Estatística para Cursos de Engenharia e Informática

Pedro Alberto Barbetta / Marcelo Menezes Reis / Antonio Cezar Bornia

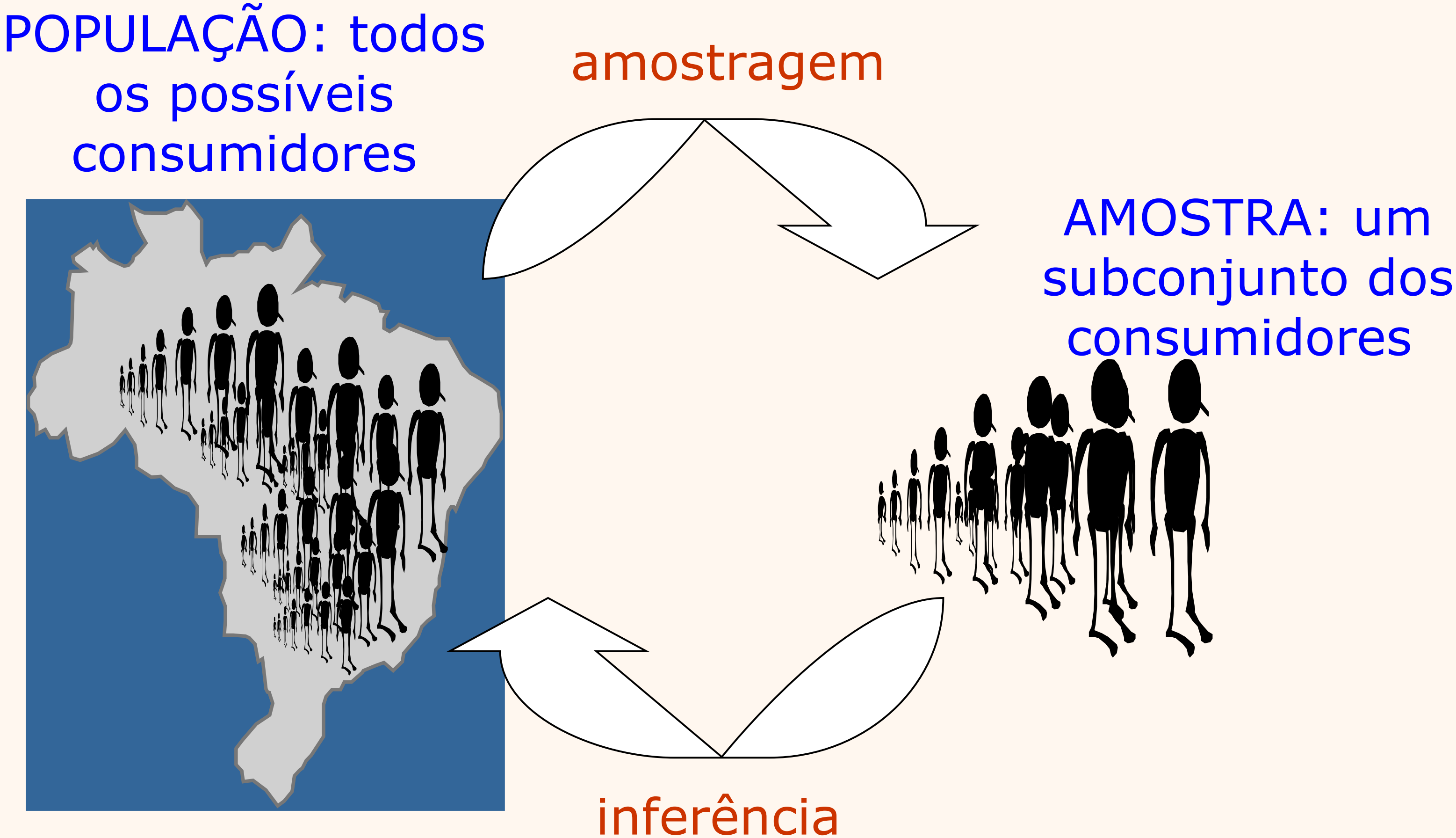
São Paulo: Atlas, 2004

Cap. 7 - Distribuições Amostrais e Estimação de Parâmetros

APOIO:
Fundação de Apoio à Pesquisa Científica e Tecnológica do Estado de Santa Catarina (FAPESC)
Departamento de Informática e Estatística – UFSC (INE/CTC/UFSC)

Amostragem e Inferência estatística

Ex.



