

CE008

Introdução à Bioestatística

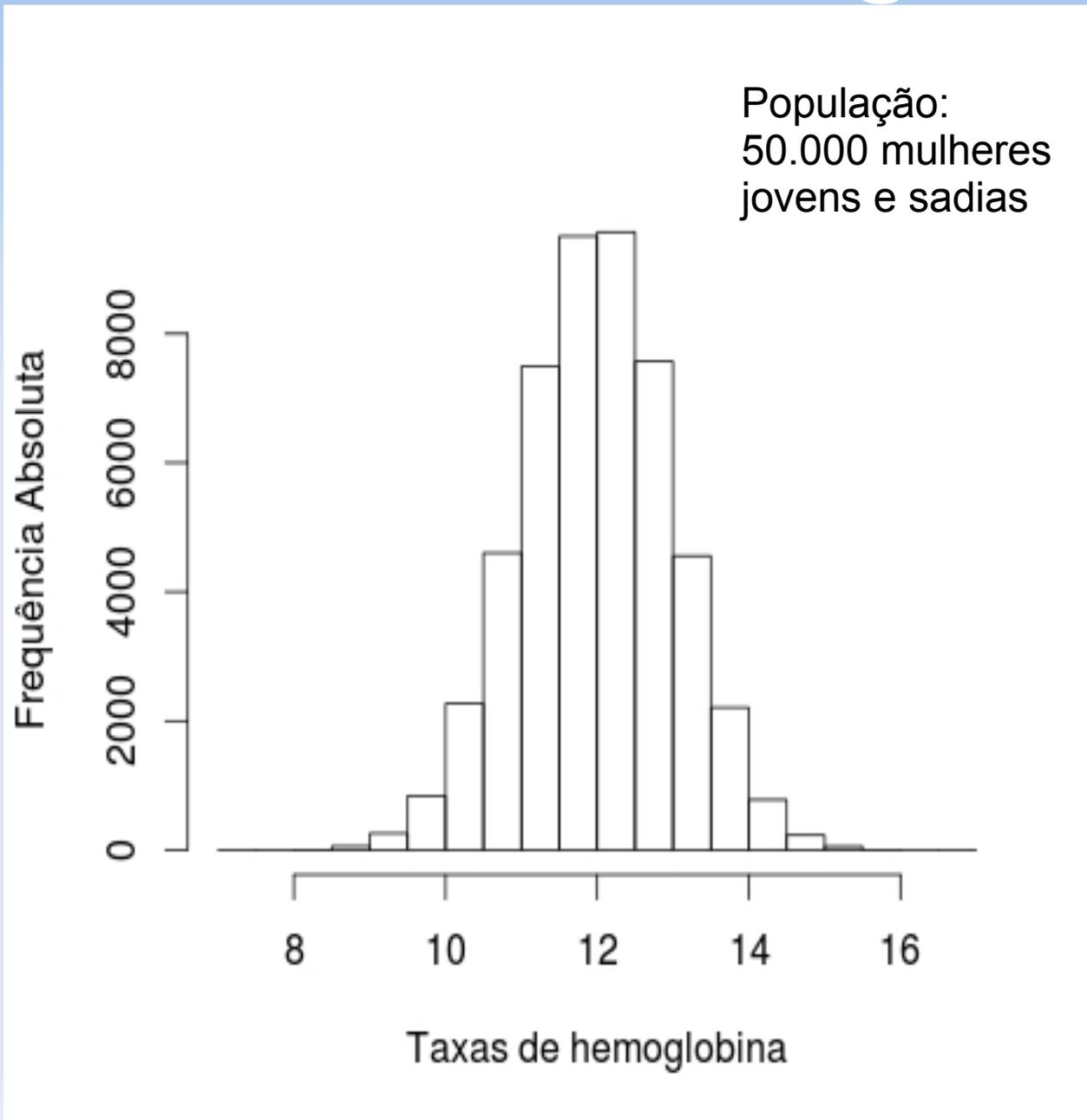
INFERÊNCIA ESTATÍSTICA

Silvia Shimakura

Estimação

- **Amostras** são usadas para **estimar** quantidades desconhecidas de uma **população**.
- **Exemplo:** prevalência de doenças, efeito de medicamentos, diferença entre grupos
- É importante saber qual é a **variação** destas estimativas de amostra para amostra.
- **Teoria de probabilidades** permite usar amostras para **estimar** quantidades de populações, e determinar a **precisão** destas estimativas.

