

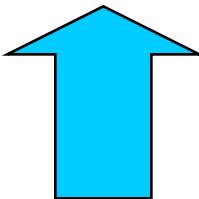


ESTIMAÇÃO DE PARÂMETROS



Estimação de Parâmetros

universo do estudo (população)



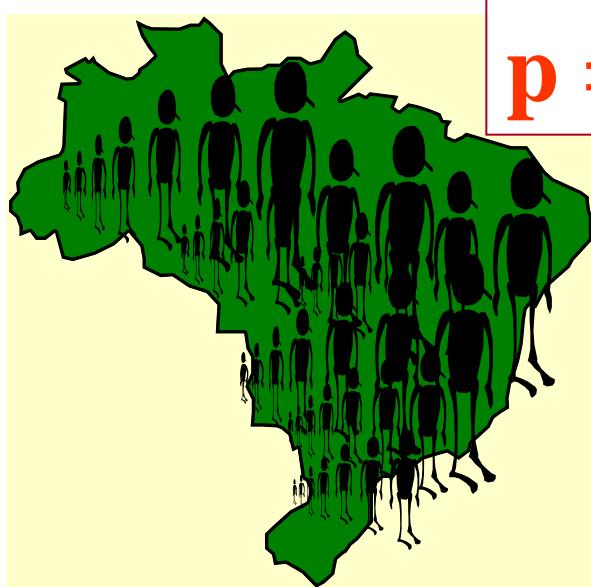
dados observados

O raciocínio indutivo da estimativa de parâmetros



Estimação de Parâmetros

POPULAÇÃO



$$p = ?$$

AMOSTRA



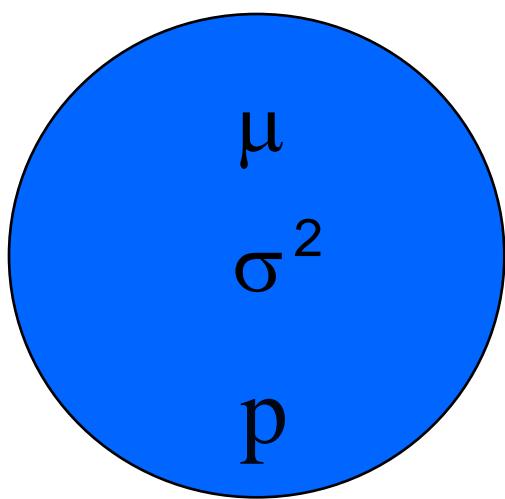
Observações: X_1 X_2 $X_3 \dots \rightarrow \hat{p}$

$$p = \hat{p} \pm \text{erro amostral}$$

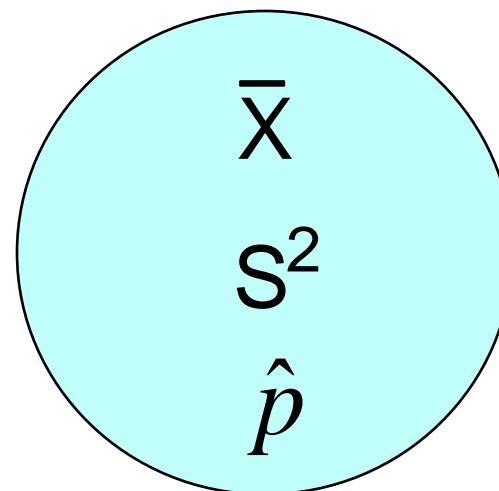


Estimação de Parâmetros

População



Amostra



(parâmetros: números
reais desconhecidos)

(estatísticas / estimadores:
variáveis aleatórias)



Objetivo

- ◆ A partir de uma amostra estimar os parâmetros populacionais.



Propriedades Desejáveis de um Estimador

- ◆ **Não-tendenciosidade:** um estimador é *não tendencioso* (*não viesado; não viciado*) se sua média (ou valor esperado) for o próprio parâmetro que se pretende estimar.



Não Tendenciosidade

$$\bar{X} = \frac{\sum X_i}{n} \quad E(\bar{X}) = \mu_{\bar{X}} = \mu$$



Não-tendencioso

Número de ocorrências
(binomial)

$$\hat{p} = \frac{X}{n} \quad E(\hat{p}) = \frac{1}{n} E(X) = \frac{1}{n} \cdot np = p$$



Não-tendencioso



Não Tendenciosidade

$$\hat{\sigma}^2 = \frac{\sum (x_i - \bar{x})^2}{n}$$

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

→ **Tendencioso**

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

$$E(s^2) = \sigma^2$$

→ **Não-tendencioso**



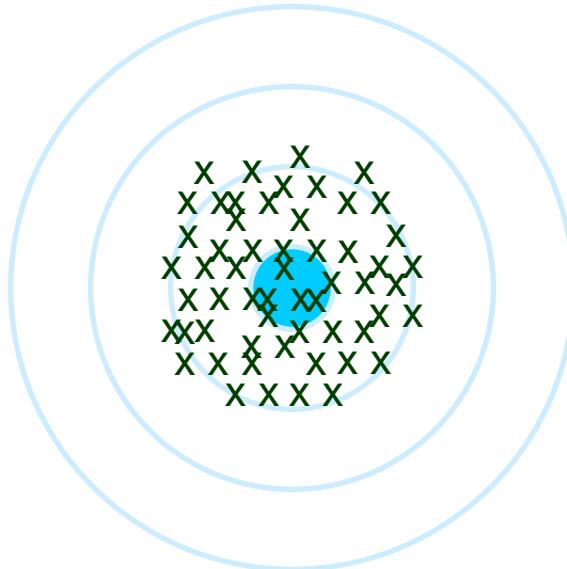
Eficiência

- ◆ Um estimador (T_1) é mais eficiente que outro (T_2) na estimação de um parâmetro θ se
 $E\{(T_1 - \theta)^2\} < E\{(T_2 - \theta)^2\}$
- ◆ Se T_1 e T_2 forem não tendenciosos, T_1 é mais eficiente que T_2 se $\text{Var}(T_1) < \text{Var}(T_2)$



