

PÓS-GRADUAÇÃO EM 2025  
PUC MINAS

Disciplina:

**TÉCNICAS DE AMOSTRAGEM E  
REGRESSÃO LINEAR**

Professora: Anaíle Mendes Rabelo



 <https://www.linkedin.com/in/anaile-rabelo/>

- Bacharel em estatística – [Universidade Federal de Ouro Preto \(UFOP\)](#)
- Pós-graduação em Engenharia de Produção – [PUC Minas](#)
- Mestre em Ciências da Computação – [PUC Minas](#)
  - [Análise de Dados, Descoberta de Conhecimento e Recuperação de Informação](#)
- Experiência como Cientista e Analista de dados – 10 anos
  - [PUC Minas – área de inteligência e pesquisa](#)
  - [Tribunal de justiça de Minas gerais – Centro Estatístico Jurídico](#)
- Professora e Coordenadora da [Pós-Graduação da PUC Minas](#)

# Apresentação da disciplina

## Ementa

### Técnicas de Amostragem

Noções básicas de distribuição e amostragem.  
Tipos de amostragem.  
Dimensionamento da amostra.  
Estimadores.  
Técnicas de amostragem.  
Amostragem aleatória e viés de Amostra.  
Viés de seleção.  
Viés e erros não amostrais de não resposta.  
Estimação da variância.  
Distribuição de amostragem.  
Bootstrap.  
Intervalos de Confiança.  
Tipos de distribuição.

### Regressão linear

Introdução a Modelos Estatísticos  
Regressão linear simples  
Regressão com Variáveis Qualitativas.  
Análise de Resíduos.  
Diagnóstico em Regressão.  
Seleção de variáveis e construção de modelos.  
Validação de Modelos.  
Transformação de Variáveis  
Regressão linear múltipla

# Agenda

| Data                                 | Conteúdo programado                              | C/H       |
|--------------------------------------|--|-----------|
| <b>03/11/2025</b><br>Segunda - feira | Introdução a Amostragem e Técnicas de amostragem | 04        |
| <b>10/11/2025</b><br>Segunda - feira | Tamanho amostral e Intervalo de confiança        | 04        |
| <b>17/11/2025</b><br>Segunda - feira | Modelos Estatísticos e Introdução a regressão    | 04        |
| <b>24/11/2025</b><br>Segunda – feira | Regressão múltipla                               | 04        |
| <b>01/12/2025</b><br>Segunda – feira | Comparação de modelos de regressão               | 04        |
| <b>15/12/2025</b><br>Segunda - feira | Apresentação dos Trabalhos                       | 04        |
| <b>Total C/H</b>                     |  | <b>24</b> |

## Atividade Final

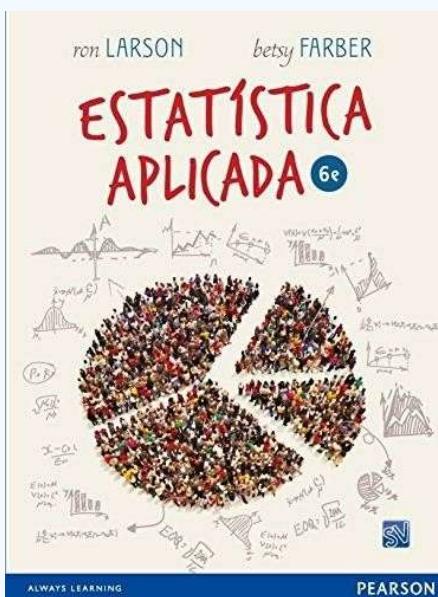
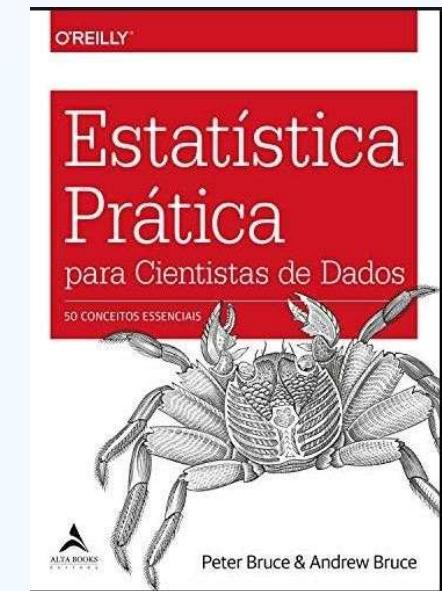
- Acessar o Kaggle escolher uma base de dados que mais te agrade.(<https://www.kaggle.com/>)
- Fazer um estudo de caso (contextualização dessa base)
- Utilizar o R ou Python para apresentar
- Predição para resolver determinado problema contendo:
  - Análise descritiva de dados
  - Utilização do modelo de regressão linear múltipla
  - Interpretação dos parâmetros e métricas dos modelos e pressupostos
- Apresentação na última aula(**10/07/2025**).

## Distribuição de Pontos

| Data (entrega) | Conteúdo programado                  | Pontos     |
|----------------|--------------------------------------|------------|
| 24/11/2025     | <b>Lista 1 – Amostragem</b>          | 20         |
| 15/12/2025     | <b>Lista 2 – Modelo de regressão</b> | 20         |
| 15/12/2025     | <b>Trabalho final</b>                | 60         |
| <b>Total</b>   |                                      | <b>100</b> |

## Referências

Estatística Prática Para Cientistas de Dados- Andrew Bruce e Peter Bruce



Estatística Aplicada - Ron Larson e Betsy Farber

# Introdução a Amostragem

# VISÃO GERAL

## Introdução às Técnicas de Amostragem

- Importância e aplicabilidade das técnicas de amostragem
- Conceitos básicos.

# TÉCNICAS DE AMOSTRAGEM NA ERA BIG DATA

---

Importância das técnicas de amostragem na era do Big Data.

