

# Curso de Data Science



# Aula 05 - Estatística pt.2

# O que você irá aprender nesta aula?

- Teorema do limite central
- Desvio padrão e Erro padrão

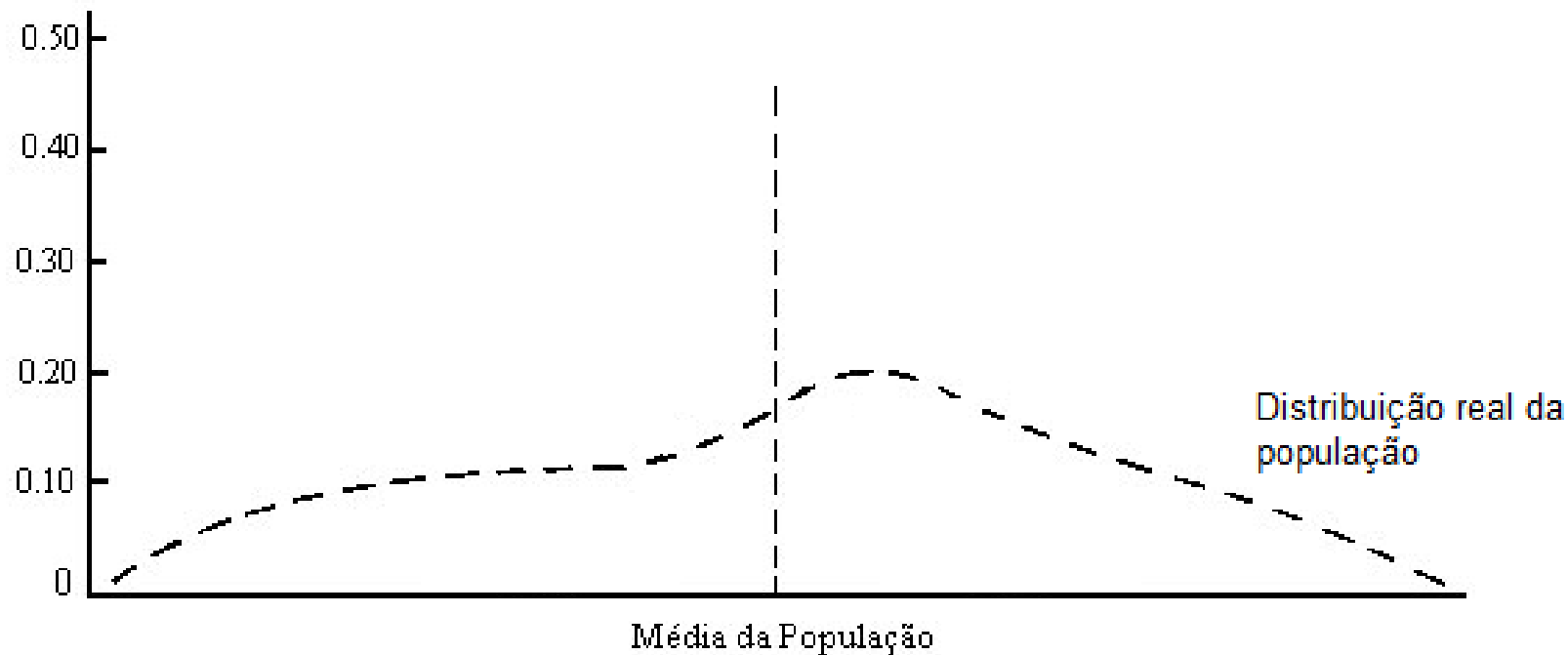
# Teorema do limite central

- Na última aula, foi apresentado que a amostragem é uma boa técnica para estimar as características da população
- O que garante, matematicamente, que a amostragem é uma boa ferramenta estatística é o Teorema do limite central.
- Ainda assim, média da amostra é, em geral, diferente da média da população (parâmetro x estatística do parâmetro)

