

  
**FIAP**

# Statistics for Machine Learning

## Aula 19: Tópicos de Amostragem e Intervalos de Confiança

---

**Prof. Jones Egydio**

[profjones.egydio@fiap.com.br](mailto:profjones.egydio@fiap.com.br)



# Objetivos

- Introduzir os conceitos de Amostragem e Intervalos de Confiança;
- Formas de representação;
- Exemplos e exercícios;
- Conclusão;
- Perguntas.

# Conceitos iniciais

- Amostragem
- Intervalos de confiança





# Conceitos

The background of the slide features a complex pattern of interlocking puzzle pieces. On the left side, there are several concentric circular gear-like shapes, also composed of puzzle pieces, creating a mechanical or interconnected visual theme.

- Censo, População e amostra
- Tipos de amostras
- Margem de erro e intervalo de confiança
- Determinação do tamanho da amostra
- Conceito de hipótese
- Construção e riscos



# Censo, população e amostra

- **Censo** é um levantamento de todas as informações relacionadas a todos os elementos de uma determinada população.

## O Censo Demográfico 2020 do IBGE



Todos os aviões de um país



Todos os automóveis de uma cidade



Censo, população e amostra

- **População** é formada por todas as observações do universo de referência.
- O **Tamanho** será denotado por  $N$ .

# Tipos de população

- **População Finita:** pode ser quantificada numericamente
- **População Infinita:** não pode ser quantificada numericamente (gerados num processo contínuo).



**População finita:** número de alunos da FIAP.



**População infinita:** número de habitantes no planeta.



## Alguns aviões do país



## Alguns automóveis da cidade



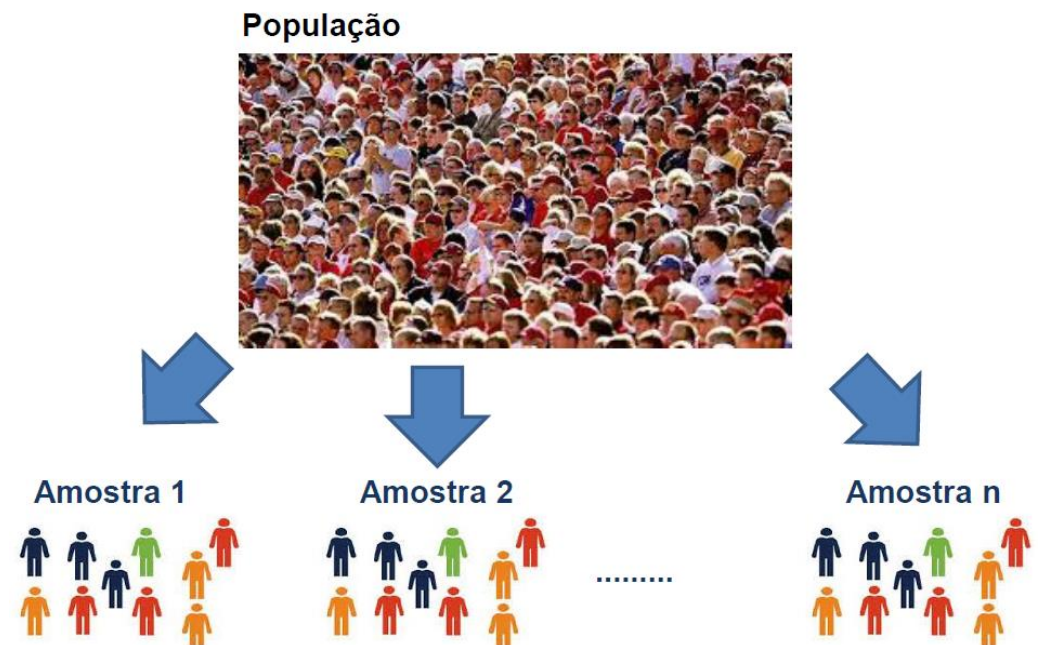
## Censo, população e amostra

- **Amostra** é formada por alguma parte da população.
- O **tamanho** será denotado por  $n$ .



# Amostragem

- **Amostragem** é o processo de obtenção de amostras de uma determinada população.



