

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)

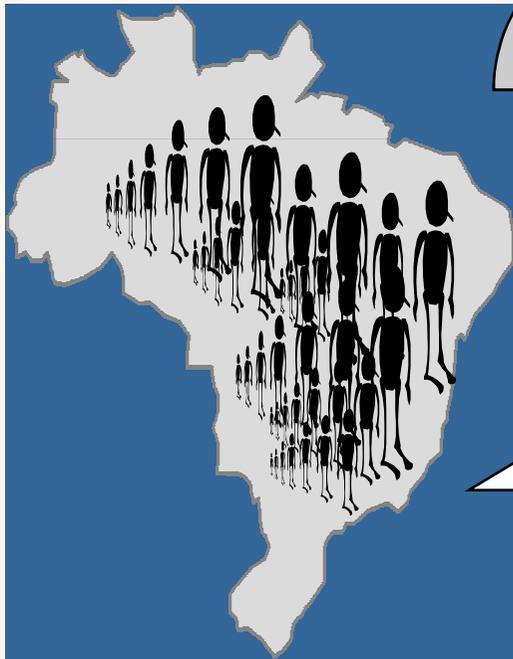
Departamento de Informática e Estatística – UFSC (INE/CTC/UFSC)

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

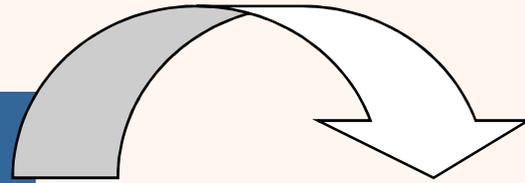
Amostragem e Inferência estatística

Ex.

POPULAÇÃO: todos
os possíveis
consumidores



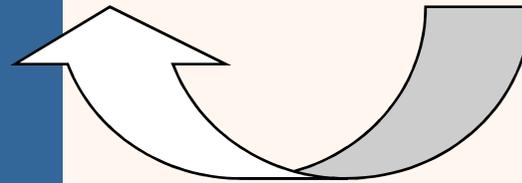
amostragem



AMOSTRA: um
subconjunto dos
consumidores



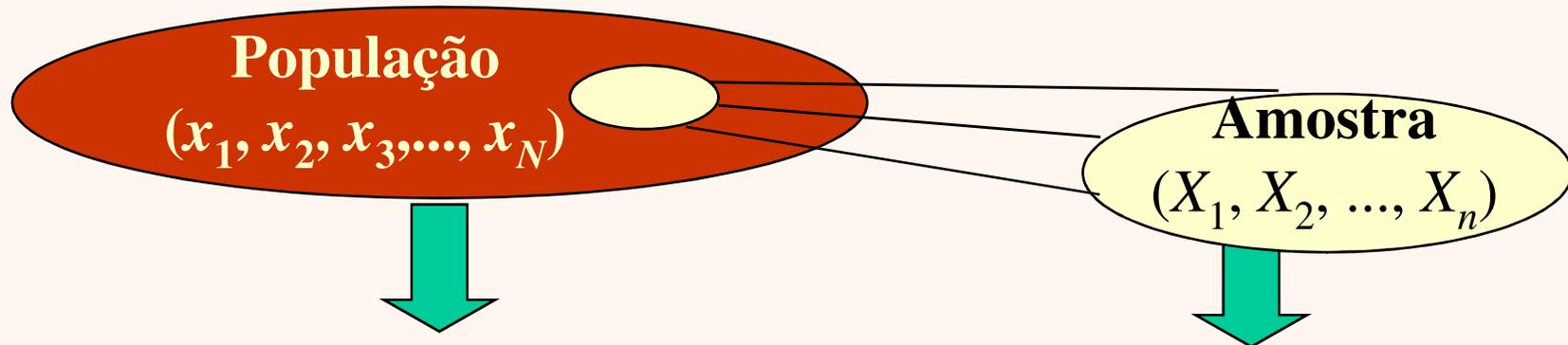
inferência



Conceitos

- **Parâmetro:** alguma medida descritiva (média, variância, proporção, etc.) dos valores x_1, x_2, x_3, \dots , associados à população.
- **Amostra aleatória simples:** conjunto de n variáveis aleatórias independentes $\{X_1, X_2, \dots, X_n\}$, cada uma com a mesma distribuição de probabilidades de uma certa variável aleatória X . Esta distribuição de probabilidades deve corresponder à distribuição de frequências dos valores da população (x_1, x_2, x_3, \dots) .
- **Estatística:** alguma medida descritiva (média, variância, proporção, etc.) das variáveis aleatórias X_1, X_2, \dots, X_n , associadas à amostra

Parâmetros e Estatísticas



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

Exemplos de parâmetros

- X variável aleatória discreta com função de probabilidade p

Média ou valor esperado: $\mu = E(X) = \sum_{j=1}^k x_j p_j$

Variância: $\sigma^2 = V(X) = \sum_{j=1}^k (x_j - \mu)^2 p_j$

ou: $V(X) = E(X^2) - \mu^2$ onde: $E(X^2) = \sum_{j=1}^k x_j^2 p_j$

Exemplos de parâmetros

- X variável aleatória contínua com função de densidade f

Média ou valor esperado: $\mu = E(X) = \int_{-\infty}^{\infty} xf(x)dx$

Variância: $\sigma^2 = V(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)dx$

ou: $V(X) = E(X^2) - \mu^2$ onde: $E(X^2) = \int_{-\infty}^{\infty} x^2 f(x) dx$

Exemplos de estatísticas

- (X_1, X_2, \dots, X_n) : uma amostra aleatória simples da população caracterizada por X .

