

Desafio 1: Distribuição da Média Amostral no Problema das Tiras de Papel

Amostragem

Data máxima de entrega: 2/05/2025

Objetivo

Este projeto tem como objetivo investigar, através de simulação computacional em R, a distribuição da média amostral (\bar{y}) obtida de uma população finita (representada por “tiras de papel” numeradas de 1 a N). Serão explorados os planos de amostragem aleatória simples com e sem reposição. Espera-se que o aluno compreenda como o tamanho da amostra e o tipo de amostragem afetam as propriedades da média amostral (seu valor esperado, variância e forma da distribuição), comparando os resultados empíricos com a teoria de amostragem.

1 Contexto Teórico

Consideramos uma população finita $U = \{Y_1, Y_2, \dots, Y_N\}$, onde os valores $Y_k = k$ para $k = 1, 2, \dots, N$. O parâmetro de interesse é a média populacional:

$$\mu = \frac{1}{N} \sum_{k=1}^N Y_k$$

Para a população $U = \{1, 2, \dots, N\}$, a média populacional é $\mu = \frac{N+1}{2}$.

A variância populacional (usando divisor N) é:

$$\sigma^2 = \frac{1}{N} \sum_{k=1}^N (Y_k - \mu)^2 = \frac{N^2 - 1}{12}$$

A variância populacional (usando divisor $N - 1$) é:

$$S^2 = \frac{1}{N-1} \sum_{k=1}^N (Y_k - \mu)^2 = \frac{N(N+1)}{12}$$

O estimador para μ é a média amostral, $\bar{y} = \frac{1}{n} \sum_{i \in s} y_i$, onde s é uma amostra de tamanho n .

Propriedades da Média Amostral (\bar{y}):

1. Amostragem Aleatória Simples Sem Reposição (AASSR):

- Valor Esperado: $\mathbb{E}_{AASSR}[\bar{y}] = \mu$ (não-viesado)
- Variância: $\text{Var}_{AASSR}(\bar{y}) = \frac{S^2}{n} \left(1 - \frac{n}{N}\right) = \frac{S^2}{n} \text{fpc}$, onde fpc é o fator de correção para população finita. Substituindo S^2 : $\text{Var}_{AASSR}(\bar{y}) = \frac{N(N+1)}{12n} \left(\frac{N-n}{N-1}\right)$ (para $N > 1$). Para $N = 1$, a variância é 0. *Nota:* A fórmula $\frac{S^2}{n} \left(1 - \frac{n}{N}\right)$ pode ser reescrita como $\frac{\sigma^2}{n} \frac{N-n}{N-1}$. Para a população $\{1, \dots, N\}$, isto é $\frac{N^2-1}{12n} \frac{N-n}{N-1}$.

2. Amostragem Aleatória Simples Com Reposição (AASCR):

- Valor Esperado: $\mathbb{E}_{AASCR}[\bar{y}] = \mu$ (não-viesado)
- Variância: $\text{Var}_{AASCR}(\bar{y}) = \frac{\sigma^2}{n}$ Substituindo σ^2 : $\text{Var}_{AASCR}(\bar{y}) = \frac{N^2-1}{12n}$

2 Etapas do Projeto

1. Definição da População

- Definir uma população de “tiras de papel” numeradas de 1 a N .
 - Exemplo: $N = 20$. A população é $U = \{1, 2, \dots, 20\}$.
 - Calcular a média populacional μ e as variâncias populacionais σ^2 e S^2 para o N escolhido.

2. Processo de Amostragem e Cálculo da Média

- Para cada tipo de amostragem (AASSR e AASCR):
 - Para cada tamanho de amostra n_{atual} (variando de 1 até um n_{max} definido, por exemplo, $n_{max} = N$ para AASSR e $n_{max} = N$ ou $1.5N$ para AASCR):
 - * Realizar $R = 2500$ repetições (sortear R amostras).
 - * Em cada repetição, calcular a média dos valores selecionados (\bar{y}_j para $j = 1, \dots, R$).

