Universidade Federal Fluminense
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DEPARTAMENTO DE ESTATÍSTICA

Estatística Básica para Engenharia
Prof. Mariana Albi

Material de Apoio
Assunto: Amostragem

Amostragem

Relembrando...

Etapas da Análise Estatística:

Definição do problema → Coleta dos dados (Amostragem) → Organização, resumo e apresentação dos dados (AED)

↓ (Probabilidade)

Análise e interpretação dos dados para solução do problema (Inferência)

Vamos nos concentrar na etapa da **AMOSTRAGEM**...

Amostragem consiste em selecionar parte de uma população para observar, de modo que seja possível **estimar** alguma coisa sobre toda a população.

Vimos nos exemplos discutidos em aula como uma amostra pode ser **tendenciosa**, no sentido de não ser representativa da população da qual foi obtida.

Várias técnicas de amostragem podem gerar amostras tendenciosas. Em geral, isto ocorre quando os dados são coletados de maneira inapropriada, sem aleatoriedade. Alguns exemplos comuns de amostras tendenciosas acontecerem em:

- **Amostra de resposta voluntária**: aquela na qual os respondentes decidem, eles mesmos, se serão incluídos na amostra.
  ex.1: Pesquisas feitas pela internet ou correio, onde os indivíduos decidem se respondem ou não a pesquisa.
  ex.2: Pesquisas feitas por telefone, como em programas de rádio ("Ligue e dê sua opinião!").
  Em pesquisas de resposta voluntária, indivíduos com opiniões mais fortes têm mais tendência a participar.

- **Amostras pequenas** (com relação a população):
  Cuidado! 1000 indivíduos pede ser um número satisfatório para algumas pesquisas, porém muito pequeno para outras, especialmente no estudo de eventos raros.
  ex.: o Fundo de Defesa da Criança americano publicou um estudo que dizia que, em uma dada região, entre os alunos da escola secundária suspensos, 67% tinham sido suspensos pelo menos 3 vezes; só esqueceram de mencionar que a amostra era de apenas 3 estudantes.
  Existem algumas técnicas para escolha de tamanho de amostra, além de métodos que lidam com situações onde poucos dados são disponíveis, porém existem muitas hipóteses a serem consideradas nestes casos.

- **Questões pessoais/polêmicas**: o entrevistado se sente coagido e prefere não responder ou mentir. **ex.:** Entrevista perguntando ao indivíduo se ele já foi traído ou sua opção sexual.

A moral da história é que devemos sempre ser críticos e levar em conta qual o processo de amostragem mais apropriado. Se os dados não forem coletados de maneira correta, não podemos confiar em nenhuma conclusão tirada.

O importante em amostragem é que os **dados** sejam **representativos da população**, isto é, que todos os elementos da população tenham a chance de serem bem representados. Os dados devem ser coletados de modo apropriado, tal como através de um processo de **seleção aleatória**. Se os dados não forem coletados de forma apropriada podem não serão úteis para nenhuma análise estatística.

Definições (Tipos de amostragem):

- **Amostra Aleatória**: amostra na qual os membros da população são selecionados de tal modo que cada membro individual tenha chance igual de ser selecionado.

- **Amostra Aleatória Simples**: uma amostra aleatória simples de tamanho n é selecionada de tal modo que toda amostra possível de mesmo tamanho n tenha a mesma chance de ser escolhida. **ex.:** Sorteio dos números da loteria.

- **Amostra Probabilística**: envolve a seleção de membros de uma população de tal modo que cada membro tenha uma chance conhecida (não necessariamente igual) de ser selecionado.

- **Amostragem Sistemática**: após a escolha de algum ponto inicial, obter os outros elementos da amostra tomando espaçamentos iguais entre eles. **ex.:** A cada 10 itens de uma linha de produção, um é testado para verificar se há defeitos.

- **Amostragem por Conveniência**

- **Amostragem Estratificada**: subdivida a população em (pelo menos 2) subgrupos (estratos) diferentes, de modo que os elementos dentro de cada subgrupo tenham as mesmas características e a seguir extraia uma amostra de cada subgrupo. **ex.:** Separa-se os consumidores por sexo e depois seleciona uma amostra de cada grupo para estudo.

- **Amostragem por Conglomerado**: divida a população em seções (conglomerados), a seguir selecione aleatoriamente alguns desses conglomerados; a amostra será composta pela união de todos os elementos dos conglomerados selecionados. **ex.:** Em uma pesquisa eleitoral, selecionam-se algumas zonas eleitorais para entrevistar todos os eleitores destas zonas selecionadas.

Cuidado! A **amostragem por conglomerado** usa todos os elementos de uma amostra de conglomerados, enquanto a **amostragem estratificada** usa uma amostra de membros de todos os estratos.

O que se faz na prática, é, dependendo da natureza do problema, combinar os vários tipos de amostragem. Por exemplo, podemos ter **amostragem aleatória estratificada simples** se utilizamos amostragem aleatória simples em cada um dos estratos, ou **amostragem aleatória estratificada proporcional** se ponderamos cada amostra pelo tamanho do estrato considerado.

Exemplo:

(Triola) Imagine uma sala de aula com 60 alunos arrumados em 6 filas com 10 alunos cada. Suponha que o professor selecione uma amostra de 10 alunos jogando um dado e selecionando a fila correspondente ao resultado do dado. O resultado é uma amostra aleatória? Se sim, é simples? É uma amostra probabilística?

É amostra aleatória, pois cada aluno tem a mesma chance de ser sorteado (1/6). Não é amostra aleatória simples, pois nem todo grupo de 10 alunos tem a mesma chance de ser sorteado (por exemplo, é impossível sortear um grupo que contenha alunos de filas diferentes). Além disso, a amostra é probabilística porque cada aluno tem uma chance conhecida de ser selecionado.

A **amostragem aleatória simples** (AAS) é a mais utilizada. Muitos dos métodos de inferência são baseados em amostragem aleatória simples.

Podemos fazer amostragem aleatória simples de 2 formas: **com ou sem reposição**.

- sem reposição: uma vez selecionada, a unidade é desconsiderada num próximo sorteio.

- com reposição: unidades já selecionadas podem ser repetidas na amostra.

Na prática, quando trabalhamos com populações grandes, não faz diferença trabalhar com sorteios com e sem reposição, por isso, em geral, a teoria é baseada em sorteios com reposição, por fornecerem propriedades probabilísticas mais simples.