

Tópico 10: Amostragem

Ben Dêivide

6 de outubro de 2021

A amostragem já é utilizada no nosso dia a dia inconscientemente. Quando preparamos uma refeição e vamos prová-la antes de servi-la, estamos na realidade fazendo uma amostragem, ao passo que a refeição representa o todo (população) e a parte retirada para provar se a refeição estava pronta para servi-la é a amostra. Dessa forma, queremos, por meio de uma amostra, saber algumas características (parâmetro) sobre a população. As medidas obtidas da amostra, com as quais nos indicam alguma característica da população, chamamos de estimador. Com base nisso, para obtermos bons estimadores, precisamos garantir antecipadamente que a amostra coletada da população represente a população, de modo a preservar as características relevantes da população.

Definição 1 (População). *O conjunto de elementos (unidades) para os quais desejamos que as conclusões de pesquisa sejam válidas, com a restrição de que esses elementos possam ser observados ou mensurados sob as mesmas condições é chamado de população, universo ou população objetivo. O tamanho da população será representado por N .*

A população pode ser formada por pessoa, famílias, estabelecimentos industriais, ou qualquer tipo de elementos, dependendo basicamente dos objetivos da pesquisa. Podemos dizer ainda que uma população pode ser finita ou infinita. Uma população é finita quando se consegue enumerar todos os elementos que a formam. Refere-se a um universo limitado em uma dada unidade de tempo. Como exemplo podemos dizer que a quantidade de automóveis produzidos em um mês, a população de uma cidade, o número de alunos de uma sala de aula são exemplos de **população finita**. Uma população é dita infinita quando não podemos enumerar todos os elementos. Refere-se a um universo não delimitado. Os resultados (cara ou coroa) obtidos em sucessivos lances de uma moeda, o conjunto dos números inteiros, reais ou naturais são exemplos de **população infinita**.

Definição 2 (Elemento, unidade ou unidade elementar). *Indivíduo ou objeto a ser medido ou observado na pesquisa, do qual é a entidade portadora de informações que pretende-se coletar é chamado de unidade elementar (UE).*

Um outro tipo de elemento da população é a unidade amostral.

Definição 3 (Unidade amostral). *Um conjunto formado por uma unidade elementar ou várias unidades elementares do qual seja de interesse para pesquisa é chamada de unidade amostral (UA).*

Exemplo 1. *Uma unidade amostral pode ser os conglomerados pelo método de amostragem por conglomerados. Exemplos de conglomerados são: quarterões, ruas, departamentos, prateleiras, caixas, lotes de produtos, etc..*

O interesse na pesquisa está em determinar características da população relevantes para o estudo. Definimos,

Definição 4 (Parâmetro). *Qualquer característica atribuída a população é chamada de parâmetro.*

Exemplo 2. *Numa pesquisa epidemiológica, a população pode ser definida como todas as pessoas (unidade elementar) da região em estudo, no momento da pesquisa. Um parâmetro de interesse pode ser a porcentagem de pessoas contaminadas.*

Como nem sempre é possível coletar informações de todas as unidades elementares da população, faz-se necessário retirarmos informações sobre uma parte da população acessível, e assim, por meio destas informações, consigamos obter informações sobre o parâmetro de interesse na pesquisa.

Definição 5 (Amostragem). *O mecanismo (técnica) que consistem em selecionar parte de uma população para observar, de modo que seja possível preservar as principais informações sobre toda a população.*

A parte da população chamamos de amostra.

Definição 6 (Amostra). *Qualquer subconjunto da população obtido por meio de um processo de seleção adequado é chamado de amostra. O tamanho da amostra será representado por n .*

Dizemos que quando uma amostra preserva as principais características de uma população, diz-se que esta é uma amostra representativa. Quanto mais próximo n estiver de N mais representativo será esta amostra de tamanho n .

As informações obtidas por meio de uma amostra são chamadas de estatísticas. Quando uma estatística estima um parâmetro da população, temos um estimador.