---

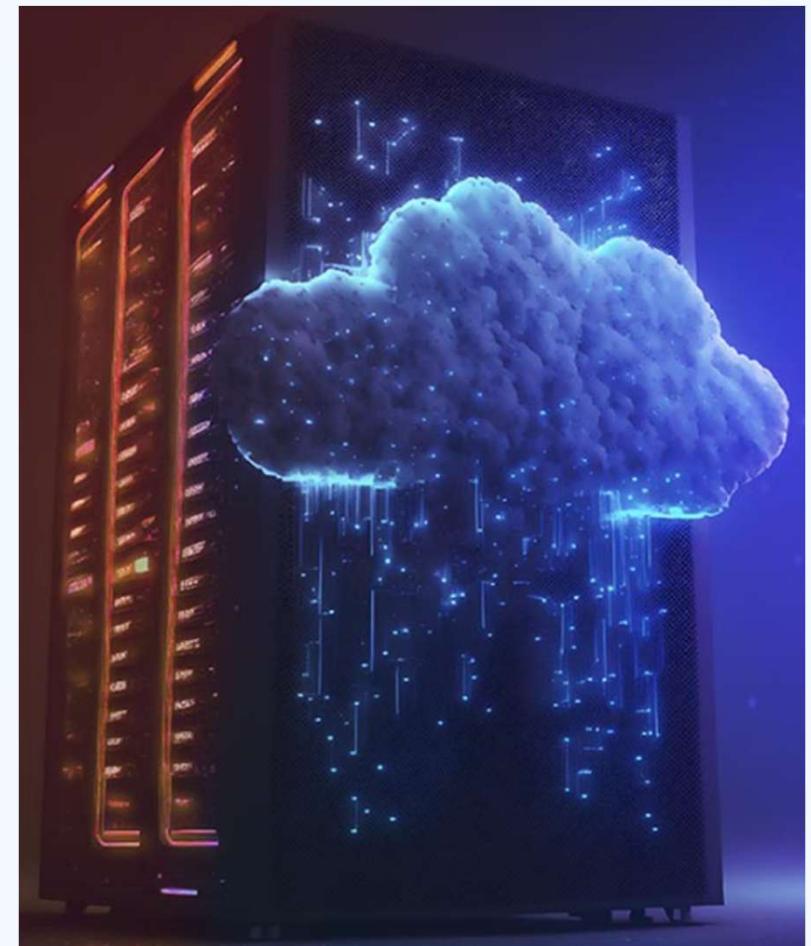
Crescimento exponencial dos bancos de dados, a obtenção de informações torna-se cada vez mais desafiadora.

---

Obtenção de informações sobre o todo a partir de uma amostra representativa.

---

Benefícios da amostragem na redução de custos, processamento e armazenamento de dados.



<https://blog.dbacorp.com.br/2023/06/02/tipos-banco-dados-nuvem/>

## EXEMPLO: AMOSTRAGEM EM FLUXO DE DADOS



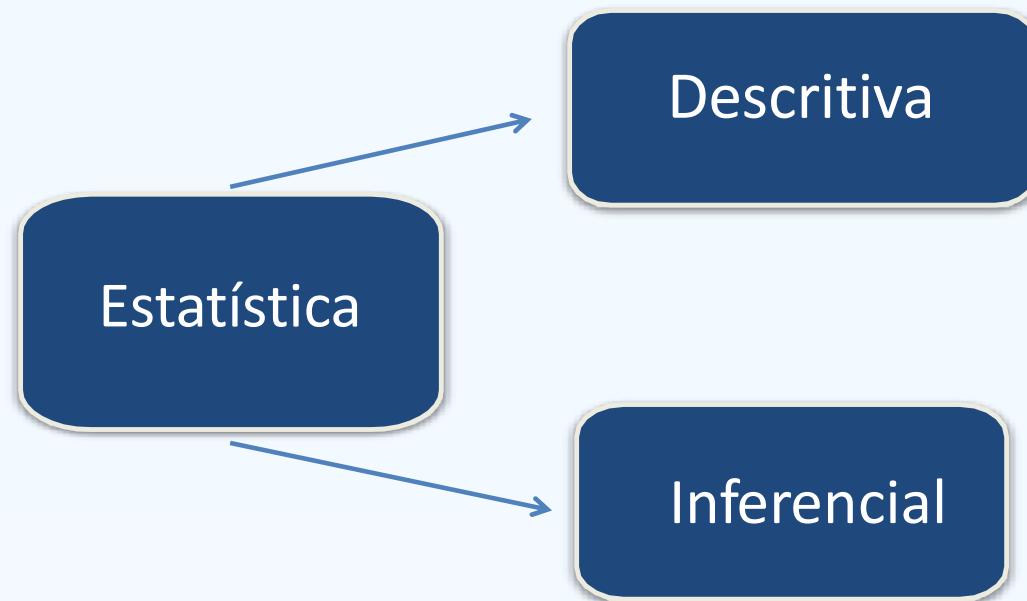
É um equívoco pensar que a era do big data significa **o fim da necessidade de amostragem.**

A proliferação de dados reforça a necessidade de amostragem como ferramenta para trabalhar eficientemente com uma variedade de dados e para minimizar o viés. Mesmo em um projeto de Big Data, os modelos preditivos são tipicamente desenvolvidos e conduzidos com amostras, que são usadas também em testes de diversos tipos (por exemplo, especificação, web treatments).

Em machine learning as técnicas de **amostragem** são **amplamente** utilizadas para seleção do **conjunto de base e treino**.

# INTRODUÇÃO: AMOSTRAGEM

Principais áreas da Estatística



Organizar, demonstrar e resumir dados

Obter respostas sobre um fenômeno com dados representativos



# DEFINIÇÃO CONCEITUAL DE ESTATÍSTICA

Estatística

Conjunto de  
Métodos

Transformar dados  
em informação

- Resumo
- Apresentação
- Análise

# ORIGEM DOS DADOS

**De onde vem  
os dados?**



**Fonte de dados Primários**

Estudo de Caso

Experimentos

Pesquisas



**Fonte de dados Secundários**

Registros administrativos

Pesquisa bibliográfica

Big Data (banco de dados corporativos ,  
etc).

Dados orgânicos



## BASE DE DADOS

- Fonte de dados típicas
  - Censo
  - Pesquisas amostrais
  - Registros administrativos

Características comum: dados planejados

- Novas fontes de dados (Big Data)
  - Transações (nota fiscal eletrônica)
  - Comunicações e mensagens
  - Imagens

- Características comuns
  - Grande volume , variedade e velocidade
  - Dados não planejados (orgânicos)

## INTRODUÇÃO A AMOSTRAGEM

Um dos **principais objetivos** da maioria dos estudos, análises ou pesquisas estatísticas é **fazer generalizações seguras – com base em amostras** – sobre as populações das quais se extraíram as amostras.