Média :

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{1}{n} \sum_{i=1}^n X_i$$

Variância:

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

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

Estatísticas

- (x_1, x_2, \dots, x_n) : amostra efetivamente observada.

Média :

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i$$

Variância:

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

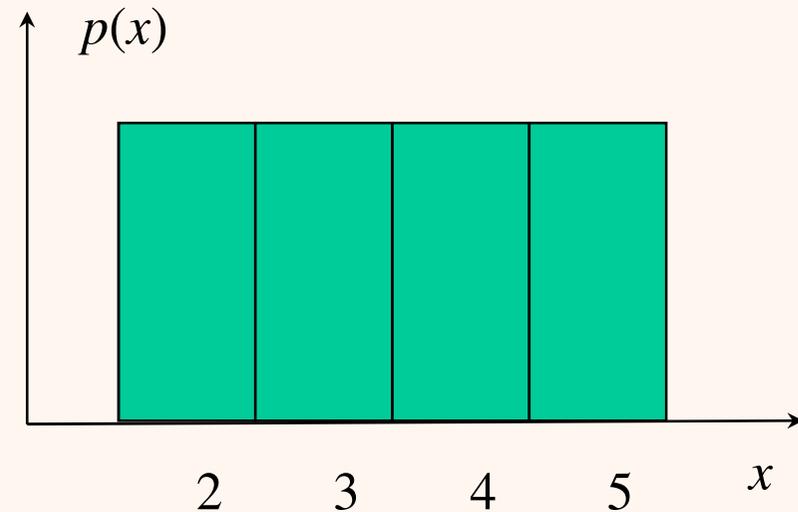
ou:

$$s^2 = \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right)$$

Ex. 7.2

- População: $\{2, 3, 4, 5\}$

- Parâmetros:



$$\mu = E(X) = 2 \times \frac{1}{4} + 3 \times \frac{1}{4} + 4 \times \frac{1}{4} + 5 \times \frac{1}{4} = 3,5$$

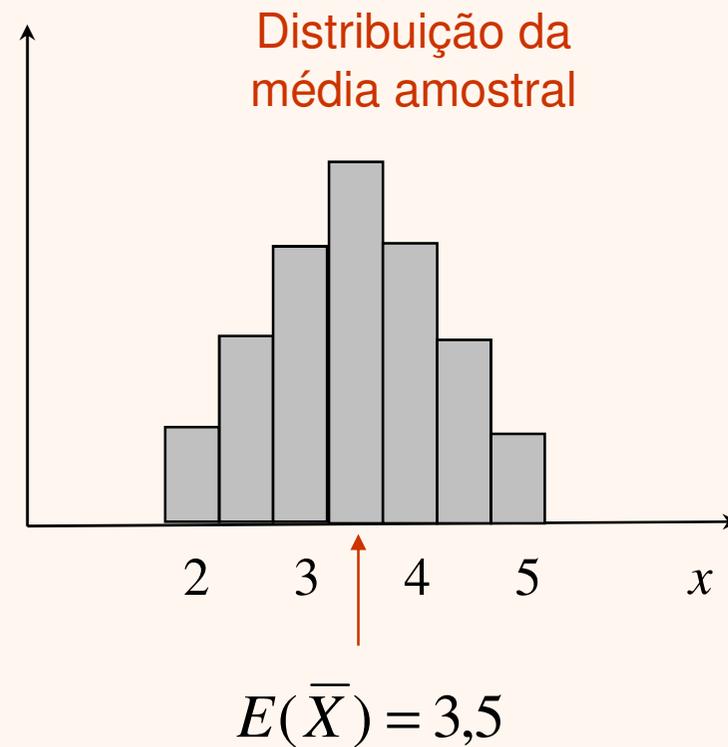
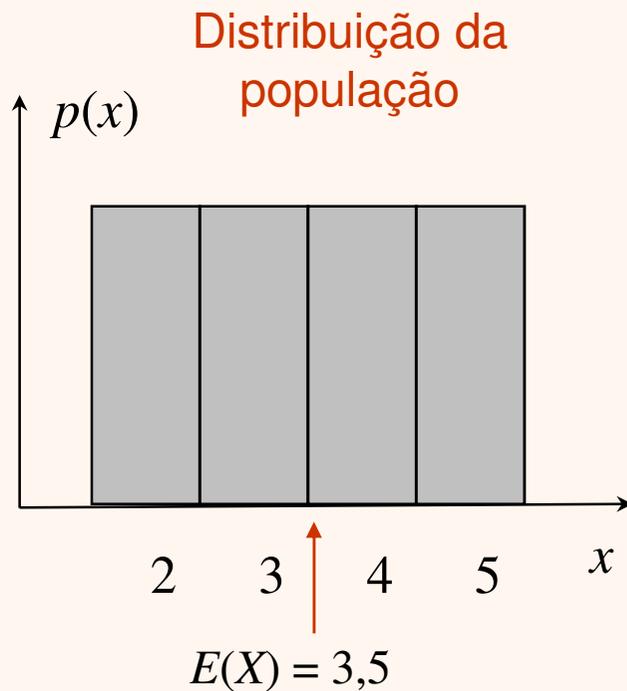
$$\sigma^2 = V(X) = \sum_{i=1}^N (x_i - \mu)^2 \cdot p_i = \frac{1}{4} [(2 - 3,5)^2 + (3 - 3,5)^2 + (4 - 3,5)^2 + (5 - 3,5)^2] = 1,25$$

