS DE CONFIANÇA

**Estimação:** É o processo que consiste no uso de dados amostrais para estimar valores de parâmetros populacionais desconhecidos, tais como média, desvio padrão, proporções, etc..

**Estimação por ponto:** único valor amostral é utilizado como a aproximação do parâmetro a ser estimado.

Ex: o resultado da média amostral é uma estimativa por ponto da média populacional

**Estimação por intervalo:** intervalo de valores, no qual se admite que o parâmetro populacional esteja contido, com determinada confiança.

Ex.: (40 < $\mu$ < 60) = 0,95, ou seja, a média populacional está dentro do intervalo de 40 a 60 com uma certeza de 95%.



### Principais intervalos de confiança

**Intervalo de Confiança para 1 Média**

$$IC(\mu)_{1-\alpha} : \bar{x} \pm e$$

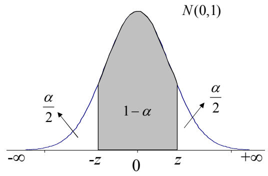
IC → intervalo de Confiança; $\mu$ → média populacional; 1-$\alpha$ → confiança; $\bar{x}$ → média da amostra e → erro da estimativa

1º Caso - Amostras grandes (n ≥ 30)

$$e = \left(z_{\alpha/2}\right)\frac{s}{\sqrt{n}}$$

2º Caso – Amostras pequenas (n < 30)

$$e = \left(t_{\alpha/2}\right)\frac{s}{\sqrt{n}} \qquad \text{GL= n-1}$$





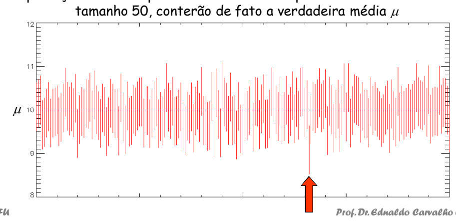

- Como Interpretar o IC para $\mu$?

Suponha uma v.a. $X$ normalmente distribuída com $\mu = 10$ e $\sigma^2 = 4$ → $X \sim N(10,4)$
Sorteia-se 50 valores aleatoriamente e calcula-se $\bar{X}$. Em seguida determina-se o IC para $\mu$ com 95% de confiança, ou seja

$$P(\bar{X} - 1{,}96\frac{2}{\sqrt{50}} < \mu < \bar{X} + 1{,}96\frac{2}{\sqrt{50}}) = 95\%$$

$$P(\bar{X} - 0{,}5544 < \mu < \bar{X} + 0{,}5544) = 95\%$$

Interpretação: 95% dos possíveis ICs obtidos a partir de uma amostra de tamanho 50, conterão de fato a verdadeira média $\mu$






Exemplo: Usando as informações sobre peso e comprimento contidas no arquivo DADOS1 (site da disciplina) fazer a estimativa por ponto e por intervalo (com confiança de 90%, 95% e 99%) para as variáveis peso e comprimento. Fazer as estimativas geral e também por sexo.

**Tamanho da amostra para estimar a Média**

$$n_o = \frac{\left(z_{\alpha/2} + z_{\beta}\right)^2 s^2}{e^2}$$

$Z_{\alpha/2}$ e $Z_{\beta}$ são valores da distribuição normal padrão associados à confiança e ao poder da estimativa, respectivamente (veremos esta definição mais adiante); e é o erro da estimativa e s é o desvio padrão.

Se o tamanho da população é desconhecido temos $n_o$ = n

Se o tamanho da população é conhecido, faz-se a correção para o tamanho da população:

$$n = \frac{n_0}{1 + \frac{n_0 - 1}{N}}$$



A tabela a seguir mostra os valores aproximados do fator $(Z_{\alpha/2} + Z_{\beta})^2$ utilizado na fórmula do dimensionamento da amostra, em função dos valores de $\alpha$ e de $\beta$

| | | Valores do β | | | |
|---|---|---|---|---|---|
| | | 0,05 | 0,1 | 0,2 | 0,5 |
| Valores do α | 0,1 | 10,8 | 8,6 | 6,2 | 2,7 |
| | 0,05 | 13,0 | 10,5 | 7,9 | 3,8 |
| | 0,02 | 15,8 | 13,0 | 10,0 | 5,4 |
| | 0,01 | 17,8 | 14,9 | 11,7 | 6,6 |



**Intervalo de Confiança para 1 Proporção**

$$IC(p)_{1-\alpha} : \hat{p} \pm e$$

IC → intervalo de Confiança; p → proporção populacional; $\hat{p}$ → proporção na amostra; 1-$\alpha$ → confiança; e → erro da estimativa

$$e = \left(z_{\alpha/2}\right)\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

Exemplo: Usando as informações sobre sexo e escores contidas no arquivo DADOS1 (site da disciplina) fazer a estimativa por ponto e por intervalo (com confiança de 90%, 95% e 99%) para a proporção de machos e para a proporção de escore 5.