BARBETTA, REIS e BORNIA – Estatística para Cursos de Engenharia e Informática. Atlas, 2004

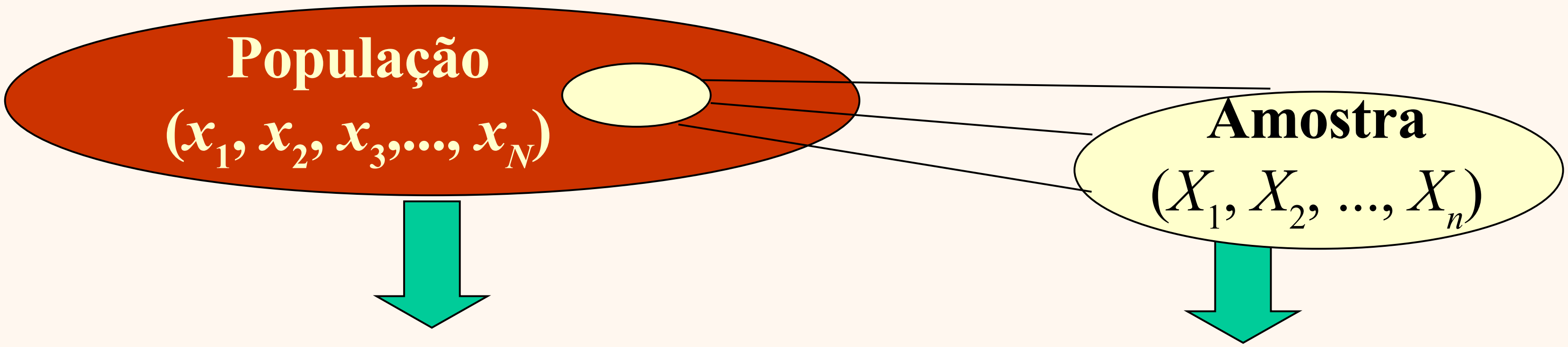
Conceitos

- **Parâmetro**: alguma medida descritiva (média, variância, proporção, etc.) dos valores x_1, x_2, x_3,..., associados à população.

- **Amostra aleatória simples**: conjunto de n variáveis aleatórias independentes $\{X_1, X_2, ..., X_n\}$, cada uma com a mesma distribuição de probabilidades de uma certa variável aleatória X. Esta distribuição de probabilidades deve corresponder à distribuição de freqüências dos valores da população (x_1, x_2, x_3, ...).

- **Estatística**: alguma medida descritiva (média, variância, proporção, etc.) das variáveis aleatórias X_1, X_2, ..., X_n, associadas à amostra.

Parâmetros e Estatísticas



	Parâmetros	Estatísticas
Proporção	$p = \dfrac{n^o \text{ de elementos com o atributo}}{N}$	$\hat{P} = \dfrac{n^o \text{ de elementos com o atributo}}{n}$
Média	$\mu = \dfrac{1}{N} \sum_{i=1}^{N} x_i$	$\bar{X} = \dfrac{1}{n} \sum_{i=1}^{n} X_i$
Variância	$\sigma^2 = \dfrac{1}{N} \sum_{i=1}^{N} \left(x_i - \mu \right)^2$	$S^2 = \dfrac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X} \right)^2$

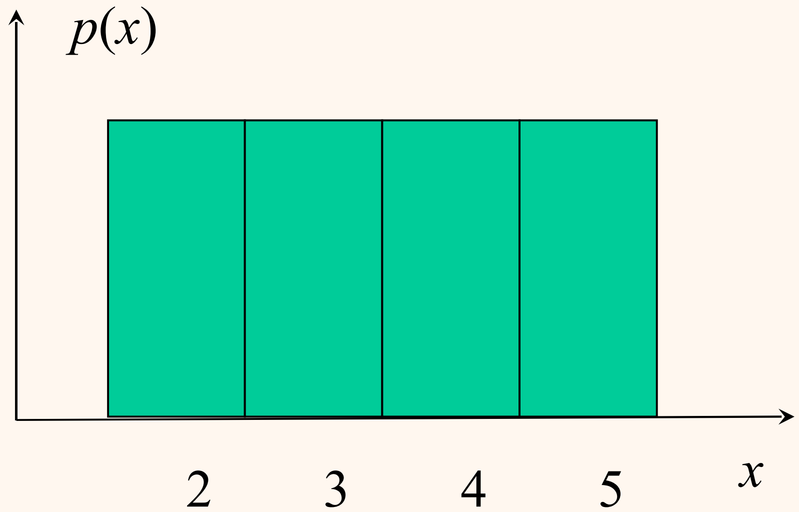
BARBETTA, REIS e BORNIA – Estatística para Cursos de Engenharia e Informática. Atlas, 2004

Estatística

- Uma *estatística* é uma variável aleatória e a sua distribuição de probabilidades é chamada de *distribuição amostral*.

