

MAT02036 - Amostragem 2

Aula 13 - Amostragem por Conglomerados - Cont. Estimação e Lab R

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 por Conglomerados em 1 Estágio Simples

Abaixo segue um resumo de estimadores do total, média e respectivas variâncias sob AC1S.

Estimador	HT
Total	$\hat{T}_{AC1S/HT} = \frac{M}{m} \sum_{i \in a} T_i = M \bar{y}_C = \sum_{i \in a} \sum_{j \in C_i} d_{ij} y_{ij}$
Média	$\bar{y}_{AC1S/HT} = \frac{\hat{T}_{AC1S/HT}}{N} = \frac{M}{N} \frac{1}{m} \sum_{i \in a} Y_i = \bar{y}_C / \bar{N}$
Variância do total	$\hat{Var}_{AC1S} \left(\hat{T}_{AC1S/HT} \right) = M^2 \left(\frac{1}{m} - \frac{1}{M} \right) \hat{S}_{ec}^2$
Variância da média	$\hat{Var}_{AC1S} \left(\bar{y}_{AC1S/HT} \right) = \frac{1}{N^2} \left(\frac{1}{m} - \frac{1}{M} \right) \hat{S}_{ec}^2$

Aula passada

Amostragem por Conglomerados em 1 Estágio Simples

Abaixo segue um resumo de estimadores do total, média e respectivas variâncias sob AC1S.

Estimador	Razão
Total	$\hat{T}_{AC1S}^R = \frac{N}{n} \sum_{i \in a} T_i = N\bar{y} = \sum_{i \in a} \sum_{j \in C_i} w_{ij}^R y_{ij}$
Média	$\bar{y}_{AC1S}^R = \frac{\hat{T}_{AC1S}^R}{N} = \frac{1}{n} \sum_{i \in a} T_i = \bar{y}_C / \bar{n} = \bar{y}$
Variância do total	$\hat{V}ar_{AC1S} \left(\hat{T}_{AC1S}^R \right) = M^2 \left(\frac{1}{m} - \frac{1}{M} \right) \frac{1}{m-1} \sum_{i \in a} N_i^2 (\bar{Y}_i - \bar{y})^2$
Variância da média	$\hat{V}ar_{AC1S} \left(\bar{y}_{AC1S}^R \right) = \frac{1}{n^2} \left(\frac{1}{m} - \frac{1}{M} \right) \frac{1}{m-1} \sum_{i \in a} N_i^2 (\bar{Y}_i - \bar{y})^2$

Continuação AC1S e Laboratório R

Amostragem por Conglomerados em 1 Estágio Simples

Exercício

Um certo país possui $M = 10$ companhias aéreas com N_i aviões cada. As milhas (*em milhares*) de cada avião (y_{ij}) num determinado período de tempo foram registradas.

Cia (i)	No. aviões (N_i)	T_i	\bar{Y}_i
1	10	40	4
2	15	75	5
3	15	75	5
4	15	60	4
5	10	60	6
6	15	90	6
7	15	75	5
8	10	70	7
9	10	40	4
10	15	90	6
Total	130	675	52

Amostragem por Conglomerados em 1 Estágio Simples

Exercício

- a. Calcule os parâmetros total (T) e média, individual (\bar{Y}) e por conglomerados (\bar{Y}_C) e a variância entre totais dos conglomerados S_{ec}^2 .
- b*. Calcule o viés dos estimadores **HT** e de **razão** para o total, \hat{T} (ou média, \bar{y}), são não viesados para os respectivos parâmetros que se destinam a estimar, T e \bar{Y} . (Obs. mostrar analiticamente ou com os dados do exercício)
- b. Assumindo o plano **AC1S** com **AASs** de conglomerados, para amostras de tamanho $m = 4$, calcule a variância do estimador natural (HT) do total, $Var_{AC1S}(\hat{T}^{HT})$, e a variância do estimador da média, $Var_{AC1S}(\bar{y}^{HT})$.
- c. Repetir o item (b) para estimador o estimador de razão.
- d. Escolha um estimador para o total (ou para a média), selecione uma amostra e estime o parâmetro com base na amostra observada.

Solução

Amostragem por Conglomerados em 1 Estágio Simples

Exercício

```
## dados do problema
i <- 1:10 # indice dos estratos
Ni <- c(10, 15, 15, 15, 10, 15, 15, 10, 10, 15) # tamanhos dos estratos
Ti <- c(40, 75, 75, 60, 60, 90, 75, 70, 40, 90) # totais dos estratos
N <- sum(Ni) # tamanho da populacao
M <- length(Ni) # no. estratos na populacao
m <- 4 # no. estratos na amostra
```

a. Sabemos que

Definição	Parâmetro
Total populacional	$T = \sum_{i=1}^M T_i = \sum_{i \in C} T_i$
Média populacional por unidade	$\bar{Y} = T/N = \frac{1}{N} \sum_{i \in C} T_i$
Média populacional por conglomerado	$\bar{Y}_C = T/M = \frac{1}{M} \sum_{i \in C} T_i$
Variância entre totais	$S_{ec}^2 = \frac{1}{M-1} \sum_{i \in C} (T_i - \bar{Y}_C)^2$