**Por exemplo:**

Como estimar o salário médio de jogadores de futebol?

# INFERÊNCIA ESTATÍSTICA

## O que é Inferência Estatística ?

É o conjunto de metodologias que apoiam na formulação de conclusões sobre as características de uma **POPULAÇÃO** a partir de uma parte dessa população, denominada de **AMOSTRA**.



Fonte: <https://audtcmsp.org.br/comunicacao/artigos/1091-amostragem-o-que-e-e-qual-sua-utilidade-para-a-auditoria>

## TERMINOLOGIAS UTILIZADAS

- **Dados:** Consistem em informações provenientes de observações, contagens, medidas ou respostas.
- **População:** É o conjunto de todos os resultados, respostas, medidas ou contagens a serem estudados.
- **Amostra:** É um subconjunto de elementos de uma população.

## CARACTERÍSTICAS DA AMOSTRA

**Representativa** → conter em proporção tudo o que a população possui qualitativa e quantitativamente;

**Imparcial** → todos os elementos da população tem igual oportunidade de fazer parte da amostra;

Uma AMOSTRA é a redução de uma população em dimensões menores, porém, **sem perda de suas características**.

O processo de definição da amostra chama-se **AMOSTRAGEM**



Inferência sobre a população

Parâmetros Populacionais

Estatísticas Amostrais

População

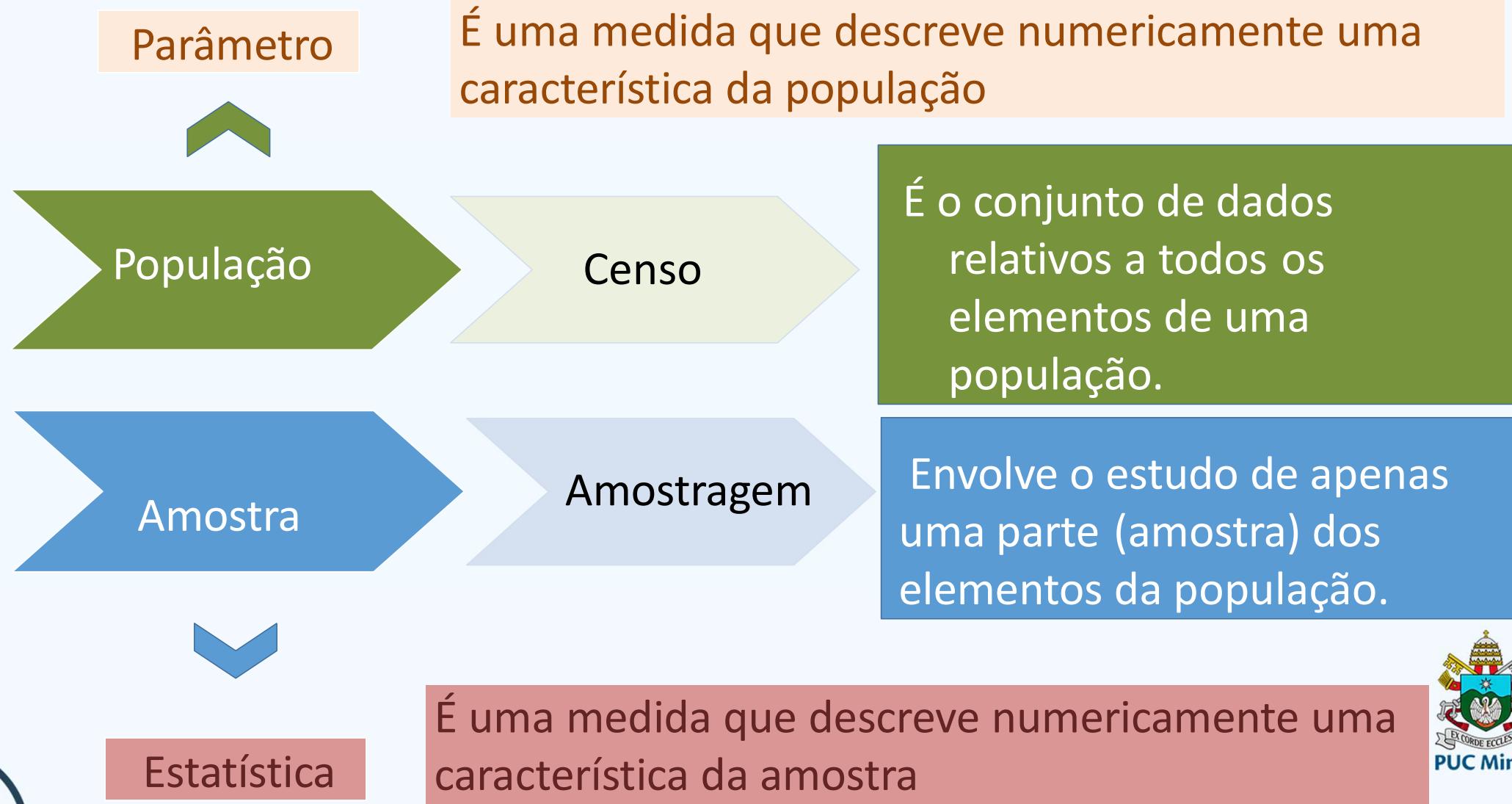


Amostragem

Amostra

Tratamento e análise de dados

## PROCESSO DE AMOSTRAGEM



É o conjunto de dados relativos a todos os elementos de uma população.

