

**”** “[...] all models are approximations. Essentially, all models are wrong, but some are useful. However, the approximate nature of the model must always be borne in mind [...]”

— **George Box**. Empirical Model-Building and Response Surfaces. 1987

**”** “Since all models are wrong the scientist must be alert to what is importantly wrong. It is inappropriate to be concerned about mice when there are tigers abroad.”

— **George Box**. Science and statistics. 1976.

# **MATD44 — Amostragem**

## Definição



**Estatística** é a **ciência** que estuda métodos e técnicas de obtenção de **informação quantitativa útil**, em situações caracterizadas pela presença de variação (**incerteza**).

# Alguns conceitos

## População

Um **conjunto** de elementos sobre os quais há interesse em investigar ao menos **uma característica** pertinente.

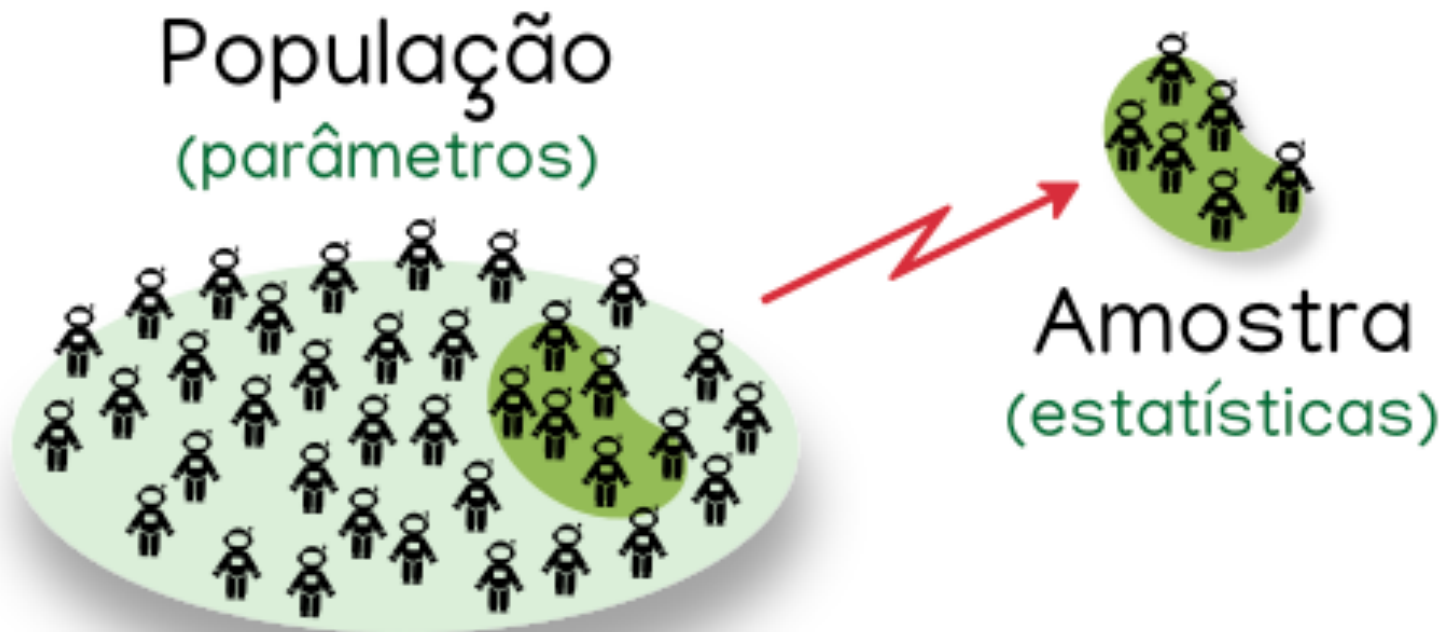
## Amostra

E um **subconjunto** da **população**.