Definição 7 (Estimador). *Um estimador é uma medida, função da amostra (estatística) que representará o parâmetro desconhecido da população.*

Exemplo 3. *Uma pesquisa deseja saber a altura dos estudantes de uma determinada universidade. Se usarmos uma medida de tendência central para representar a altura dos estudantes, poderíamos escolher a média. Logo, um parâmetro para a população é μ a média populacional. Geralmente μ ou qualquer outro parâmetro é desconhecido, então como estimador para μ podemos considerar \bar{X} , a média amostral. Observe que este estimador é função da amostra. Dizemos que o resultado de um estimador é a estimativa.*

A estimação nada mais é do que criarmos mecanismos para extrapolar as medidas amostrais para que estas possam representar os parâmetros desconhecidos. Como o estimador é função da amostra, se esta não é representativa, com certeza estaremos cometendo algum erro em afirmar que esse estimador poderia representar o parâmetro desconhecido. Daí a importância de obtermos metodologias para coletar amostras representativas da população.

Definição 8 (Erro amostral). *O erro amostral é a diferença entre a estimativa e o parâmetro que se quer estimar.*

Esse erro reflete a tendência da amostra. Querendo ou não, o fato de já tomarmos decisão com base em uma amostra, já estamos cometendo erro. Contudo, o que a estatística tenta mostrar, por meio da amostragem, é que podemos tomar conclusões com base em uma amostra sobre a população de modo a cometer o mínimo de erro

amostral possível. Assim, dizemos que o objetivo da pesquisa amostral será conhecer características sobre a população, pesquisando (estudando) a amostra, de modo a cometer o mínimo de erro amostral possível.

Se a pesquisa envolve a observação de todas as unidades da população, o método de pesquisa é denominado **censo** ou **pesquisa exaustiva**. Se é conduzida sobre uma amostra da população, o método de pesquisa é denominado **levantamento por amostragem**. O censo somente é aplicável na situação que a população seja finita e suas unidades sejam identificáveis e disponíveis para a coleta da amostra. Por essa razão, o levantamento por amostragem é muito mais frequentemente utilizado. Por que fazer amostragem ao invés de um censo?

- Vantagens: Pesquisa por amostragem em relação ao censo:
 - É mais barata;
 - É mais rápida;
 - É mais fácil de ser controlada por envolver operações menores.
- Desvantagens: Pesquisa por amostragem em relação ao censo:
 - O censo pode ser mais vantajoso quando a população é pequena e/ou as informações são de fácil obtenção;
 - Os resultados da pesquisa por amostragem contém erros amostrais;
 - Se a população for muito heterogênea o erro pode ser muito grande, dependendo do método de amostragem que seja utilizado.

Numa fase inicial dos levantamentos amostrais é necessário **formular o problema** e aventar **hipóteses** sobre o objeto de estudo ou expectativas sobre os possíveis resultados. Ainda nessa fase inicial, o investigador deve **definir a população de estudo**, parte identificável e acessível da população objeto, os **objetivos** e as **variáveis observadas**. Numa segunda etapa é realizado o planejamento, elaborado o **plano de amostragem** ou determinando o caminho a ser percorrido para atingir os objetivos propostos. O plano de amostragem devem ter bem definidos:

1. Unidade amostral: indivíduos ou grupos de indivíduos (conglomerados);
2. Sistema de referência: lista completa das unidades amostrais;
3. Tamanho da população (N), é definido pelo número de indivíduos da população objetivo;
4. Tamanho da amostra (n), definido pelo número de indivíduos selecionados na amostra, tal que $n < N$.

Os planos ou métodos de amostragem podem ser classificados em amostragem probabilística e amostragem não probabilística. Dizemos que se um método de amostragem é objetivo e estabelece uma probabilidade conhecida a cada unidade da população objetivo ser incluída na amostra, esse é denominado amostragem probabilística ou amostragem aleatória; caso contrário, é denominado não probabilística ou amostragem não aleatória. Os principais esquemas amostrais podem ser observados na Figura 1.