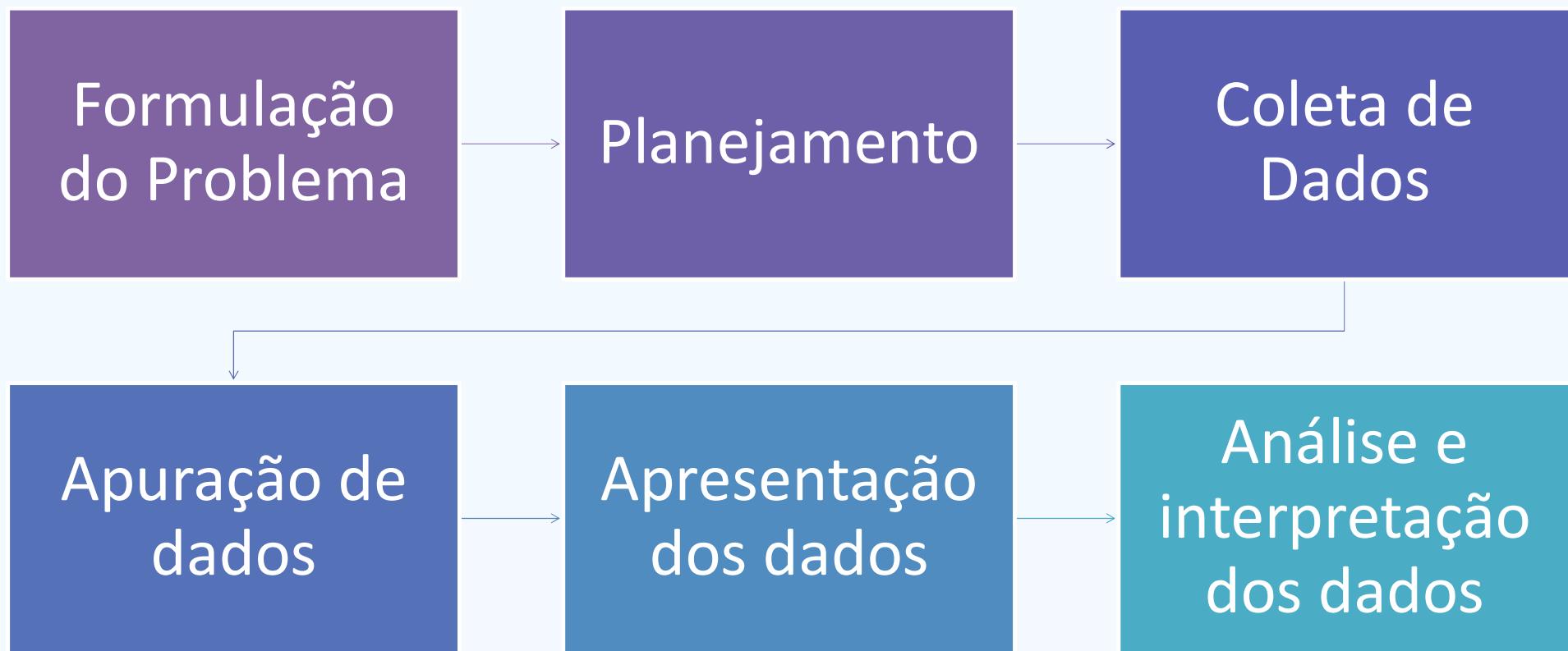
Envolve o estudo de apenas uma parte (amostra) dos elementos da população.

É uma medida que descreve numericamente uma característica da amostra

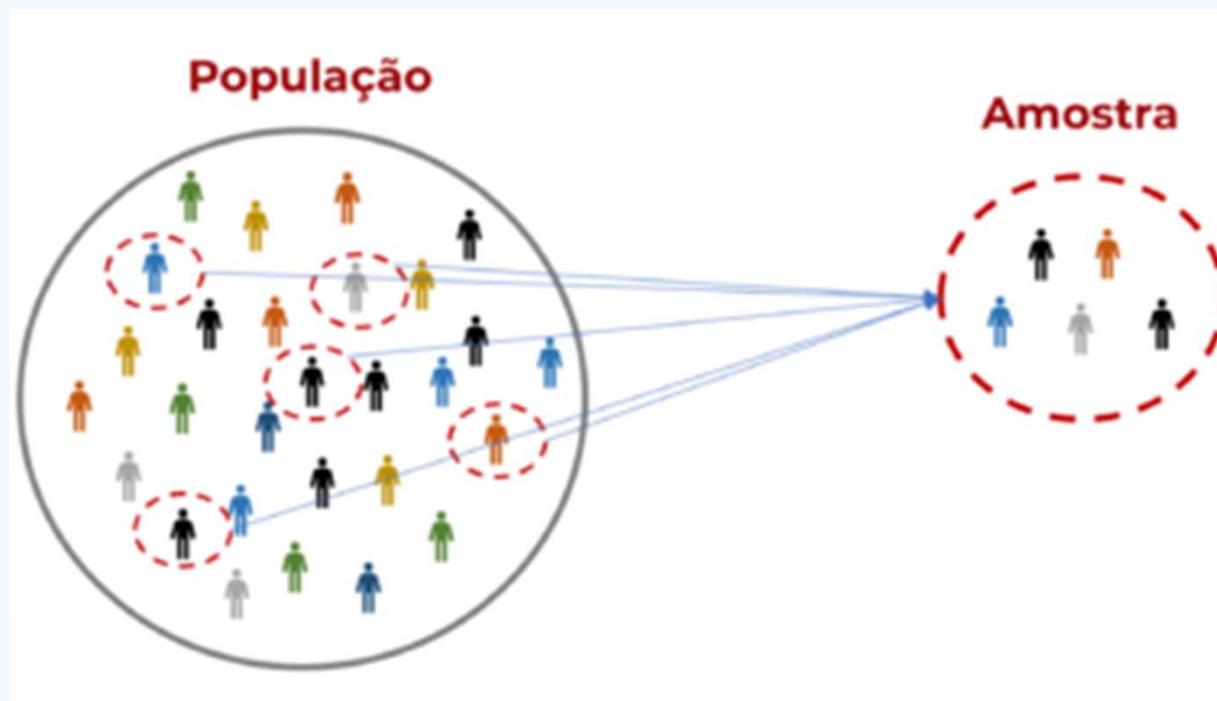
## VANTAGENS

| Amostragem   | Censo  |
|--|--|
| Tempo de coleta menor, o que torna as pesquisas mais atualizadas.              | A população pode ser tão pequena que o custo e o tempo de um censo sejam pouco maiores que para uma amostra. |
| Quando existe pesquisas de testes destrutivos, o censo não pode ser utilizado. | Se a amostra é especialmente grande em relação ao tamanho da população às vezes se justifica o censo.        |
| Se a população é infinita o censo é impossível.                                | Se é exigida a precisão completa, então o censo é o único método aceitável.                                  |
| Custos menores.  |  |

## FASES DO MÉTODO



# OBTENDO UMA AMOSTRA REPRESENTATIVA



<https://solvis.com.br/pesquisa-de-satisfacao-como-calcular-a-amostragem-de-respostas/>

Amostra  
Representativa

Tamanho da  
Amostra

Técnicas de  
amostragem

# Técnicas de Amostragem

## TÉCNICAS DE AMOSTRAGEM

- Amostra deve ser representativa;
- Técnicas de amostragem apropriadas devem ser usadas para garantir que as inferências sobre a população sejam válidas.

