MAT02025 - Amostragem 1

AAS: propriedades dos estimadores

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

Relembrando

Amostragem aleatória simples

A **amostragem aleatória simples**[1] (AAS) é um processo para selecionar n unidades de N de modo que cada uma das ${}_N C_n$ amostras distintas tenha uma **chance igual de ser extraída**.

[1] Também conhecida como **amostragem casual simples** ou **amostragem acidental irrestrita**

Definições e notação

- As letras **maiúsculas** referem-se às características da população e as **minúsculas** às da amostra.
  - Para totais e médias, temos as seguintes definições.

Característica	População	Amostra
Total	$Y_T = \sum_{i=1}^{N} Y_i = Y_1 + Y_2 + \ldots + Y_N$	$y_T = \sum_{i=1}^{n} Y_i = Y_1 + Y_2 + \ldots + Y_n$
Média	$\overline{Y} = \frac{\sum_{i=1}^{N} Y_i}{N} = \frac{Y_1+Y_2+\ldots+Y_N}{N}$	$\overline{y} = \frac{\sum_{i=1}^{n} Y_i}{n} = \frac{Y_1+Y_2+\ldots+Y_n}{n}$

Definições e notação

- A estimativa das três primeiras quantidades será discutida na primeira parte desta área.
- O símbolo " ^ " denota uma estimativa de uma característica da população feita a partir de uma amostra.
- De acordo com a relação acima, temos:

	Representação da estimativa
Média da população $\overline{Y}$	$\hat{\overline{Y}} = \overline{y} = \sum_{i=1}^{n} Y_i/n$
Total da população Y_T	$\hat{Y}_T = N\overline{y} = N\sum_{i=1}^{n} Y_i/n$
Índice da população R	$\hat{R} = \sum_{i=1}^{n} Y_i/\sum_{i=1}^{n} X_i$

Propriedades dos estimadores

Propriedades dos estimadores

- Como vimos, um método de estimativa é **imparcial (não viciado, não enviesado, não tendencioso)** se o valor médio da estimativa, **tomado em todas as amostras possíveis de dado tamanho** n, for exatamente igual ao valor verdadeiro da população.
- Se o método tiver que ser imparcial, irrestritamente, este resultado deve ser válido para qualquer população de valores finitos, Y_i, e para qualquer que seja n.
- Para investigar se $\overline{y}$ é imparcial com a amostragem aleatória simples, calculamos o valor de $\overline{y}$ para todas as amostras ${}_N C_n$ e encontramos a média das estimativas.
- O símbolo E denota essa média sobre todas as amostras possíveis.

Propriedades dos estimadores

Teorema 9.1

A média amostral $\overline{y}$ é uma estimativa sem tendência para $\overline{Y}$.

Propriedades dos estimadores

Demonstração. Por definição,

$$
\begin{aligned}
\mathrm{E}\left(\overline{y}\right) &= \sum_{i=1}^{{}_N C_n} \left[\overline{y}_i \times \pi_i\right] \\
&= \sum_{i=1}^{{}_N C_n} \left[\overline{y}_i \times \frac{1}{{}_N C_n}\right] \ (\pi_i = 1/{}_N C_n) \\
&= \frac{1}{N!/[n!(N-n)!]} \sum_{i=1}^{{}_N C_n} \left[\frac{1}{n} \sum_{j=1}^{n} Y_j(i)\right] \\
&= \frac{1}{n[N!/n!(N-n)!]} \sum_{i=1}^{{}_N C_n} \left[Y_1(i) + Y_2(i) + \ldots + Y_n(i)\right]. \qquad (1)
\end{aligned}
$$

Propriedades dos estimadores

- Para avaliar essa soma, encontramos em quantas amostras qualquer valor específico Y_j aparece.
- Uma vez que existem $(N-1)$ outras unidades disponíveis para o resto da amostra e $(n-1)$ outros locais para preencher a amostra, o número de amostras contendo Y_j é

$$ {}_{N-1}C_{n-1} = \frac{(N-1)!}{(n-1)!(N-n)!}. $$

Propriedades dos estimadores

- Portanto,

$$\begin{aligned}
\sum_{i=1}^{{}_NC_n}[Y_1(i)+Y_2(i)+\ldots+Y_n(i)] &= {}_{N-1}C_{n-1}\times Y_1 + {}_{N-1}C_{n-1}\times Y_2 + \ldots {}_{N-1}C_{n-1}\times Y_N \\
&= {}_{N-1}C_{n-1}\times (Y_1+Y_2+\ldots+Y_N) \\
&= {}_{N-1}C_{n-1}\times \sum_{j=1}^{N} Y_j \\
&= \frac{(N-1)!}{(n-1)!(N-n)!}\times \sum_{j=1}^{N} Y_j. \qquad (2)
\end{aligned}$$