Ex. 7.2

- População: {2, 3, 4, 5}



- Parâmetros:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i = \frac{1}{4}(2+3+4+5) = 3{,}5$$

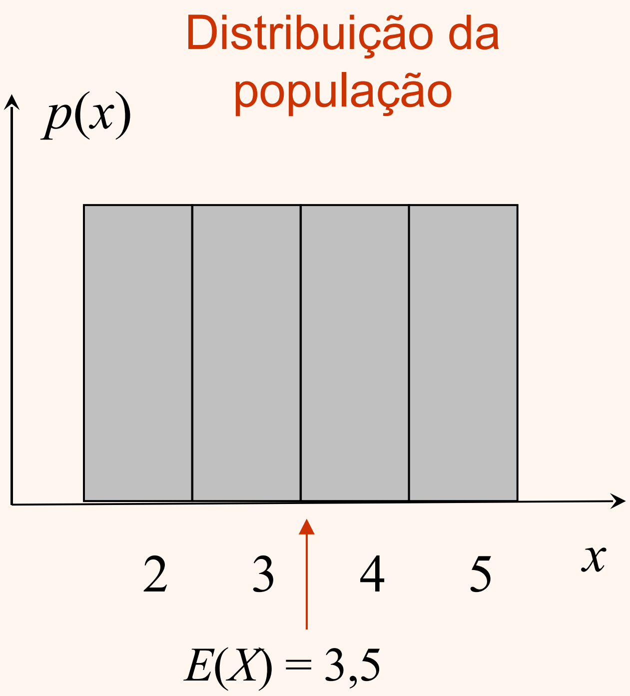
$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2 = \frac{1}{4}\left[(2-3{,}5)^2+(3-3{,}5)^2+(4-3{,}5)^2+(5-3{,}5)^2\right] = 1{,}25$$

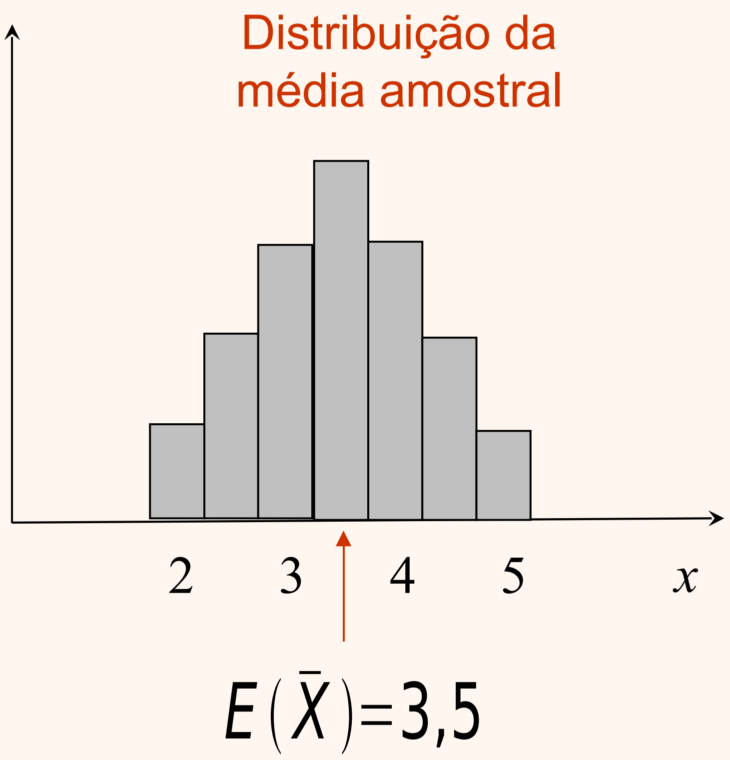
Distribuição da média amostral (Ex. 7.2)

- Amostragem aleatória simples de tamanho $n = 2$.
  - Construção da distribuição amostral da média:

Amostras possíveis	$\bar{X}$	Probabilidade
(2, 2)	2,0	$1/16$
(2, 3), (3, 2)	2,5	$2/16$
(2, 4), (3, 3), (4, 2)	3,0	$3/16$
(2, 5), (3, 4), (4, 3), (5, 2)	3,5	$4/16$
(3, 5), (4, 4), (5, 3)	4,0	$3/16$
(4, 5), (5, 4)	4,5	$2/16$
(5, 5)	5,0	$1/16$

