

# *Estimação de Parâmetros*

## *Erro padrão da média amostral*

- Fórmula para calcular o erro padrão da média amostral

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

- Amostras grandes
  - 68,2% dos valores estão dentro do intervalo  $\bar{X} \pm S_{\bar{X}}$
  - 95,4% dos valores estão dentro do intervalo  $\bar{X} \pm 2S_{\bar{X}}$
  - 99,7% dos valores estão dentro do intervalo  $\bar{X} \pm 3S_{\bar{X}}$
- Assume-se as mesmas premissas da aproximação da distribuição binomial por uma normal

**EXEMPLO 9.2, continuação** - Observando uma amostra aleatória simples de  $n = 100$  crianças do primeiro ano letivo, nas escolas municipais, em que se estava servindo uma merenda especial, encontraram-se as seguintes estatísticas relativas à variável *ganho de peso ao longo do ano*.

Ganho médio de peso das crianças da amostra:  $\bar{X} = 6,0$  kg;

Desvio padrão dos pesos das crianças na amostra:  $S = 2,0$  kg.

Com o objetivo de estimar o parâmetro  $\mu =$  *ganho médio de peso da população*, podemos calcular uma estimativa para o erro padrão da média amostral

$$S_{\bar{x}} = \frac{S}{\sqrt{n}} = \frac{2,0}{\sqrt{100}} = 0,2 \text{ kg}$$

e o erro amostral máximo provável (95% de confiança) de

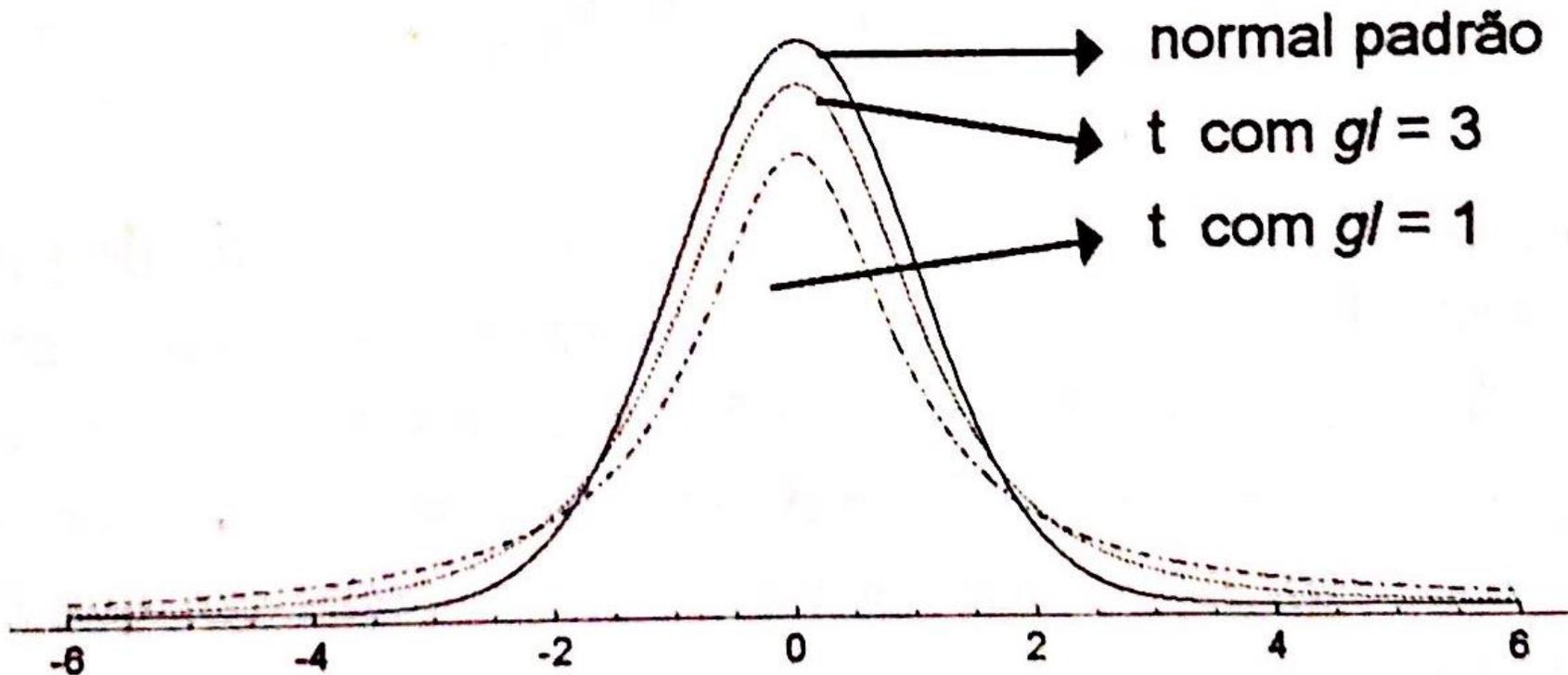
$$2 \cdot (0,2) = 0,4 \text{ kg}$$

donde resulta o seguinte intervalo de 95% de confiança para  $\mu$ :

