

Aula 1 -Introdução: O papel da Inferência Estatística EST 187 - Inferência

Denise Duarte
denise@est.ufmg.br

Departamento de Estatística
Universidade Federal de Minas Gerais (UFMG)

2020/01

- O progresso da ciência é geralmente associado à experimentação:
 - 1 Um experimento é realizado e dados são coletados;
 - 2 Com base nos dados observados, tiramos algumas "conclusões sobre o fenômeno que estamos interessados em estudar.
- O processo de extrair conclusões a respeito de um fenômeno com base em observações é chamado de *inferência indutiva*.
- A inferência indutiva é um processo sujeito à aleatoriedade, uma vez que não é possível termos absoluta certeza sobre nossas generalizações/conclusões.
- A incerteza associada às nossas inferências é medida em termos de distribuições de probabilidades.

Definição

População: *Em estatística, uma população é um conjunto de itens ou eventos semelhantes que interessam para alguma questão ou experimento.*

- **Exemplo 1:** Pesquisa de intenção de voto para prefeito de Belo Horizonte
população: eleitores de BH.
- **Exemplo 2:** Estabelecer prazo de garantia de uma televisão modelo A de uma marca X .
população: todas as TV's modelo A da marca X .
- Podemos representar os valores possíveis de uma medida específica em uma população com uma variável aleatória.
- Assumimos que a "lei" que governa a população em estudo é descrita por uma variável aleatória e sua respectiva distribuição de probabilidade.

- Do ponto de vista da estatística, o problema da inferência é considerado da seguinte forma:
 - O objetivo da investigação é buscar informações sobre uma certa população;
 - Geralmente é impossível (ou impraticável!) examinar toda a população, mas apenas uma parcela (isto é, uma amostra) da mesma;
 - Com base na informação limitada (contida na amostra), buscamos fazer inferências para toda a população em estudo.

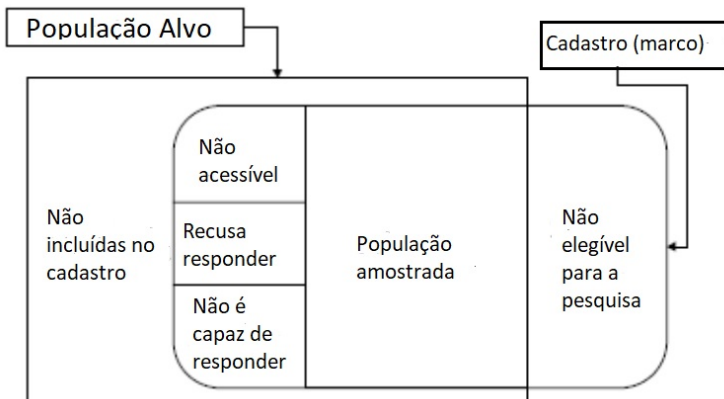
Definição

Amostra: *Uma amostra é qualquer subconjunto finito da população.*

- Para que as conclusões sobre o fenômeno em estudo sejam válidas, a amostra deve ser representativa da população.
- A escolha "subjetiva" das observações que irão compor a amostra, em geral, produz uma amostra tendenciosa, e pode levar a conclusões completamente equivocadas.
- Amostras tendenciosas podem ser evitadas se o processo de amostragem for aleatório, ou seja, deve haver sorteio.

Razões de trabalhar com amostras:

- **Envelhecimento da pesquisa:** tempo excessivo para coleta dos dados;
- **Econômicas:** restrições orçamentárias de projetos;
- **Éticas:** experimentos laboratoriais com seres vivos;
- **Inacessibilidade de informações:** análise do sangue de uma pessoa ou experimento para determinar o tempo de funcionamento de uma lâmpada.



Tipos de estudos

Descritivos: Descrevem determinada situação, sem preocupar-se com comparações, e as conclusões são meramente descritivas.

Exemplo: Perfil de pacientes com determinada doença, etc.

Comparativos: Há o interesse na comparação do comportamento de pelos menos dois grupos.

Exemplo: Comparação de fumantes e não fumantes quanto ao desenvolvimento de câncer de pulmão.

Tipos de estudos

Normalmente obtemos dados de duas fontes distintas:

1. **Estudos observacionais:** observamos ou medimos características específicas, mas não tentamos modificar os sujeitos objeto do estudo.
Ex.: Pesquisa de opinião.
2. **Experimentos:** aplicamos algum tratamento e passamos, então, a observar seu efeito sobre os sujeitos.
Ex.: Teste clínico para avaliar o efeito de um medicamento.

Tipos diferentes de estudos observacionais

- **Estudo transversal:** os dados são observados, medidos e coletados em um ponto no tempo.
- **Estudo retrospectivo:** os dados são coletados do passado, voltando-se no tempo (através de registros, entrevistas, etc).
 - Voltamos no tempo para coletar dados sobre a característica resultante que está sendo considerada.
 - **Ex.:** Dados de dois grupos, um de motoristas que sofreram acidentes de carro e outro de motoristas que não sofreram acidentes.