Distribuição da média amostral (Ex. 7.2)





Média e variância da média amostral (Ex. 7.2)

$$E(\bar{X}) = 2\left(\frac{1}{16}\right) + 2{,}5\left(\frac{2}{16}\right) + 3\left(\frac{3}{16}\right) + 3{,}5\left(\frac{4}{16}\right) + 4\left(\frac{3}{16}\right) + 4{,}5\left(\frac{2}{16}\right) + 5\left(\frac{1}{16}\right) = 3{,}5$$

$$V(\bar{X}) = (2 - 3{,}5)^2 \frac{1}{16} + (2{,}5 - 3{,}5)^2 \frac{2}{16} + \ldots + (5 - 3{,}5)^2 \frac{1}{16} = 0{,}625$$

Distribuição amostral da média

Amostragem aleatória simples

População: N elementos

X : variável quantitativa

Parâmetros:

$\boldsymbol{\mu} = E(X)$, $\boldsymbol{\sigma}^2 = V(X)$

Amostra:

$(X_1, X_2, ..., X_n)$

X pode ser vista como uma variável aleatória se considerar a distribuição de freqüências da população como uma distribuição de probabilidades – a *distribuição da população*.

Estatísticas:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

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

Média e variância da média amostral

- Seja a **população** com média μ e variância σ^2.

$$E(\bar{X}) = \mu$$

$$V(\bar{X}) = \frac{\sigma^2}{n}$$

se a amostragem for *com* reposição,
ou N muito grande ou infinito

$$V(\bar{X}) = \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1}$$

se a amostragem for *sem* reposição e
N não muito grande, $N < 20n$

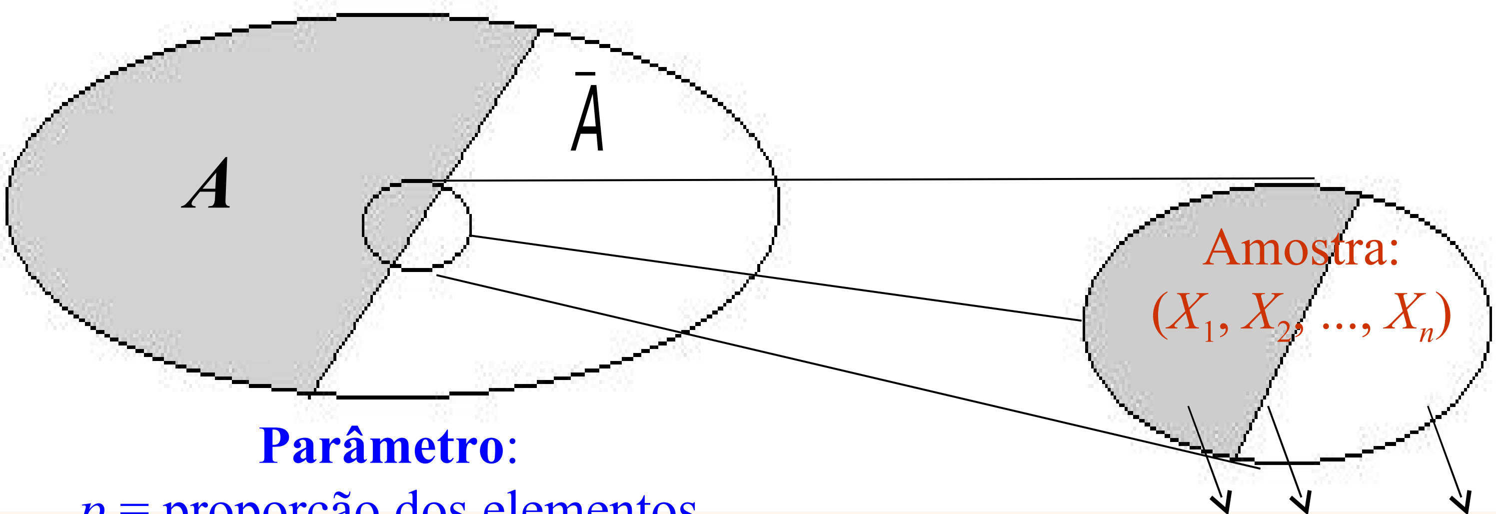
BARBETTA, REIS e BORNIA – Estatística para Cursos de Engenharia e Informática. Atlas, 2004

Distribuição da média amostral

- (*Teorema limite central*) Se o tamanho da amostra for razoavelmente *grande*, então a distribuição amostral da média pode ser aproximada pela ***distribuição normal***.