3. Avaliação e Análise

- Para cada combinação de tipo de amostragem e tamanho de amostra n_{atual} :
 - Analisar a distribuição das $R = 2500$ médias amostrais (\bar{y}_j):
 - * Calcular a média das médias amostrais: $\bar{\bar{y}} = \frac{1}{R} \sum_{j=1}^R \bar{y}_j$. Comparar com μ .
 - * Calcular a variância empírica das médias amostrais: $\text{Var}_{emp}(\bar{y}) = \frac{1}{R-1} \sum_{j=1}^R (\bar{y}_j - \bar{\bar{y}})^2$. Comparar com a variância teórica ($\text{Var}_{AASSR}(\bar{y})$ ou $\text{Var}_{AASCR}(\bar{y})$).
 - * Gerar histogramas da distribuição das médias amostrais.
 - * Opcional: Gerar QQ-plots para avaliar a normalidade.

3 Sugestões de Implementação no R

- Utilize as funções `sample()` para realizar a amostragem, especificando `replace = FALSE` para AASSR e `replace = TRUE` para AASCR.
- Organize os resultados em listas ou data frames para facilitar a análise e a geração de gráficos e tabelas.
- O pacote `ggplot2` pode ser útil para criar histogramas comparativos e outros gráficos.
- Para tabelas formatadas, considere os pacotes `knitr` (com a função `kable()`) e `kableExtra`.

4 Tópicos de Investigação

Com base nos resultados das simulações, discuta os seguintes pontos:

- **Não-viesamento:** A média das médias amostrais ($\bar{\bar{y}}$) aproxima-se da média populacional μ para ambos os tipos de amostragem e diferentes tamanhos de amostra n_{atual} ?
- **Variância:**
 - Como a variância empírica das médias amostrais ($\text{Var}_{emp}(\bar{y})$) se compara com a variância teórica ($\text{Var}_{AASSR}(\bar{y})$ e $\text{Var}_{AASCR}(\bar{y})$)?
 - Como o aumento do tamanho da amostra n_{atual} afeta a variância das médias amostrais? Compare a taxa de diminuição entre AASSR e AASCR.
 - Qual o impacto do fator de correção para população finita (fpc) na variância da média amostral em AASSR, especialmente quando n_{atual} é uma fração considerável de N ?
- **Forma da Distribuição:**

- Como a forma da distribuição das médias amostrais (visualizada por histogramas) muda com o aumento de n_{atual} ?
- A distribuição das médias amostrais tende à normalidade (Teorema Central do Limite)? Isso ocorre de forma similar para AASSR e AASCR?
- **Comparação entre AASSR e AASCR:**
 - Para um mesmo tamanho de amostra n_{atual} , qual método de amostragem geralmente resulta em menor variância para a média amostral? Por quê?
 - Existem situações (tamanhos de amostra n_{atual} em relação a N) onde as diferenças entre AASSR e AASCR são mais ou menos pronunciadas?

Análise do Viés da Média Amostral e Variância da Média Amostral

Análise da Forma da Distribuição das Médias Amostrais

Comparação Direta entre AASSR e AASCR

(Discussão sobre as diferenças observadas nas métricas e distribuições)

5 Conclusões Finais

(Resumir as principais descobertas do estudo sobre o comportamento da média amostral sob diferentes cenários de amostragem, relacionando com a teoria.)

6 Entrega

- Relatório em PDF contendo:
 - Introdução teórica breve.
 - Descrição da população e do processo de simulação da amostragem.
 - Métricas descritivas (tabelas de médias, variâncias).
 - Gráficos das distribuições das médias amostrais (histogramas, etc.).
 - **Discussão detalhada dos resultados** em relação aos tópicos de investigação.
- Código-fonte implementado em R bem comentado (este arquivo .Rmd ou um script .R separado).

7 Referências Sugeridas (Opcional)

- Cochran, W. G. (1977). *Sampling Techniques* (3rd ed.). Wiley.
- Lohr, S. L. (2019). *Sampling: Design and Analysis* (3rd ed.). Chapman and Hall/CRC.
- Särndal, C. E., Swensson, B., & Wretman, J. (2003). *Model Assisted Survey Sampling*. Springer.