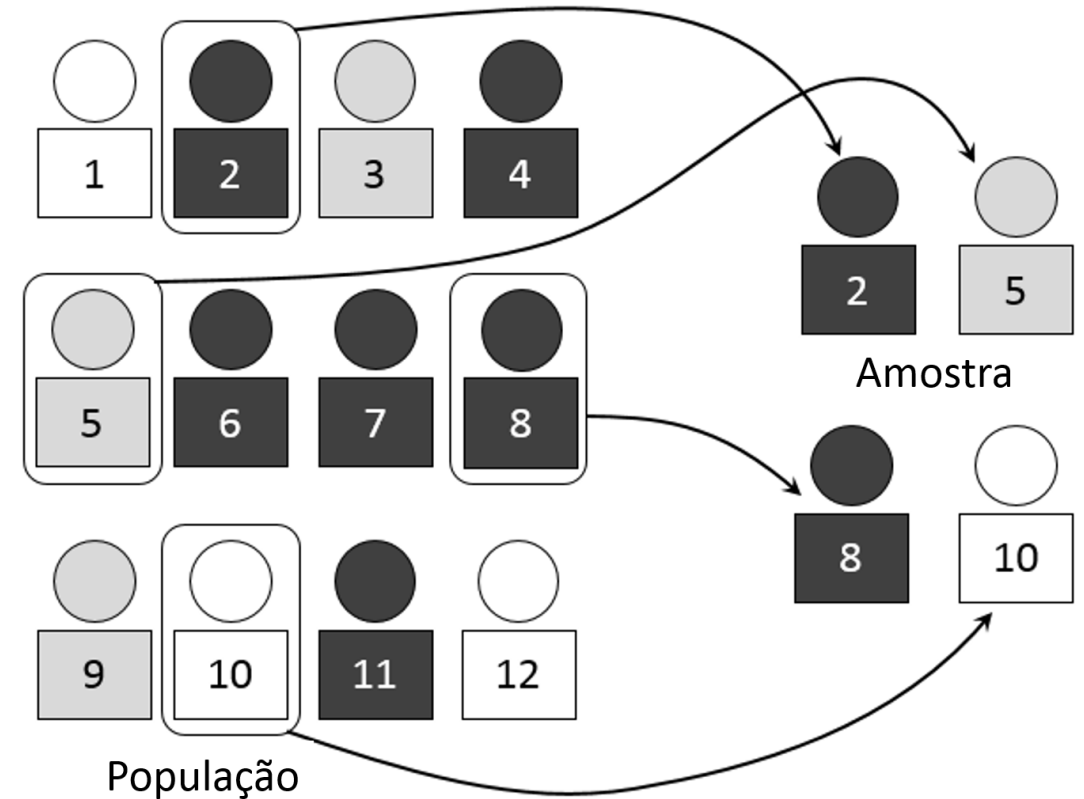
# Tipos de amostra

- **Amostra aleatória simples**
- **Amostra aleatória sistemática**
- **Amostra aleatória estratificada**



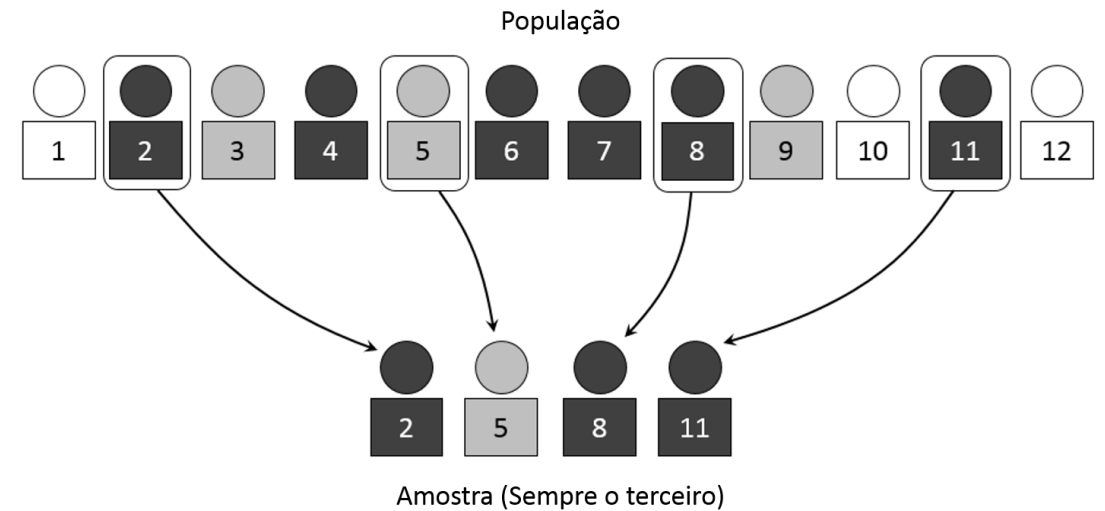
# Amostra aleatória simples

- É uma amostra de  $n$  elementos extraída de uma população de  $N$  elementos, em que os elementos são selecionados ao acaso.