Distribuição amostral da proporção

População: $N = N_A + N_{\bar{A}}$ elementos



Parâmetro:
p = proporção dos elementos
que têm o atributo A

0 *ou* 1
(0 = sem o atributo;
1 = com o atributo)

Distribuição da população (caso de proporção)

x	$p(x)$
0	$1-p$
1	p

Média e variância:

$$\mu = p$$

$$\sigma^2 = p(1-p)$$

Média e variância da proporção amostral

$$E(\hat{P}) = p$$

$$V(\hat{P}) = \frac{p(1-p)}{n}$$

se a amostragem for *com* reposição, ou N muito grande ou infinito

ou:

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

se a amostragem for *sem* reposição e N não muito grande, $N < 20n$

Distribuição da proporção amostral

- Se o tamanho da amostra for razoavelmente *grande*, então a distribuição amostral da proporção pode ser aproximada pela *distribuição normal*.

- OBS. Se *n* for pequeno, a distribuição exata é binomial ou hipergeométrica (dependendo se a amostragem for *com* ou *sem* reposição)

Estimação de Parâmetros

universo do estudo (população)

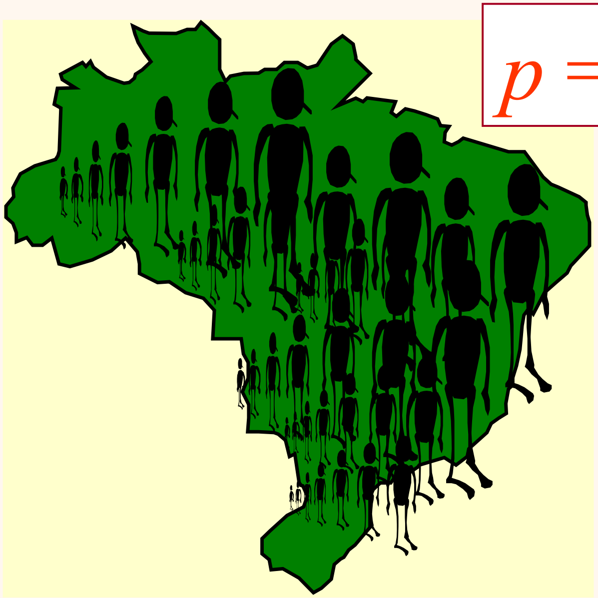
↑

dados observados

*O raciocínio **indutivo** da estimação de parâmetros*

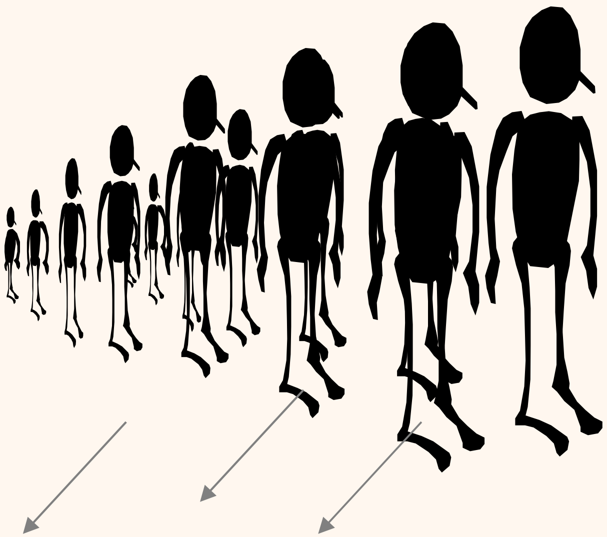
Estimação de Parâmetros

POPULAÇÃO



$p = ?$

AMOSTRA



Observações: $\mathbf{X_1}$ $\mathbf{X_2}$ $\mathbf{X_3}$ ... $\Rightarrow$ $\hat{p}$

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

BARBETTA, REIS e BORNIA – Estatística para Cursos de Engenharia e Informática. Atlas, 2004

Estimação de parâmetros: intervalo de confiança para proporção

p = proporção na população (parâmetro que se quer estimar)

$\hat{p}$ = proporção na amostra (pode ser calculada com base na amostra)

$\sigma_{\hat{p}}$ = erro-padrão da proporção, que para amostra aleatória simples com reposição (ou sem reposição, mas com N >> n), pode ser estimado por:

$$s_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Estimação de parâmetros: intervalo de confiança para proporção