Nas aulas em sala explanamos sobre os métodos de amostragem probabilística. Iremos falar sobre os métodos de amostragem não-probabilística.

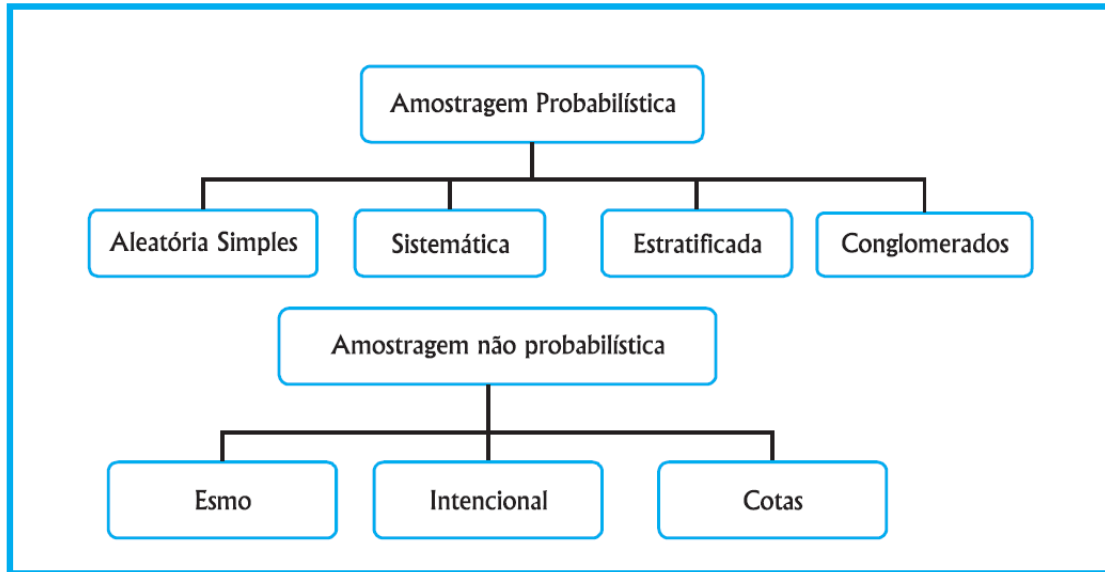


Figura 1: Métodos de amostragem probabilística e não-probabilística.

O primeiro método é a **amostragem a esmo**. Imagine uma caixa com 1.000 parafusos. A enumeração destes parafusos ficaria muito difícil, e a amostragem aleatória simples se tornaria inviável. Então, em situações deste tipo, supondo que a população de parafusos seja homogênea, escolhemos a esmo a quantidade relativa ao tamanho a amostra. Quanto mais homogênea for a população, mais podemos supor a equivalência com uma amostragem simples ao acaso. Desta forma, os parafusos escolhidos para compor a amostra de um determinado tamanho sem nenhuma norma ou a esmo. Por isso, do nome ao método de amostragem.

Outro método é a **amostragem intencional** que corresponde àquela em que o amostrador deliberadamente escolhe certos elementos para pertencer à amostra, por julgar tais elementos bem representativos da população. Um exemplo deste tipo de amostragem corresponde à situação em que deseja saber a aceitação em relação a uma nova marca de um determinado produto a ser inserida no mercado de uma cidade. Somente entrarão para compor a amostra pessoas que façam uso do produto e que tenham condições financeiras de comprar esta nova marca (classe social de maior poder aquisitivo).

O último método falado é a **amostragem por cotas**. Nesse tipo de amostragem, a população é dividida em grupos, e seleciona-se uma cota proporcional ao tamanho de cada grupo. Entretanto, dentro de cada grupo não é feito sorteio, e sim os elementos são procurados até que a cota de cada grupo seja cumprida. Em pesquisas eleitorais, a divisão de uma população em grupos (considerando, por exemplo, o sexo, o nível de escolaridade, a faixa etária e a renda) pode servir de base para a definição dos grupos, partindo da suposição de que estas variáveis definem grupos com comportamentos diferenciados no processo eleitoral.