# Amostra aleatória sistemática

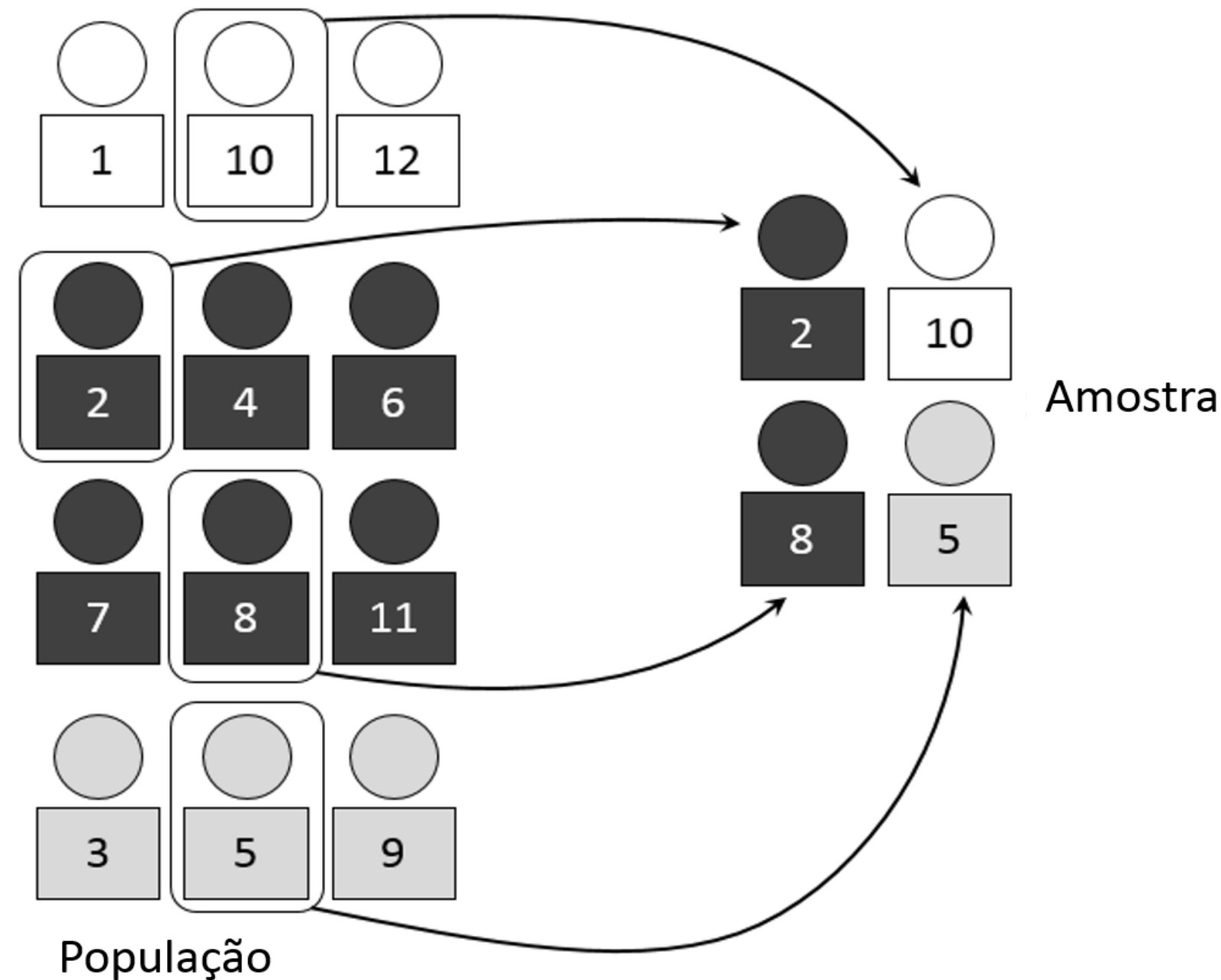
- É um esquema amostral em que os elementos são selecionados por um critério em que deve ser aplicado em toda população.

