



# **3 – ESTATÍSTICA INFERENCIAL**

# ÁREAS DA ESTATÍSTICA

1

## Estatística Descritiva

(usada no processo de análise)

Descreve como os dados  
estão organizados.

2

## Probabilidade

(usada nos algoritmos de machine  
learning)

Mede a variabilidade de  
fenômenos casuais de  
acordo com a sua ocorrência.

3

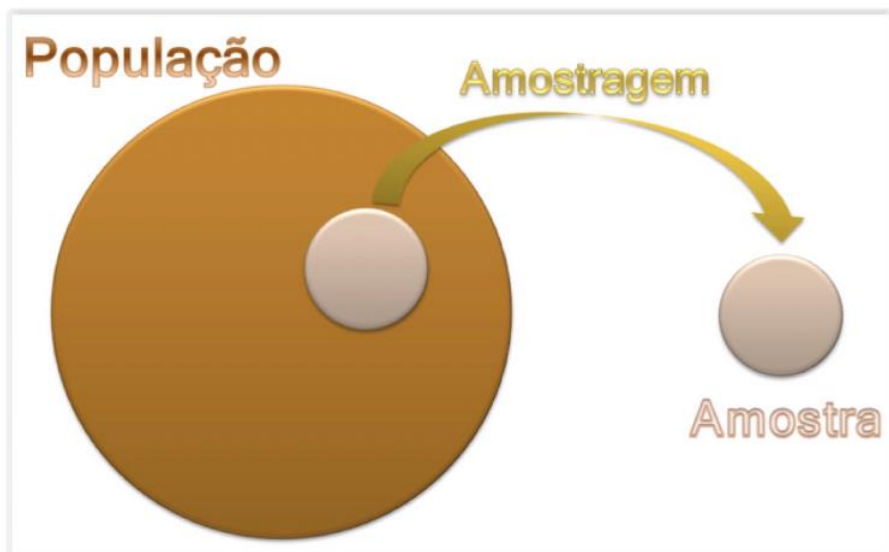
## Estatística Inferencial

(usada para fazer inferencias sobre uma  
população/amostra)



# ESTATÍSTICA INFERENCIAL

- É o processo de estimar informações sobre uma população a partir dos resultados observados em uma amostra.



A estatística inferencial tem como objetivo a extrapolação dos resultados (obtidos com a estatística descritiva) para a população.

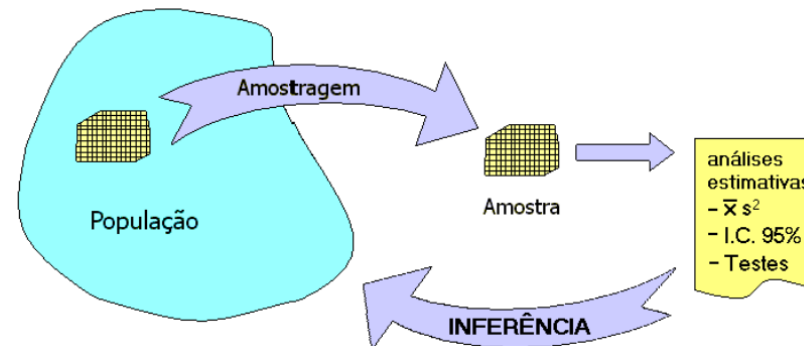
- **População:** é o conjunto de todos os elementos ou resultados sob investigação.
- **Amostra:** é qualquer subconjunto da população.

# PRINCÍPIOS DA AMOSTRAGEM

- **Amostragem** é o processo de determinação de uma amostra a ser pesquisada, a amostra é uma parte de elementos selecionada de uma população.
- A teoria da amostragem estuda as relações existentes entre uma população e as amostras extraídas dessa população. É útil para avaliação de grandezas desconhecidas da população, ou para determinar se as diferenças observadas entre duas amostras são devidas ao acaso ou se são verdadeiramente significativas.