Estudo feito com dados imprecisos geram resultados questionáveis.

Amostra tendenciosa é aquela que não é representativa da população da qual foi extraída

Dados coletados de forma descuidada podem ser inúteis, de maneira, que nenhum processamento estatístico consiga salvá-los

Técnicas de Amostragem

Probabilísticas

Não Probabilísticas

- Seleção é aleatória
- Probabilidade de ocorrência iguais
- Mais Recomendada
- **Permite inferência**
  
- Escolha deliberada dos elementos da amostra.
- Pode depender dos critérios e julgamento do pesquisador.
- Não temos a lista completa dos indivíduos que formam a população.
- **Não permite inferência**

# TÉCNICAS DE AMOSTRAGEM

Probabilística

- Amostragem aleatória Simples
- Amostragem aleatória Sistemática
- Amostragem Estratificada
- Amostragem por Conglomerados

Não Probabilística

- Amostragem por conveniência ou acessibilidade
- Amostragem por Cotas
- Amostragem Bola de Neve

# Técnicas de Amostragem Probabilísticas

# AMOSTRAGEM ALEATÓRIA SIMPLES



Fonte: <https://bookdown.org/luisfca/docs/tipos-de-amostragem.html>

# Amostragem Aleatória Simples

## Vantagens

- Simplicidade
- Permite medir a precisão das estimativas
- Propriedades conhecidas e testadas.

## Desvantagens

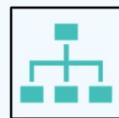
- Precisa de cadastro da população
- Custo elevado
- Não usa informações auxiliares

# AMOSTRAGEM ALEATÓRIA SIMPLES

## Aplicações:



Opinião dos alunos sobre alguma política estudantil



Clima organizacional de uma empresa



Satisfação dos alunos com o ensino remoto



Satisfação dos clientes de uma rede bancária

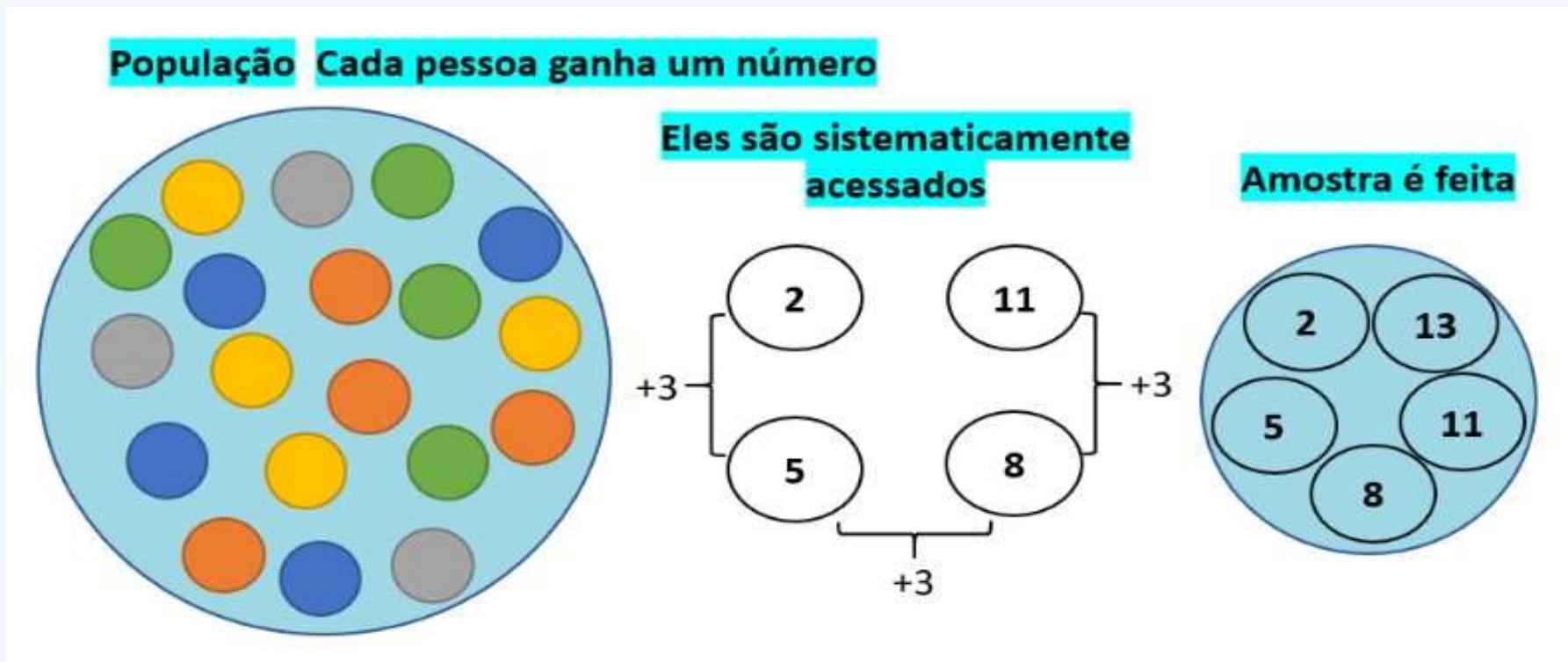


Produtos retirados de um lote para inspeção



Sorteio de pessoas para serem mesários nas eleições.

# AMOSTRAGEM SISTEMÁTICA



Fonte: <https://bookdown.org/luisfca/docs/tipos-de-amostragem.html>

## AMOSTRAGEM SISTEMÁTICA

### Vantagens:

- Simples
- Cadastro pode ser construído junto com a amostra
- Fácil de estimar quantidade populacionais

### Desvantagens:

- Difícil para estimar a precisão da estimação
- Custo Elevado
- Periodicidade no cadastro pode impactar nas estimativas

# AMOSTRAGEM SISTEMÁTICA

## Aplicações:



AVALIAÇÃO DE  
QUALIDADE DE  
PEÇAS EM UMA  
LINHA DE  
PRODUÇÃO.