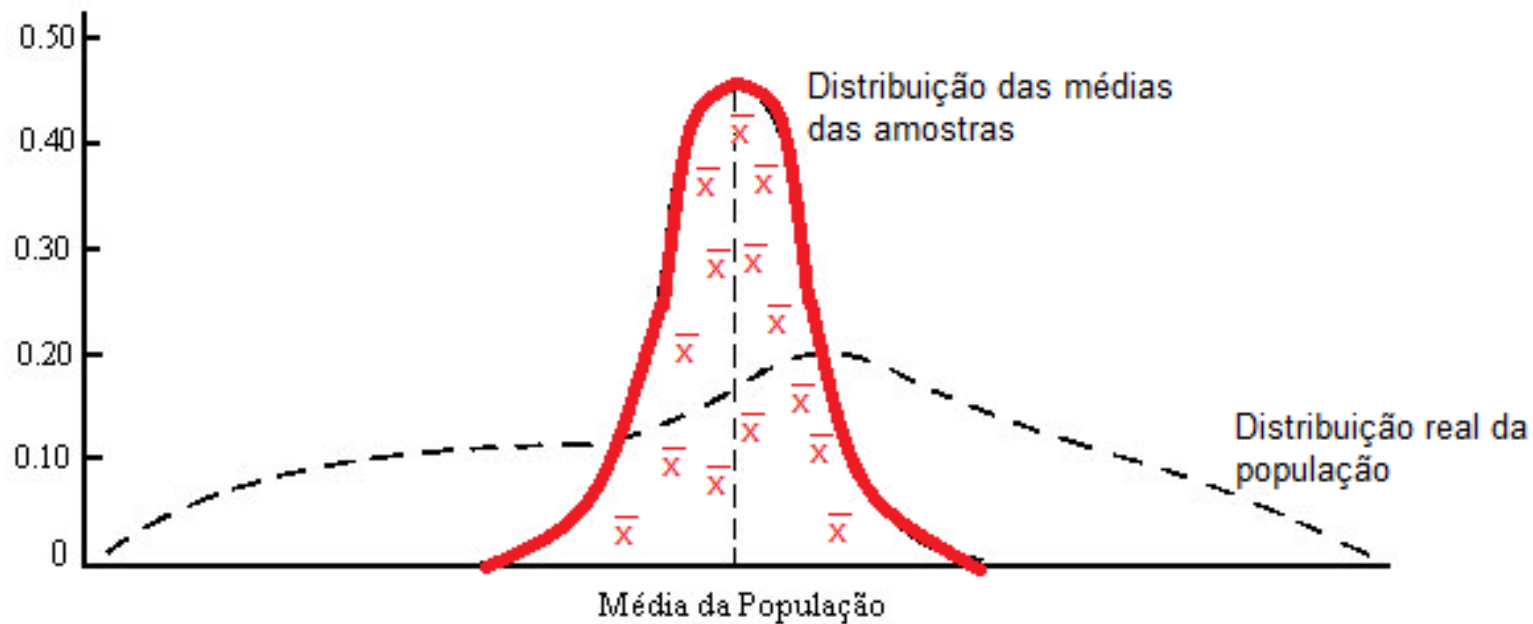
# Teorema do limite central

- O Teorema do limite central afirma que a **média dos valores** de um grupo de amostras será **normalmente distribuída** em torno da **média da população**
- Isso é verdade, não importa a distribuição original da população (contra intuitivo...)
- Exemplo 1 no Jupyter Notebook da aula

