

# MAT02025 - Amostragem 1

## Amostragem probabilística

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, 2021



# Teoria da amostragem

# O papel da teoria da amostragem

- ▶ O objetivo da teoria da amostragem é **aperfeiçoar processos de seleção de amostras e de avaliação** que proporcionem, aos menores custos possíveis, estimativas suficientemente precisas para os propósitos em vista.
- ▶ Para aplicar este princípio, devemos ser capazes de prever, para qualquer procedimento de amostragem que esteja sendo considerado, a **precisão** e o **custo esperados**.

# O papel da teoria da amostragem

- ▶ No que diz respeito à precisão, não podemos prever exatamente o quão grande um erro estará presente em uma estimativa em qualquer situação específica, **pois isso exigiria um conhecimento do verdadeiro valor para a população.**
- ▶ Em vez disso, a precisão de um procedimento de amostragem é avaliada examinando a **distribuição de frequência gerada para a estimativa**<sup>1</sup> se o **procedimento for aplicado repetidamente**<sup>2</sup> à mesma população.
- ▶ Esta é a técnica padrão pela qual a precisão é avaliada na teoria estatística.

---

<sup>1</sup>Também chamada de **distribuição amostral**, ou ainda, **distribuição de aleatorização**.

<sup>2</sup>Paradigma clássico: princípio da repetitibilidade.

# O papel da teoria da amostragem

- ▶ Uma simplificação adicional é introduzida: com amostras de tamanhos comuns na prática, muitas vezes há boas razões para supor que as estimativas da amostra são aproximadamente distribuídas de acordo com o modelo normal.
- ▶ Por exemplo, seja  $\hat{\theta}$  um estimador para um parâmetro  $\theta$ , então, sob certas condições, é razoável supor que  $\hat{\theta} \sim N(\mu_{\hat{\theta}}, \sigma_{\hat{\theta}}^2)$ , ou seja,

$$f_{\hat{\theta}}(\hat{\theta}) = \frac{1}{\sqrt{2\pi\sigma_{\hat{\theta}}^2}} \exp \left\{ \frac{1}{2\sigma_{\hat{\theta}}^2} (\hat{\theta} - \mu_{\hat{\theta}})^2 \right\}.$$

# O papel da teoria da amostragem

- ▶ Com uma estimativa normalmente distribuída, toda a forma da distribuição de frequência é conhecida se conhecermos a média e o desvio padrão (ou a variância).
- ▶ Uma parte considerável da teoria do levantamento por amostragem está preocupada em encontrar fórmulas para essas médias e variâncias.

# O papel da teoria da amostragem

- ▶ Existem duas diferenças entre a teoria padrão de levantamentos por amostragem e a teoria clássica de amostragem conforme ensinada em livros sobre estatística.
- ▶ Na **teoria clássica**, as medições que são feitas nas unidades de amostragem na população são geralmente assumidas seguir uma distribuição de frequência, por exemplo, a distribuição normal, de forma matemática conhecida, à parte de certos parâmetros populacionais, como a média e a variância cujos valores têm a ser estimado a partir dos dados da amostra<sup>3</sup>.

## Exemplo

A altura da população segue uma distribuição normal com média  $\mu$  e variância  $\sigma^2$  desconhecidas.

---

<sup>3</sup>Inferência baseada no modelo.

# O papel da teoria da amostragem

- ▶ Na teoria e levantamentos por amostragem, por outro lado, a atitude tem sido assumir apenas informações muito limitadas sobre essa distribuição de frequência.
- ▶ Em particular, sua forma matemática não é considerada conhecida, de modo que a abordagem pode ser descrita como livre de modelo ou livre de distribuição<sup>4</sup>.

---

<sup>4</sup>Inferência baseada no delineamento.



# O papel da teoria da amostragem

- ▶ Uma segunda diferença é que **as populações** no trabalho de levantamento **contêm um número finito de unidades**.
- ▶ Os resultados são um pouco mais complicados quando a amostragem é de uma população finita em vez de infinita.
- ▶ Para fins práticos, essas diferenças nos resultados para populações finitas e infinitas podem frequentemente ser ignoradas.
- ▶ Casos em que não seja assim serão apontados.

# Amostragem probabilística

# Amostragem probabilística

- ▶ Dentre os vários processos existentes para a obtenção de amostras, a amostragem probabilística caracteriza-se por garantir, *a priori*, que todo elemento pertencente à população de estudo possua probabilidade, **conhecida e diferente de zero**, de pertencer à amostra sorteada.
- ▶ A identificação, direta ou indireta, dos elementos e o sorteio deles fundamentam as propriedades matemáticas desse tipo de processo.

## Amostragem probabilística: propriedades

1. Podemos definir o conjunto de amostras distintas,  $S_1, S_2, \dots, S_v$ , que o procedimento é capaz de selecionar se aplicado a uma população específica.
  - ▶ Isso significa que podemos dizer exatamente quais unidades de amostragem pertencem a  $S_1$ , a  $S_2$ , e assim por diante.
  - ▶ Por exemplo, suponha que a população contenha seis unidades, numeradas de 1 a 6. Um procedimento comum para escolher uma amostra de tamanho 2 fornece três candidatos possíveis:  $S_1 = (1, 4)$ ;  $S_2 = (2, 5)$ ;  $S_3 = (3, 6)$ . Observe que nem todas as amostras possíveis de tamanho 2 precisam ser incluídas.