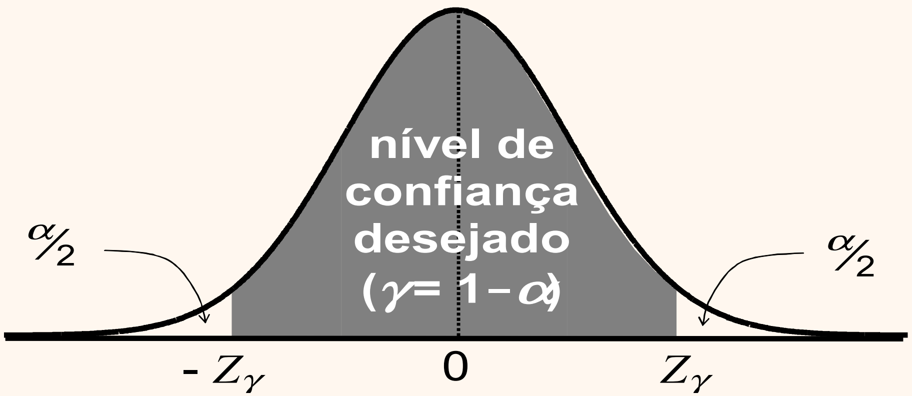
- Com dados de uma amostragem aleatória simples com reposição (ou sem reposição, mas com N >> n), tem-se um intervalo de confiança para *p*, com nível de confiança γ :

$$IC(p,\gamma) = \hat{p} \pm z_{\gamma} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Verificar a expressão acima a partir da distribuição (aproximada) da proporção amostral (ver livro).

Estimação de parâmetros: intervalo de confiança para proporção

$$IC(p,\gamma) = \hat{p} \pm z_{\gamma}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$



γ	0,800	0,900	**0,950**	0,980	0,990	0.995	0,998
z_γ	1,282	1,645	**1,960**	2,326	2,576	2,807	3,090

BARBETTA, REIS e BORNIA – Estatística para Cursos de Engenharia e Informática. Atlas, 2004

Estimação de parâmetros: intervalo de confiança para média

μ = média na população (parâmetro que se quer estimar)

$\bar{x}$ = média na amostra (pode ser calculada com base na amostra)

$\sigma_{\bar{x}}$ = erro-padrão da média.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Estimação de parâmetros: intervalo de confiança para média

Caso o desvio padrão (populacional) seja conhecido:

$$IC(\mu, \gamma) = \bar{x} \pm z_{\gamma} \frac{\sigma}{\sqrt{n}}$$

Estimação de parâmetros: intervalo de confiança para média

Caso o desvio padrão (populacional) **não** seja conhecido:

uso da distribuição t de Student.