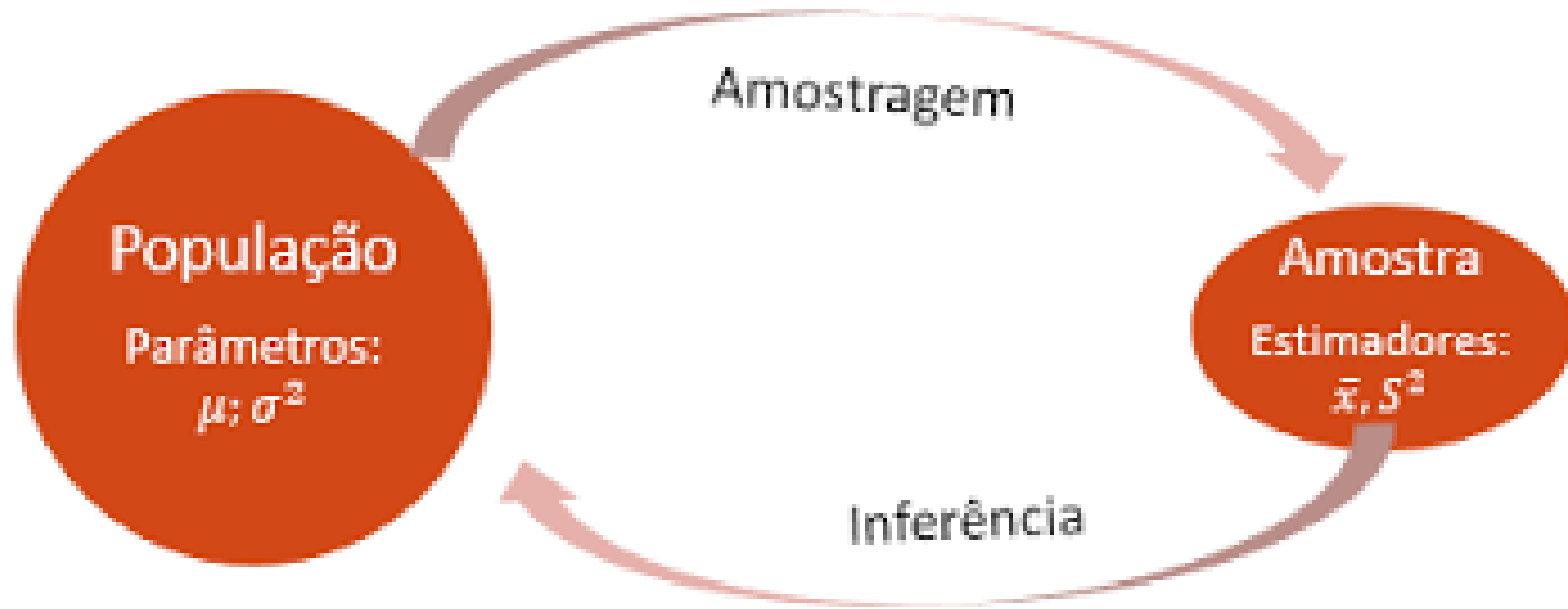




## Amostra aleatória estratificada

- É um processo de amostragem em que divide-se uma população em extratos e retira-se cada elemento aleatoriamente de cada extrato.





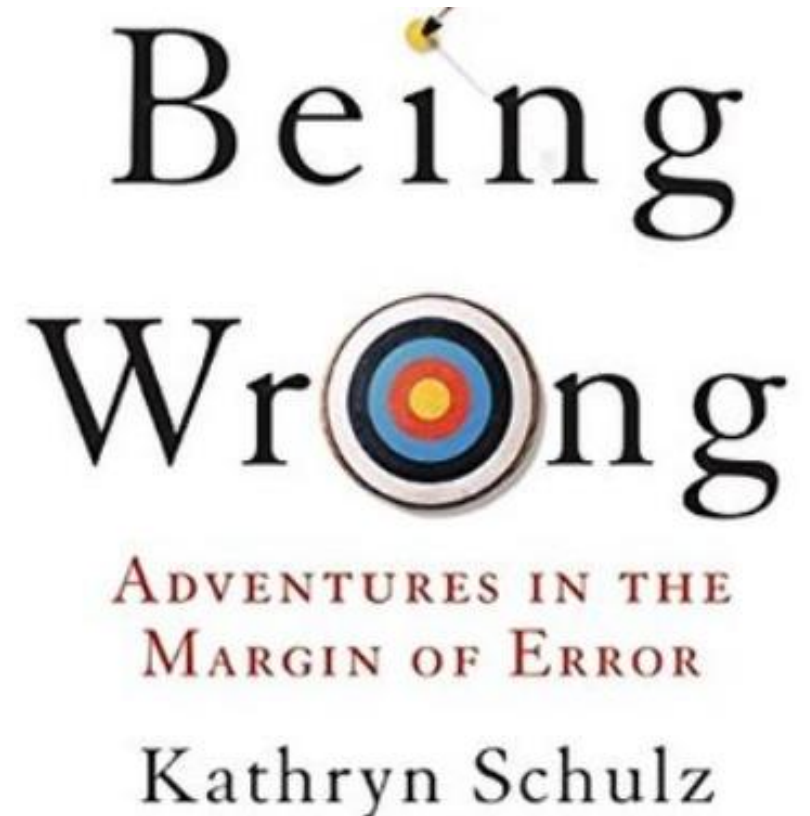
## População e amostra

- Fazemos amostragem para estimar parâmetros.
- Depois utilizamos estes parâmetros para realizar inferências sobre a população.