Distribuição das taxas de hemoglobina



- Média=12
- Desvio-padrão=1
- **Na prática a média e o desvio-padrão são desconhecidos!!!**
- Censo inviável ou impossível.
- Conclusões são baseadas numa amostra.

Estimação de médias

- O que acontece quando retiramos uma amostra de uma população e estimamos a média populacional usando a média amostral?
- A estimativa é precisa?
- E se retiramos outra amostra a estimativa da média populacional será diferente?

Amostragem 1

- Uma amostra de tamanho 6 é selecionada da população de taxas de hemoglobina.

Amostra 1	11,75	11,26	11,80	12,95	11,62	10,86
Média 1	11,71					

Amostragem 2

- Seleccionando-se outras 6 mulheres...temos um resultado diferente...

Amostra 1	11,75	11,26	11,80	12,95	11,62	10,86
-----------	-------	-------	-------	-------	-------	-------

Média 1	11,71
---------	--------------

Amostra 2	11,43	12,60	10,86	10,93	12,24	13,76
-----------	-------	-------	-------	-------	-------	-------

Média 2	11,97
---------	--------------

- **A média amostral varia de uma amostra para outra!**

PERGUNTAS

- É possível estimar a média populacional e determinar a precisão da estimativa?
- Existe um comportamento sistemático das médias amostrais?