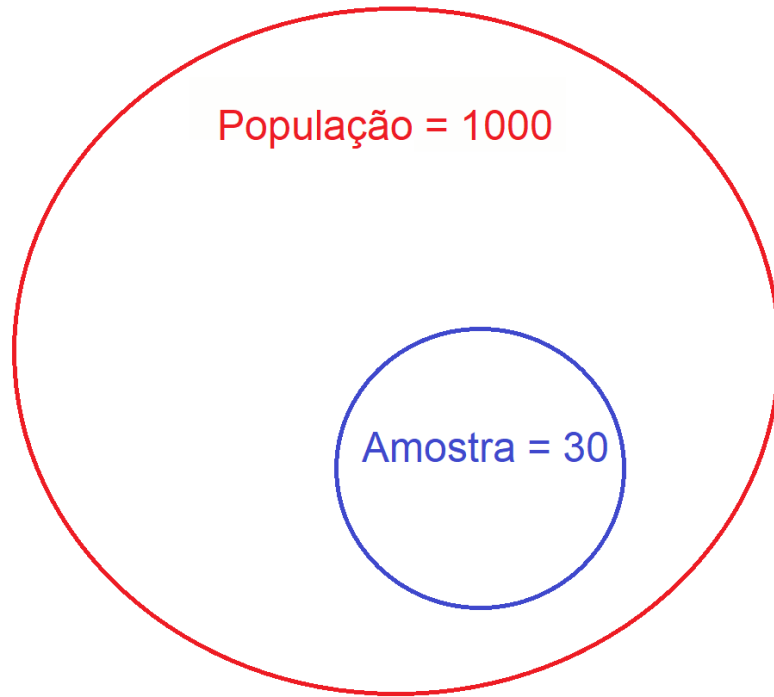
# Teorema do limite central



# Teorema do limite central



# Desvio padrão e Erro padrão



$N$  = # de membros da população  
 $P$  = parâmetro da população  
 $\sigma$  = desvio padrão da população

$n$  = # de membros da amostra  
 $\hat{p}$  = estatística da amostra  
 $SE$  = erro padrão da amostra



# Desvio padrão e Erro padrão

$\sigma$  = desvio padrão da população

Medida do quanto os valores dos parâmetros de cada indivíduo da população estão distantes da média populacional

Analogia: média de idade da população é 30 anos, idade do Roberto é 24 anos.

SE = erro padrão da amostra

Medida do quanto a média da amostra está distante da média populacional

Analogia: média de idade da população é 30 anos, mas da nossa amostra é 33 anos.

## Desvio padrão e Erro padrão

$\sigma$  = desvio padrão da população      SE = erro padrão da amostra

$$SE = \frac{\sigma}{\sqrt{n}}$$

## Desvio padrão e Erro padrão

Exemplo 2: Um teste de QI é idealizado de forma a ter uma pontuação média de 100, com desvio padrão de 15 pontos.

Se uma amostra de 10 pontuações tem média de 104, calcule o erro padrão da amostra.

$$SE = \frac{\sigma}{\sqrt{n}}$$

## Desvio padrão e Erro padrão

Exemplo 2: Resolução

$$SE = 15/\text{sqrt}(10)$$

$$SE = 4.743$$

$$SE = \frac{\sigma}{\sqrt{n}}$$

## Desvio padrão e Erro padrão

Exemplo 2: Conclusão

$$\hat{p} = 104$$

$$SE = 4.743$$

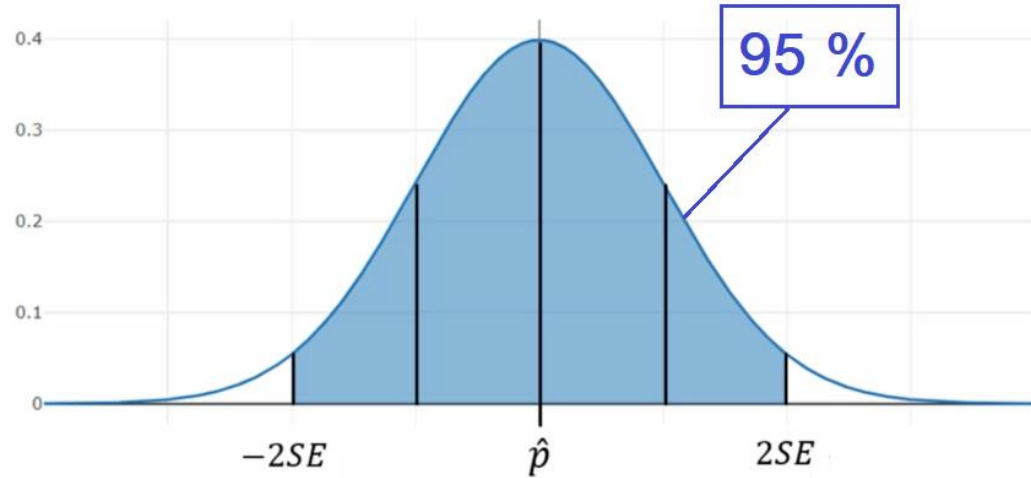
O que isso nos diz sobre a inferência do parâmetro populacional, à partir da média amostral e do erro padrão amostral?