## Termos básicos da Amostragem:

- População
- Unidade
- Amostra
- Variável



# TIPOS DE AMOSTRAGEM

## Machine Learning

### Métodos Aleatórios

1. Amostragem Aleatória Simples
  2. Amostragem Sistemática
  3. Amostragem Estratificada
  4. Amostragem por Aglomerados
  5. Amostragem Multi-etapas
  6. Amostragem Multifásica
- + usadas

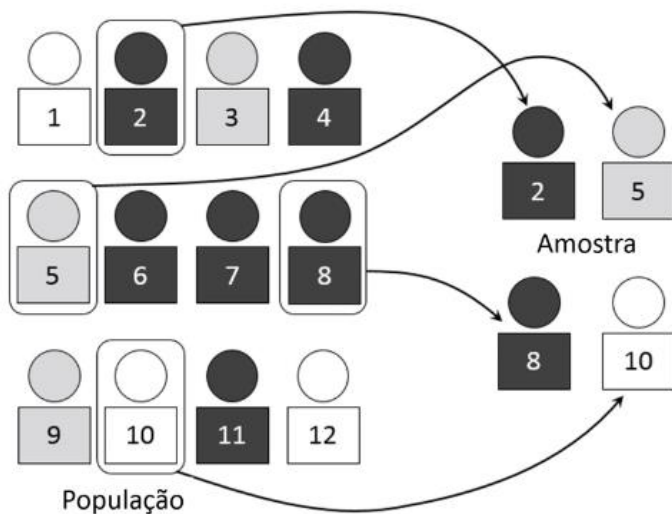
### Métodos Não Aleatórios

1. Amostra Intencional
2. Amostra "Snowball"
3. Amostra por quotas
4. Amostra por conveniência

# TIPOS DE AMOSTRAGEM

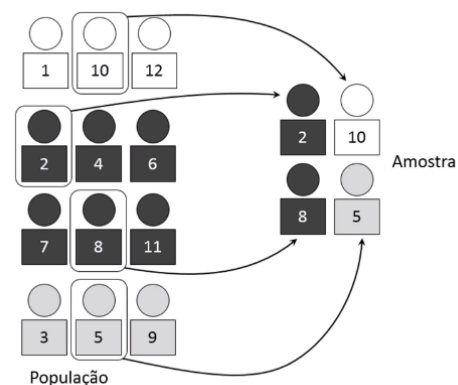
## 1. Amostragem Probabilística ou Aleatória

- Amostragem Aleatória Simples (dados de treino e teste)
- Amostragem Aleatória Simples sem reposição
- Amostragem Aleatória Simples com reposição

