MAT02025 - Amostragem 1

Conceitos básicos de probabilidade e inferência estatística: uma revisão

Rodrigo Citton P. dos Reis
citton.padilha@ufrgs.br

Universidade Federal do Rio Grande do Sul
Instituto de Matemática e Estatística
Departamento de Estatística

Porto Alegre, 2022

População e amostra, estatística, parâmetro, estimador, estimativa e distribuição amostral

População

População ou universo

Esse termo é usado em estatística com um sentido bem mais amplo do que na linguagem coloquial.
É entendido aqui como o **conjunto de todos os elementos** que apresentam uma ou mais características **em comum**.

- **Exemplo 1:** a população de colegiais de oito anos de Belo Horizonte.
  - Estes colegiais têm em comum a idade e o local onde vivem.
- **Exemplo 2:** a população de indústrias brasileiras.
  - Estas indústrias têm em comum o fato de que foram criadas no Brasil.

Censo e amostra

- Quando o estudo é realizado com toda a população de interesse, chamemos este estudo de **censo**.
- Por motivos de tempo, custo, logística, entre outros, geralmente não é possível realizar um censo.
  - Nestes casos, estudamos apenas uma parcela da população, que chamamos de **amostra**.

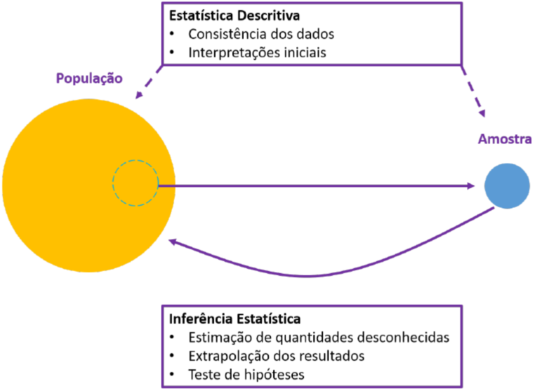
Amostra

É qualquer fração de uma população.

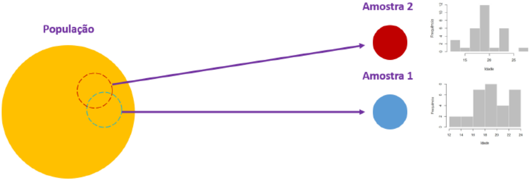
Inferência estatística

- A **Inferência Estatística** é um conjunto de técnicas que objetiva estudar a população por meio de evidências fornecidas por uma amostra.
- É a amostra que contém os elementos que podem ser observados e, a partir daí, **quantidades de interesse** podem ser medidas.

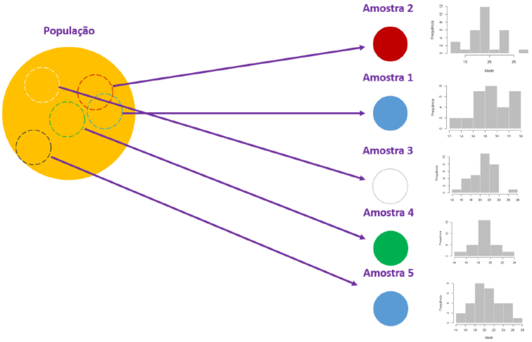
População e amostras



População e amostras



População e amostras



Uma questão . . .

- Uma questão que surge agora é: apesar de diferentes, podemos ter respostas próximas ou iguais nas diversas amostras?
  - A reposta é afirmativa e estará subjacente às ideias que desenvolvermos a partir desta aula.
- Podemos dizer que devido à **natureza aleatória**, geralmente envolvida no procedimento amostral, não podemos garantir que repetições de amostras produzam sempre resultados idênticos.
  - Assim, **todas as quantidades associadas à amostra terão caráter aleatório** e, portanto, devem receber tratamento probabilístico.
- Nesta aula, formalizaremos alguns conceitos relacionados a um ramo da Inferência Estatística denominado **estimação**.
  - Estudaremos combinações dos valores de amostras aleatórias, objetivando a obtenção de informações a respeito de características de interesse na população.
- **Notação:** vamos representar uma amostra de tamanho n, a ser retirada da população, por $(X_1, X_2, \ldots, X_n)$.

Parâmetro

Definição

As quantidades da população, em geral desconhecidas, sobre as quais temos interesse, são denominadas **parâmetros** e, usualmente, representadas por letras gregas θ, μ, σ, α, β, λ, entre outras.

