

Estatística e Probabilidade

Profa. Carolina Silva Pena

O que é Estatística?

A Estatística pode ser definida como um conjunto de técnicas que permite **organizar, descrever, analisar e interpretar** dados provenientes de experimentos ou estudos realizados em qualquer área do conhecimento.

Podemos dividir a Estatística em três áreas:

- Estatística Descritiva;
 - Conjunto de técnicas destinadas a descrever e resumir os dados.
- Probabilidade;
 - Teoria matemática utilizada para modelar a incerteza presente em fenômenos aleatórios.
- Inferência Estatística.
 - É o processo de estimação dos parâmetros de uma população a partir de informações retiradas de uma amostra dessa população.

Amostragem

- Por que amostrar?
 - Em geral não se tem acesso a toda a população (população grande);
 - Economia;
 - Rapidez.
- Por que fazer censo?
 - Os resultados obtidos não estão sujeitos a margem de erro;
 - No Brasil, o Censo é realizado a cada 10 anos pelo IBGE. A população é contada em todo o território do Brasil e os resultados são utilizados para definir políticas públicas.

Definições

- População: conjunto formado por todos os indivíduos ou objetos sobre o qual desejamos obter informações;
- Amostra: qualquer subconjunto da população;
- Amostragem: processo de seleção da amostra;
- Parâmetro: característica populacional de interesse;
- Estimativa: valor numérico obtido a partir da amostra para o parâmetro;
- Erro amostral: diferença entre o verdadeiro valor do parâmetro populacional e sua estimativa (amostra).

Tipos de amostragem

Amostragem não-probabilística

- Os elementos da amostra são escolhidos de maneira deliberada, de acordo com critérios estabelecidos pelo pesquisador.

Amostragem probabilística

- Cada elemento da população possui probabilidade não nula e conhecida de fazer parte da amostra.

Amostragem não-probabilística

- Amostragem por acessibilidade ou por conveniência
 - Seleção dos elementos aos quais se tem acesso;
 - É considerado o menos rigoroso de todos os métodos de amostragem.
- Amostragem intencional
 - Seleciona-se um subconjunto da população com base em característica desse subgrupo.
 - Requer conhecimento da população e também do subgrupo selecionado.
 - Exemplo: entrevistar representantes de uma comunidade para avaliar a percepção sobre a segurança pública da região.

Amostragem não-probabilística

Amostragem por cotas

- É considerada a mais rigorosa entre as amostragens não-probabilísticas;
- Etapas:
 - Escolher características que representem a variabilidade da população;
 - Determinar a proporção da população dentro de cada classe formada pela interseção dessas características;
 - Fixar cotas para a amostra de acordo com a proporção da população dentro de cada classe.
- É utilizada quando não existe um cadastro da população que possibilite a realização do sorteio necessário a amostragem aleatória, mas existe informação sobre o perfil populacional.
- Esse tipo de amostragem é utilizada em pesquisas eleitorais.

Amostragem probabilística

A grande vantagem da amostragem probabilística é que podemos calcular o erro associados às nossas estimativas! Podemos ainda calcular um tamanho de amostra a partir da especificação dos seguintes fatores:

- ❶ Nível de confiança: Quanto maior o valor adotado, maior o tamanho da amostra necessário.
- ❷ Erro máximo permitido: quanto menor o erro máximo permitido, maior será o tamanho da amostra necessário.
- ❸ Variabilidade do fenômeno que está sendo investigado: quanto maior a variabilidade, maior o tamanho da amostra necessário.
- ❹ Tipo de procedimento amostral: alguns procedimentos amostrais exigem um tamanho de amostra menor do que outros para o mesmo erro máximo associado.

Amostragem Aleatória Simples (AAS)

Na AAS todos os elementos da população tem a mesma probabilidade de fazer parte da amostra. Exemplo: rotula-se os elementos da população e sorteia-se n deles para fazer parte da amostra.

- População infinita ou amostragem com reposição

$$n = \left(\frac{Z_{\alpha/2}\sigma}{E} \right)^2$$

onde:

- n = número de indivíduos na amostra;
- $Z_{\alpha/2}$ = valor crítico que corresponde ao grau de confiança desejado;
- σ = desvio-padrão populacional da variável estudada;
- E = margem de erro.

Amostragem sistemática

Se os elementos da população já se encontram ordenados segundo algum critério, pode-se selecionar um elemento qualquer e escolher um “fator” que definirá qual o próximo elemento escolhido.

Exemplo: $N = 20$, $n = 4$, $s = 20/4 = 5$, $m = 3$.

- Etapas:

- Calcular o fator de sistematização $s = \frac{N}{n}$;
- Sortear um número m entre 1 e s .

O primeiro elemento da amostra será o de número m ; O segundo elemento da amostra será o de número $m+s$; O terceiro elemento da amostra será o de número $m+2s$.

O n -ésimo elemento da amostra será o de número $m + s \times (n - 1)$.

Amostragem aleatória estratificada

Este tipo de amostragem deve ser utilizado quando a população for heterogênea com relação a variável de interesse.

O objetivo da amostragem aleatória estratificada é formar estratos homogêneos e então realizar amostragem aleatória simples dentro de cada estrato. Entre os extratos deve haver heterogeneidade.

- Exemplo: Realizar uma pesquisa no curso de Engenharia utilizando as turmas como estratos.

Amostragem por conglomerados

Consideramos conglomerados os grupos de elementos com as seguintes características:

- Dentro de cada conglomerado há uma grande heterogeneidade, ou grande variabilidade;
- Entre os conglomerados há uma grande homogeneidade, ou pequena variabilidade.

Exemplo: Se estamos interessados no salário médio dos operários da indústria automobilística, podemos selecionar uma montadora e, dentro dela, estudar os salários.