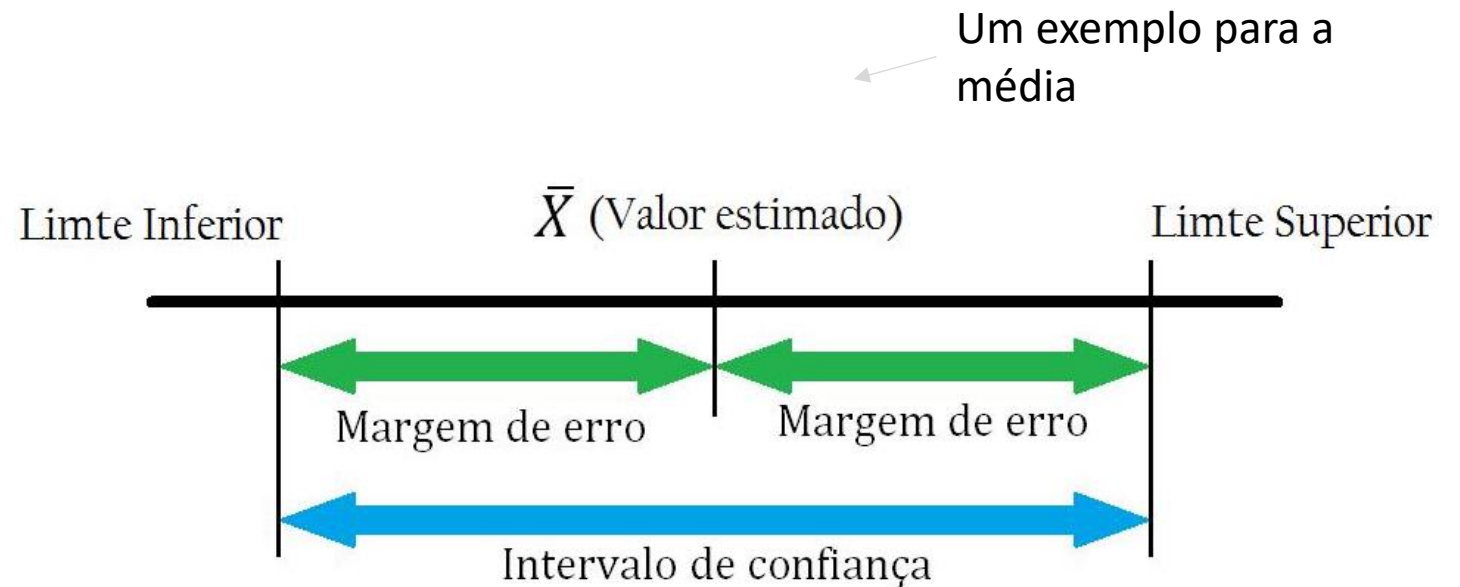
# Margem de erro

- Exemplo: “esta pesquisa pode variar 2 pontos percentuais para mais ou para menos.”
- A margem de erro auxilia a mensurar quanto uma medida amostral está próxima da **verdadeira medida da população** estudada.



# Intervalo de confiança

- É um intervalo estimado por meio de um parâmetro de interesse. O intervalo de confiança apresentará um conjunto de estimativas prováveis (dentro de um nível de confiança) para um resultado.





# Determinação do tamanho da amostra

- Qual o tamanho da amostra que precisamos retirar da população?
- Precisaremos do seguinte:
- Nível de confiança (90%, 95% ou 99%)
- Margem de erro (depende da variável em estudo)



$$n = \frac{Z^2 \cdot S^2 \cdot N}{(ME)^2 \cdot (N - 1) + Z^2 \cdot S^2}$$

## Determinação do tamanho da amostra

Tamanho da amostra quando a variável de estudo é contínua

- **ME**: margem de erro
- **N**: tamanho da população
- **Z**: valor da distribuição normal para o nível de confiança adotado
- **S**: variância amostral obtida por uma amostra piloto.

$$n = \frac{(1,96)^2 \cdot (250.000) \cdot (15.000)}{(100)^2 \cdot (15000 - 1) + (1,96)^2 \cdot (250.000)} = 95$$

## Determinação do tamanho da amostra

**Exemplo:** renda média de uma cidade

- **N: 15.000** habitantes
- **ME: 100** reais
- **Z: 1,96** – valor da distribuição normal para nível de confiança de 95%
- **S:** desvio padrão da renda da amostra piloto – **500** reais.

$$n = \frac{Z^2 \cdot 0,25 \cdot N}{(ME)^2 \cdot (N - 1) + Z^2 \cdot 0,25}$$

## Determinação do tamanho da amostra

Tamanho da amostra quando a variável de estudo é discreta.

- ME: margem de erro
- N: tamanho da população
- Z: valor da distribuição normal para o nível de confiança adotado



$$n = \frac{(1,96)^2 \cdot 0,25 \cdot (8.000)}{(0,03)^2 \cdot (8.000 - 1) + (1,96)^2 \cdot 0,25} = 941,62 = 942$$