**Tamanho da amostra para estimar a Proporção**

$$n_o = \frac{(z_{\alpha/2} + z_{\beta})^2 \hat{p}\hat{q}}{e^2}$$

Se o tamanho da população é desconhecido temos $n_o = n$

Se o tamanho da população é conhecido, faz-se a correção para o tamanho da população:

$$n = \frac{n_0}{1 + \frac{n_0 - 1}{N}}$$

**OBSERVAÇÃO:** Podemos construir intervalos de confiança para diferenças entre médias; diferenças entre proporções; variância; relação entre variâncias. Estes intervalos de confiança são construídos a partir das distribuições amostrais e suas fórmulas podem ser encontradas em livros de estatística básica e de inferência estatística.



## TESTE DE HIPÓTESES

Valores observados na amostra, dão suporte a uma conjectura sobre a população?

Para responder esta questão precisamos aplicar o teste de hipóteses estatísticas. Esta área é chamada de Decisão Estatística.

O objetivo da decisão estatística é utilizar ferramentas para verificar a validade de uma determinada hipótese. Para isso utilizam-se dados da amostra.

Formulamos uma hipótese inicial de trabalho sobre um determinado parâmetro populacional ($\mu$, $\sigma$, p, etc.) ou sobre o comportamento dos dados (O atributo X segue a distribuição normal; o atributo X segue a binomial; etc.)

**HIPÓTESE:** De uma maneira geral, uma hipótese estatística é uma afirmação ou conjectura sobre um parâmetro da distribuição de uma variável aleatória.



**Exemplo:**
O tratamento A apresenta melhores resultados do que o tratamento B.

A proporção de caras em lançamentos de moedas é de 0,50.

A variável segue a distribuição normal.

Sempre testamos uma hipótese inicial (chamada na estatística de hipótese nula (Ho)) que será uma igualdade, contra uma hipótese alternativa ($H_1$) que será uma desigualdade.

Exemplos:
1) O tratamento A apresenta melhores resultados do que o tratamento B

Ho: $p_A = p_B$ (proporção de curados de A é igual a de B)
H1: $p_A > p_B$ (proporção de curados de A é maior que de B)



2) A proporção de caras em lançamentos de moedas é de 0,50.

Ho: p = 0,5 (proporção de caras é de 0,50)

$H_1$: $p \neq 0,5$ (proporção de caras é diferente de 0,50)

3) Os dados seguem a distribuição normal

Ho: A variável segue a distribuição normal

$H_1$: A variável não segue a distribuição normal

O teste de hipótese pode ser:

bilateral (quando na $H_1$ usamos diferente)

unilateral à direita ( quando usamos maior na hipótese alternativa)

unilateral a esquerda ( quando H1 for menor).



**PROBABILIDADES ENVOLVIDAS EM UM TESTE DE HIPÓTESES:**

### Erro tipo I

É a probabilidade de se rejeitar a hipótese Ho quando esta é verdadeira. Chamamos de $\alpha$ a probabilidade do erro tipo I.

O valor de $\alpha$ é, geralmente, um valor estipulado pelo pesquisador (geralmente 5% ou 1%) e é também chamado de nível de significância.

### Erro tipo II

É a probabilidade de não rejeitar Ho quando ela é falsa, a qual é indicada por $\beta$.



### Poder de um teste

É a probabilidade de rejeitar Ho quando esta é falsa.

Poder = $1 - \beta$

### Coeficiente de confiança

É a probabilidade com que se aceitará Ho , quando Ho é verdadeira.

$C = 1 - \alpha$

|  | $H_0$ é verd. | $H_0$ é falso |
|---|---|---|
| Aceita $H_0$ | $1 - \alpha$ | $\beta$ |
| Rejeita $H_0$ | $\alpha$ | $1 - \beta$ |



**PASSOS PARA A FORMULAÇÃO DE UMA REGRA DE DECISÃO:**

1) Estabelecer a hipótese nula Ho
2) Estabelecer a hipótese alternativa H1
3) Escolher o nível de significância ($\alpha$)
4) Selecionar a estatística adequada
5) Calcular a estatística
6) Comparar estatística calculada com a da distribuição normal ou comparar p-valor com significância
7) Concluir



## TESTE DE HIPÓTESES PARA NORMALIDADE

A estatística apresenta diversas possibilidades de se testar normalidade dos dados.

No BIOESTAT: D'Agostino, Kolmogorov-Smirnov, Lilliefors, Shapiro-Wilk

No ACTION: Anderson-Darling, Kolmogorov-Smirnov, Shapiro-Wilk.

As estatísticas de testes podem ser encontardas em livros de estatísticas ou na ajuda dos programas.

A formulação do teste:

Ho: A variável tem distribuição normal

$H_1$: A variável não tem distribuição normal



Significância do teste ➔ geralmente 5%

Selecionar o teste, aplicar e verificar resultados

Se valor-p < 0,05 rejeita-se Ho, caso contrário não rejeita-se Ho.

Exemplo de Aplicação: Verificar a normalidade da PROTEINA e da CCS para os dados usados no AVALIATIVO1-PARTE2.