## Amostragem probabilística: propriedades

2. A cada amostra possível,  $S_j$ , é atribuído uma probabilidade conhecida de seleção,  $\pi_j$ .
3. Seleccionamos uma das  $S_j$  por um processo aleatório em que cada  $S_j$  recebe sua adequada probabilidade  $\pi_j$  de ser seleccionada.
  - ▶ No exemplo, podemos atribuir **probabilidades iguais** às três amostras. Em seguida, o sorteio em si pode ser feito escolhendo um **número aleatório**<sup>5</sup> entre 1 e 3. Se este número for  $\ell$ ,  $S_\ell$  é a amostra retirada.

---

<sup>5</sup>Veja o material suplementar sobre números aleatórios no Moodle.

## Amostragem probabilística: propriedades

4. O método para calcular a estimativa da amostra deve ser conhecido e deve levar a uma estimativa única para qualquer amostra específica.
  - ▶ Podemos declarar, por exemplo, que a estimativa deve ser a média das medidas nas unidades individuais da amostra.

## Amostragem probabilística: propriedades

- ▶ Para qualquer procedimento de amostragem que satisfaça essas propriedades, podemos calcular a distribuição de frequência das estimativas que ele gera, **se aplicado repetidamente à mesma população**.
- ▶ Sabemos com que frequência qualquer amostra particular  $S_j$  será selecionada, e sabemos como calcular a estimativa a partir dos dados em  $S_j$ .
- ▶ Portanto, é evidente que uma teoria de amostragem pode ser desenvolvida para qualquer procedimento desse tipo, embora os detalhes do desenvolvimento possam ser complexos.
- ▶ O termo **amostragem probabilística** se refere a um método desse tipo.

## Amostragem probabilística: propriedades

- ▶ Na prática, raramente extraímos uma amostra probabilística escrevendo  $S_j$  e  $\pi_j$  conforme descrito acima.
- ▶ Isso é insuportavelmente trabalhoso com uma grande população, onde um procedimento de amostragem pode produzir bilhões de amostras possíveis.
- ▶ O sorteio é mais comumente feito especificando-se as **probabilidades de inclusão para as unidades individuais**, e sorteando as unidades, uma por uma ou em grupos, até que o tamanho e tipo de amostra desejados sejam construídos.
- ▶ Para os propósitos de uma teoria, é suficiente saber que podemos escrever o  $S_j$  e  $\pi_j$  se quiséssemos e tivéssemos tempo ilimitado.



## Exercício

## Exercício

Considere, a título de ilustração, uma população composta dos elementos  $(A, B, C, D, E, F)$  (**Ana, Bruno, Carlos, Dorcina, Emília, Fernando**), nos quais se observou a característica  $X$  (**idade**). Então,  $N = 6$  e  $X$  é uma variável discreta (idade em anos). Logo,  $i = 1, 2, 3, 4, 5, 6$ . Os valores podem ser vistos na tabela a seguir

Elementos	$i$	$X_i$
A	1	2
B	2	4
C	3	6
D	4	8
E	5	10
F	6	12

## Exercício

Utilizando um sorteio **com reposição** de uma amostra de **dois elementos** dessa população, responda:

1. Liste as possíveis amostras. Qual o nome é dado a esta lista?
2. Qual a probabilidade de cada amostra ser selecionada? É preciso realizar alguma suposição para atribuição destas probabilidades?
3. Calcule a média amostral de  $X$  ( $\bar{x}_j = \frac{1}{2} \sum_{i \in S_j} X_i$ ) para cada amostra.
4. Faça o gráfico da distribuição de frequências da média amostral.

# Exercício

$S_j$	Amostras	$\pi_j$	$(x_1, x_2)$	$\bar{x}_j$
1				
2				
3				
4				
5				
$\vdots$				

# Amostragem não-probabilística

# Alternativas à amostragem probabilística

Alguns tipos comuns de amostragem não-probabilística:

1. **Amostragem convencional:** a amostra é restrita a uma parte da população que é facilmente acessível. Ex: uma amostra de carvão de um vagão aberto pode ser retirada da parte superior de 6 a 9 polegadas.
2. **Amostragem acidental ou a esmo:** a amostra é selecionada **ao acaso**. Ex: ao escolher 10 coelhos de uma grande gaiola em um laboratório, o investigador pode pegar aqueles em que suas mãos repousam, sem um planejamento consciente.
3. **Amostragem por cotas:** no caso de uma população pequena, mas heterogênea, o amostrador inspeciona o conjunto da população e escolhe uma pequena amostra de unidades “típicas” - isto é, unidades que estão próximas de sua impressão da média da população.

# Alternativas à amostragem probabilística

- 4. Amostragem de voluntários:** a amostra é constituída essencialmente por voluntários, em estudos em que o processo de medição é desagradável ou incômodo para a pessoa que está sendo medida.

## Para casa

- ▶ Repita o exercício da aula, mas agora utilizando um sorteio **sem reposição** de uma amostra de **dois elementos** daquela população.



# Próxima aula

- ▶ Distribuição normal, viés e EQM.

# Por hoje é só!

Bons estudos!