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 | $\frac{1}{16}$ |
| (2, 3), (3, 2) | 2,5 | $\frac{2}{16}$ |
| (2, 4), (3, 3), (4, 2) | 3,0 | $\frac{3}{16}$ |
| (2, 5), (3, 4), (4, 3), (5, 2) | 3,5 | $\frac{4}{16}$ |
| (3, 5), (4, 4), (5, 3) | 4,0 | $\frac{3}{16}$ |
| (4, 5), (5, 4) | 4,5 | $\frac{2}{16}$ |
| (5, 5) | 5,0 | $\frac{1}{16}$ |

Distribuição da média amostral (Ex. 7.2)



Média e variância da média amostral (Ex. 7.2)

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

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

Distribuição amostral da média

Amostragem
aleatória simples

População: N elementos

X : variável quantitativa

Parâmetros:

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

Amostra:
 (X_1, X_2, \dots, 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:

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

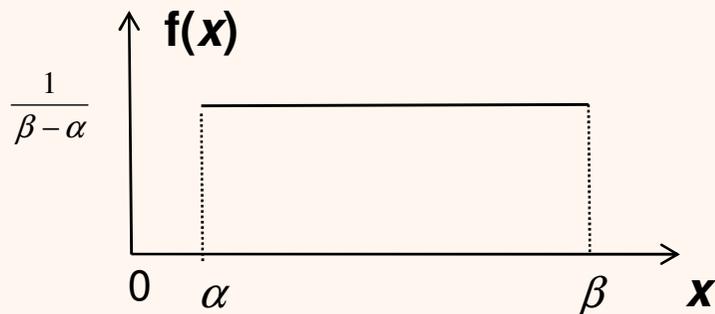
$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

Exemplo de motivação

- Lembrando:

X uniforme em $[\alpha, \beta]$.

$$f(x) = \begin{cases} \frac{1}{\beta - \alpha}, & \text{para } x \in [\alpha, \beta] \\ 0, & \text{para } x \notin [\alpha, \beta] \end{cases}$$



$$E(X) = \frac{\alpha + \beta}{2} \quad V(X) = \frac{(\beta - \alpha)^2}{12}$$

- População X uniforme em $[0, 1]$.

$$E(X) = \frac{1}{2}$$

$$V(X) = \frac{1}{12}$$

- Simular dados numa planilha eletrônica e calcular estatísticas.

Média e variância da média amostral

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

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

$$V(\bar{X}) = \frac{\sigma^2}{n} \quad \text{se a amostragem for } \textit{com} \text{ reposição,}$$

ou N muito grande ou infinito

$$V(\bar{X}) = \frac{\sigma^2}{n} \cdot \frac{N - n}{N - 1} \quad \text{se a amostragem for } \textit{sem} \text{ reposição e}$$