$$SE = \frac{\sigma}{\sqrt{n}}$$

# Desvio padrão e Erro padrão

## Exemplo 2: Conclusão

Podemos afirmar, com um nível de confiança de 95%, que a média da população encontra-se no intervalo de confiança de mais ou menos 2 erros padrões da média amostral.



## Desvio padrão e Erro padrão

Exemplo 2: Conclusão

$$\hat{p} = 104, SE = 4.743, 2*SE = 9.486$$

$$\text{Intervalo: } 104 - 9.486 = 94.514$$

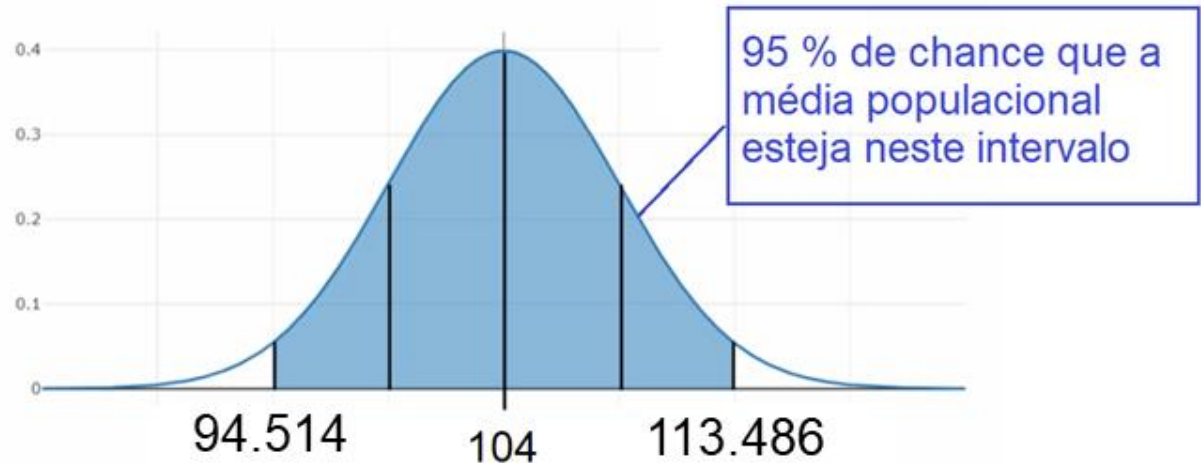
$$104 + 9.486 = 113.486$$

No nosso caso, afirmamos com 95% de nível de confiança que a média populacional se encontra entre 94.514 e 113.486

# Desvio padrão e Erro padrão

## Exemplo 2: Conclusão

Erro padrão da amostra “conecta” a amostra com população!





# Desvio padrão e Erro padrão

Lembrete:

POPULAÇÃO:

PARÂMETRO: característica da população (pontuação da população)

AMOSTRA:

ESTATÍSTICA: característica da amostra (pontuação da amostra)

## Então, nesta aula vimos:

- Teorema do limite central
- Desvio padrão e Erro padrão

# Muito obrigado!