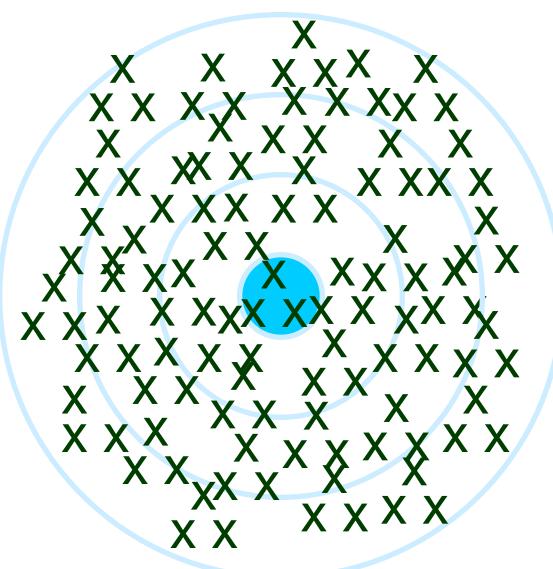
Não tendenciosidade e Eficiência

T₁ é mais eficiente que T₂



T₁

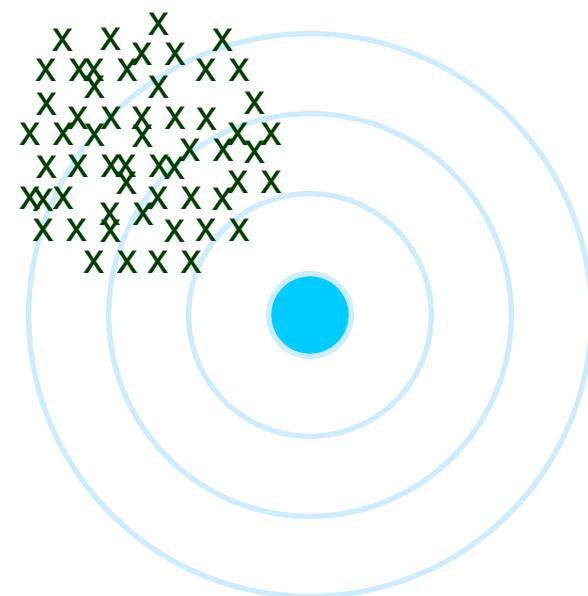
Não-tendenciosos



T₂

T₃

Tendencioso¹⁰





Construção de estimadores

- ◆ **Estimador de mínimos quadrados:** função que minimiza o quadrado dos erros na amostra.
- ◆ Ex. Dada uma amostra de medidas: X_1, X_2, \dots, X_n , supondo o modelo $X_i = \mu + e_i$
onde:
 μ = medida verdadeira,
 e_i = erro associado à i-ésima medida, $i = 1, 2, \dots, n$



Estimador de mínimos quadrados

- ◆ $X_i = \mu + e_i$
- ◆ Valor de μ que minimiza

$$SQE = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (X_i - \mu)^2$$

$$\frac{\partial}{\partial \mu} \left[\sum_{i=1}^n (X_i - \mu)^2 \right] = 0 \quad \Rightarrow \quad \hat{\mu} = \bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$



Construção de estimadores

- ◆ **Estimador de máxima verossimilhança (EMV):** supondo que os dados provém de um certo modelo probabilístico, o EMV de um parâmetro θ é a estatística $T = T(X_1, X_2, \dots, X_n)$ que produz um valor, o qual torna a amostra observada mais verossímil possível.

Ex. Expto: 3 lançamentos de uma moeda;

Modelo: $P(\text{cara}) = p;$

$P(\text{coroa}) = 1 - p$

Amostra: {cara, cara, coroa}



Estimador de máxima verossimilhança (EMV)

◆ Ex. Expto: 3 lançamentos de uma moeda;

Modelo: $P(\text{cara}) = p$;

$$P(\text{coroa}) = 1 - p$$

Amostra: {cara, cara, coroa}

EMV: Valor de p que maximiza a função de verossimilhança $L(p) = p^2 \cdot (1-p)$

$$\frac{d}{dp} [L(p)] = 0 \Rightarrow \hat{p} = \frac{2}{3} \quad (\text{proporção amostral})$$



Estimador de máxima verossimilhança (EMV)

- ◆ Em geral, numa distribuição binomial, o EMV de p é a proporção amostral

$$\hat{p} = \frac{\bar{X}}{n}$$

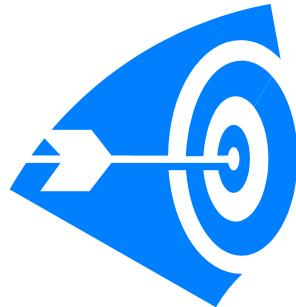
- ◆ Numa distribuição normal, os EMV de μ e σ são, respectivamente:

$$\hat{\mu} = \bar{X} = \frac{\sum_i X_i}{n}$$
$$\hat{\sigma}^2 = \frac{\sum_i \{(X_i - \bar{X})^2\}}{n}$$



Estimação de Parâmetros

Por ponto: estima-se apenas um valor para o parâmetro $p = \hat{p}$



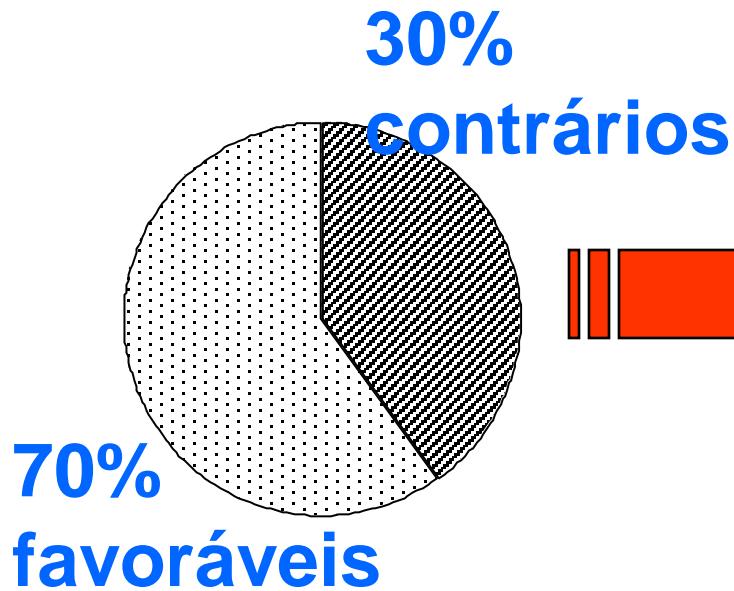
Por intervalo: estima-se um intervalo de valores onde deve-se encontrar o parâmetro (intervalo de confiança).

