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, 2023

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 predizer 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**[1] se o **procedimento for aplicado repetidamente**[2] à mesma população.

Esta é a técnica padrão pela qual a precisão é avaliada na teoria estatística.

[1] Também chamada de **distribuição amostral**, ou ainda, **distribuição de aleatorização**.

[2] 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 θ, então, sob certas condições, é razoável supor que $\hat{\theta} \dot{\sim} N\left(\mu_{\hat{\theta}}, \sigma^2_{\hat{\theta}}\right)$, ou seja,

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

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[3].

Exemplo

"A altura da população de estudantes da UFRGS segue uma distribuição normal com média μ e variância σ^2 desconhecidas".

[3] Inferência baseada no modelo.

O papel da teoria da amostragem

- Na **teoria de 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**[4].

O papel da teoria da amostragem

- Essa atitude é natural para **grandes levantamentos** nos quais **muitas medições diferentes** com diferentes distribuições de frequência são feitas nas unidades.
- Em levantamentos em que são feitas apenas algumas medições por unidade, estudos de suas distribuições de frequência podem justificar a suposição de formas matemáticas conhecidas, permitindo a aplicação dos resultados da teoria clássica (inferência baseada no modelo).

[4] 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

amostragem probabilística

Os procedimentos de amostragem considerados neste curso têm em comum as seguintes propriedades.

1. Podemos definir o conjunto de amostras distintas, $S_1, S_2, \ldots, S_v$, 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, π_j.
3. Selecionamos uma das S_j por um processo aleatório em que cada S_j recebe sua adequada probabilidade π_j de ser selecionada.
   - No exemplo, podemos atribuir **probabilidades iguais** às três amostras. Em seguida, o sorteio em si pode ser feito escolhendo um **número aleatório**[5] entre 1 e 3. Se este número for ℓ, S_ℓ é a amostra retirada.
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.

[5] Veja o material suplementar sobre números aleatórios no Moodle.

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 π_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 π_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	π_j	(x_1, x_2)	$\bar{x}_j$
1				
2				
3				
4				
5				
$\vdots$				

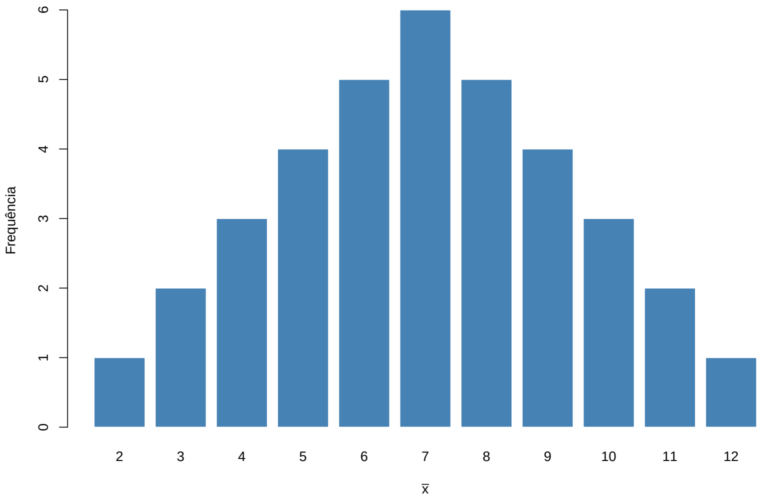
Exercício

S_j	Amostras	π_j	(x_1, x_2)	$\bar{x}_j$
1	(A, A)	1/36	(2, 2)	2
2	(B, A)	1/36	(4, 2)	3
3	(C, A)	1/36	(6, 2)	4
4	(D, A)	1/36	(8, 2)	5
5	(E, A)	1/36	(10, 2)	6
6	(F, A)	1/36	(12, 2)	7
7	(A, B)	1/36	(2, 4)	3
8	(B, B)	1/36	(4, 4)	4
9	(C, B)	1/36	(6, 4)	5
10	(D, B)	1/36	(8, 4)	6
11	(E, B)	1/36	(10, 4)	7
12	(F, B)	1/36	(12, 4)	8
13	(A, C)	1/36	(2, 6)	4
14	(B, C)	1/36	(4, 6)	5
15	(C, C)	1/36	(6, 6)	6
16	(D, C)	1/36	(8, 6)	7
17	(E, C)	1/36	(10, 6)	8
18	(F, C)	1/36	(12, 6)	9
19	(A, D)	1/36	(2, 8)	5
20	(B, D)	1/36	(4, 8)	6
21	(C, D)	1/36	(6, 8)	7

Exercício

22	(D, D)	1/36	(8, 8)	8
23	(E, D)	1/36	(10, 8)	9
24	(F, D)	1/36	(12, 8)	10
25	(A, E)	1/36	(2, 10)	6
26	(B, E)	1/36	(4, 10)	7
27	(C, E)	1/36	(6, 10)	8
28	(D, E)	1/36	(8, 10)	9
29	(E, E)	1/36	(10, 10)	10
30	(F, E)	1/36	(12, 10)	11
31	(A, F)	1/36	(2, 12)	7
32	(B, F)	1/36	(4, 12)	8
33	(C, F)	1/36	(6, 12)	9
34	(D, F)	1/36	(8, 12)	10
35	(E, F)	1/36	(10, 12)	11
36	(F, F)	1/36	(12, 12)	12

Exercício

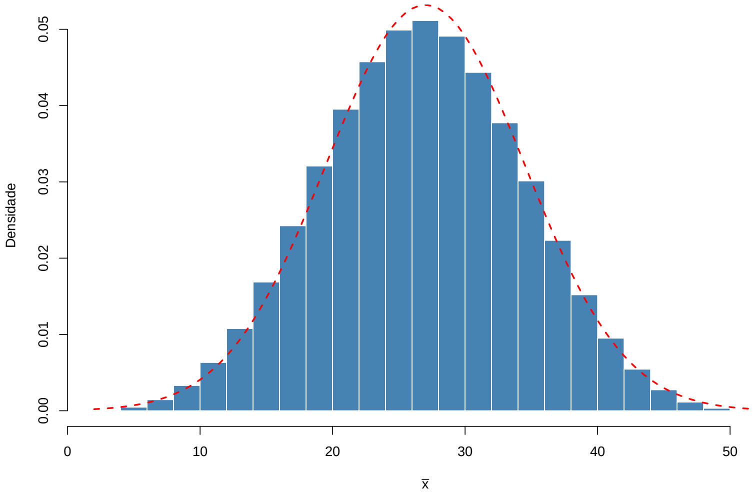


Exercício

Extras:

5. Onde você já viu este esquema de sorteio?
6. Discuta de onde vem a distribuição de $\bar{x}$.
7. Imagine $N = 26$ e amostras de tamanho 4.

Exercício



Amostragem não-probabilística

Amostragem não-probabilística

Ver os slides do **Prof. Wagner Hugo Bonat** do **Departamento de Estatística** da **Universidade Federal do Paraná** (Moodle).

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!