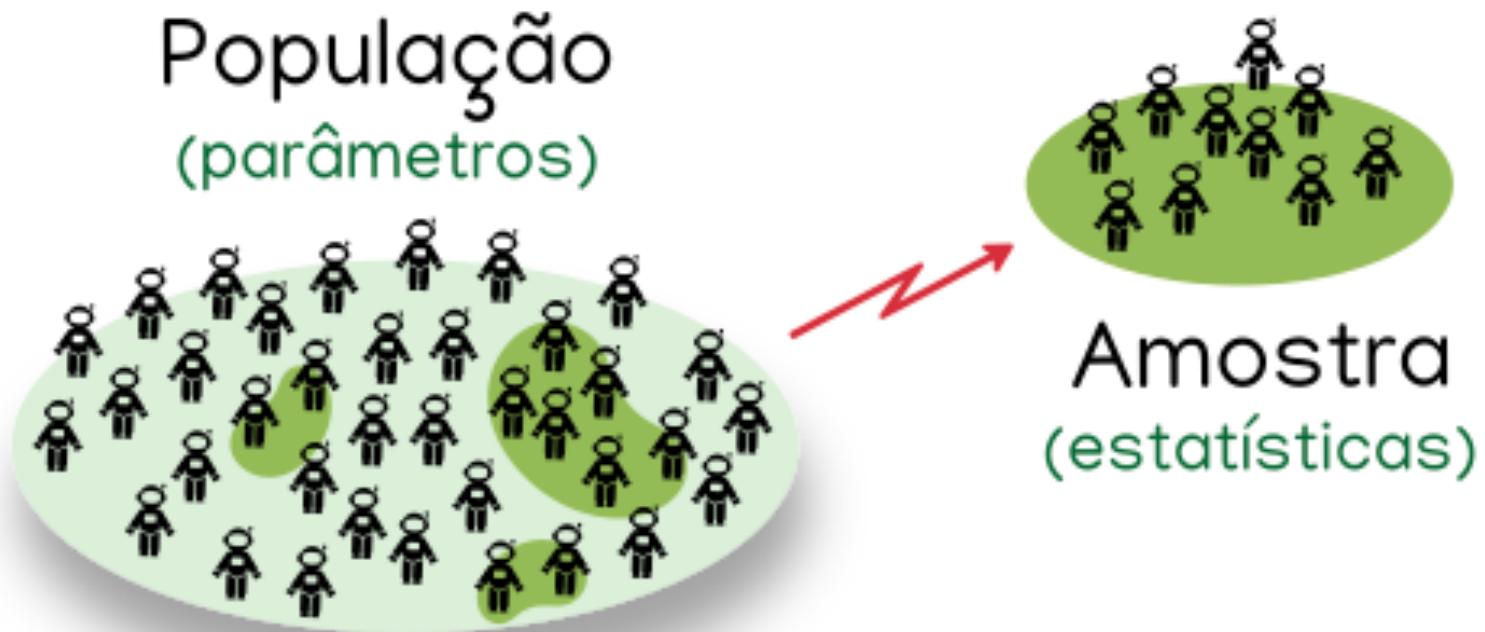
## Estatística Descritiva

Como o próprio nome sugere, é a parte da **Estatística** que estuda métodos e técnicas para **descrever analiticamente** um conjunto de dados (População ou Amostra).