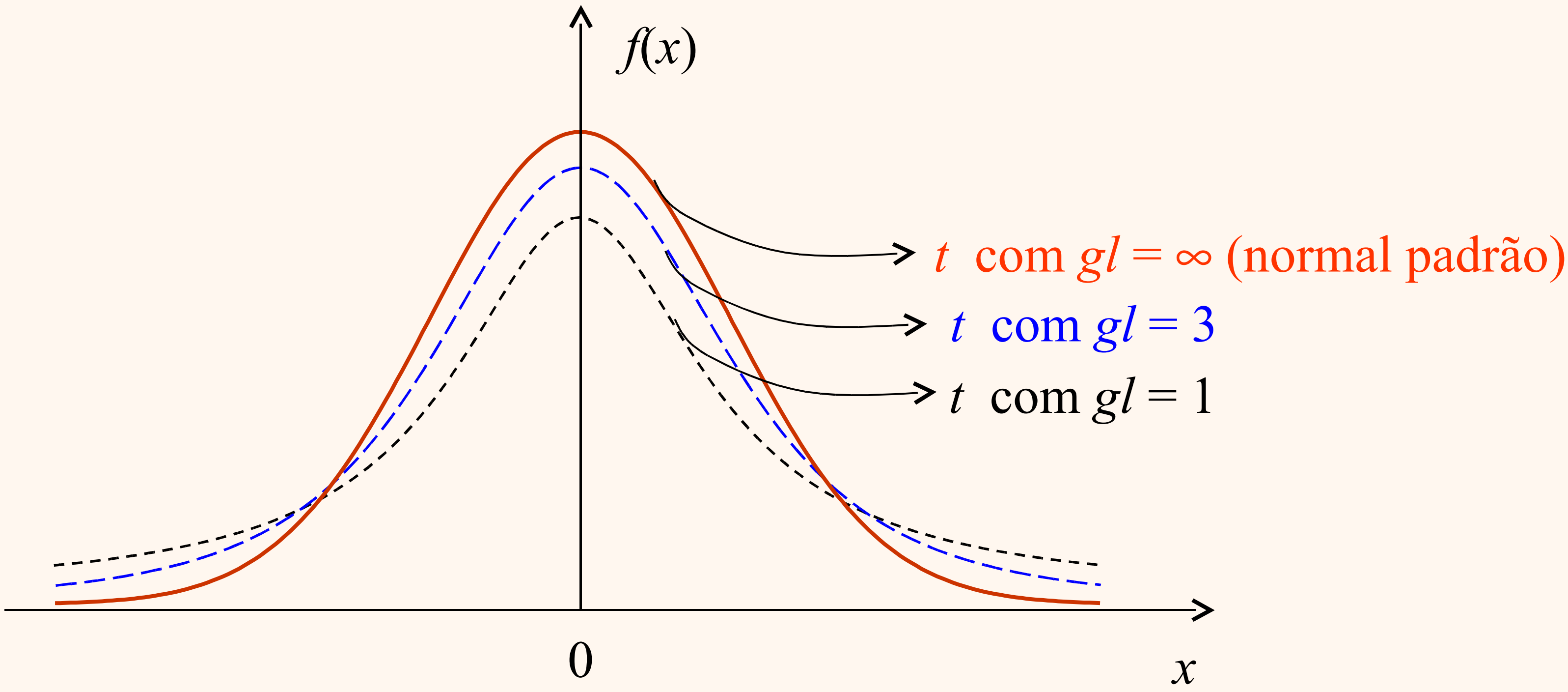
A distribuição *t* de *Student*

- Supondo a população com distribuição normal, a estatística

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

tem distribuição de probabilidades conhecida como *distribuição t de Student*, com $gl = n - 1$ graus de liberdade.

A distribuição *t* de *Student*



BARBETTA, REIS e BORNIA – Estatística para Cursos de Engenharia e Informática. Atlas, 2004

Estimação de parâmetros: intervalo de confiança para média

Caso o desvio padrão (populacional) **não** seja conhecido:

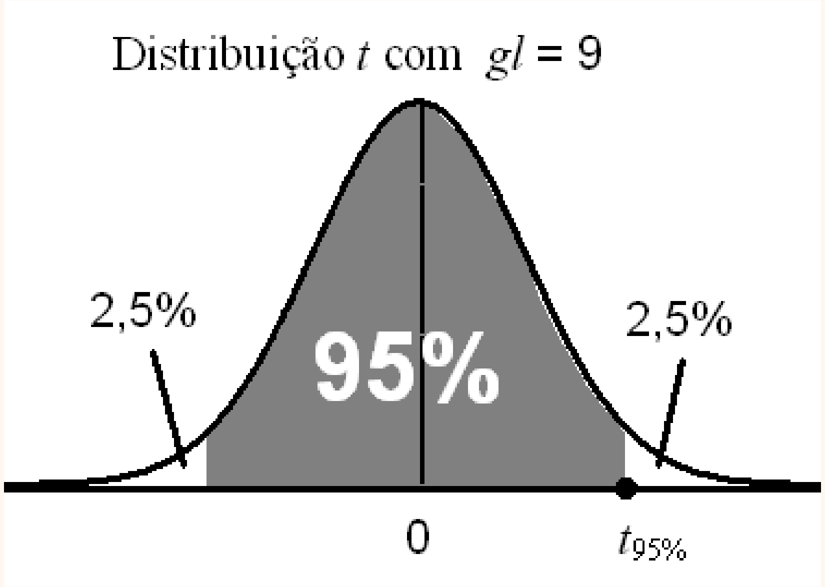
$$IC(\mu,\gamma)=\bar{x}\pm t_{\gamma}\frac{s}{\sqrt{n}}$$