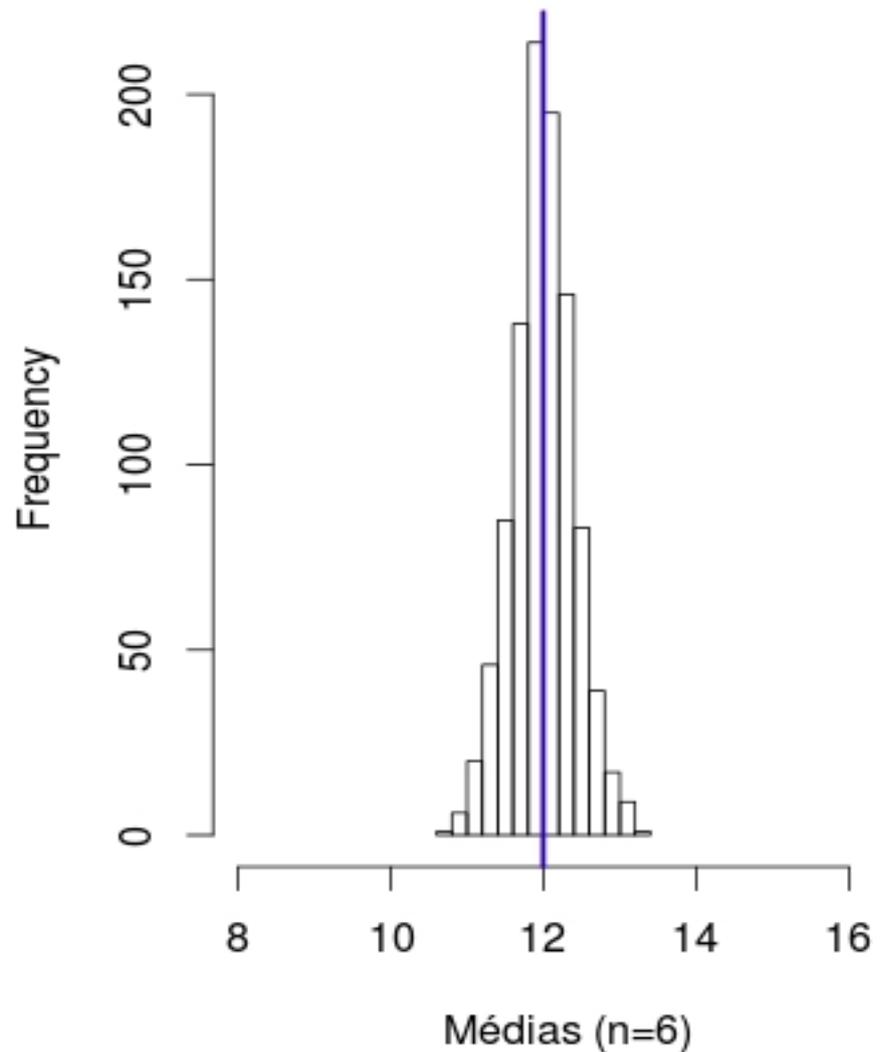
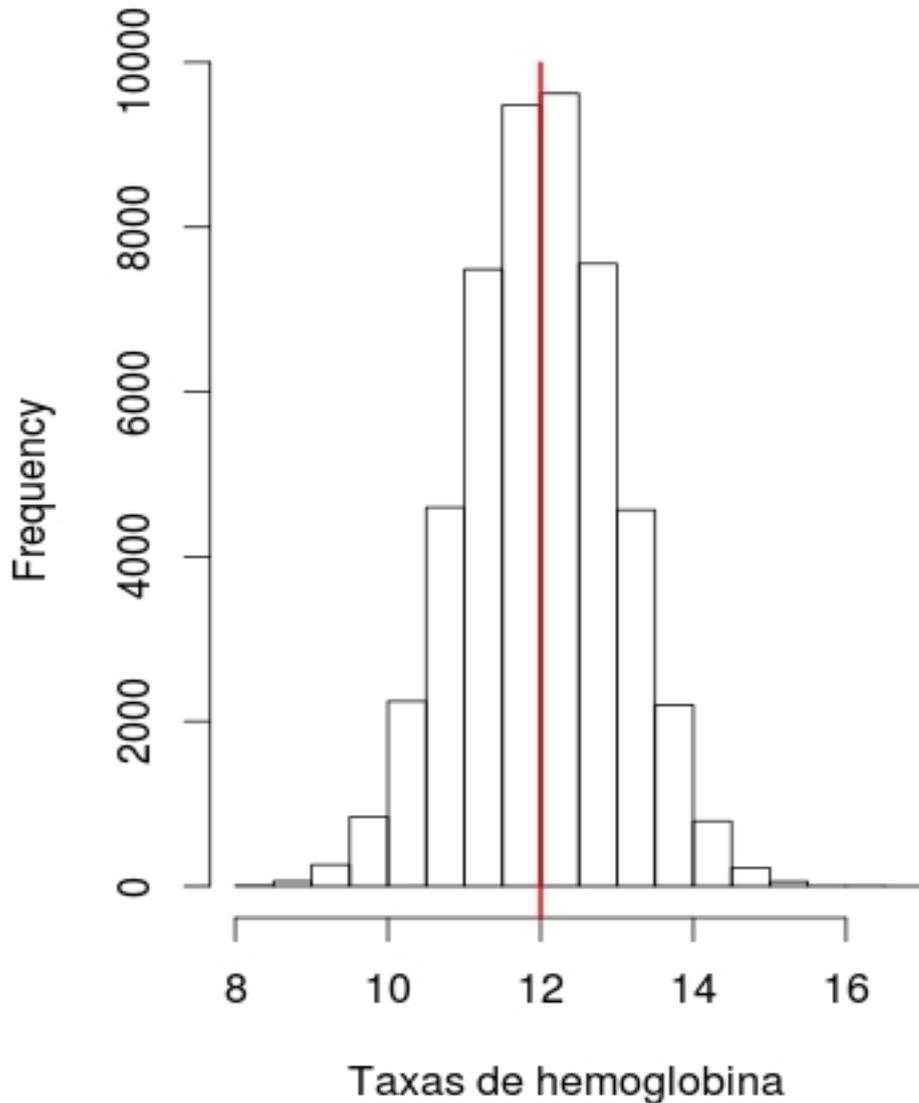
RESPOSTA

- Vamos tentar responder as perguntas com um exercício de simulação.
- Seleccionamos 1000 amostras de 6 mulheres e calculamos as médias amostrais.

Amostra	1	2	3	4	5	6	7	8	9	10
	11,78	11,48	10,91	11,35	11,95	10,95	12,32	12,18	12,41	10,58
	11,46	10,71	11,11	10,42	10,14	11,35	12,25	12,20	14,35	12,74
	13,41	13,06	11,31	13,57	12,01	11,83	11,33	11,50	12,29	10,42
	12,33	11,11	12,66	11,47	13,05	9,81	11,50	11,21	12,31	12,59
	11,02	12,69	11,33	11,75	12,07	12,72	12,29	10,05	13,49	12,21
	12,19	11,62	11,42	12,93	13,12	12,84	10,42	13,61	11,12	11,47
Média	12,03	11,78	11,46	11,92	12,06	11,58	11,69	11,79	12,66	11,67

- As médias amostrais (\bar{X}) variam de acordo com alguma distribuição de probabilidade conhecida?