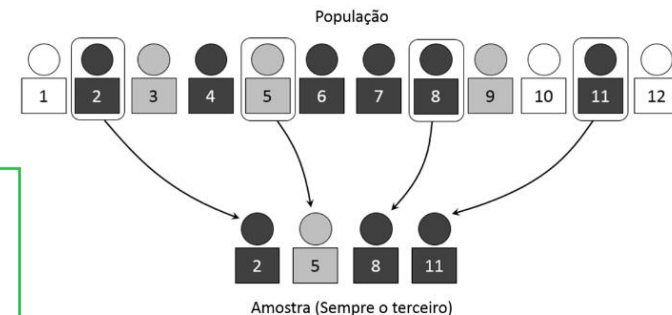


## 3. Amostragem

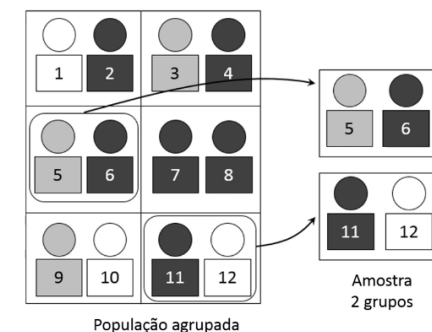
### Estratificada



## 2. Amostragem Sistemática

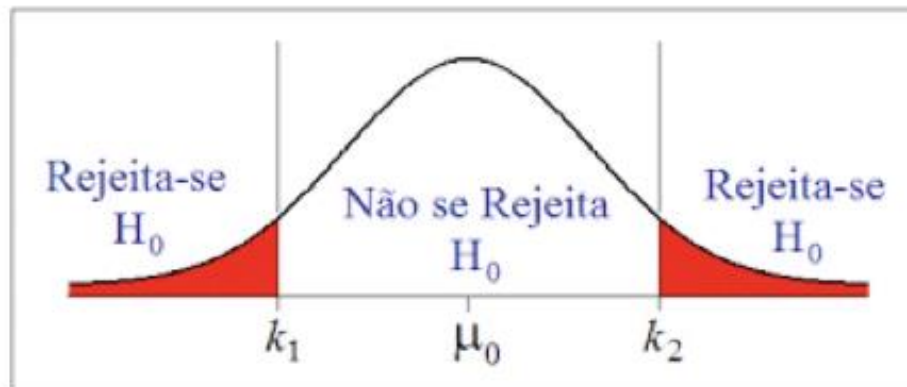


## 4. Amostragem Por Conglomerados



# TESTE DE HIPÓTESE

- Uma hipótese estatística é uma suposição sobre um determinado parâmetro da população, como media, desvio padrão, coef de correlação, etc. Um teste de hipótese é um procedimento para decisão sobre a veracidade ou falsidade de uma determinada hipótese.
- Um **Teste de Hipótese Estatística** é um procedimento de decisão que nos possibilita decidir entre  $H_0$  (hipótese nula) ou  $H_a$  (hipótese alternativa), com base nas informações contidas na amostra.



# TESTE DE HIPÓTESE

## ➤ $H_0$

A **hipótese nula** afirma que o parâmetro da população (como a media, o desvio padrão, e assim por diante) é igual a um valor hipotético. A hipótese nula, é muitas vezes, uma alegação inicial baseado em análises anteriores ou conhecimentos especializados.

## ➤ $H_a$

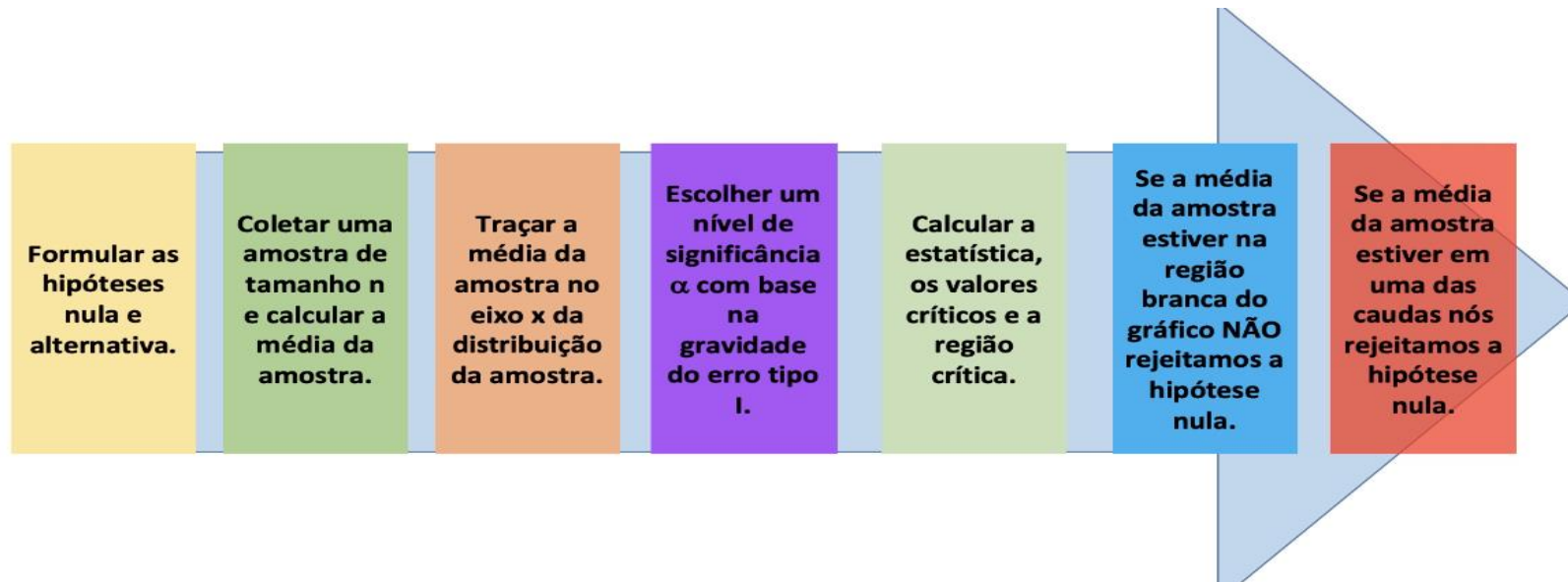
A **hipótese alternativa** afirma que um parâmetro da população é menor, maior ou diferente do valor hipotético na hipótese nula. A hipótese alternativa é aquela que você acredita que pode ser verdadeira ou espera provar ser verdadeira.