## Determinação do tamanho da amostra

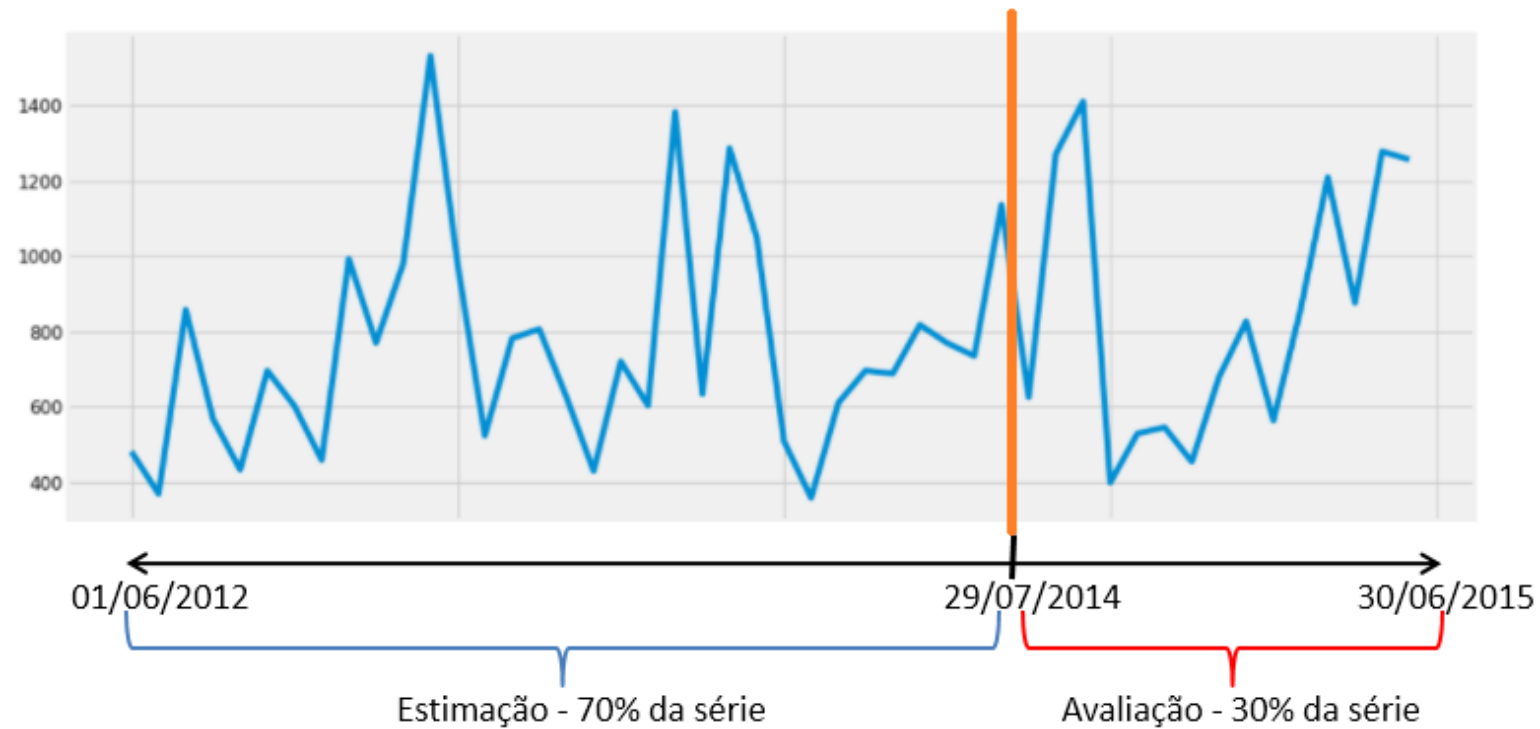
**Exemplo:** pesquisa intenção de voto.

- **N: 8.000** habitantes
- **ME: 3%**
- **Z: 1,96** – valor da distribuição normal para nível de confiança de 95%