$$p = \hat{p} \pm \text{erro amostral}$$

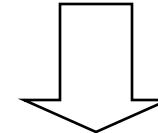


Relação entre p e \hat{p}

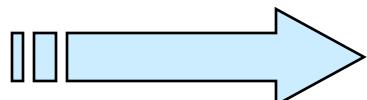
População



Amostra aleatória com
 $n = 400$ indivíduos

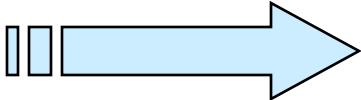
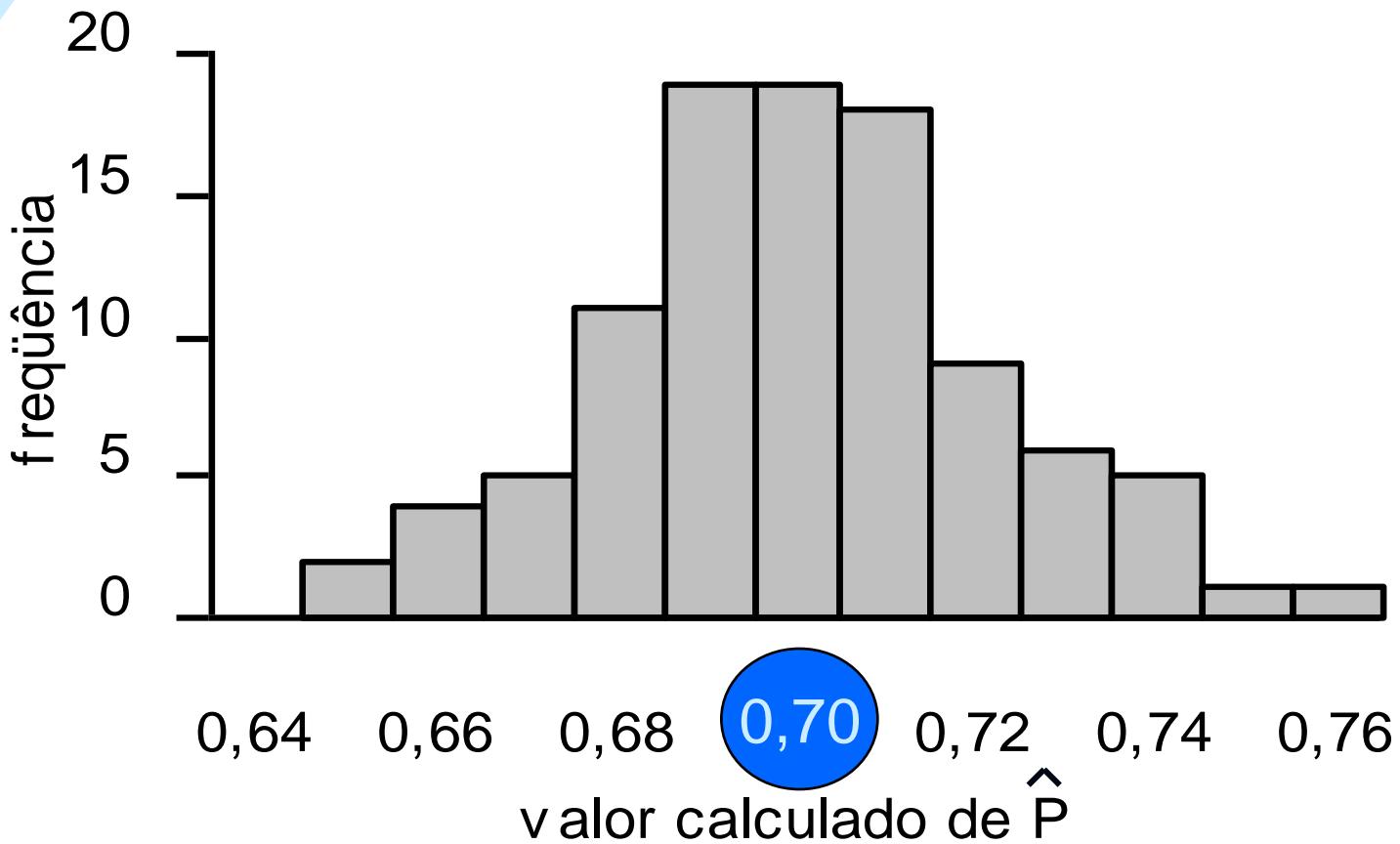