## Conceitos fundamentais

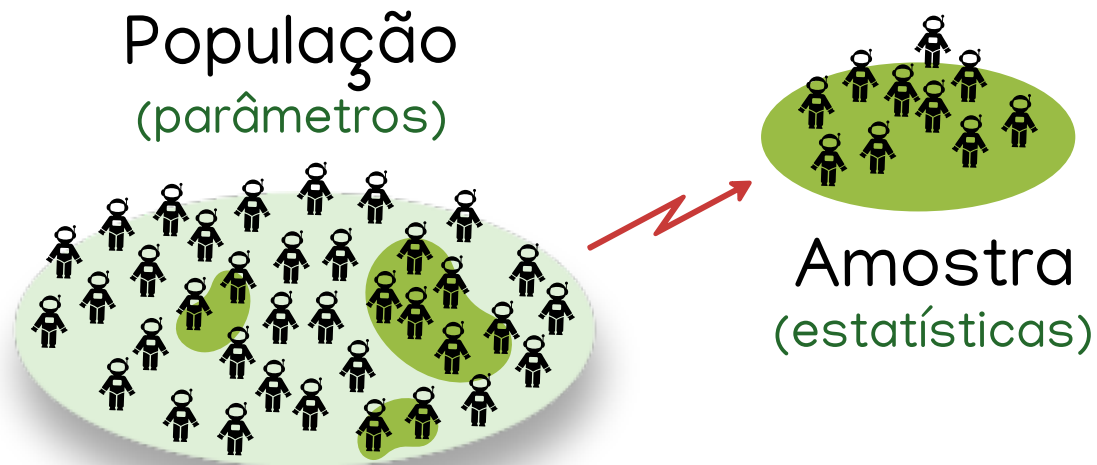


## Conceitos fundamentais



## Parâmetro

Quando técnicas de **estatística descritiva** são aplicados à dados que correspondem a uma **população**, as quantidades observadas são chamadas **parâmetros** e descrevem **características desta população**. Quando aplicadas em uma **amostra**, chamamos de **estatísticas**.



## Por que precisamos de amostragem?

A descrição de qualquer ferramenta estatística começa com algo do tipo  
“Seja  $x_1, x_2, \dots, x_n$  uma amostra aleatória da população...”

Com base nessa amostra, a análise estatística é realizada.

De fato, a estatística só tem utilidade porque pode fornecer inferências estatísticas para toda a população usando os dados amostrais.



## Como obter estes $x_1, x_2, \dots, x_n$ ?

Se estes " $x_1, x_2, \dots, x_n$ " forem adequados, obtemos inferências confiáveis.

Se estes " $x_1, x_2, \dots, x_n$ " forem inadequados, obtemos inferências deficientes.

O sucesso integral das ferramentas estatísticas depende dos resultados e os resultados dependem da qualidade da amostra utilizada na análise.

## Amostragem

Ela fornece metodologias para escolher " $x_1, x_2, \dots, x_n$ ".

As metodologias garantem que os " $x_1, x_2, \dots, x_n$ " sejam "adequados" de acordo com os requisitos das ferramentas estatísticas a serem utilizadas..

O sucesso integral das ferramentas estatísticas depende dos resultados e os resultados dependem da qualidade da amostra utilizada na análise.

### Amostragem

---

Amostragem é a área da estatística que estuda técnicas e procedimentos para retirar e analisar uma amostra com o objetivo de fazer inferência a respeito da população de onde essa foi retirada

## Exemplos de áreas de aplicação em populações humanas:

**Marketing** - Pesquisas de mercado - Pesquisas de satisfação do cliente

**Sociologia** - Pesquisas de opinião

**Ciência Política** - Pesquisas eleitorais

**Economia** - Levantamentos econômicos

## Exemplos de áreas de aplicação em populações não-humanas:

**Agronomia** - Levantamentos de produção agrícola

**Ecologia** - Levantamentos de impacto ambiental - Levantamentos de recursos naturais

**Engenharia** - Implementação de planos para monitoramento de qualidade - "acceptance sampling"

**Biologia** - Levantamentos de abundância de espécies

## Importância social e econômica

Em vários países existe pelo menos um órgão do governo encarregado de produzir informações “oficiais” sobre a nação.

**Brasil:** - IBGE - [www.ibge.gov.br](http://www.ibge.gov.br)

**EUA:** - US Census Bureau - [www.census.gov](http://www.census.gov)

**Canadá:** - Statistics Canada - [www.statcan.ca](http://www.statcan.ca)

**Colômbia:** - DANE - [www.dane.gov.co](http://www.dane.gov.co)

**Chile:** - INE - [www.ine.gov.cl](http://www.ine.gov.cl)

## Importância social e econômica

Esses órgãos fazem uso intensivo de técnicas estatísticas de amostragem.

- Pesquisas amostrais coletam informações de uma fração da população total
- Informações sobre a população inteira são coletadas no **censo**.

Alguns levantamentos são baseados em necessidades, por exemplo, pesquisas economicas ou agricolas são conduzidas para avalia a riqueza de uma nação

- Gastos estimados com levantamentos amostrais nos EUA: 5 bilhões de dólares por ano.
- Importância acadêmica:** gerar técnicas que atendam às necessidades da sociedade.
- Importância prática:** atender novas demandas do mercado de trabalho.