Distribuição população x média



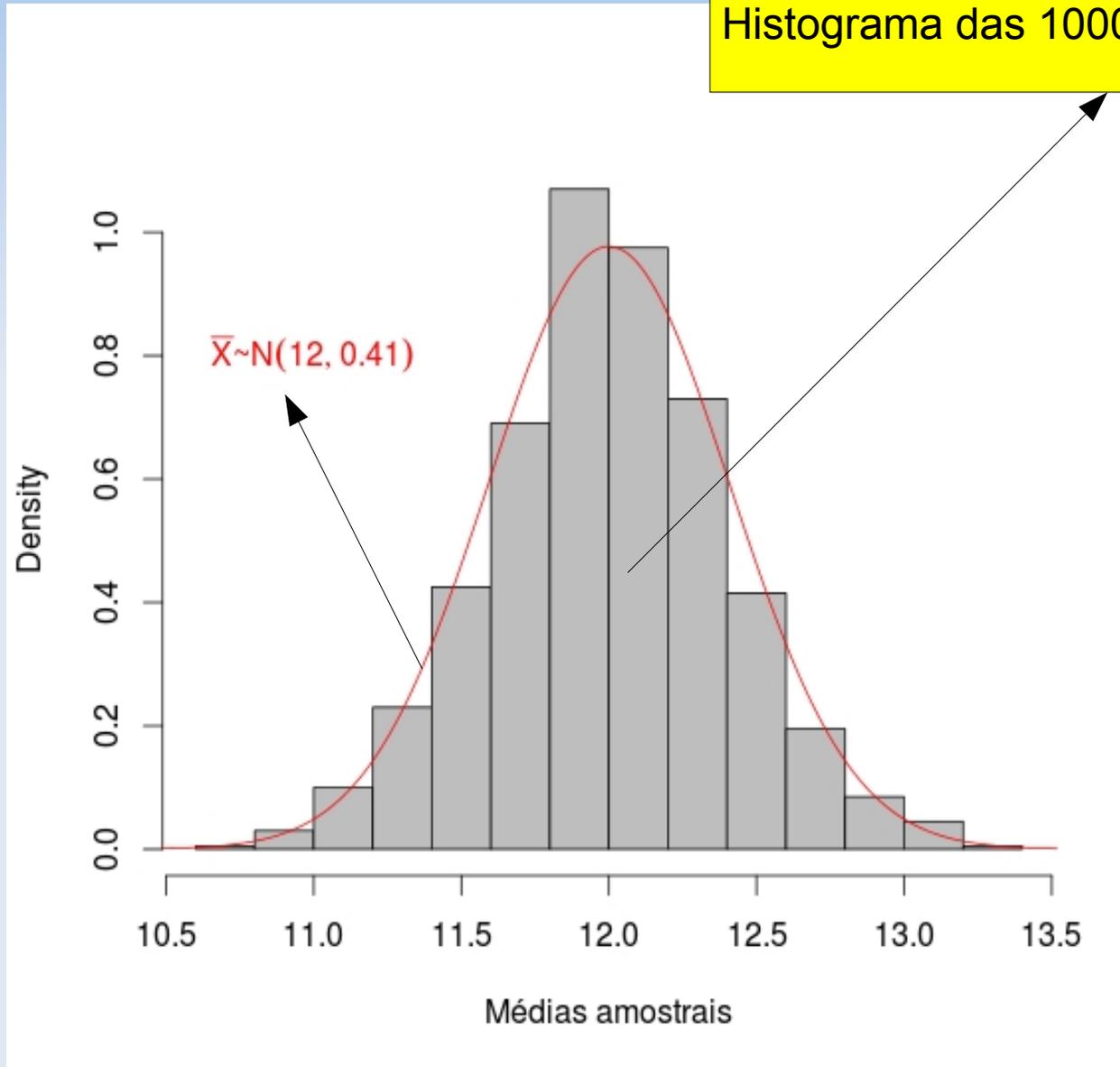
Erro padrão da média amostral

- As 1000 médias podem ser usadas para estimar os parâmetros da distribuição de \bar{X}
- Média das 1000 médias amostrais = 12
- Desvio-padrão das 1000 médias amostrais = 0,41 < 1
- **Teorema Central do Limite**: a distribuição das médias amostrais é Normal com média igual à média da população e desvio-padrão