Calcula-se \hat{p}

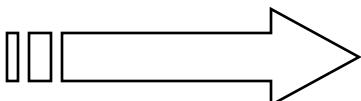


Simularam-se 100 amostras

Relação entre p e \hat{p}



Em geral, erro amostral < 0,05



Em geral, o intervalo $\hat{p} \pm 0,05$ contém p



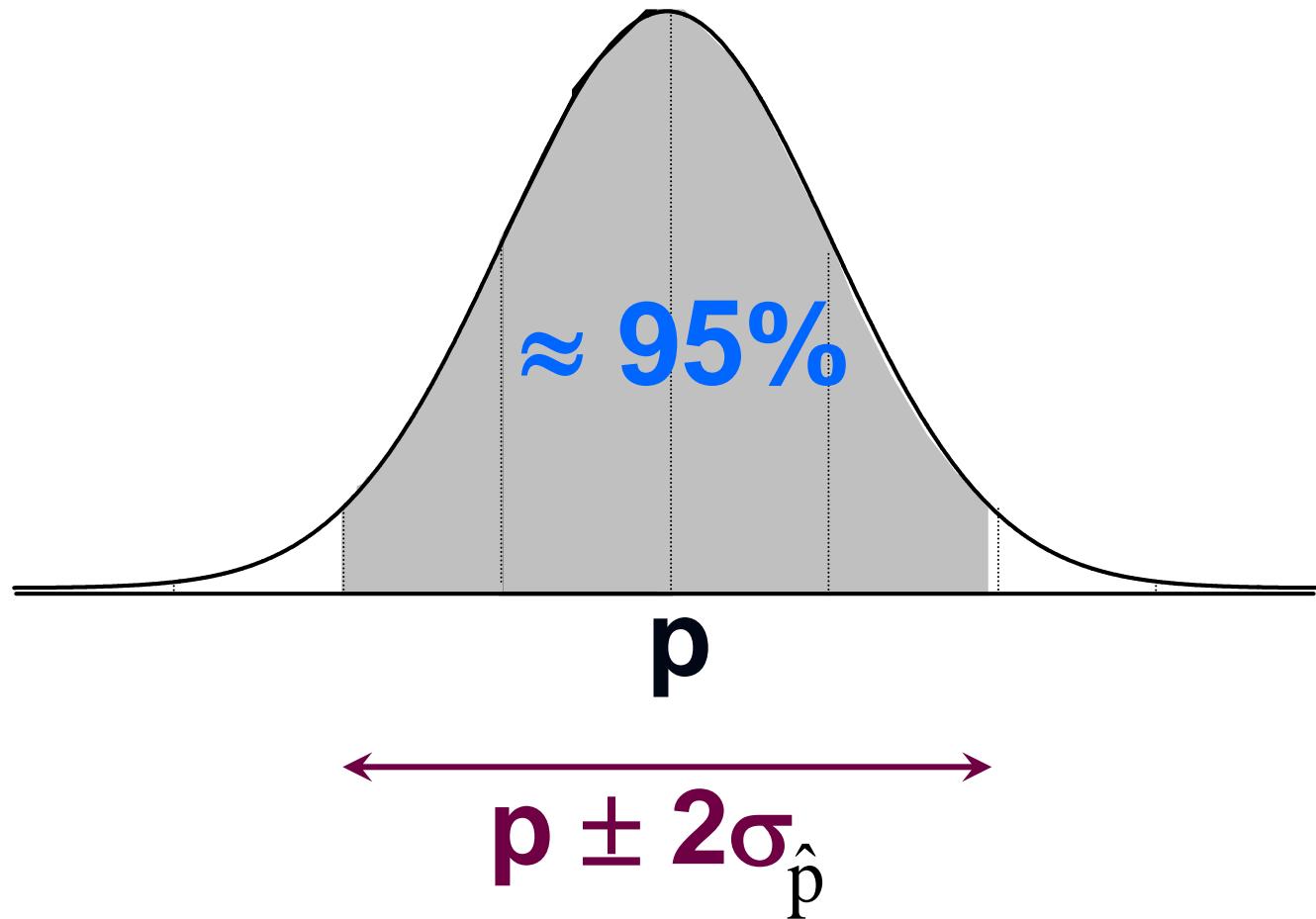
Relação entre p e \hat{p}

- ◆ Desde que a amostra seja aleatória e razoavelmente grande ($n > 30$), tem-se:
 - Os possíveis valores de \hat{p} seguem uma distribuição (aproximada) normal com média e desvio padrão dados por