$$6,0 \pm 0,4 \text{ kg.}$$

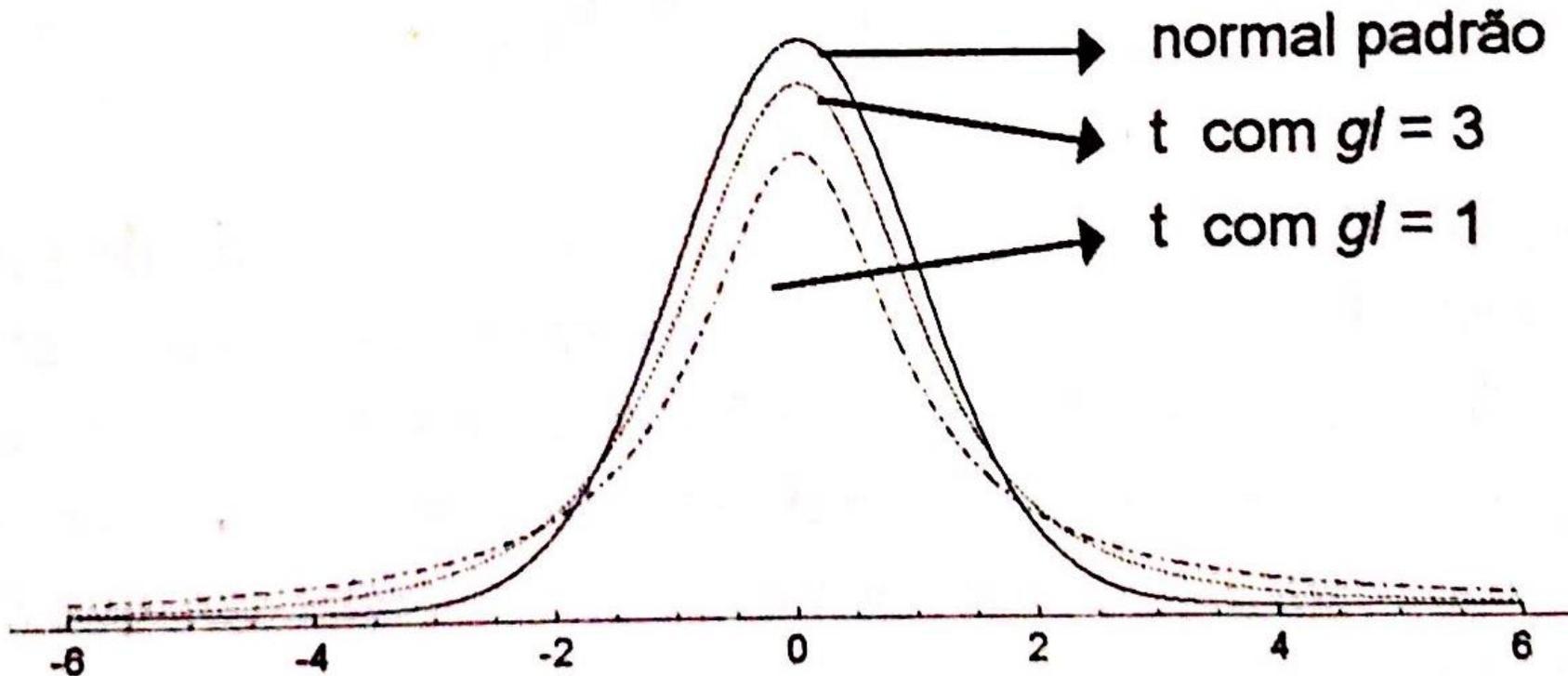
# *Amostras pequenas*

- Aproximamos a distribuição de amostras pequenas por uma distribuição t
  - Mais dispersa do que a distribuição normal



# Distribuição $t$

- Esta distribuição é apresentada em função do parâmetro
  - Grau de liberdade ( $gl$ )
    - No caso da estimativa da média,  $gl = n-1$



# Tabela da distribuição t

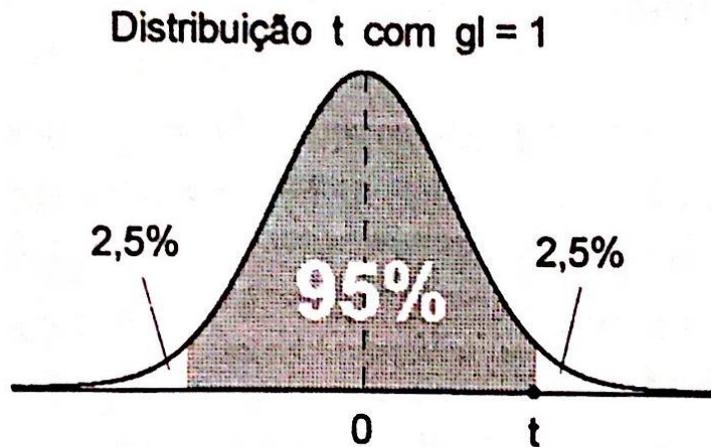
- Encontrar gl e o nível de confiança
  - $gl=1$  e confiança de 95%

$gl$	Área na cauda superior
	... 0,025 ...
1	12,706
...	

→  $t = 12,706$

# Tabela da distribuição t

- Encontrar gl e o nível de confiança
  - gl=1 e confiança de 95%



gl	Área na cauda superior
...	... 0,025 ...
1	12,706
...	

→ t = 12,706

- Intervalo de confiança da média  $\mu$ 
  - $\bar{X} \pm t.S_{\bar{X}}$

**EXEMPLO 9.3** Para verificar a eficácia de um programa de prevenção de acidentes de trabalho, fez-se um estudo experimental, implementando este programa em dez empresas da construção civil, escolhidas ao acaso, numa certa região. Os dados abaixo referem-se aos *percentuais de redução de acidentes de trabalho*, na amostra destas dez empresas em que o programa foi implementado.

Amostra	Estatísticas
20 15 23 11 29 5 20 22 18 17	Média: $\bar{X} = 18$
	Desvio padrão: $S = 6,65$