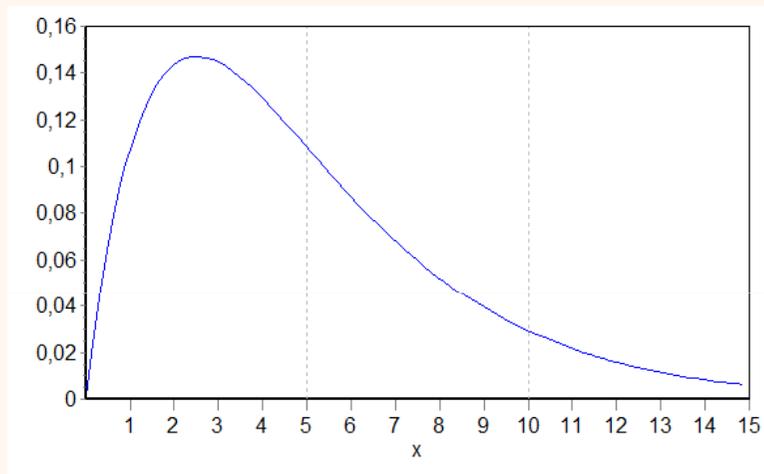
N não muito grande, $N < 20n$

Distribuição da média amostral

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

Distribuição da média amostral

- População



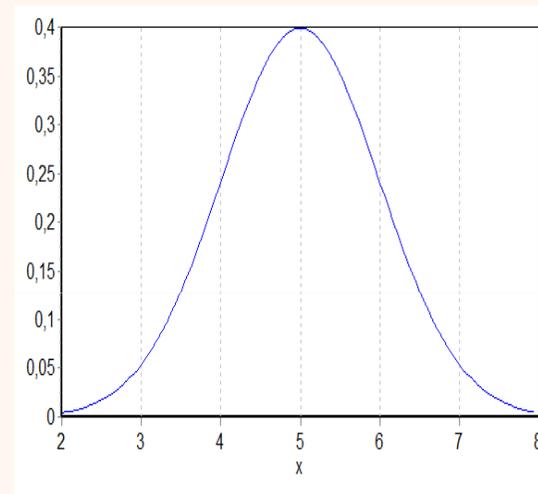
$$E(X) = \mu$$

$$V(X) = \sigma^2$$

$$DP(X) = \sigma$$

- Amostra

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



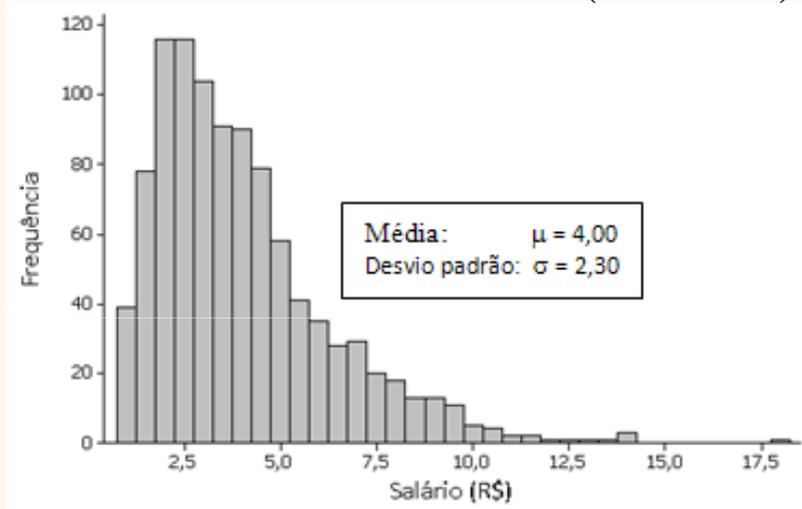
$$E(\bar{X}) = \mu$$

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

$$DP(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$

População de tamanho N

População dos salários dos empregados de certo setor da economia (N = 1.000)