$$\mu_{\hat{p}} = p \quad \sigma_{\hat{p}} = \sqrt{\frac{p.(1-p)}{n}}$$

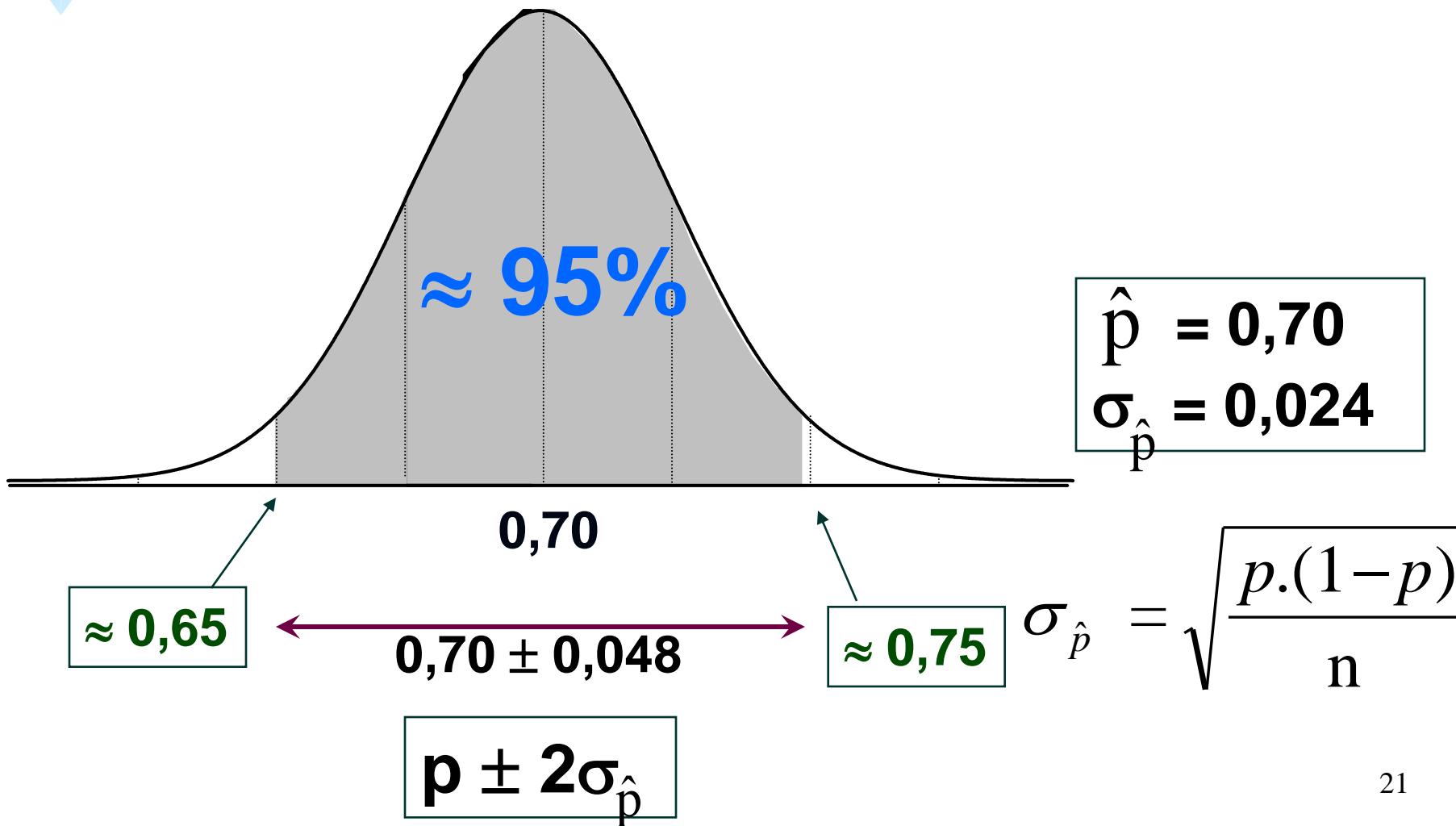


Distribuição de \hat{p}





No exemplo distribuição de \hat{p}





Estimação de uma proporção p

Na prática, estima-se o erro padrão da proporção por

$$S_{\hat{P}} = \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}}$$



Estimação de uma proporção p

Intervalo de 95% de confiança para p:

$$p = \hat{p} \pm 2 \times s_{\hat{p}}$$

Onde:

$$S_{\hat{P}} = \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}}$$

Ex. n = 400 acusando 268 favoráveis $\Rightarrow \hat{p} = 0,67$

$$S_{\hat{P}} = \sqrt{\frac{0,67 \cdot (1 - 0,67)}{400}} = 0,0235$$

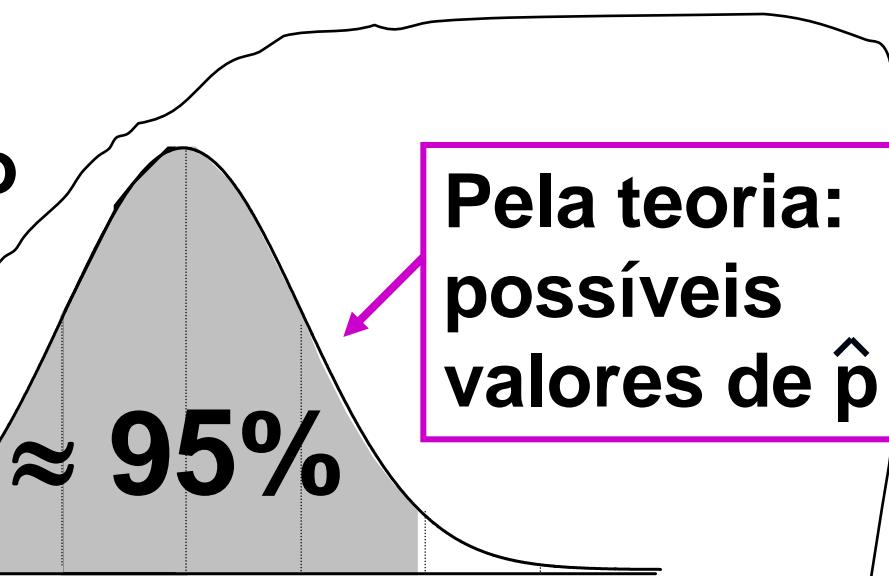
I. C.: **0,67 ± 0,047**

ou: **67,0% ± 4,7%**

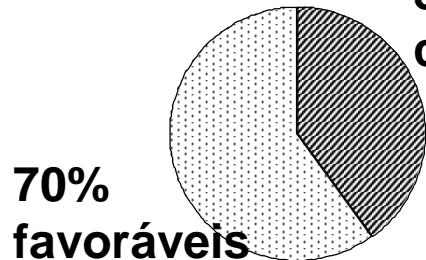


Estimação de p (exemplo)

Com 95% de confiança a verdadeira proporção de favoráveis está no intervalo
 $67,0\% \pm 4,7\%$
(ou de 62,3% a 71,7%)



População



0,67

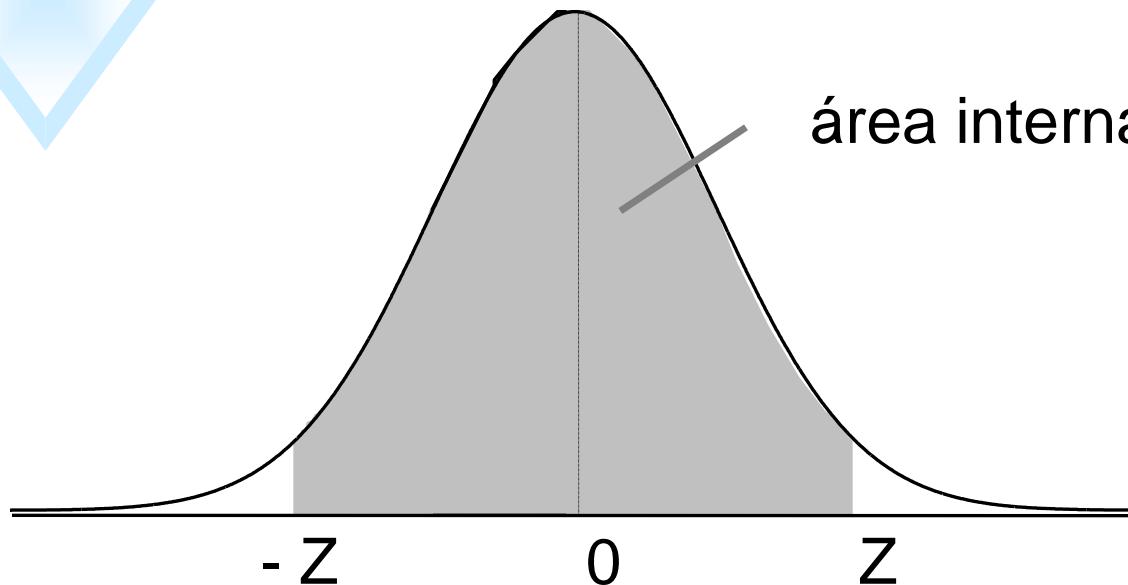
0,70

$0,70 \pm 0,048$

$\hat{p} \pm 2\sigma_{\hat{p}}$



Nível de confiança



PARTE DE UMA TABELA NORMAL PADRÃO

Área	0,800	0,900	0,950	0,980	0,990	0,995	0,998
z	1,282	1,645	1,960	2,326	2,576	2,807	3,090

Intervalo
de conf.

$$\hat{p} \pm z \cdot S_{\hat{p}}$$

Repetir ex. anterior,
usando 99% de conf.