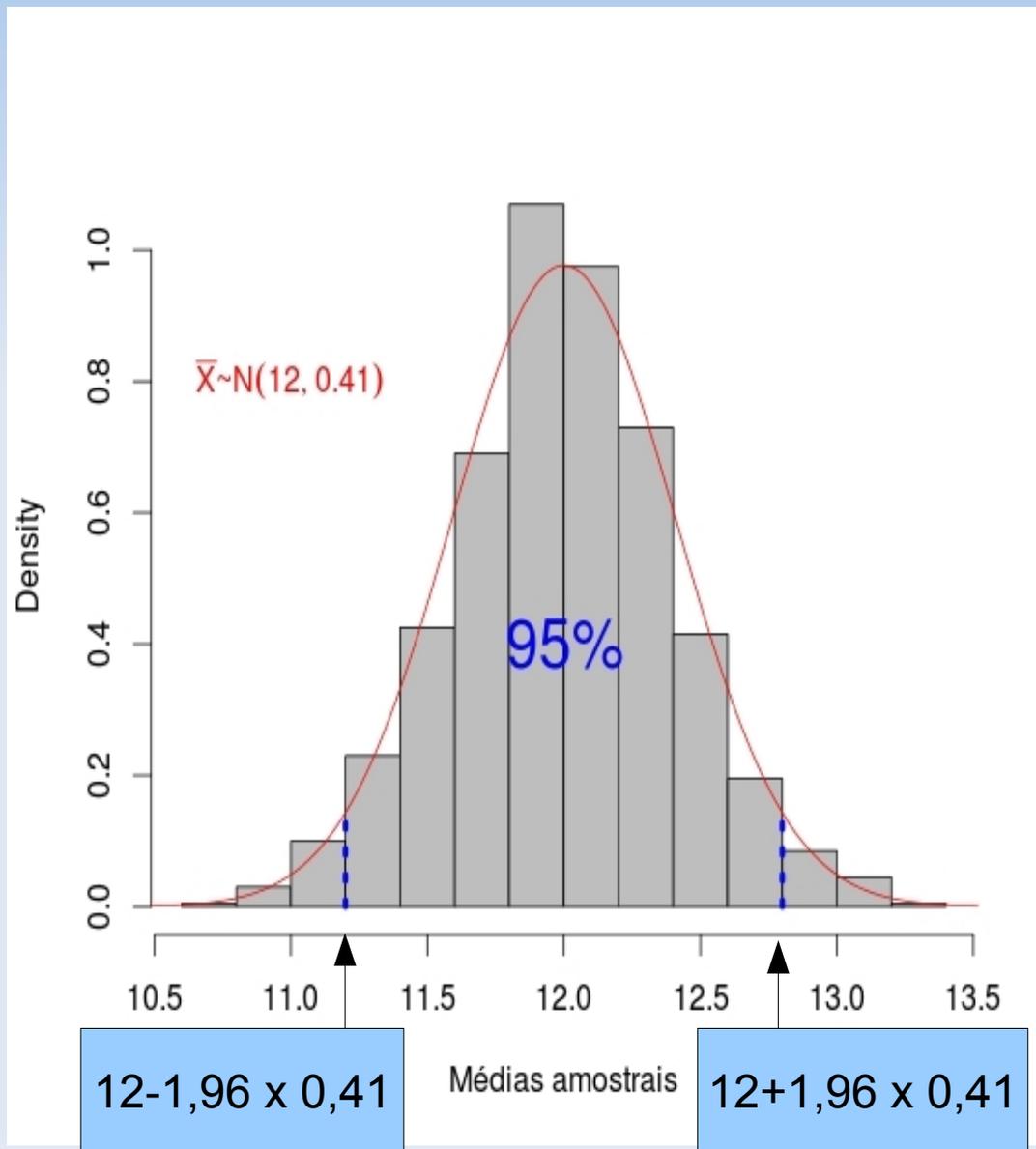
$$\sigma / \sqrt{n} = 1 / \sqrt{6} = 0,41$$

Teorema Central do Limite

Histograma das 1000 médias amostrais



Consequência do TCL



- 95% das médias amostrais estão entre $(12 \pm 1,96 \times 0,41)$

$$P(12 - 1,96 \times 0,41 \leq \bar{X} \leq 12 + 1,96 \times 0,41) = 0,95$$



$$P(\bar{X} - 1,96 \times 0,41 \leq 12 \leq \bar{X} + 1,96 \times 0,41) = 0,95$$

- 95% dos intervalos $(\bar{X} \pm 1,96 \times 0,41)$ cobrem a média populacional 12