Parâmetro

Exemplos

- a média de idade dos estudantes da UFRGS.
- a média de altura da população de Porto Alegre.
- a probabilidade de uma lâmpada ser produzida de maneira defeituosa em uma certa linha de fabricação.
- a média de peso de pacotes de queijo ralado.

Estimador e estimativa

Definição

A combinação dos elementos da amostra, construída com a finalidade de representar, ou estimar, um parâmetro de interesse na população, denominamos **estimador**. Em geral, denotamos os estimadores por símbolos com o acento circunflexo: $\hat{\theta}$, $\hat{\mu}$, $\hat{\sigma}$, etc. Aos valores numéricos assumidos pelos estimadores denominamos **estimativas pontuais** ou simplesmente **estimativas**.

Estimador e estimativa

Exemplos de estimadores

- $\hat{\mu}_1 = \bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, **a média amostral**, é um estimador para a média populacional.
- $\hat{\mu}_2 = X_1$, **a primeira observação da amostra**, é um estimador para a média populacional.
- $\hat{\mu}_3 = \frac{X_{(1)}+X_{(n)}}{2}$, **a média aritmética entre o valor mínimo ($X_{(1)}$) e o valor máximo ($X_{(n)}$) da amostra**, é um estimador para a média populacional.
- $\hat{\sigma}_1^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$ é um estimador para a variância populacional (este é chamado de **estimador "natural"** da variância populacional).
- $\hat{\sigma}_2^2 = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$ é um estimador para a variância populacional (este estimador é chamado de **variância amostral**).

Estimador e estimativa

Exemplos de estimativas

Suponha que a seguinte amostra de altura de pessoas (em metros) foi observada (1,65; 1,57; 1,72; 1,66; 1,71). Considerando os estimadores do exemplo anterior, temos as seguintes **estimativas**:

- $\hat{\mu}_1 = \frac{1{,}65+1{,}57+1{,}72+1{,}66+1{,}71}{5} = 1,662.$
- $\hat{\mu}_2 = 1,65.$
- $\hat{\mu}_3 = \frac{1{,}57+1{,}72}{2} = 1,645.$
- $\hat{\sigma}_1^2 = \frac{(1{,}65-1{,}662)^2+(1{,}57-1{,}662)^2+(1{,}72-1{,}662)^2+(1{,}66-1{,}662)^2+(1{,}71-1{,}662)^2}{5} \approx 0,0029.$
- $\hat{\sigma}_2^2 == \frac{(1{,}65-1{,}662)^2+(1{,}57-1{,}662)^2+(1{,}72-1{,}662)^2+(1{,}66-1{,}662)^2+(1{,}71-1{,}662)^2}{4} \approx 0,0036.$

Estimador e estimativa

- Notamos que um estimador, digamos $\hat{\theta}$, é uma **função das variáveis aleatórias** constituintes da amostra. Logo, um estimador **também é uma variável aleatória**.
- A correspondente distribuição de probabilidade formará a base das argumentações probabilísticas utilizadas na extrapolação da informação da amostra para os parâmetros da população.
- Diferentes amostras (observações diferentes e/ou diferentes tamanhos) produzirão diferentes estimativas para o mesmo parâmetro.
  - Amostra 1: (1,65; 1,57; 1,72; 1,66; 1,71)
    - $\hat{\mu}_1^1 = \bar{X}_5 = 1,662$.
  - Amostra 2: (1,78; 1,63; 1,82; 1,54; 1,78)
    - $\hat{\mu}_1^2 = \bar{X}_5 = 1,71$.
  - Amostra 3: (1,78; 1,63; 1,82; 1,54; 1,78; 1,72; 1,66; 1,71)
    - $\hat{\mu}_1^3 = \bar{X}_8 = 1,705$.

Propriedades dos estimadores

- Vimos nos exemplos anteriores que mais de uma função da amostra pode ser proposta para estimar um parâmetro de interesse.
- Para facilitar a escolha entre tais estimadores, torna-se importante verificar se possuem algumas **propriedades** que serão definidas a seguir.

Vício

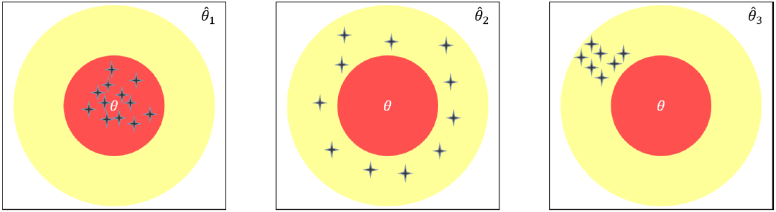
Um estimador $\hat{\theta}$ é **não viciado** ou **não viesado** para um parâmetro θ se

$$\mathrm{E}\left[\hat{\theta}\right] = \theta.$$

Em outras palavras, um estimador é não viciado se o seu valor esperado coincide com o parâmetro de interesse.