Estimação de uma proporção p Tamanho N da população conhecido

Faz-se uma correção no cálculo do erro padrão:

$$S_{\hat{P}} = \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}} \sqrt{\frac{N - n}{N - 1}}$$

OBS. Se $N \gg n$, pode-se usar a expressão anteriormente empregada.



Exemplo 1

- ◆ Numa amostra aleatória simples de 120 domicílios, realizada num certo bairro da cidade, observou-se que apenas 33,3% possuíam instalações sanitárias adequadas. Considerando que existam 460 domicílios no bairro, encontre um intervalo de 95% de confiança para a proporção de domicílios com instalações sanitárias adequadas.



Exemplo 1

$$\hat{p} \pm z \cdot S_{\hat{p}}$$

$$S_{\hat{p}} = \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}} \sqrt{\frac{N - n}{N - 1}}$$

$$S_{\hat{p}} = \sqrt{\frac{0,333 \cdot (1 - 0,333)}{120}} \sqrt{\frac{460 - 120}{120 - 1}} = 0,0307$$

$$2 \times s_{\hat{p}} = 2 \times 0,037 = 0,074$$

Intervalo de 95% de confiança para p: 33,3% ± 7,4%



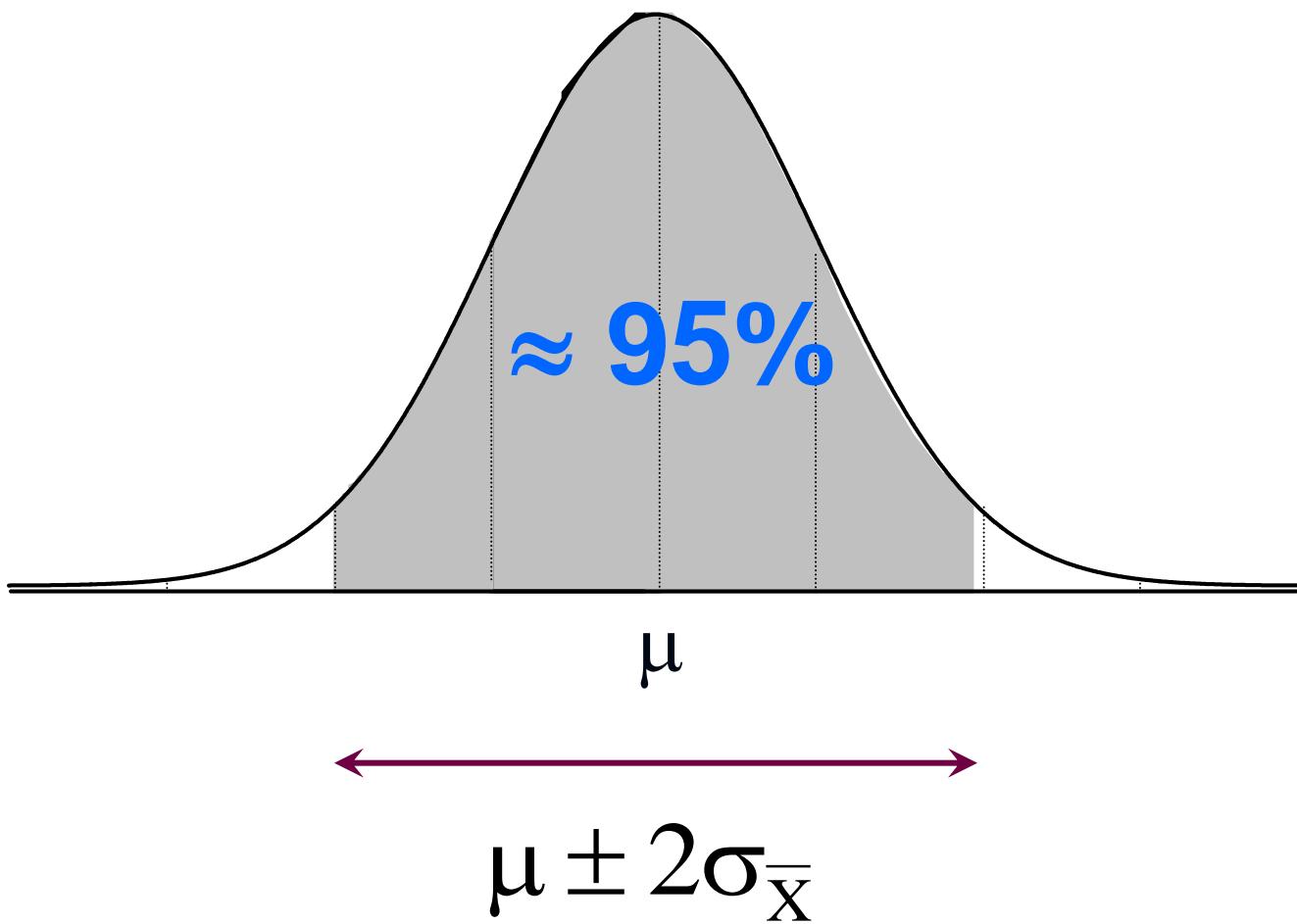
Estimação de uma média μ

- ◆ Estimador pontual: o melhor estimador para a média populacional μ de uma variável é a média amostral da variável.
- ◆ Desde que a amostra seja aleatória e razoavelmente grande ($n > 30$), tem-se:
 - ◆ Os valores da média amostral seguem uma distribuição aproximadamente **normal**, cujos parâmetros são:

$$\mu_{\bar{X}} = \mu \quad \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$