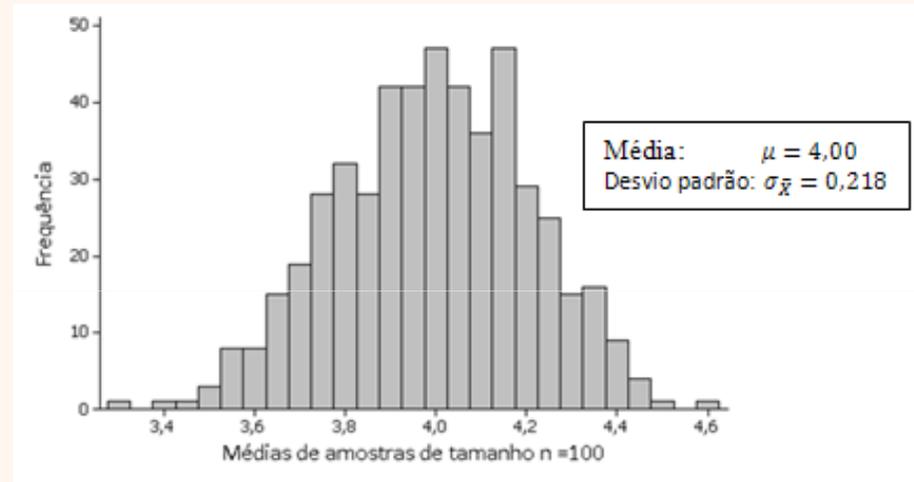


$$E(X) = \mu$$

$$V(X) = \sigma^2$$

$$DP(X) = \sigma$$

Distribuição de freqüências de médias de amostras de tamanho n = 100



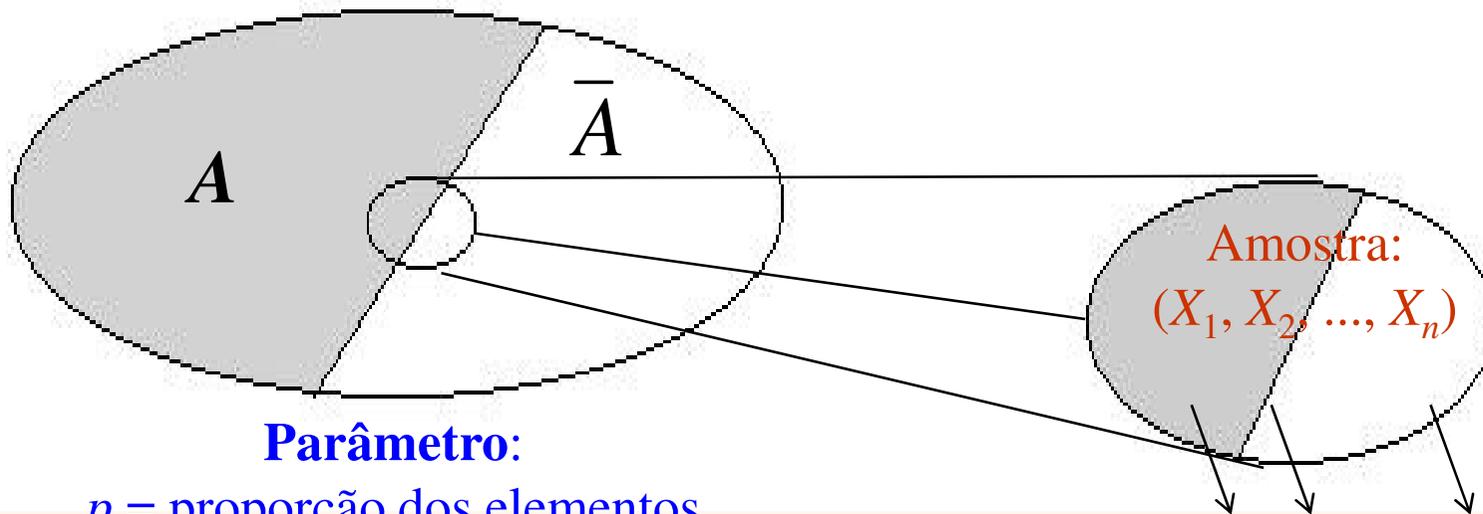
$$E(\bar{X}) = \mu$$

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

$$DP(\bar{X}) = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

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

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

$$V(\hat{P}) = \frac{p(1-p)}{n} \quad \text{se a amostragem for } \textit{com} \text{ reposição, ou } N \text{ muito grande ou infinito}$$

ou:

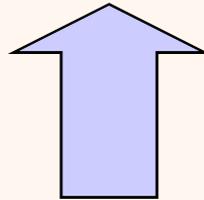
$$V(\hat{P}) = \frac{p(1-p)}{n} \cdot \frac{N-n}{N-1} \quad \text{se a amostragem for } \textit{sem} \text{ reposição e } N \text{ não muito grande, } N < 20n$$

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)

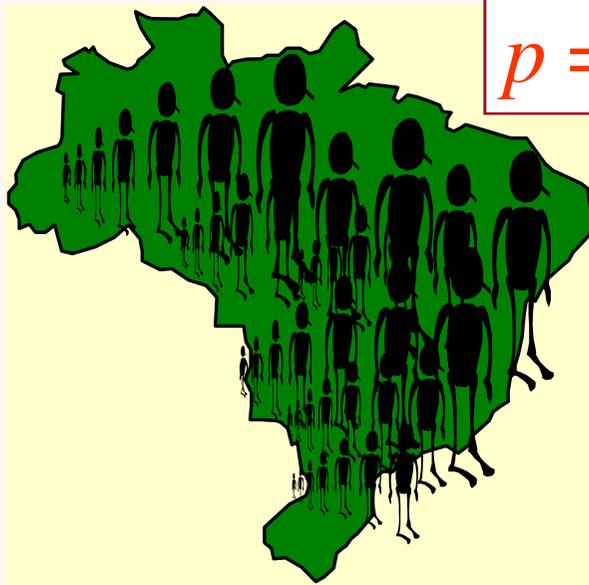


dados observados

O raciocínio indutivo da estimação de parâmetros

Estimação de Parâmetros

POPULAÇÃO



$$p = ?$$

AMOSTRA



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

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

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 $N \gg n$), pode ser estimado por:

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

Estimação de parâmetros: intervalo de confiança para proporção

Seja uma variável aleatória normal padrão, Z . Vale:

$$P\{-1,96 \leq Z \leq 1,96\} = 0,95$$

$$\text{Então: } P\left\{-1,96 \leq \frac{\hat{P} - p}{\sigma_{\hat{P}}} \leq 1,96\right\} = 0,95 \quad \text{sendo: } \sigma_{\hat{P}} = \sqrt{\frac{p(1-p)}{n}}$$

$$\text{ou: } P\left\{\hat{P} - (1,96)\sigma_{\hat{P}} \leq p \leq \hat{P} + (1,96)\sigma_{\hat{P}}\right\} = 0,95$$

$$\text{com probabilidade de 95\%: } |\hat{P} - p| \leq (1,96)\sigma_{\hat{P}}$$

*intervalo de confiança para p ,
com nível de confiança de 95%:*

$$IC(p, 95\%) = \hat{p} \pm (1,96)\sigma_{\hat{p}}$$

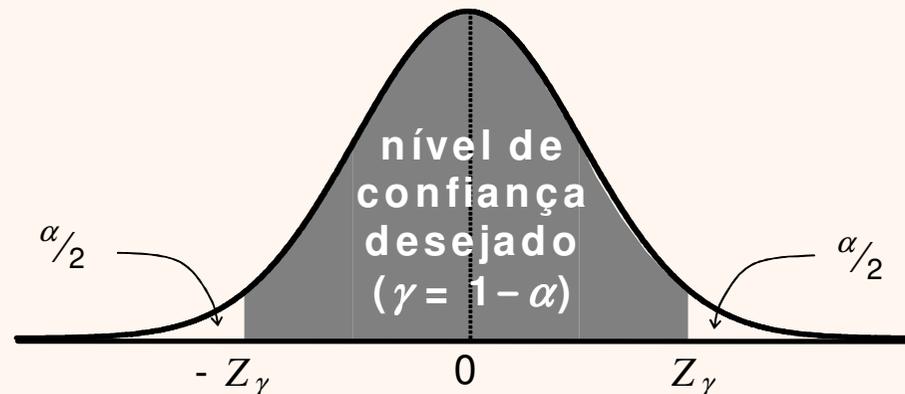
Estimação de parâmetros: intervalo de confiança para proporção

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

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

Estimação de parâmetros: intervalo de confiança para proporção

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



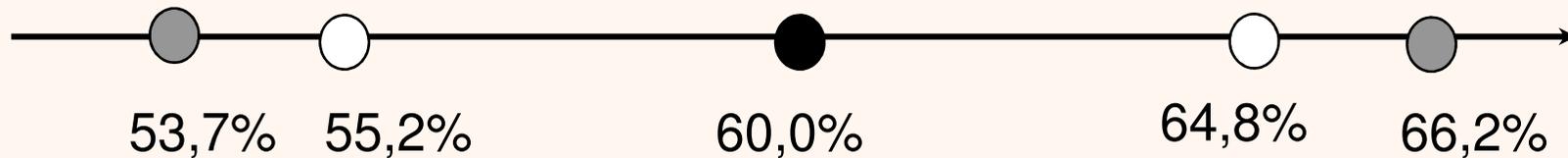
| | | | | | | | |
|--------------|-------|-------|--------------|-------|-------|-------|-------|
| γ | 0,800 | 0,900 | 0,950 | 0,980 | 0,990 | 0,995 | 0,998 |
| z_{γ} | 1,282 | 1,645 | 1,960 | 2,326 | 2,576 | 2,807 | 3,090 |

Resultados do Exemplo 7.3:

Intervalo de 99% de confiança para p
(60,0 ± 6,3%)



Intervalo de 95% de confiança para p
(60,0 ± 4,8%)



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 μ

Seja uma variável aleatória normal padrão, Z . Vale:

$$P\{-1,96 \leq Z \leq 1,96\} = 0,95$$

$$P\{-z_\gamma \leq Z \leq z_\gamma\} = \gamma \quad \text{Então:} \quad P\left\{-z_\gamma \leq \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq z_\gamma\right\} = \gamma$$

$$\text{ou:} \quad P\left\{\bar{X} - z_\gamma \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_\gamma \frac{\sigma}{\sqrt{n}}\right\} = \gamma$$

com probabilidade de 95%:

*intervalo de confiança para μ ,
com nível de confiança de γ :*

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

Estimação de parâmetros: intervalo de confiança para média

Caso o desvio padrão (populacional) 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.

