MAT02036 - Amostragem 2

Aula 20 - Amostragem Sistemática - Parâmetros e Estimação

Markus Stein

Departamento de Estatística, IME/UFRGS

2022/2

Housekeeping

- Aproveitem o momento presencial para tirar dúvidas
- Se estivéssemos no ensino remoto ou à distância
  - vocês poderiam estar somente ouvindo, sem interação
  - ou assistindo vídeos e material em outro momento
- Depois das aulas, rever material da aula passada
  - fazer exercícios
  - se preparar para a próxima aula

Aula passada 💾

Amostragem Sistemática Simples

- *O método*: selecionar cada K-ésima unidade da população;
    - N - tamanho da população;

$$N = nK + c, \quad 0 \backslash \mathrm{leq\ c} < \mathrm{K}$$

    - K - **intervalo de selação**;
    - $n = \lfloor N/K \rfloor$ tamanho da amostra;
    - c - é o resto da divisão N/K;
    - r - **valor inicial**, número inteiro de 1 a K,

$$r \sim Uniforme - Discreta(K);$$

- Na **AS** a amostra $s_r = \{i : i = r + lK \leq N; \;\; l = 0, \ldots, n\}$, satisfaz

$$p(s) = \begin{cases} 1/K, \text{ se } \; s = s_r \text{ para } r = 1, \;\; 2, \;\; \ldots, \;\; K \\ 0, \;\; \text{caso contrário} \end{cases}$$

Aula passada

Amostragem Sistemática Simples

- A probabilidade de inclusão na amostra de uma unidade i qualquer é dada por:

$$\pi_i = \frac{1}{K}, \; i = 1, \ldots, \; N$$

- A probabilidade de inclusão das unidades $i \neq j$ na amostra é dada por:

$$\pi_{ij} = \begin{cases} 1/K, & \text{se } i \neq j \in s_r \; \text{ para } \; r = 1, \ldots, K \\ 0, & \text{caso contrário} \end{cases}$$

- As variáveis indicadoras associadas às amostras possíveis s_r:

$$I(r) = \begin{cases} 1, & \text{se a amostra é } s_r \; \text{ para } 1 \leq r \leq K \\ 0, & \text{caso contrário} \end{cases}$$

- O valor esperado de $I(r)$ é

$$E_{AS}[I(r)] = 1/K \,, \; r = 1, \ldots, K,$$

Estimação

Amostragem Sistemática Simples

Estimação de totais na **AS**

- O estimador tipo Horvitz-Thompson do total sob **AS**,
  - o peso amostral das unidades da amostra é sempre igual a $d_i = 1/\pi_i = K$, então

$$\widehat{T}_{AS} = Kt_r = K \sum_{i \in s_r} y_i$$

  em que $t_r = \sum_{i \in s_r} y_i$ é a soma amostral dos valores observados da variável y.

- Já sabemos que este estimador é não viciado para o total populacional.

$$\begin{aligned} E_{AS}(\widehat{T}_{AS}) &= E_{AS}\left[Kt_r\right] = KE_{AS}\left[\sum_{r=1}^{K} I(r)t_r\right] \\ &= K\sum_{r=1}^{K} E_{AS}\left[I(r)\right]t_r = K\sum_{r=1}^{K} \frac{1}{K}t_r = \sum_{r=1}^{K} t_r = T \end{aligned}$$

Amostragem Sistemática Simples

Estimação de totais na **AS**

Exemplo

Considere a população composta de $N = 19$ unidades, cujos dados da variável de interesse y, da qual se deseja retirar uma amostra sistemática simples com intervalo de seleção com $K = 4$ para estimar o total populacional. Verifique numericamente que o estimador $\widehat{Y}_{AS}$ é não viciado.

s_1	s_2	s_3	s_4
99	54	96	54
85	88	55	83
62	85	96	55
91	92	67	68
54	79	76	

Para mostrar que o estimador é não viciado, basta verificar que a média dos seus valores possíveis é igual ao parâmetro populacional a ser estimado.

Amostragem Sistemática Simples

Estimação de totais na **AS**

```
K=4      # Intervalo de seleção        #print(paste("Intervalo de sele
pop=matrix(c(99,54,96,54,85,88,55,83,62,85,96,55,91,92,67,68,54,79,76
## Calculando a estimativa do total para cada uma das AS possíveis
Yhatr=NULL
for(l in (1:K)) Yhatr[l] = K*sum(pop[,l],na.rm=T)
EYhat=mean(Yhatr)     # Calculando a média das estimativas de total
EYhat
```

```
## [1] 1439
```

```
## Calculando o total populacional Y
Y=sum(pop,na.rm=T)
Y
```

```
## [1] 1439
```

Amostragem Sistemática Simples

Estimação de médias na **AS**

A média populacional $\overline{Y} = Y/N$ pode ser vista como uma razão entre o total da variável y e o total de unidades populacionais: $$ \overline Y = \frac Y N = \displaystyle \frac {\sum{r=1}^{K} t_r} {|sum{r=1}^{K} n_r} $$ Para estimá-la deve-se considerar duas situações: quando o tamanho \(N\) da população é conhecido, ou quando esse valor é desconhecido. Um estimador não viciado para a média quando \(N\) é conhecido é dado por:

$$ \overline y{AS} = |frac {|widehat Y{AS}} N = \frac {K tr}{N} (#eq:eqsis4) $$ *Este estimador é não viciado para \(\overline Y\), pois foi visto que `(\widehat Y*{AS})é não viciado para o total populacional(Y). Também é interessante observar que este estimador não é igual à média amostral, a menos no caso em que(N = nK)`.

(#exm:exmsis6) Considerando a mesma população do Exemplo \@ref(exm:exmsis5), pode-se verificar que $\bar{y}_{AS}$ é não viciado para a média populacional $\overline{Y}$. Neste caso, basta calcular a estimativa da média para cada coluna (amostra sistemática possível) da Tabela \@ref(tab:tabsis2), calcular o valor médio das estimativas e comparar com média populacional. Pode-se utilizar o R para realizar a tarefa.

Para casa 🏠

- Continuar exercícios.
- Ler o capítulo 3 da apostila da Profa. Vanessa.
- Ler o capítulo 8 do livro 'Amostragem: Teoria e Prática Usando R'.
- Rever os slides.
- Preparação para avaliação parcial 2

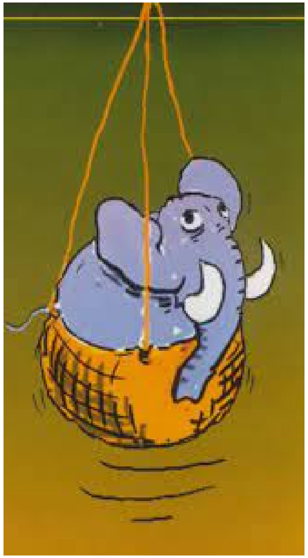
Próxima aula 📊

- Acompanhar o material no moodle.

Amostragem Sistemática

- Estimação.

Muito obrigado!



Fonte: imagem do livro *Combined Survey Sampling Inference: Weighing of Basu's Elephants*.

Referências

- Amostragem: Teoria e Prática Usando o R
- **Elementos de Amostragem**, Bolfarine e Bussab.
- Cochran(1977)

Resumo da notação