Distribuição de \bar{X}





Estimação do erro padrão da média

$$S_{\bar{X}} = \frac{S}{\sqrt{n}}$$

**para N desconhecido
ou muito grande**

$$S_{\bar{X}} = \frac{S}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

**para N conhecido
e não muito grande**



Intervalo de confiança para a média μ

$$\bar{X} \pm t \cdot S_{\bar{X}}$$

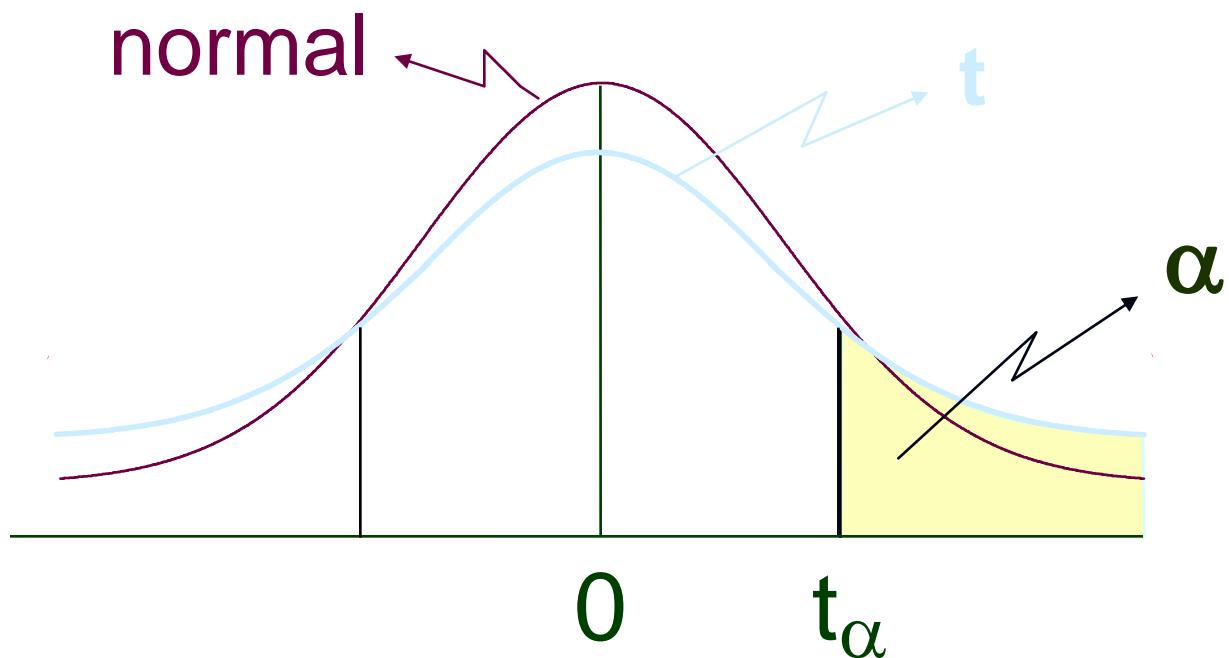
\bar{X} **média da amostra**

$S_{\bar{X}}$ **erro padrão da média**

t abscissa da distrib. t com gl = n - 1



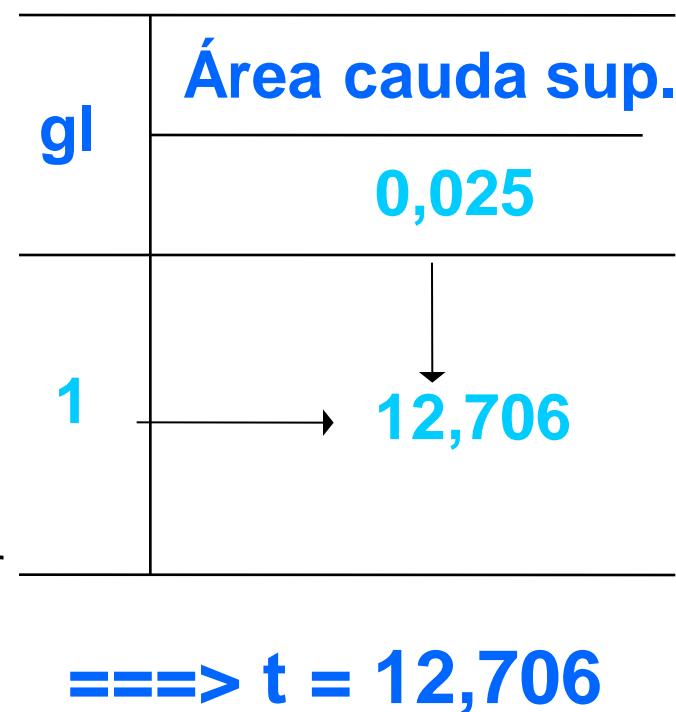
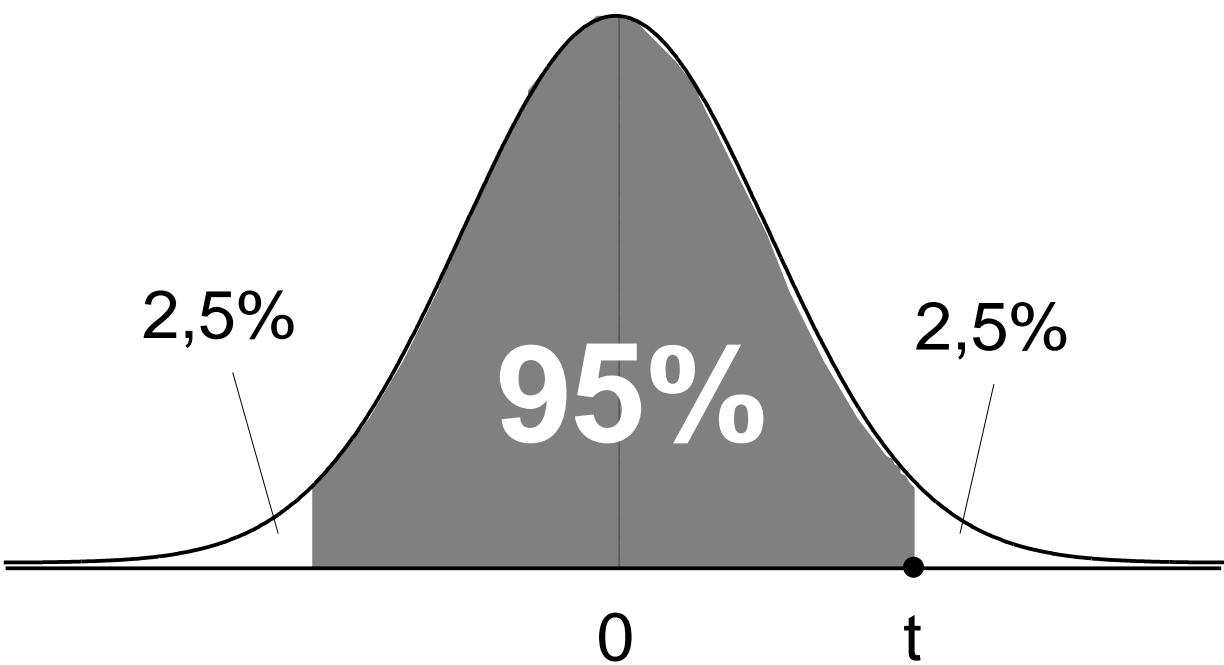
A distribuição t de Student