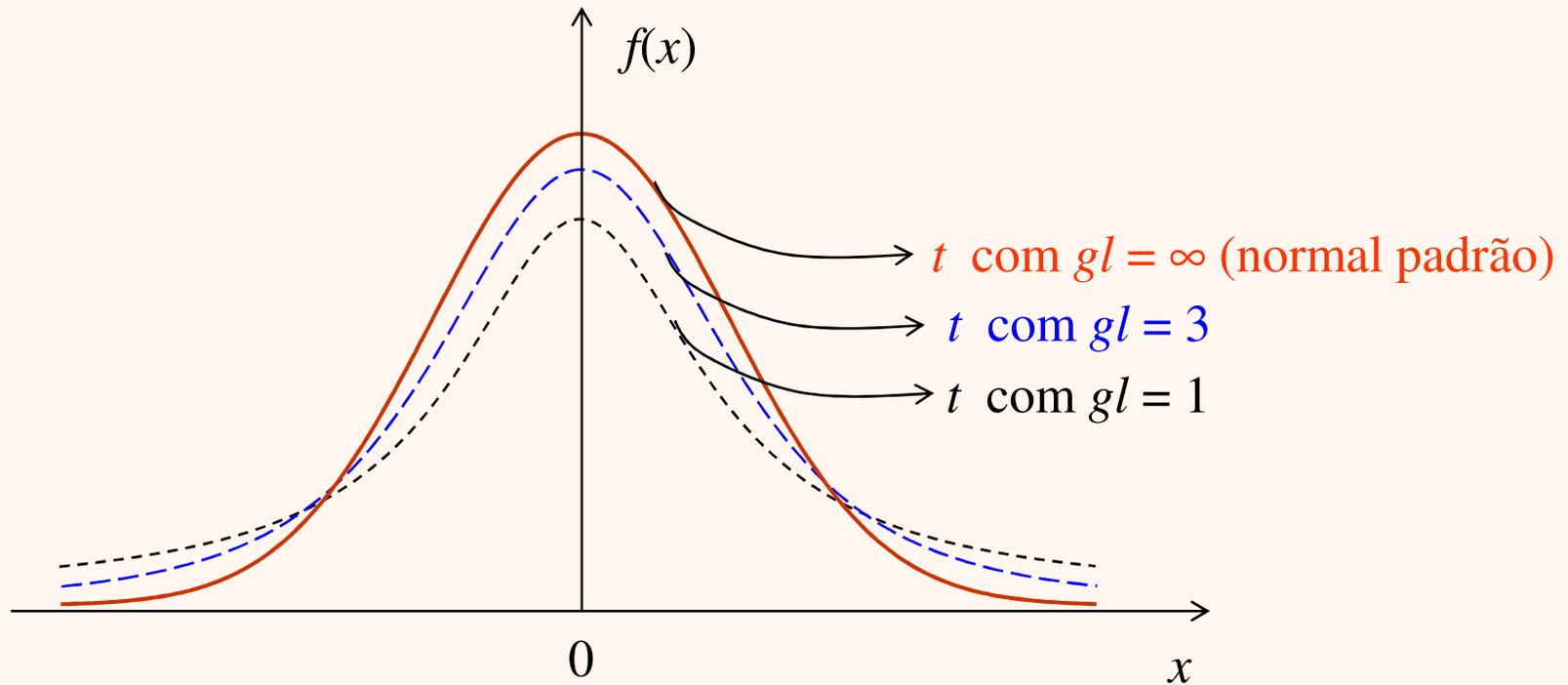
A distribuição *t* de *Student*

- Supondo a população com distribuição normal, a estatística

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

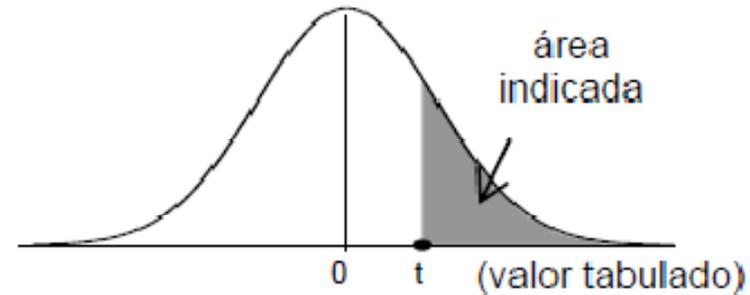
tem distribuição de probabilidades conhecida como *distribuição t de Student*, com $gl = n - 1$ graus de liberdade.

A distribuição t de *Student*



A distribuição *t* de *Student*

TABELA IV Distribuição *t* de *Student*



| <i>gl</i> | Área na cauda superior | | | | | | | | |
|-----------|------------------------|-------|-------|-------|-------|-------|--------|-------|--------|
| | 0,25 | 0,10 | 0,05 | 0,025 | 0,01 | 0,005 | 0,0025 | 0,001 | 0,0005 |
| 1 | 1,000 | 3,078 | 6,314 | 12,71 | 31,82 | 63,66 | 127,3 | 318,3 | 636,6 |
| 2 | 0,816 | 1,886 | 2,920 | 4,303 | 6,965 | 9,925 | 14,09 | 22,33 | 31,60 |
| 3 | 0,765 | 1,638 | 2,353 | 3,182 | 4,541 | 5,841 | 7,453 | 10,21 | 12,92 |
| 4 | 0,741 | 1,533 | 2,132 | 2,776 | 3,747 | 4,604 | 5,598 | 7,173 | 8,610 |
| 5 | 0,727 | 1,476 | 2,015 | 2,571 | 3,365 | 4,032 | 4,773 | 5,894 | 6,869 |
| 6 | 0,718 | 1,440 | 1,943 | 2,447 | 3,143 | 3,707 | 4,317 | 5,208 | 5,959 |
| 7 | 0,711 | 1,415 | 1,895 | 2,365 | 2,998 | 3,499 | 4,029 | 4,785 | 5,408 |
| 8 | 0,706 | 1,397 | 1,860 | 2,306 | 2,896 | 3,355 | 3,833 | 4,501 | 5,041 |
| 9 | 0,703 | 1,383 | 1,833 | 2,262 | 2,821 | 3,250 | 3,690 | 4,297 | 4,781 |
| 45 | 0,680 | 1,301 | 1,679 | 2,014 | 2,412 | 2,690 | 2,952 | 3,281 | 3,520 |
| 50 | 0,679 | 1,299 | 1,676 | 2,009 | 2,403 | 2,678 | 2,937 | 3,261 | 3,496 |
| <i>z</i> | 0,674 | 1,282 | 1,645 | 1,960 | 2,326 | 2,576 | 2,807 | 3,090 | 3,291 |