# TESTE DE HIPÓTESE

Como estamos analisando dados da amostra, erros podem ocorrer:

- **Erro Tipo I** é a probabilidade de rejeitarmos a hipótese nula quando ela é efetivamente verdadeira.
- **Erro Tipo II** é a probabilidade de rejeitarmos a hipótese alternativa quando ela é efetivamente verdadeira.

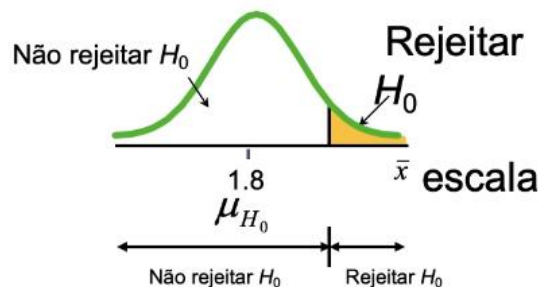


# TESTE DE HIPÓTESE UNILATERAL

- O teste Unilateral ou Unicaudal é usado quando a hipótese alternativa é expressa como:  $<$  ou  $>$

$$\begin{aligned} H_0: \mu &= 1.8 \\ H_A: \mu &> 1.8 \end{aligned}$$

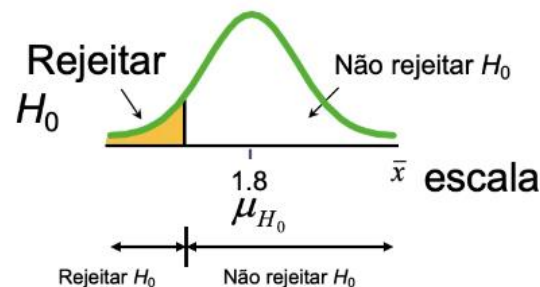
**Teste Cauda Superior:** nós assumimos que  $\mu = 1.8$  a menos que a média da amostra seja maior que the 1.8



Teste Unilateral Direito

$$\begin{aligned} H_0: \mu &= 1.8 \\ H_A: \mu &< 1.8 \end{aligned}$$

**Teste Cauda Inferior:** nós assumimos que  $\mu = 1.8$  a menos que a média da amostra seja menor que 1.8



Teste Unilateral Esquerdo

Resumindo:

Teste Unilateral Esquerdo:  
(Inferior)

$H_0: \mu = \text{valor numérico}$

$H_A: \mu < \text{valor numérico}$

Teste Unilateral Direito:  
(Superior)