Generalizando

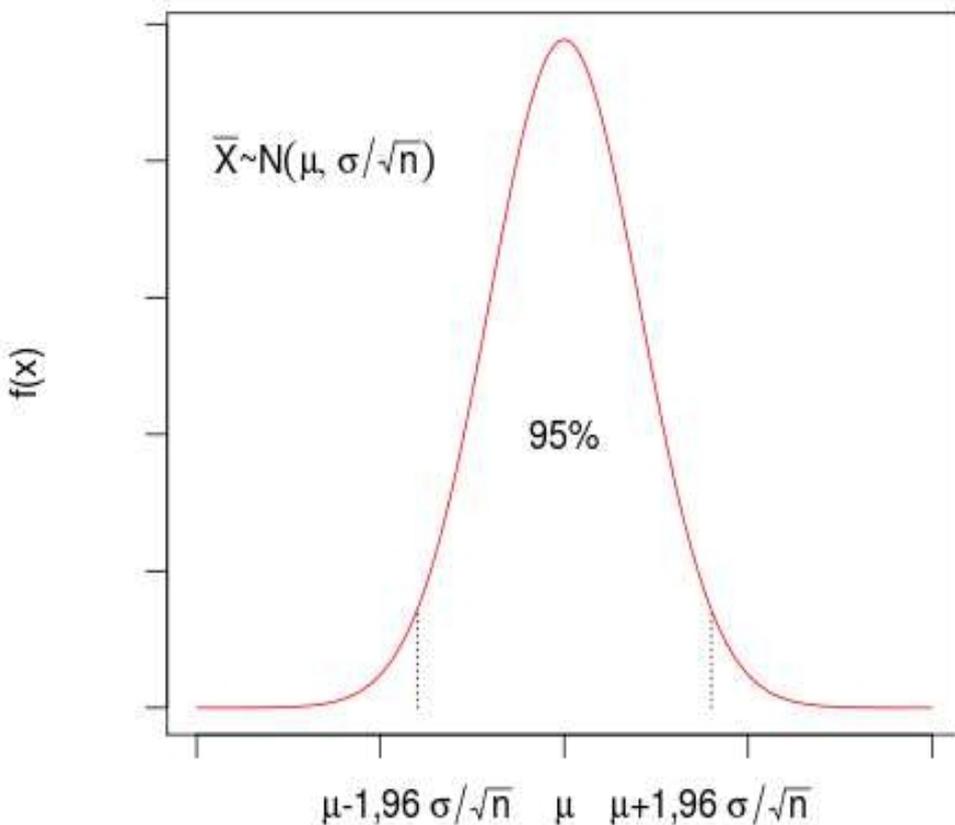
- 95% das médias amostrais estão entre $(\mu \pm 1,96 \sigma/\sqrt{n})$

$$P(\mu - 1,96 \sigma/\sqrt{n} \leq \bar{X} \leq \mu + 1,96 \sigma/\sqrt{n}) = 0,95$$



$$P(\bar{X} - 1,96 \sigma/\sqrt{n} \leq \mu \leq \bar{X} + 1,96 \sigma/\sqrt{n}) = 0,95$$

- 95% dos intervalos $(\bar{X} \pm 1,96 \sigma/\sqrt{n})$ cobrem μ



Teorema central do limite

- Usando este resultado, podemos construir intervalos para estimar a média μ
- IC de 95% para a média populacional μ

$$\left(\bar{X} - 1,96 \frac{\sigma}{\sqrt{n}}, \bar{X} + 1,96 \frac{\sigma}{\sqrt{n}} \right)$$

- Como $\bar{X} \sim N(\mu, \sigma/\sqrt{n})$ então:

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$$

t-Student

- Como $\bar{X} \sim N(\mu, \sigma/\sqrt{n})$ então: $Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$
- Na prática **σ não é conhecido**
- **σ estimado usando s :** $T = \frac{\bar{X} - \mu}{s/\sqrt{n}} \sim t_{n-1}$

t-Student

- IC de 95% para a média populacional μ quando σ é desconhecido

$$\left(\bar{X} - t_{n-1} \frac{s}{\sqrt{n}}, \bar{X} + t_{n-1} \frac{s}{\sqrt{n}} \right)$$