O objetivo é estimar o parâmetro  $\mu =$  *média da redução percentual de acidentes de trabalho, devido ao programa preventivo, em todas as empresas da construção civil da região*. Podemos obter uma estimativa para o erro padrão da média, como segue:

$$S_{\bar{x}} = \frac{6,65}{\sqrt{10}} = 2,10$$

Usando nível de 95% de confiança, graus de liberdade  $gl = 9$  (pois,  $n = 10$  e  $gl = n - 1$ ), obtemos na Tabela V (apêndice) o valor  $t = 2,262$ , donde podemos calcular o erro máximo provável,  $t \cdot S_{\bar{x}} = (2,262) \cdot (2,10) = 4,75 \approx 4,8$ . Então, temos o seguinte intervalo de 95% de confiança para o parâmetro  $\mu$ :

$$18,0 \pm 4,8 \text{ pontos percentuais}^9.$$

## *Tamanho da população conhecido*

- Conhecendo o tamanho da população,
  - Pode-se refinar a estimativa de erro padrão ?
    - Sim. Mas a correção só é relevante se a população for pequena
- A fórmula do fator de correção é

$$\sqrt{\frac{N - n}{N - 1}}$$

## *Tamanho da população conhecido*

- Erro padrão da proporção (corrigido por N)

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

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

**EXEMPLO 9.4 (a)** Vamos refazer o Exemplo 9.3, considerando que existam  $N = 30$  empresas na região. Neste caso:

$$S_{\bar{x}} = \frac{S}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}} = (2,10) \cdot \sqrt{\frac{30-10}{30-1}} = (2,10) \cdot (0,83) = 1,74$$

$$t.S_{\bar{x}} = (2,262) \cdot (1,74) \approx 3,9$$

Resultando no seguinte intervalo de 95% de confiança para a média  $\mu$ :

$$18,0 \pm 3,9 \text{ pontos percentuais.}$$

**EXEMPLO 9.4 (a)** Vamos refazer o Exemplo 9.3, considerando que existam  $N = 30$  empresas na região. Neste caso:

**(b)** E se a população fosse constituída de  $N = 400$  empresas?

Neste caso:

$$S_{\bar{x}} = \frac{S}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}} = (2,10) \cdot \sqrt{\frac{400-10}{400-1}} = (2,10) \cdot (0,99) = 2,08$$

$$t \cdot S_{\bar{x}} = (2,262) \cdot (2,08) = 4,7$$

E o intervalo de 95% de confiança para a média  $\mu$ :

$18,0 \pm 4,7$  pontos percentuais.

## Exercício de Avaliação – 10 (Para Casa)

- 9) A tabela seguinte mostra os valores das médias e desvios padrão da renda familiar, de uma amostra de 120 famílias, do Bairro Saco Grande II, dividida em três localidades. Os dados foram obtidos do Anexo 4A (Cap. 4).

Localidade	Tamanho da amostra	Renda familiar (sal. mín.)	
		média	desvio padrão
Monte Verde	40	8,1	4,3
Pq. da Figueira	42	5,8	2,6
Encosta do Morro	37	5,0	4,5

Construa um intervalo de confiança, ao nível de 95% de confiança, para a renda familiar média de cada localidade. Interprete as estimativas.

# *Exercício de Avaliação – 11 (Na Sala)*

Resolver o exercício do Jupyter notebook Estimação de Media.

Instruções:

1. Fazer o download do Jupyter Notebook (ou do arquivo .r) e enviar por e-mail para o professor.

# *Trabalho Final*

1. Estabeleça um objetivo de análise de dados.
2. Escolha uma ou mais bases de dados do seu interesse.
3. Faça análises estatísticas sobre esta base de dados e apresente argumentos, achados ou conclusões não triviais baseados nos parâmetros calculados por você.
4. Utilize pelo menos 2 das fórmulas apresentadas.
5. Trabalhos utilizando uma base de dados entre 5 mil e 100 mil linhas, valem no máximo 8. Entre 100 mil e 500 mil linhas, 9. E mais do que 500 mil, 10.
6. A base de dados deve ser pública ou acessível a mim.
7. Entregue um relatório com 1 ou 2 páginas contendo o objetivo da análise, um link para a(s) base(s) de dados analisada(s) e os seus achados (podendo conter gráficos se julgar necessário).
8. Anexe o seu código R ao relatório.