Estimação de parâmetros: intervalo de confiança para média

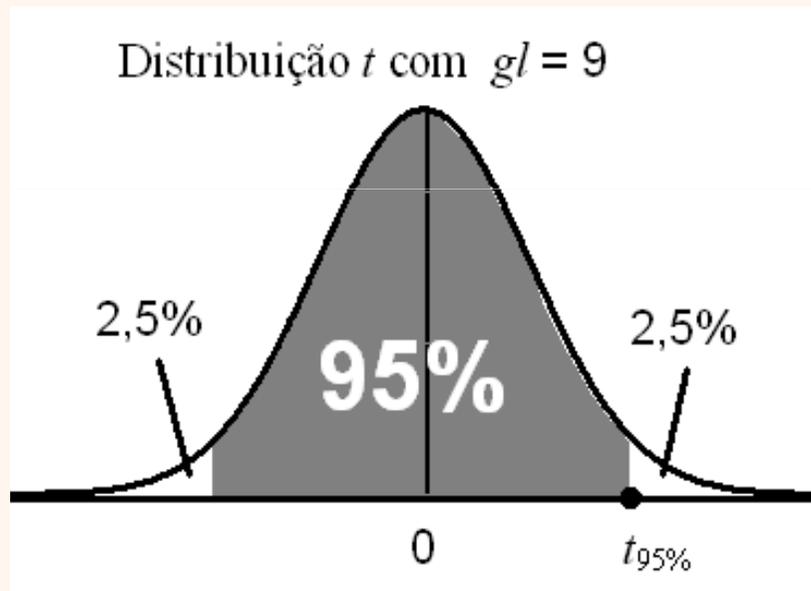
Caso o desvio padrão (populacional) **não** seja conhecido:

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

s = desvio padrão calculado na amostra

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

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



| gl | Área na cauda superior |
|------|------------------------|
| ... | ... 0,025 ... |
| 9 | 2,262 |
| ... | |

$t_{95\%} = 2,262$

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

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

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

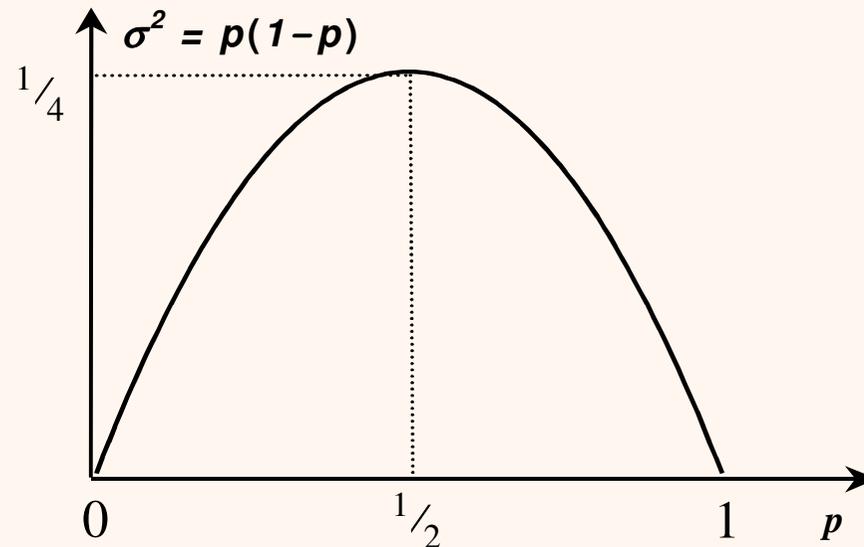
- No caso de estimação de p , a população é caracterizada por uma variável 0-1, portanto:

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

Assim:

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

Então,
$$n = \frac{z_\gamma^2}{4E_0^2}$$



Ver discussão no livro.

é suficiente para qualquer p .

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, \dots): | $n_0 = \frac{z_\gamma^2}{4E_0^2}$ |
| Tamanho da amostra | |
| População infinita: | $n = n_0$ (arredondamento para o inteiro superior) |
| População de tamanho N : | $n = \frac{N \cdot n_0}{N + n_0 - 1}$ (arredondamento para o inteiro superior) |