Amostragem por Conglomerados em 1 Estágio Simples

Exercício

```
## (a) parametros  
Ty <- sum( Ti)  
Ty
```

```
## [1] 675
```

```
Ybarra <- Ty / N  
Ybarra
```

```
## [1] 5.192308
```

```
Ybarrac <- Ty / M  
Ybarrac
```

```
## [1] 67.5
```

```
S2ec <- (1/(M-1)) * sum((Ti - Ybarrac)^2)  
S2ec
```

Amostragem por Conglomerados em 1 Estágio Simples

Exercício

b*. Usando o R podemos gerar todas as $\binom{M}{m} = \binom{10}{4} = 210$ amostras possíveis, calcular a estimativa baseada em cada possível amostra e calcular a média das estimativas de todas as amostras. Sabemos que os estimadores *HT* são

$$t_{y,AC1S/HT} = \hat{T}_{AC1S/HT} = \frac{M}{m} \sum_{i \in a} T_i \text{ e } \bar{y}_{AC1S/HT} = \frac{\hat{T}_{AC1S/HT}}{N}$$

```
## (b*) Distribuicao amostral dos estimadores do total e da média
n_amostras <- choose(M, m) # no. possiveis amostras
Ti_amostras <- combn(Ti, m) # lista possiveis amostras dos totais
## estimadores HT
ty_HT <- (M / m) * apply( Ti_amostras, 2, sum) # estimativas HT totais
mean(ty_HT - Ty) # vies HT total
```

```
## [1] 0
```

```
ybarra_HT <- ty_HT / N # estimativas HT media
mean(ybarra_HT - Ybarra) # vies HT media
```

Amostragem por Conglomerados em 1 Estágio Simples

Exercício

E para o estimador de razão

$$\hat{T}_{AC1S}^R = \frac{N}{n} \sum_{i \in a} T_i = N\bar{y} \text{ e } \bar{y}_{AC1S}^R = \frac{\hat{T}_{AC1S}^R}{N}$$

```
## estimadores de RAZAO
Ni_amostras <- combn(Ni, m)      # lista possiveis tamanhos de congl
n <- apply(Ni_amostras, 2, sum)   # tamanhos de amostras possiveis
ty_R <- (N / n) * apply(Ti_amostras, 2, sum) # estimativas de razao
mean(ty_R - Ty)                  # vies de razao total
```

```
## [1] 0.1511544
```

```
ybarra_R <- ty_R / N              # estimativas de razao
mean(ybarra_R - Ybarra)           # vies de razao media
```

```
## [1] 0.001162726
```

Amostragem por Conglomerados em 1 Estágio Simples

Exercício

b. Sabemos que

$$Var_{AC1S} \left(\hat{T}_{AC1S/HT} \right) = M^2 \left(\frac{1}{m} - \frac{1}{M} \right) S_{ec}^2$$

e

$$Var_{AC1S} \left(\bar{y}_{AC1S/HT} \right) = \frac{M^2}{N^2} \left(\frac{1}{m} - \frac{1}{M} \right) S_{ec}^2$$

```
## (b) distribuicao amostral do total e media HT
```

```
Vart_HT <- M^2 * (1/m - 1/M) * S2ec # total
```

```
Vart_HT
```

```
## [1] 4687.5
```

```
Varybarra_HT <- Vart_HT / N^2 # media
```

```
Varybarra_HT
```

```
## [1] 0.2773669
```

Amostragem por Conglomerados em 1 Estágio Simples

Exercício

c. As variâncias para o estimador de razão são

$$Var_{AC1S}(\hat{T}_{AC1S}^R) = M^2 \left(\frac{1}{m} - \frac{1}{M} \right) \frac{1}{M-1} \sum_{i \in C} N_i^2 (\bar{Y}_i - \bar{Y})^2$$

e

$$Var_{AC1S}(\bar{y}_{AC1S}^R) = \frac{Var_{AC1S}(\hat{T}_{AC1S}^R)}{N^2}$$

```
## (c) distribuicao amostral do total e media RAZAO
Vart_R <- M^2 * ( 1/m - 1/M ) * 1/(M-1) * sum( Ni^2 * ( Ti / Ni - Yba
Vart_R
```

```
## [1] 2191.198
```

```
Varybarra_R <- Vart_R / N^2 # media
Varybarra_R
```

Amostragem por Conglomerados em 1 Estágio Simples

Exercício

d.

```
## (d) selecionar amostra de conglomerados e estimacao
amostra <- sample(1:M, m)
Niamostra <- Ni[amostra]
Tiamostra <- Ti[amostra]
ty <- (sum(Ni) / sum(Niamostra)) * sum(Tiamostra)      # estimador d
ty
```

```
## [1] 689
```

Estimamos que a distância total percorrida por todos os aviões do país pesquisado foi de $\hat{T} = \sum_{i \in a} T_i = 689$ milhares de milhas.

Para casa

- Continuar o Exercício.
- Ler o capítulo 12 do livro 'Amostragem: Teoria e Prática Usando R'.
- Rever os slides.

Próxima aula

- Acompanhar o material no moodle.

Amostragem por Conglomerados

- Coeficiente de correlação intraclass
- Tamanho de amostra e Intervalos de confiança

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