

# Disciplina: 5170 - Probabilidade e Estatística

## Aula 02

**Prof. George Lucas M. Pezzott**  
**glmpezzott2@uem.br**

Departamento de Estatística - UEM  
Sala: 222 - Bloco: E-90

Segunda-feira - 19:30 ~ 21:10 - D67 - Sala 108  
Sexta-feira - 21:20 ~ 23:00 - D67 - Sala 108

# 1 - ANÁLISE EXPLORATÓRIA DE DADOS

## Algumas Definições

- **População:** É uma coleção completa de todos os elementos a serem estudados e que possuem certa característica comum.

### Exemplos de população:

- Brasileiros acima de 16 anos;
- Crianças com menos de 1 ano no Brasil;
- Peças produzidas em uma linha de produção de uma fábrica;
- Número diário de veículos que passam em uma praça de pedágio;
- Número de mensagens que chegam no correio eletrônico por minuto;
- Os **dados** consistem em informações provenientes de observações coletadas (por exemplo, medidas, sexo, respostas de um questionário, etc).

## Algumas Definições

- **Censo:** É um conjunto de dados obtidos de **todos** os membros da população.
- **Amostra:** É um subconjunto finito de membros selecionados de uma população.

### Exemplo

Em um levantamento recente sobre o hábitos de adultos nos EUA, perguntou-se a 3002 se eles liam notícias na internet pelo menos uma vez por semana. 600 responderam que sim. Identifique a população, a amostra e descreva o conjunto de dados.

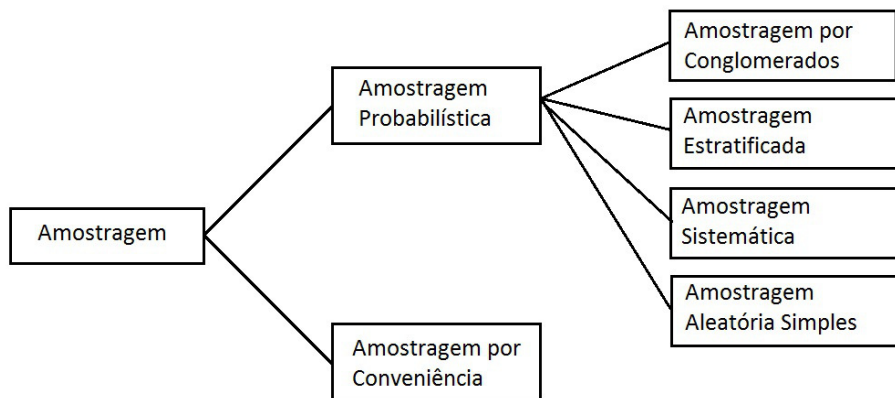
- **Amostragem:** Processo de seleção elementos de uma população (amostra).
- **Amostragem probabilística:**
  - Cada elemento da população tem uma chance conhecida de ser selecionado.
  - Nesta amostragem, sempre ocorre algum sorteio
  - Este tipo de amostragem é importante nos processos de inferências pois os métodos estatísticos são construídos sob suas propriedades.

versus

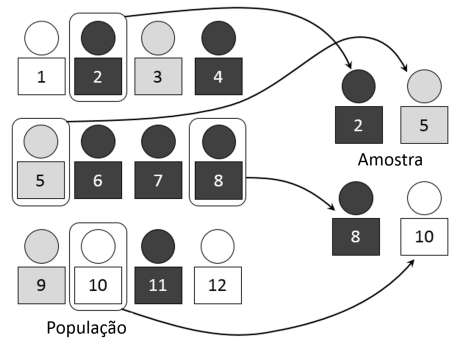
- **Amostragem por conveniência:** Coleta-se elementos de fácil acesso para o pesquisador.

## Algumas Definições

## Técnicas de Amostragem



- **Amostragem aleatória simples:** Cada indivíduo tem mesma chance de ser amostrado.



# Amostragem aleatória simples

- Para a seleção de uma amostra aleatória simples, precisamos ter uma lista completa dos elementos da população.
- Este tipo de amostragem consiste em selecionar a amostra através de sorteios, sem restrição.

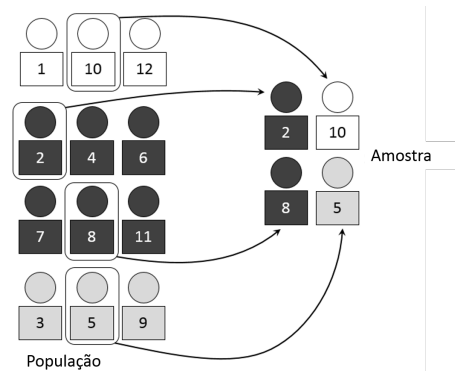
## Propriedade

Qualquer conjunto da população, com o mesmo número de elementos, tem a mesma probabilidade de fazer parte da amostra.

- Em uma população de tamanho  $N$ , cada amostra de tamanho  $n$  tem probabilidade  $1 / \binom{N}{n}$  de ser sorteada.