AVALIAÇÃO DE  
LOTES DE  
PRODUTOS QUE  
CHEGAM EM  
SEQUÊNCIA



PESQUISA DE  
BOCA DE URNA



PLANTAS EM UM  
POMAR OU  
LAVOURA

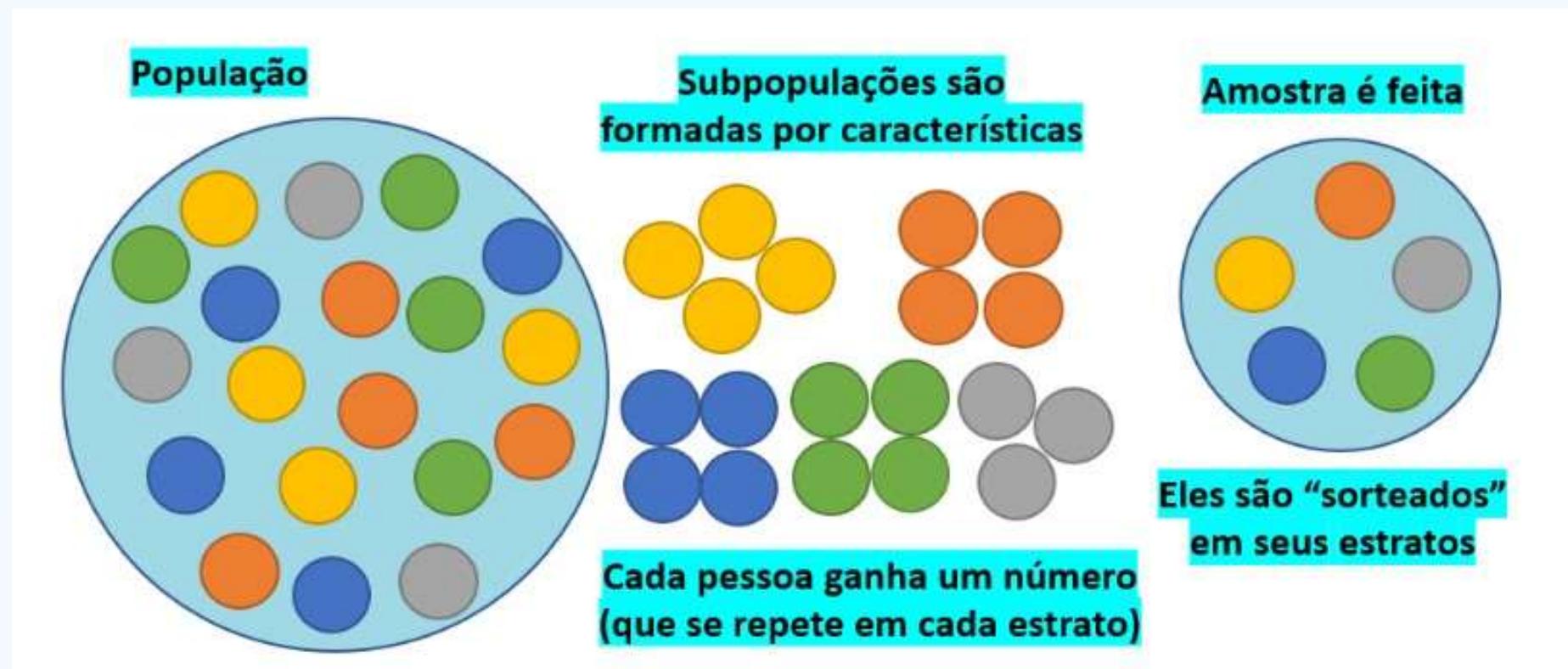


ANIMAIS QUE  
PASSAM POR UM  
CORREDOR



PESSOAS EM UM  
TEATRO /  
CINEMA.

# AMOSTRAGEM ESTRATIFICADA



Fonte: <https://bookdown.org/luisfca/docs/tipos-de-amostragem.html>

# AMOSTRAGEM ESTRATIFICADA

## *Vantagens*

Plano intuitivo;

Estratos são divisões naturais da população.

Melhora a eficiencia amostral - reduz a variância

Permite medir a precisão das estimativas

## *Desvantagens*

Precisa de informação prévia sobre os estratos

Pode ser necessário usar métodos de coleta diferentes em cada estrato

Pelo menos duas etapas para coleta de dados

## AMOSTRAGEM ESTRATIFICADA

### *Exemplos*

- Taxa de desemprego nas capitais
- Pesquisas eleitorais
- Clima organizacional em diferentes setores da empresa
- Pesquisas de opinião em população fechada ou estratificadas
- Amostragem de animais por rebanho

## AMOSTRAGEM ESTRATIFICADA

Amostragem  
estratificada

Proporcional

A proporcionalidade do tamanho de cada estrato da população é mantida na amostra.

Uniforme

Selecionamos o mesmo número de elementos em cada estrato. É o processo usual quando se deseja comparar os diversos estratos.

# AMOSTRAGEM POR CONGLOMERADOS



Fonte: <https://bookdown.org/luisfca/docs/tipos-de-amostragem.html>

# AMOSTRAGEM POR CONGLOMERADOS

## Vantagens:

- Pode facilitar a coleta dos dados
- Conglomerados são grupos naturais da população
- Permite medir a precisão das estimativas
- Mais barato

## Desvantagens

- Precisa de informação prévia sobre os conglomerados
- Pode ser necessário usar métodos de coleta diferentes em cada conglomerado
- Pelo Menos duas etapas para coleta de dados
- Mais complicado de estimar parâmetros de interesse

# AMOSTRAGEM POR CONGLOMERADOS

## Exemplos

Renda das famílias de uma cidade

Pesquisa eleitoral que são feitas de casa em casa

Qualidade de vida dos trabalhadores da indústria

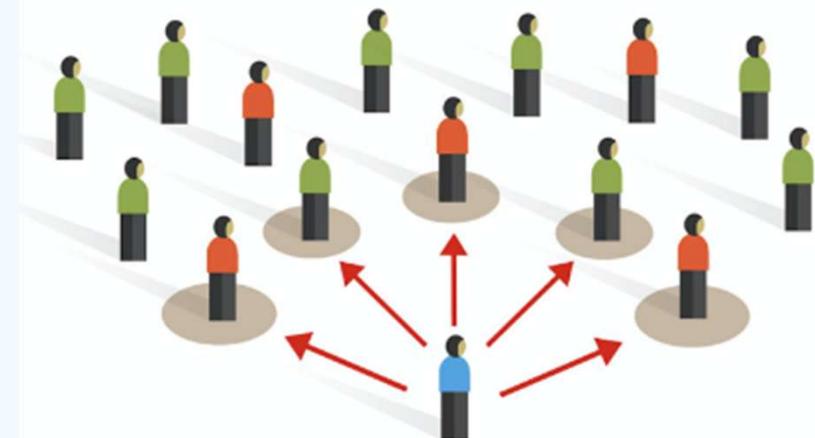
Pesquisas por telefone (DDD e códigos que separam regiões)

# Técnicas de Amostragem Não Probabilísticas

## AMOSTRAGEM POR CONVENIENCIA

Amostragem sem probabilidade  
previamente  
especificada ou conhecida

Unidades são escolhidas  
prioritariamente pela sua  
disponibilidade.



