

MAT02036 - Amostragem 2

Aula 15 - Amostragem por Conglomerados - IC e tamanho de amostra

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

Coeficiente de Correlação Intraclasses

Se os **tamanhos** dos conglomerados forem todos *iguais*,

$N_i = \bar{N}$, $\forall i = 1, \dots, M$, então, de acordo com @Cochran1977, página 242, tem-se:

$$EPA(AC1S^R; AAS) \doteq 1 + (\bar{N} - 1)\rho$$

onde:

$$\rho = \frac{\sum_{i \in C} \sum_{j \in C_i} \sum_{k \neq j \in C_i} (y_{ij} - \bar{Y}) (y_{ik} - \bar{Y})}{(\bar{N} - 1) (M\bar{N} - 1) S_y^2} \doteq 1 - \frac{S_d^2}{S_y^2}$$

é o **coeficiente de correlação intraconglomerado** ou **intraclasses**; e S_d^2 é a medida da variância *dentro* dos conglomerados, dada por:

Aula passada

Coeficiente de Correlação Intraclass

Exemplo: (cont. Exemplo slides Aula 12)

Considere a população de tamanho $N = 6$ agrupada em $M = 3$ conglomerados, de três maneiras diferentes:

$$YA = ((7, 8); (9, 10); (12, 14))$$

$$YB = ((7, 10); (12, 8); (9, 14))$$

$$YC = ((7, 14); (12, 8); (9, 10))$$

Aula passada

Coeficiente de Correlação Intraclass

A expressão para o $EPA(AC1S^R; AAS)$ resulta do uso das expressões de acordo com @Cochran1977, página 241:

$$Var_{AC1S}(\bar{y}_{AC1S}^R) \doteq \left(\frac{1}{m\bar{N}} - \frac{1}{M\bar{N}} \right) S_y^2 [1 + (\bar{N} - 1)\rho]$$

$$Var_{AAS}(\bar{y}) = \left(\frac{1}{m\bar{N}} - \frac{1}{M\bar{N}} \right) S_y^2$$

- Numa **amostra** retirada **com reposição**, o **coeficiente de correlação intraclass** pode ser **estimado** por:

$$r = \frac{s_{ec}^2 \frac{s_{dc}^2}{\bar{N}}}{s_{ec}^2 + s_{dc}^2}$$

Intervalos de Confiança na AC1S

Intervalos de Confiança na AC1S

- Para a **média**, se os conglomerados são de tamanhos iguais, a normalidade de \bar{y}_{AC1S} segue da média amostral na **AAS**.

Para M e M suficientemente grandes (?), então pelo TCL

$$\frac{\bar{y}_{AC1S} - \bar{Y}}{\sqrt{Var_{AC1S}(\bar{y}_{AC1S})}} \approx Normal(0; 1)$$

em que $Var_{AC1S}(\bar{y}_{AC1S}) = \frac{Var_{ec}}{m}$.

Logo, um intervalo de confiança de nível $1 - \alpha$ para \bar{Y} é dado por:

$$IC_{AC1S}(\bar{Y}; 1 - \alpha) = \left[\bar{y}_{AC1S} \mp z_{\alpha/2} \sqrt{\widehat{Var}_{AC1S}(\bar{y}_{AC1S})} \right]$$

em que $\widehat{Var}_{AC1S}(\bar{y}_{AC1S}) = \frac{s_{ec}^2}{m}$.

Intervalos de Confiança na AC1S

- Para o total:

$$IC_{AC1S}(T; 1 - \alpha) = \left[\hat{T}_{AC1S} \pm z_{\alpha/2} \sqrt{\widehat{Var}_{AC1S}(\hat{T}_{AC1S})} \right]$$

- Para a proporção:

$$IC_{AC1S}(P; 1 - \alpha) = \left[\hat{P}_{AC1S} \pm z_{\alpha/2} \sqrt{\widehat{Var}_{AC1S}(\hat{P}_{AC1S})} \right]$$

Exemplo - Dados de companhias aéreas - aula passada

Considere os dados das companhias aéreas, construa o IC 95% para a média e para o total.

Tamanho de amostra na AC1S

Tamanho de amostra na AC1S

- Para planejar uma **AC1S**, é necessário **determinar o número de conglomerados m** que serão sorteados para fazer parte da amostra.
 - Vimos que pode-se utilizar fórmulas que dependem do **erro relativo e_r** ou do **erro absoluto e** .
- Erro relativo

Seja $CV = \frac{Var_{ecT}}{\bar{Y}_c}$ então na **AASc** de conglomerados

$$m = \frac{z_{\alpha/2} CV^2}{e_r^2},$$

e na **AASs** de conglomerados

$$m = \frac{M z_{\alpha/2} CV^2}{z_{\alpha/2} CV^2 + (M - 1) e_r^2}.$$

Tamanho de amostra na AC1S

- Erro absoluto


Para casa

- Fazer a lista 2 de exercícios.
- Ler o capítulo 2 da apostila da Profa. Vanessa.
- Rever os slides.

Próxima aula

- Acompanhar o material no moodle.

Amostragem por Conglomerados

- Estimação de proporção na **AC1S**.
- Laboratório de 

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