A distribuição t de Student

Distribuição t com gl = 1





Exemplo 2

- ◆ Encontrar o valor de t para:
- ◆ a) nível de confiança de 99%, com 19 graus de liberdade
(amostra com 20 elementos) **2,861**
- ◆ b) nível de confiança de 95%, com 19 graus de liberdade
(amostra com 20 elementos) **2,093**
- ◆ c) nível de confiança de 95%, com ∞ graus de liberdade
1,960



Exemplo 3

- ◆ O tempo de resposta de um algoritmo de otimização, implementado numa certa máquina, foi observado 20 vezes. Desta amostra, obteve-se as seguintes estatísticas: média de 82,0 minutos e desvio padrão de 10,0 minutos. Apresente um intervalo de confiança para o tempo médio de resposta do algoritmo nesta máquina.



Exemplo 3

$n = 20$ (19 graus de liberdade) $s = 10$ minutos

$\bar{x} = 82$ minutos Nível de 95% de confiança: $t = 2,093$

$$I.C. = \bar{x} \pm t \times \frac{s}{\sqrt{n}} = 82 \pm 2,093 \times \frac{10}{\sqrt{20}} \quad I.C. = 82 \pm 4,7$$

$$L.I. = 82 - 4,7 = 77,3 \text{ minutos}$$

$$L.S. = 82 + 4,7 = 86,7 \text{ minutos}$$

Há 95% de probabilidade de μ estar entre 77,3 e 86,7 minutos.



Exercício 1

Com o objetivo de avaliar o grau médio de impureza de determinada matéria prima, fez-se 10 observações independentes deste material, encontrando os seguintes valores (%) de impurezas:

3,3 3,7 3,5 4,1 3,4 3,5 4,0 3,8 3,2 3,7

(média = 3,62; desvio padrão = 0,294)

Construa um intervalo de 95% de confiança para a verdadeira média.



Tamanho da amostra para estimar uma Média

Considere o intervalo de confiança para μ :

$$\bar{X} \pm t \times s_{\bar{X}} \quad \text{onde} \quad s_{\bar{X}} = \frac{s}{\sqrt{n}}$$


erro máximo (E) E_0 : erro amostral
máximo especificado
a priori.



Tamanho da amostra para estimar uma Média

Cálculo inicial:

$$n_0 = \left(\frac{t \times s}{E_0} \right)^2$$

$n = n_0$ **se N é muito grande ou desconhecido**

$n = \frac{N \cdot n_0}{N + n_0}$ **se N não for muito grande e for conhecido**



Tamanho de Amostras

- ◆ Avaliação “*a priori*” do desvio padrão:
- ◆ **Estudos passados**
- ◆ **Amostra piloto**
- ◆ **Fixando-se um valor teórico**



Exemplo 4

- ◆ Qual o tamanho da amostra para estimar o tempo médio processamento de encomendas em uma agência franqueada dos correios, admitindo-se um erro máximo de 1 minuto? Considere que numa amostra piloto com 20 observações, encontrou-se um desvio padrão de 10,0 minutos.



Exemplo 4

Nível de confiança: 95%

$E_0 = 1$ minuto

Amostra piloto: $n = 20$, $s = 10$ minutos

95% de confiança para $20-1 = 19$ graus de liberdade, $t = 2,093$

$$n_0 = \left(\frac{t \times s}{E_0} \right)^2 = \left(\frac{2,093 \times 10}{1} \right)^2 = 438,0649 \simeq 439$$

Mínimo de 439 elementos!



Tamanho da amostra para estimar uma Proporção

Cálculo inicial: $n_0 = \frac{z^2 \times p \times (1 - p)}{E_0^2}$

$n = n_0$ **se N é muito grande ou desconhecido**

$n = \frac{N \cdot n_0}{N + n_0}$ **se N não for muito grande e for conhecido**



Tamanho da amostra para estimar uma Proporção

- ◆ Estimação “*a priori*” da proporção p:
 - ◆ **Estudos passados**
 - ◆ **Amostra piloto**
 - ◆ **Fixando-se um valor teórico (0,5) => estimativa EXAGERADA.**



Exemplo 5

- ◆ Qual deve ser o tamanho de uma amostra aleatória simples para avaliar a preferência por um candidato de certo partido em um grande município, admitindo erro amostral máximo de 3%, com 95% de confiança:
 - ◆ a) Supondo que usualmente o partido obtém 40% de preferência popular?
 - ◆ b) Utilizando uma estimativa exagerada.



Exemplo 5

a) $p = 0,4$ $1-p = 0,6$ $E_0 = 0,03$ Nível de confiança = 0,95
 $Z = 1,96$

$$n_0 = \frac{z^2 \times p \times (1-p)}{E_0^2} = \frac{1,96^2 \times 0,4 \times 0,6}{0,03^2} = 1024,43 \cong 1025$$

b) $p = 0,5$ $1-p = 0,5$ $E_0 = 0,03$ Nível de confiança = 0,95
 $Z = 1,96$

$$n_0 = \frac{z^2 \times p \times (1-p)}{E_0^2} = \frac{1,96^2 \times 0,5 \times 0,5}{0,03^2} = 1067,11 \cong 1068$$