$H_0: \mu = \text{valor numérico}$

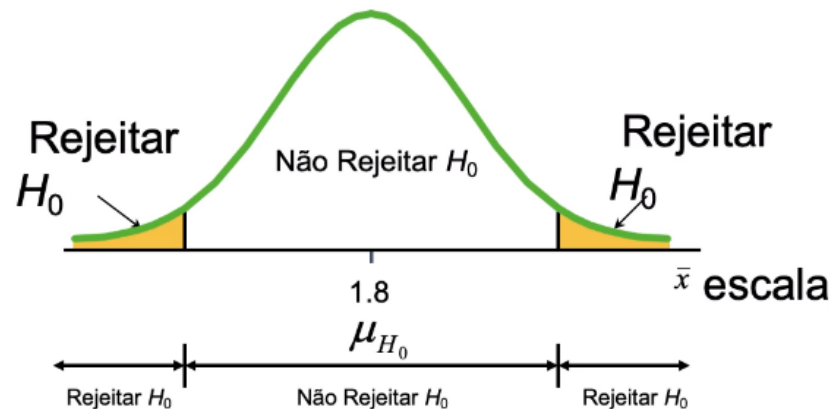
$H_A: \mu > \text{valor numérico}$

# TESTE DE HIPÓTESE BILATERAL

- O teste Bilateral é usado sempre que a hipótese alternativa é expressa como  $\neq$  de.

$$H_0: \mu = 1.8$$
$$H_A: \mu \neq 1.8$$

Nós assumimos que  $\mu = 1.8$  a menos que a média da amostra seja  $\neq$  que 1.8



A curva acima representa a distribuição da amostragem da média de utilização de banda larga. Assume-se que a média da população é 1.8 GB, de acordo com a hipótese nula  $H_0: \mu = 1.8$ .

Resumindo:

$$H_0: \mu = \text{Valor numérico.}$$

$$H_A: \mu \neq \text{Valor numérico.}$$

Por existirem duas regiões de rejeição no gráfico (regiões em amarelo), este é chamado teste de hipótese **bilateral** ou **bicaudal**.

Como a hipótese nula é expressa como  $\neq$  ela pode ser maior ou menor que, por isso o teste é **bilateral**.

# ERROS TIPO I E TIPO II

- Como estamos tomando **amostra** como base, estamos expostos ao **risco** de conclusões **erradas** sobre a **população**, por conta de **erros de amostragem**.
- A **hipótese nula** pode ser verdadeira, caso tenhamos coletado uma amostra que não seja representativa da população. Ou talvez, a amostra tenha sido muito pequena.

Para testar a  $H_0$ , é preciso definir uma regra de decisão com o objetivo de estabelecer uma zona de rejeição da hipótese, ou seja, definir um nível de significância,  $\alpha$ , sendo os mais consensuais os alfas 0.10, 0.05 e 0.01.

Grau de Confiança	Nível de Significância
90%	0,10
95%	0,05
99%	0,01

Nível de significância

## Resumindo:

Condição		A Hipótese Nula é Verdadeira	A Hipótese Nula é Falsa
D E C I S Ã O	Decidimos rejeitar a hipótese nula.	Erro Tipo I (Rejeição de uma hipótese nula verdadeira)	Decisão correta
	Não rejeitamos a hipótese nula.	Decisão correta	Erro Tipo II (Não rejeição de uma hipótese nula falsa)

## Exemplo

# ERROS TIPO I E TIPO II

A eficácia de certa vacina após um ano é de 25% (isto é, o efeito imunológico se prolonga por mais de 1 ano em apenas 25% das pessoas que a tomam). Desenvolve-se uma nova vacina, mais cara e deseja-se saber se esta é, de fato, melhor.

Que hipóteses devem ser formuladas?  
Que erros podemos encontrar?

**Hipótese Nula**  $H_0 : p = 0,25$   
**Hipótese Alternativa**  $H_A : p > 0,25$

**Erro Tipo I : aprovar a vacina quando, na realidade, ela não tem nenhum efeito superior ao da vacina em uso.**

**Erro Tipo II : rejeitar a nova vacina quando ela é, de fato, melhor que a vacina em uso.**

A probabilidade de se cometer um **Erro Tipo I** depende dos valores dos parâmetros da população e é designada por  $\alpha$  (alfa - nível de significância).

Dizemos então que o nível de significância alfa de um teste, é a probabilidade máxima com que desejamos correr o risco de um **Erro Tipo I**.

O valor alfa é tipicamente predeterminado e escolhas comuns são  $\alpha = 0.05$  e  $\alpha = 0.01$

A probabilidade de se cometer um **Erro Tipo II** é designada por  $\beta$ .

End!