Fonte: <https://www.linkedin.com/pulse/tipos-de-amostragens-dados-usando-python-parte-3-aline-marques/?originalSubdomain=pt>

## AMOSTRAGEM POR CONVENIÊNCIA

- ▶ **Vantagens:**
  - ▶ Simplicidade
  - ▶ Facilidade
  - ▶ Baixo custo.

- ▶ **Desvantagens:**
  - ▶ Impossível avaliar a "representatividade" da amostra.
  - ▶ Não é possível fazer inferência estatística

# AMOSTRAGEM POR CONVENIENCIA

## Aplicações



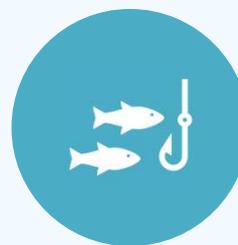
► PESQUISAS QUE USAM VOLUNTÁRIOS PARA COMPOR A AMOSTRA.;



► TESTAGEM DE EFICÁCIA DE VACINA;

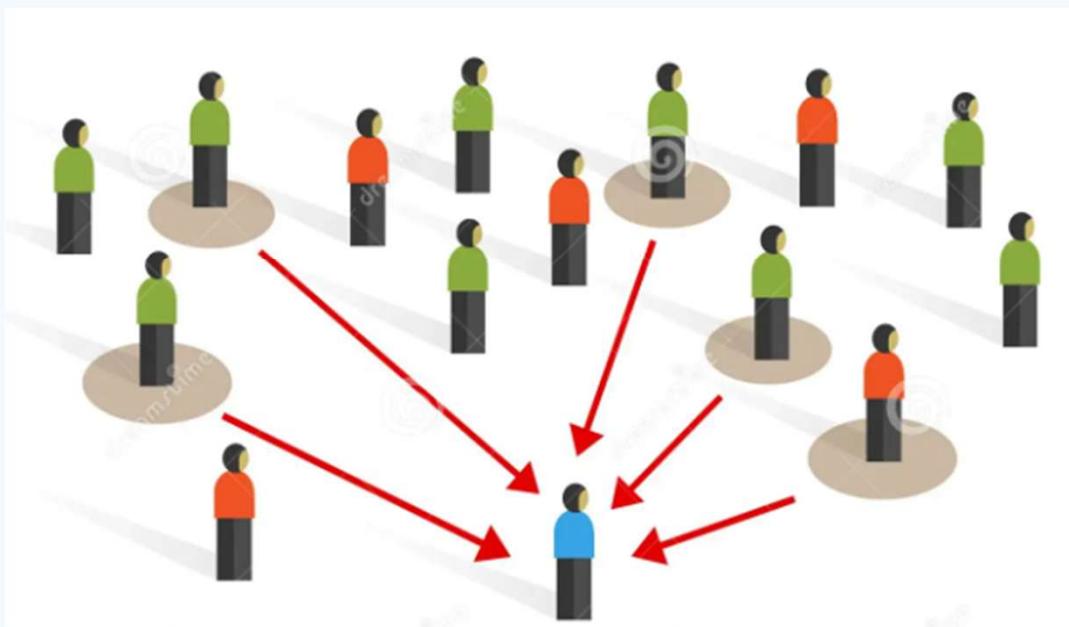


► ANIMAIS SELVAGENS;



► PEIXES CAPTURADOS POR REDES DE PESCA.

## AMOSTRAGEM POR AUTOSELEÇÃO



Fonte: <https://pt.dreamstime.com/amostra-de-amostragem-resposta-voluntaria-colhida-um-grupo-pessoas-amostrando-m%C3%A9todo-estat%C3%ADstico-n%C3%A3o-probabilidade-t%C3%A9cnica-image169045693>

**Os participantes voluntariamente solicitam participar da pesquisa.**

Ex:

- Pesquisas online (*google survey* ou *survey monkey*)
- Estudos sobre novos medicamentos

## AMOSTRAGEM POR AUTOSELEÇÃO

### Vantagens:

Mais fácil de se coletar.

### Desvantagem:

- Impossível avaliar a "representatividade" da amostra
- Não é possível fazer inferência estatística

## AMOSTRAGEM INTENCIONAL OU JULGAMENTO

**Amostra sem probabilidade** previamente especificada ou **conhecida**.

A escolha é feita por uma pessoa (**expert**) que conhece profundamente o tema de estudo.

O expert julga quais são os elementos mais representativos para compor a amostra .