# Tipos de amostra: série de tempo

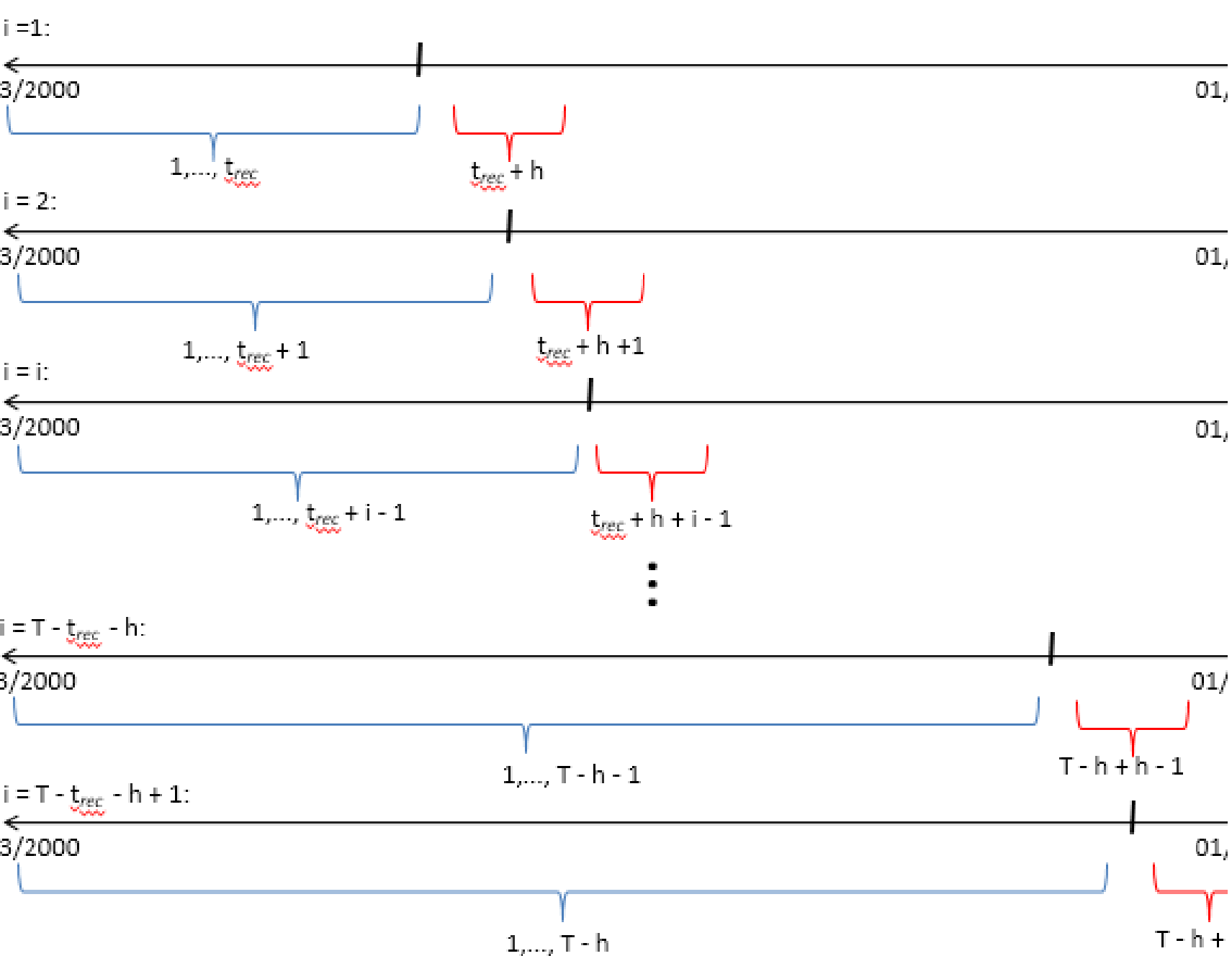
- Janela fixa
- Janela recursiva  
(*recursive/expanding window*)
- Janela deslizante  
(*rolling/sliding window*)