Propriedades dos estimadores

- Combinando (2) com (1), temos

$$
\begin{aligned}
\mathrm{E}\left(\overline{y}\right) &= \frac{n!(N-n)!}{nN!} \times \frac{(N-1)!}{(n-1)!(N-n)!} \times \sum_{j=1}^{N} Y_j \\
&= \frac{n(n-1)!(N-n)!}{nN(N-1)!} \times \frac{(N-1)!}{(n-1)!(N-n)!} \times \sum_{j=1}^{N} Y_j \\
&= \frac{1}{N} \sum_{j=1}^{N} Y_j = \overline{Y}. \qquad (3)
\end{aligned}
$$

Propriedades dos estimadores

Corolário 9.1

$\hat{Y}_T = N\overline{y}$ é um estimador imparcial do valor total populacional Y_T.

- **Para casa:** utilize o **Teorema 9.1** apresentado na aula para demonstrar o **Corolário**.

Um outro método de demonstração

Outro método de demonstração

- **Cornfield (1944)**[1] sugeriu um método para verificar os principais resultados das amostras aleatórias simples, sem reposição, que nos permite utilizar os resultados padronizados da teoria das populações infinitas.
- Seja a_i uma variável aleatória, e que assume o valor 1 se a i-ésima unidade for selecionada para a amostra, e assume valor 0 em caso contrário[2].

[1] Cornfield, J. (1944). *On Samples from Finite Populations*. Journal of the American Statistical Association, 39(226), 236.

[2] Alguns autores utilizam a notação $Y_i = a_i Y_i^{obs} + (1 - a_i) Y_i^{aus}$, em que obs e aus represetam **observado** e **ausente**, respectivamente.

Outro método de demonstração

Exemplo

Considere mais uma vez a população de tamanho $N = 6$, e $Y_1 = 2,\ Y_2 = 4,\ Y_3 = 6,\ Y_4 = 8,\ Y_5 = 10,\ Y_6 = 12$, e uma amostra aleatória simples, sem reposição, de tamanho $n = 2$. Então,

			Indicadores de seleção					
Sorteio	Probabilidade	Amostra	a_1	a_2	a_3	a_4	a_5	a_6
1	$1/15$	$(2, 4)$	1	1	0	0	0	0
2	$1/15$	$(2, 6)$	1	0	1	0	0	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
15	$1/15$	$(10, 12)$	0	0	0	0	1	1

Outro método de demonstração

- A **média amostral** pode ser expressa por

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{N} a_i Y_i,$$

em que o somatório abrange todas as N unidades da população.

Outro método de demonstração

- Nessa expressão, os a_i **são variáveis aleatórias**, e os Y_i constituem **um conjuntos de números fixos**.
- Note que

$$
\begin{aligned}
\Pr(a_i = 1) &= \frac{\#\text{amostras que incluem o i-ésimo elemento}}{\#\text{amostras que podem ser sorteadas}} \\
&= \frac{{}_{N-1}C_{n-1}}{{}_{N}C_{n}} \\
&= \frac{(N-1)!}{(n-1)!(N-n)!} \times \frac{n!(N-n)!}{N!} = \frac{n}{N}.
\end{aligned}
$$

- E consequentemente, $\Pr(a_i = 0) = 1 - \frac{n}{N}$.

Outro método de demonstração

- Dessa forma, a_i se distribui como uma **variável binomial**, em uma única tentativa[3], com $P = n/N$. Portanto,

$$\mathrm{E}\left(a_i\right) = P = \frac{n}{N}, \quad \mathrm{Var}\left(a_i\right) = \frac{n}{N}\left(1 - \frac{n}{N}\right).$$

[3] Ou seja, uma variável com distribuição Bernoulli.

Outro método de demonstração

- Assim,

$$E\left(\overline{y}\right) = E\left[\frac{1}{n}\sum_{i=1}^{N} a_i Y_i\right] = \frac{1}{n}\sum_{i=1}^{N} E\left[a_i\right] Y_i = \frac{1}{n}\sum_{i=1}^{N} P Y_i = \overline{Y}.$$

Para casa

- Refaça as demonstrações da aula de hoje.
- Leia o artigo **"On Samples from Finite Populations"**[4], disponível no Moodle.

[4] Cornfield, J. (1944). *On Samples from Finite Populations*. Journal of the American Statistical Association, 39(226), 236.

Próxima aula

- Variâncias dos estimadores.

Por hoje é só!

Bons estudos!