<https://pedrounb.blogspot.com/2012/05/calculo-do-tamanho-de-amostras.html>

## AMOSTRAGEM INTENCIONAL OU JULGAMENTO

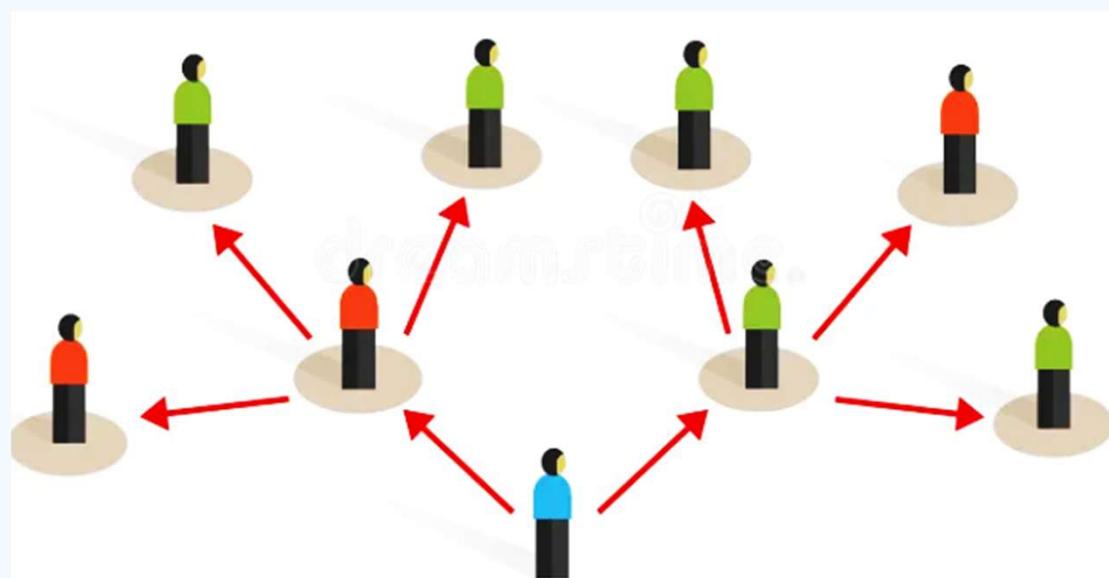
- ▶ Vantagens:
  - ▶ Simplicidade
  - ▶ Facilidade
  - ▶ Baixo custo.
  
- ▶ Desvantagens:
  - ▶ Impossível avaliar a "representatividade" da amostra.
  - ▶ Nenhum procedimento estatisticamente justificável permite análise probabilística ou inferencial dos resultados.
  - ▶ Depende do julgamento do expert

## AMOSTRAGEM INTENCIONAL OU JULGAMENTO

### Aplicações

- Um repórter escolhe senadores influentes para saber a opinião do senado sobre certo tema.
- Quando a obtenção da amostra pode ser vista como algo raro.
- Motivos de falhas corporativas entrevistando executivos.

## AMOSTRAGEM BOLA DE NEVE



Fonte: <https://pt.dreamstime.com/ilustra%C3%A7%C3%A3o-stock-amostra-da-bola-de-neve-os-m%C3%A9todos-amostra-na-pesquisa-qualitativa-image84516121>

Identificação de **poucas unidades** que o pesquisador tem acesso.

Essas unidades **indicam outras** para compor a amostra

## AMOSTRAGEM BOLA DE NEVE

### Vantagens:

- Simplicidade e facilidade.
- Baixo custo.
- Pode ser a única opção para identificar populações vulneráveis ou informações sensíveis (HIV, drogas, assédio, violência, etc).

### Desvantagens:

- Impossível avaliar a "representatividade" da amostra.
- Nenhum procedimento estatisticamente justificável permite análise probabilística ou inferencial dos resultados.

## AMOSTRAGEM BOLA DE NEVE

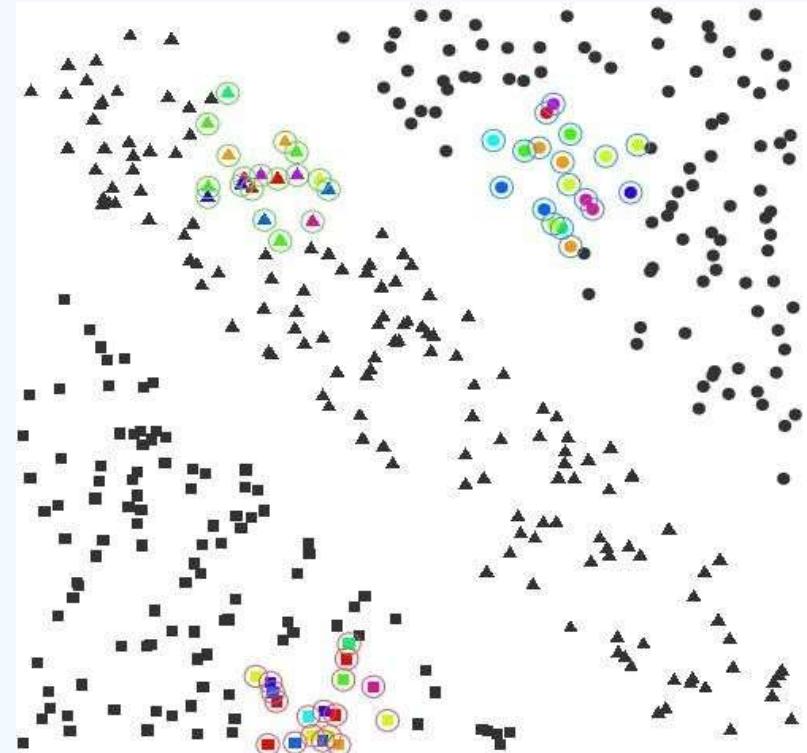
- Obter amostras de populações vulneráveis (grupos).
- Condições de saúde de imigrantes sem visto de permanência.
- Avaliação da saúde de pacientes HIV positivos.
- Avaliação das condições psicológicas de famílias com filhos portadores de necessidades especiais.
- Uso medicinal de substâncias ainda não legalizadas.
- Animais rastreados que retornam aos seus bandos.

## AMOSTRAGEM POR COTAS

A população é dividida em subgrupos (conhecidos).

Avalia a proporção de cada subgrupo na população.

Amostra proporcional sem seleção probabilística em cada subgrupo.



Fonte: [https://www.linkedin.com/posts/wagner-bonat\\_como-saber-se-minha-amostra-%C3%A9-representativa-activity-7156657335363383298-BHOW/?originalSubdomain=pt](https://www.linkedin.com/posts/wagner-bonat_como-saber-se-minha-amostra-%C3%A9-representativa-activity-7156657335363383298-BHOW/?originalSubdomain=pt)

## AMOSTRAGEM POR COTAS

- ▶ Vantagens:
  - ▶ Simplicidade e facilidade.
  - ▶ Baixo custo.
  - ▶ Tenta manter algum nível de representatividade.

- ▶ Desvantagens:
  - ▶ Nenhum procedimento estatisticamente justificável permite análise probabilística ou inferencial dos resultados.

## AMOSTRAGEM POR COTAS

### Aplicações

Pesquisas eleitorais.

Pesquisas mercadológicas.

Pesquisas de opinião pública em geral.

Coleta de água em diferentes pontos de um rio urbano.