<https://eng.uber.com/forecasting-introduction/>



## Tipos de amostra: série de tempo

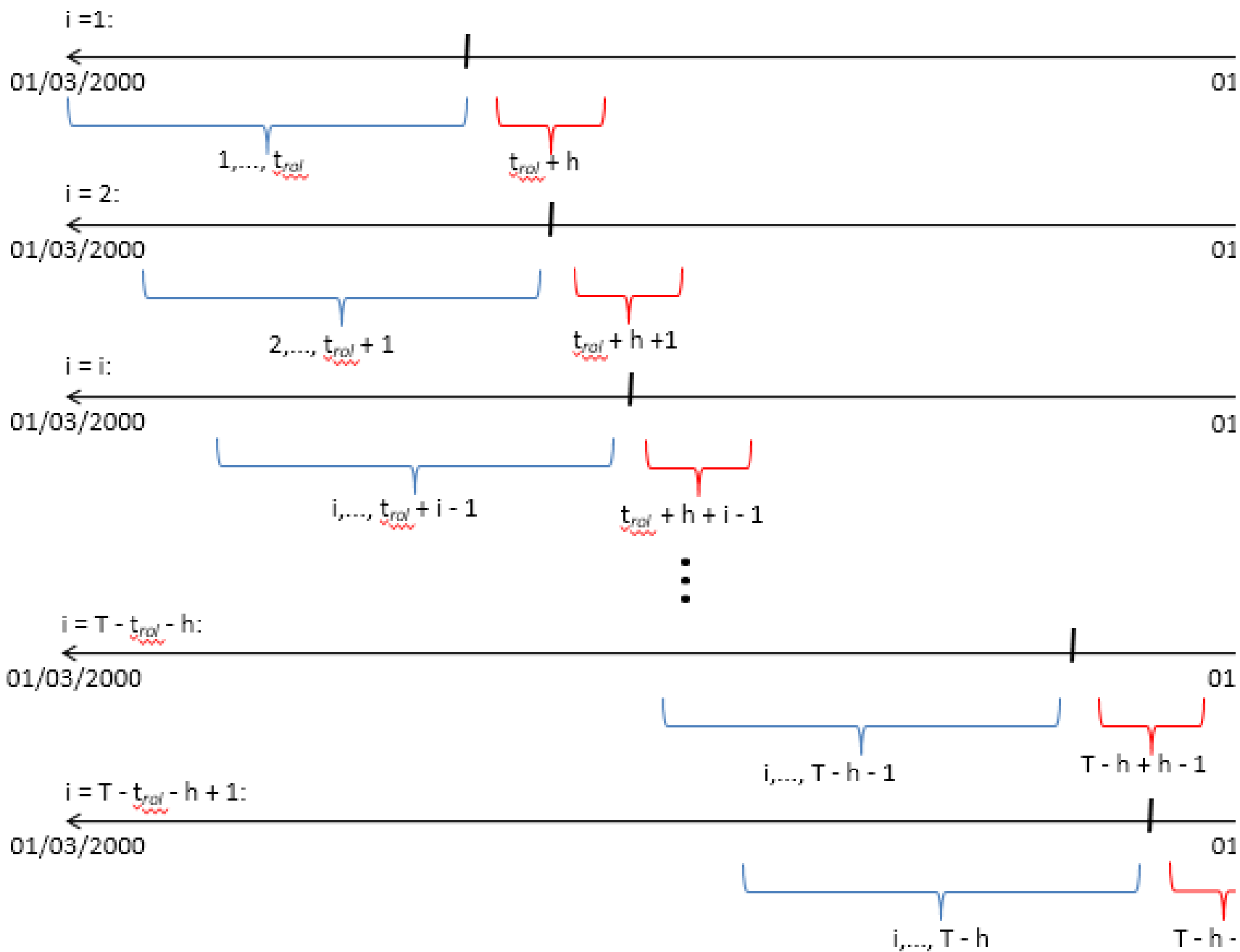
- Janela fixa



## Tipos de amostra: série de tempo

- Janela recursiva (*recursive/expanding window*)



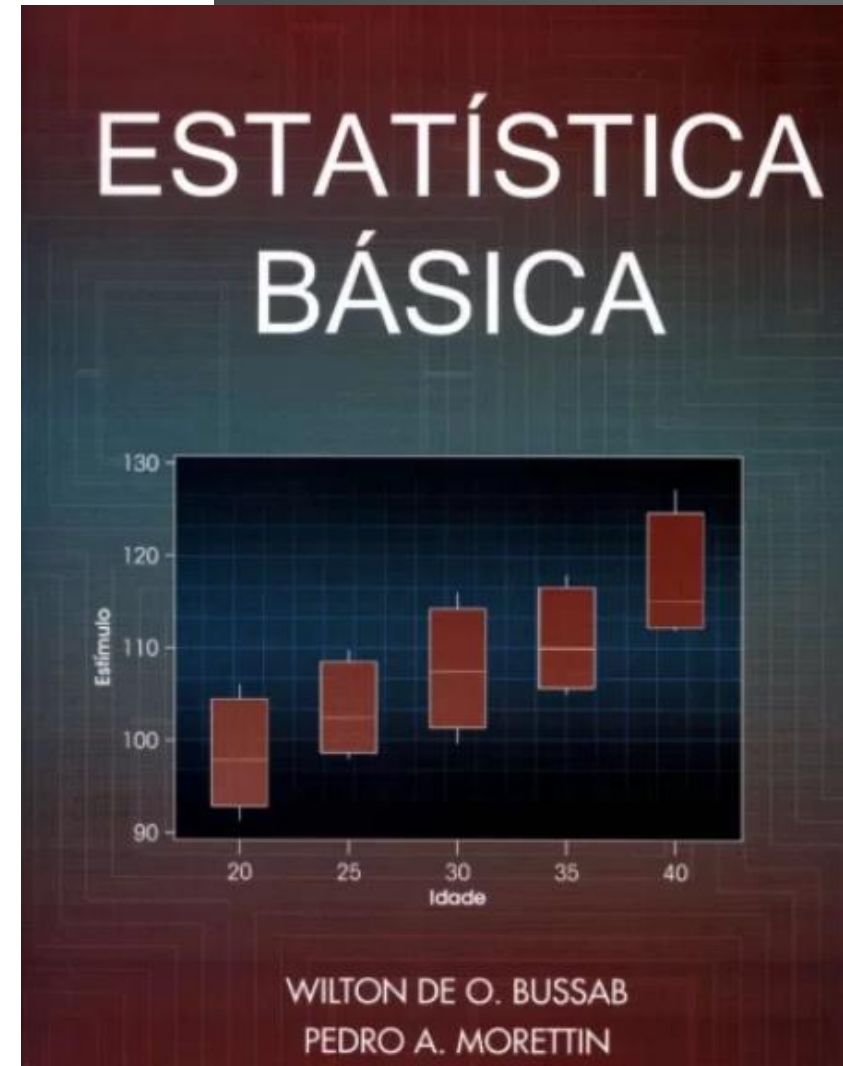


## Tipos de amostra: série de tempo

- Janela deslizante (*rolling/sliding window*)

# Onde estudar mais!!

- Vídeos
- Margem de erro: <https://pt.khanacademy.org/math/ap-statistics/estimating-confidence-ap/introduction-confidence-intervals/v/confidence-intervals-and-margin-of-error>
- Intervalo de confiança: <https://pt.khanacademy.org/math/statistics-probability/confidence-intervals-one-sample/estimating-population-proportion/v/confidence-interval-example>
- Hipótese: <https://pt.khanacademy.org/math/statistics-probability/significance-tests-one-sample/more-significance-testing-videos/v/hypothesis-testing-and-p-values>



Obrigado!