Tipos diferentes de estudos observacionais

- **Estudo prospectivo:** os dados são coletados ao longo do tempo de grupos que compartilham fatores comuns. Também conhecido como estudo **longitudinal** ou de **coorte**.
 - Avançamos no tempo acompanhando grupos, um com um fator potencialmente causativo e outro sem esse fator.
 - **Ex.:** Um grupo de motoristas que usam telefone celular enquanto dirigem e outro que não usa.

Métodos de Amostragem

Amostragem probabilística: todos os elementos da população têm uma probabilidade conhecida de pertencer à amostra.

- **Amostragem Aleatória Simples (AAS):** Consiste na seleção de n unidades amostrais de tal forma que cada amostra tenha a mesma probabilidade de ser escolhida. A seleção pode ser feita com e sem reposição.

Exemplo: O professor sorteia 12 alunos de uma turma com 60 alunos a partir de uma urna com 60 bolas numeradas.

- **Amostragem Sistemática:** escolhemos algum ponto inicial e selecionamos cada k -ésimo elemento da população (amostra selecionada de forma cíclica). São exemplos os prédios de uma rua, uma linha de produção, etc.

Métodos de Amostragem

- **Amostragem Estratificada:** A população pode ser dividida em subgrupos (estratos) que compartilham as mesmas características (tais como sexo ou faixa etária) e extraímos uma amostra de cada subgrupo.

Obs.: Quando a variável de interesse é heterogênea na população e pode ser dividida em grupos homogêneos. Utilizada para reduzir a variação nos resultados.

Exemplo: Em uma turma com 60 alunos, primeiro, o professor divide a sala em 6 grupos com 10 alunos em cada, depois sorteia com AAS 2 alunos de cada grupo.

Métodos de Amostragem

- **Amostragem por conglomerados:** Dividimos a população em conglomerados e sorteamos os conglomerados por meio de um processo aleatório.

Exemplo: Imagine uma sala de aula com 60 alunos arrumados em seis filas de 10 alunos cada. Suponha que professor selecione uma amostra de 10 alunos jogando um dado e selecionando a fila correspondente ao resultado da jogada.

Obs:

- Geralmente, podemos definir os conglomerados por fatores geográficos, como bairros e quarteirões.
- Dentro de um conglomerado podemos avaliar todos os elementos ou sortear os elementos avaliados.
- A utilização da amostragem por conglomerados possibilita uma redução significativa do custo no processo de amostragem.

Amostragem não probabilística: é assim denominada sempre que não conhecemos a probabilidade de um elemento da população pertencer à amostra.

- **Amostras conveniência e amostras de voluntários:** simplesmente usamos resultados que sejam muito fáceis de obter ou de quem se dispõe a participar.
- **Amostragem por julgamento:** o amostrador deliberadamente escolhe certos elementos para pertencer à amostra por julgar tais elementos bem representativos da população.
- **Amostragem por cotas:** a população é dividida em grupos e, na sequência, é determinada uma cota proporcional ao tamanho de cada grupo. Entretanto, dentro de cada grupo não é feito sorteio, mas sim os elementos são procurados até que a cota de cada grupo seja cumprida.

Exemplo: Pesquisa eleitoral.

- Podemos estudar características de interesse de uma determinada população através do estudo das respectivas distribuições de probabilidades das variáveis aleatórias que representam tais características;
- Sabemos que as distribuições de probabilidade, em geral, dependem de parâmetros para se tornarem completamente especificadas;
- Um mesmo modelo probabilístico pode ser utilizado para se modelar o comportamento de populações completamente diferentes;
- Desta forma, o processo de estudar características de uma população envolve duas questões fundamentais:
 - Qual modelo probabilístico é o mais adequado para representar as características de interesse da população em estudo?;
 - Definido o modelo a ser adotado, qual é o conjunto de parâmetros associado ao modelo assumido que melhor representa esta população?

- A escolha do modelo probabilístico mais adequado passa, impreterivelmente, pelas técnicas de descrição de dados;
- A determinação do conjunto mais adequado de parâmetros para um dado modelo probabilístico é papel da inferência estatística;
- Na prática, dispomos apenas de um conjunto de observações extraídas da população, isto é, uma amostra, e a partir dela desejamos fazer inferências sobre a população;
- A inferência estatística é dividida em duas grandes áreas:
 - 1 estimação de parâmetros;
 - 2 testes de hipóteses sobre os parâmetros;
- A primeira trata do conjunto de técnicas adequadas para a estimação de parâmetros associados ao modelo probabilístico adotado para representar a população em estudo;
- O segundo considera o conjunto de técnicas utilizadas para a realização de comparações envolvendo duas ou mais populações de interesse.