## Levantamentos amostrais

- Qualquer procedimento que faça uso de uma amostra para obter informação sobre uma população, de onde essa foi retirada, poderia ser chamado de um levantamento amostral.
- No entanto, nem todo procedimento usado para a composição de uma amostra é suficiente para justificar inferência estatística.
- É preciso que o procedimento tenha como base um método probabilístico de amostragem.

### Vantagens de um levantamento amostral (comparando com um censo)

- Baixo custo
- Rapidez
- Qualidade da informação

# Conceitos básicos iniciais

## População alvo

É o conjunto de todos os elementos sobre os quais alguma informação procurada

## Cadastro

É o meio pelo qual os elementos de uma população alvo são identificados

## População pesquisada

É o conjunto de todos os elementos da população alvo que podem ser selecionados para participar da amostra.



## Conceitos básicos iniciais

### Amostra

É um subconjunto da população pesquisada, onde a informação desejada é observada.

### Variável de interesse

Uma característica de interesse, relativa a cada elemento da população pesquisada, mas que é observada apenas na amostra.

### Unidade amostral (u.a.)

É a unidade que é de fato selecionada para compor a amostra.

**O cadastro é fundamental para a implementação de um plano amostral probabilístico.**

# Cadastros



- No contexto de amostragem, **cadastro** é a designação dada a qualquer material ou instrumento usado para identificar e obter acesso aos elementos que compõem a população-alvo.

## Tipos de cadastro:

- (a) Cadastros compostos por uma listagem de elementos da população-alvo;
  - (b) Cadastros compostos por uma listagem de conjuntos de elementos da população-alvo.
- A disponibilidade de um cadastro do tipo a) permite a utilização de planos amostrais com seleção direta de elementos.
  - É comum, porém, encontrar situações em que apenas um cadastro do tipo b) está disponível. Nesse caso, para observar um elemento, é preciso fazer uso de planos amostrais em estágios. É possível ainda, selecionar uma unidade do cadastro e observar todos os elementos que a compõem.

## Exemplos de unidades amostrais

**Objetivo:** Determinar a renda total de todas as pessoas no agregado familiar.

**Unidade amostral:** Agregado familiar.

**Objetivo:** Determinar a renda de uma pessoa específica no agregado familiar.

**Unidade amostral:** Renda da pessoa específica no agregado familiar.

**Objetivo:** Estudar as condições de saúde.

**Unidade amostral:** A pessoa na qual as leituras do nível de açúcar no sangue, pressão sanguínea e outros fatores serão obtidas para classificação.

## Exemplo de população alvo

**Objetivo:** As instalações médicas de um hospital devem ser avaliadas por meio dos pacientes

**População:** Número total de pacientes registrados no hospital durante o período de pesquisa

**Objetivo:** Estudar a produção de trigo em um distrito.

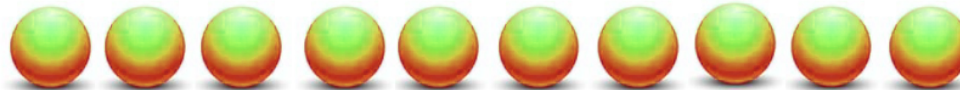
**População:** Todos os campos que cultivam trigo nesse distrito

# Tamanho populacional

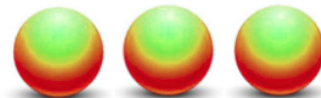
**Tamanho da população:** Número total de unidades de amostragem na população. Geralmente denotado por  $N$ .

O tamanho da população pode ser finito ou infinito ( $N$  é grande).

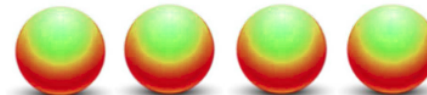
População de bola,  $N=10$



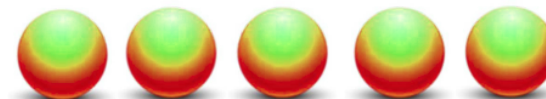
Amostra de tamanho 3



Amostra de tamanho 4



Amostra de tamanho 5



# Censo e Amostra

No contexto de pesquisas por amostragem, uma coleção de unidades como pessoas, cidades, países etc. é chamada de população finita.

Um censo é uma amostra de 100% e é uma contagem completa da população.