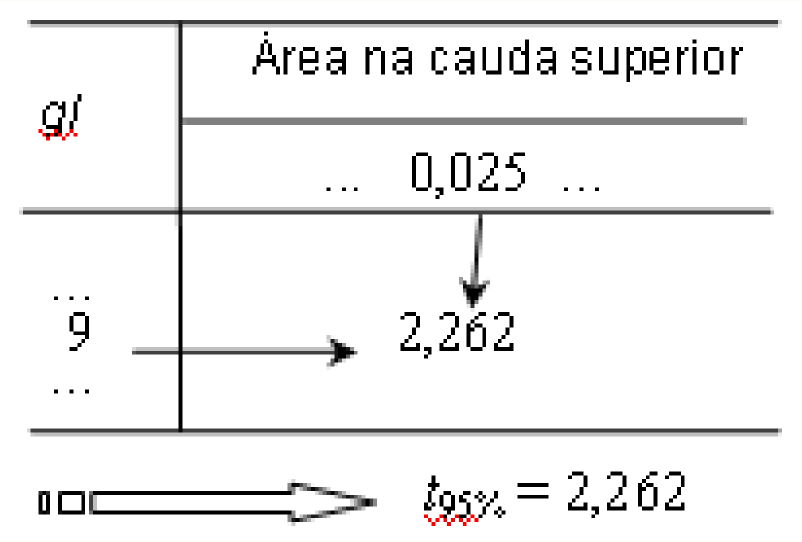
s = desvio padrão calculado na amostra

BARBETTA, REIS e BORNIA – Estatística para Cursos de Engenharia e Informática. Atlas, 2004

Como usar a Tabela *t* (Tab. IV do Apêndice)

- Ilustração com *gl* = 9 e
  nível de confiança de 95%.





Tamanho de amostra

- Na fase do planejamento da pesquisa, muitas vezes precisamos calcular o tamanho *n* da amostra, para garantir uma certa precisão desejada, a qual é descrita em termos do *erro amostral máximo tolerado* (E_0,) e do nível de confiança (γ) a ser adotado no processo de estimação.

- Suponha amostragem aleatória simples

Tamanho de amostra

- No caso de estimação de μ, podemos exigir

$$\left|\bar{X} - \mu\right| \leq E_0$$

ou:

$$z_\gamma \frac{\sigma}{\sqrt{n}} \leq E_0$$

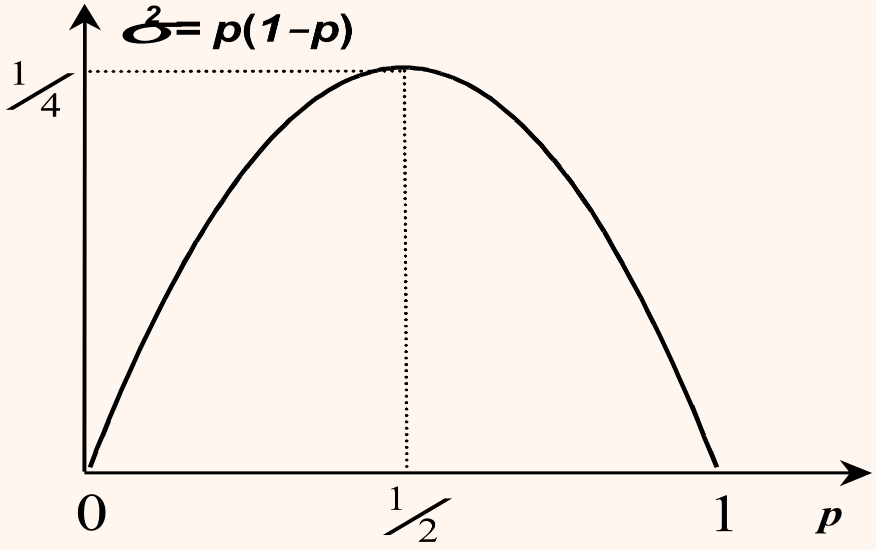
ou:

$$n \geq \frac{z_\gamma^2 \sigma^2}{E_0^{\,2}}$$

Tamanho de amostra

- No caso de estimação de *p*, a população é caracterizada por uma variável 0-1, portanto:

$$\sigma^2 = p.(1-p) \leq \frac{1}{4}$$



Assim:

$$n \geq \frac{z_\gamma^2 \, p(1-p)}{E_0^{\,2}} \geq \frac{z_\gamma^2}{4E_0^{\,2}}$$

Ver discussão no livro.

Tamanho mínimo de uma amostra aleatória simples

Parâmetro de interesse	Valor inicial do tamanho da amostra
uma média (μ):	$$n_0 = \frac{z_\gamma^2 \sigma^2}{E_0^{\,2}}$$
uma proporção (p):	$$n_0 = \frac{z_\gamma^2 p(1-p)}{E_0^{\,2}}$$
várias proporções (p_1, p_2, ...):	$$n_0 = \frac{z_\gamma^2}{4E_0^{\,2}}$$

Tamanho da amostra		
População infinita:	$$n = n_0$$	(arredondamento para o inteiro superior)
População de tamanho N:	$$n = \frac{N . n_0}{N + n_0 - 1}$$	(arredondamento para o inteiro superior)

BARBETTA, REIS e BORNIA – Estatística para Cursos de Engenharia e Informática. Atlas, 2004