Propriedades dos estimadores

- Implícita à definição de **vício**, está a ideia de podermos retirar diversas amostras da população de interesse.
- Na figura abaixo, os estimadores $\hat{\theta}_1$ e $\hat{\theta}_2$ são não viciados para θ, enquanto que $\hat{\theta}_3$ é um estimador viciado para θ.



Distribuições amostrais

- Vimos que estimadores são funções de variáveis aleatórias e, portanto, eles também são variáveis aleatórias.
- Neste momento, vamos estudar a distribuição de probabilidade de alguns dos estimadores mais utilizados.
  - As distribuições de probabilidade de estimadores são chamadas **distribuições amostrais**.
- A partir das distribuições amostrais descreveremos a incerteza com respeito às nossas estimativas.
- Devido às suas propriedades (ausência de viés, consistência e eficiência), foco será dado aos estimadores $\bar{X}_n$ para a média populacional e S^2 para a variância populacional.

Distribuições amostrais

- Consideremos, inicialmente, o caso em que conseguimos calcular facilmente a função de probabilidade dos estimadores de interesse.

Exemplo

- Um jogo consiste em lançar uma moeda honesta 3 vezes.
- Para cada lançamento, se sair cara você **ganha** um ponto, caso saia coroa, você **perde** um ponto.
  - Podemos modelar essa situação através de uma variável X que, em uma população, pode assumir os valores -1 e 1, com probabilidades iguais.

x	-1	1
$\Pr(X = x)$	$1/2$	$1/2$

Distribuições amostrais

- Se observarmos a amostra $(-1, 1, -1)$, temos

$$\bar{x} = \frac{-1+1-1}{3} = -1/3;$$

$$s^2 = \frac{(-1-(-1/3))^2 + (1-(-1/3))^2 + (-1-(-1/3))^2}{(3-1)} = 4/3$$

- Se considerarmos outra amostra $(-1, 1, 1)$, temos

$$\bar{x} = \frac{-1+1+1}{3} = 1/3;$$

$$s^2 = \frac{(-1-1/3)^2 + (1-1/3)^2 + (1-1/3)^2}{(3-1)} = 4/3$$

Distribuições amostrais

- Assim, considerando todas as possíveis amostras, teríamos a seguinte tabela:

(X_1, X_2, X_3)	probabilidade	$\bar{X}$	S^2
$(-1, -1, -1)$	$1/8$	-1	0
$(-1, -1, +1)$	$1/8$	$-1/3$	$4/3$
$(-1, +1, -1)$	$1/8$	$-1/3$	$4/3$
$(-1, +1, +1)$	$1/8$	$1/3$	$4/3$
$(+1, -1, -1)$	$1/8$	$-1/3$	$4/3$
$(+1, -1, +1)$	$1/8$	$1/3$	$4/3$
$(+1, +1, -1)$	$1/8$	$1/3$	$4/3$
$(+1, +1, +1)$	$1/8$	1	0

Distribuições amostrais

- Com base na tabela anterior, podemos construir as distribuições amostrais de $\bar{X}$ e S^2:

$\bar{x}$	-1	$-1/3$	$1/3$	1
$\Pr(\bar{X} = \bar{x})$	$1/8$	$3/8$	$3/8$	$1/8$

e

s^2	0	$4/3$
$\Pr(S^2 = s^2)$	$1/4$	$3/4$

Distribuições amostrais

- Ainda,

$$\mathrm{E}\left[\bar{X}\right] = (-1) \times 1/8 + (-1/3) \times 3/8 + 1/3 \times 3/8 + 1 \times 1/8 = 0$$

e

$$\mathrm{E}\left[S^2\right] = 0 \times 1/4 + 4/3 \times 3/4 = 1$$

como já esperado, pois $\bar{X}$ e S^2 são **não viciados** para a média e variância populacional, respectivamente.

Para casa

- Revisar os tópicos discutidos nesta aula.
    - Ler os capítulos 5, 6, 10 e 11 do Livro "Estatística Básica"[1] (disponível no Sabi+).

[1] Morettin, P. A. e Bussab, W. O. **Estatística Básica**, Saraiva, 2010.

Próxima aula

- Introdução ao delineamento de pesquisa.

Por hoje é só!

Bons estudos!