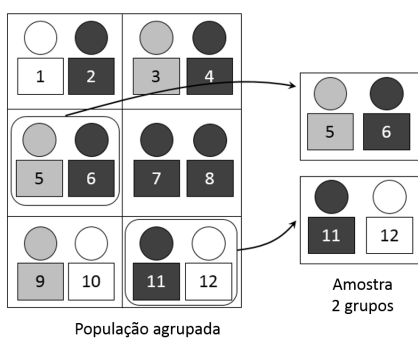
- **Amostragem estratificada:** Divide-se a população em grupo segundo algum estrato (pelo menos dois), amostra-se aleatoriamente indivíduos dentro de cada grupo.



# Amostragem estratificada

- Os estratos devem ser internamente mais homogêneos do que a população toda, com respeito às variáveis em estudo.
- **Amostragem estratificada proporcional:** a proporcionalidade do tamanho de cada estrato da população é mantida na amostra. Por exemplo, se um estrato abrange 20% da população, ele também deve abranger 20% da amostra.
- **Amostragem estratificada uniforme:** selecionamos o mesmo número de elementos em cada estrato. É o processo usual quando se deseja comparar os diferentes estratos.

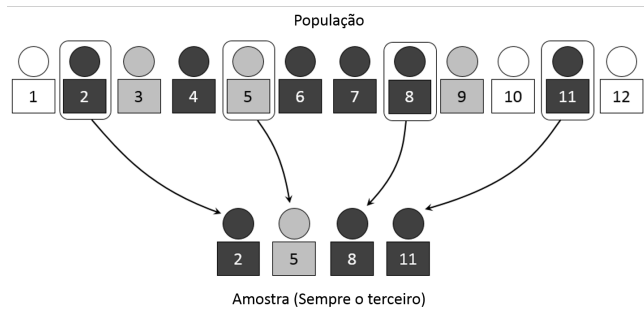
- **Amostragem por conglomerado:** Divide-se a população em blocos (heterogêneos) e amostra-se os blocos aleatoriamente e todos os indivíduos dentro do bloco são entrevistados.



# Amostragem por conglomerados

- Ao contrário da amostragem estratificada, a amostragem de conglomerados tende a produzir uma amostra que gera resultados menos precisos quando comparada com uma amostra aleatória simples de mesmo tamanho. Contudo, seu custo financeiro tende a ser bem menos, especialmente em amostragens de grandes populações.
- É comum tomar amostras de elementos dos conglomerados extraídos no primeiro estágio de seleção (amostragem de conglomerados em dois estágios).

- **Amostragem sistemática:** Selecionar um elemento da população a cada  $k$ .



# Amostragem sistemática

- Um processo mais simples é sortear o primeiro elementos e extrair os demais sistematicamente. Mais especificamente:
  - ① calcula-se o intervalo de seleção, dado por  $l = N/n$ , desprezando as decimais;
  - ② sorteia-se o primeiro elemento do conjunto  $\{1, 2, \dots, l\}$ ; e
  - ③ completa-se a amostra, extraíndo um elemento a cada  $l$  elementos.

## Apresentação dos Dados

## O que fazer com os dados coletados?

- **Etapas iniciais:** Estatística Descritiva

### Análise Exploratória

Consiste em medidas resumo (média, mediana, moda, variância, coeficiente de variação, etc), tabelas e gráficos.

## Apresentação dos Dados

## Considerações

Suponha que  $X$  representa uma variável (altura de indivíduo, tempo de resposta de um sistema, número de chegadas em uma fila, ....).

Desejamos conhecer o “comportamento” da variável  $X$ .

Para isso, considere as seguintes definições:

- **Tamanho da Amostra:**  $n$
- **Amostra Aleatória:** É uma sequência de realizações da variável de interesse  $X$  sem observá-la.  
 $\mathbf{X} = (X_1, X_2, \dots, X_n);$
- **Amostra** ou **conjunto de observações** ou **dados:** valores observados  $\mathbf{x} = (x_1, x_2, \dots, x_n)$  de  $\mathbf{X} = (X_1, X_2, \dots, X_n);$



### Exemplo

Seja  $X$  a nota de um aluno do curso Estatística -1 na primeira avaliação.

Seja  $Y$  a nota de um aluno do curso Estatística -1 na segunda avaliação.

## Apresentação dos Dados

## Considerações

Tabela: Medidas descritivas das notas dos alunos da disciplina Estatística -1.

Medidas	X	Y
$n$	49	45
Média	8,06	4,43
Mediana	8,70	4,40
Moda	10	0,75; 2; 3,75 5,5; 6,75
Mínimo	2,5	0
Máximo	10	10
$Q_1$	7,25	2,25
$Q_3$	9,50	6,50
Variância	3,83	5,85
Desvio-padrão	1,96	2,42

- **Variável:** Qualquer característica de interesse associada aos elementos de uma população.

### Classificação de variáveis

<b>Qualitativa</b>	{	<b>Nominal</b>	(Ex: cor, sexo, estado civil)
		<b>Ordinal</b>	(Ex: grau de escolaridade, classe social, porte de empresa)
<b>Quantitativa</b>	{	<b>Discreta</b>	(Ex: n <sup>o</sup> de pessoas, n <sup>o</sup> de defeitos)
		<b>Contínua</b>	(Ex: peso, temperatura, preço, tempo,...)

Para variáveis quantitativas, podemos calcular:

- **Medidas de Posição:** moda, média, mediana (medidas de tendência central), quantis (quartis, decis, percentis).
- **Medidas de Dispersão:** amplitude, intervalo interquartil, variância, desvio-padrão, coeficiente de variação.