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 Intraclasse

Se os **tamanhos** dos conglomerados forem todos *iguais*, $N_i = \overline{N},\ \forall\, i = 1, \ldots, M$, então, de acordo com @Cochran1977, página 242, tem-se:

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

onde:

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

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

$$S_{dc}^2 = \frac{1}{M} \sum_{i \in C} \frac{1}{\overline{N} - 1} \sum_{j \in C_i} \left(y_{ij} - \overline{Y_i}\right)^2 = \frac{1}{M} \sum_{i \in C} S_i^2$$

com $S_i^2 = \frac{1}{\overline{N}-1} \sum_{j \in C_i} \left(y_{ij} - \overline{Y_i}\right)^2$.

Aula passada

Coeficiente de Correlação Intraclasse

Exemplo: (cont. Exemplo slides `Aula 12`)

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

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

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

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

Aula passada

Coeficiente de Correlação Intraclasse

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

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

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

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

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

Intervalos de Confiança na AC1S

Intervalos de Confiança na AC1S

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

Para M e m "suficientemente grandes" (❓), então pelo TCL

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

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

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

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

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

Intervalos de Confiança na AC1S

- Para o total:

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

- Para a proporção:

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

Exemplo - Dados de companhias aéreas - aula passada

Considere os dados das companias 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** e_r

Ao utilizar o **erro relativo**, deseja-se que a estimativa do parâmetro não difira mais do que $100 \times e_r$% do seu verdadeiro valor, com $100 \times (1-\alpha)$% de confiança. Isto é:

$$P\left(\left|\frac{\overline{y}_{AES} - \overline{\overline{Y}}}{\overline{\overline{Y}}}\right| \leq e_r\right) = P\left(\frac{\left|\overline{y}_{AES} - \overline{\overline{Y}}\right|}{\sqrt{Var(\overline{y}_{AES})}} \leq \frac{e_r\left|\overline{\overline{Y}}\right|}{\sqrt{Var(\overline{y}_{AES})}}\right) = 1 - \alpha.$$

AASc de conglomerados $CV = \frac{Var_{ec_T}}{\overline{\overline{Y}}_c}$,

$$m = \frac{z^2_{\alpha/2}\, CV^2}{e_r^2}.$$

AASs de conglomerados $CV = \frac{S^2_{ec}}{\overline{\overline{Y}}_c}$,

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

Tamanho de amostra na AC1S

- **Erro absoluto** $e = e_r \left|\overline{Y}\right|$

Ao utilizar o erro absoluto, deseja-se que a estimativa do parâmetro não difira mais do que e unidades do seu verdadeiro valor, com $100 \times (1-\alpha)$% de confiança. Isto é:

$$P\left(\left|\overline{y}-\overline{Y}\right| \leq e\right) = P\left(\frac{\left|\overline{y}-\overline{Y}\right|}{\sqrt{Var(\overline{y})}} \leq \frac{e}{\sqrt{Var(\overline{y})}}\right) = 1-\alpha.$$

- Para média e proporção:

Na **AASc**, $m = \frac{z^2_{\alpha/2}\ Var_{ec_T}}{\overline{N} e^2}$;

Na **AASs**, $m = \frac{M\ z^2_{\alpha/2}\ S^2_{ec}}{z^2_{\alpha/2}\ S^2_{ec} + \overline{N}(M-1)\ e^2}$.

- Para o total:

Na **AASc**, $m = \frac{M^2 z^2_{\alpha/2}\ Var_{ec_T}}{e^2}$;

Na **AASs**, $m = \frac{M^3\ z^2_{\alpha/2}\ S^2_{ec}}{M^2 z^2_{\alpha/2}\ S^2_{ec} + (M-1)\ e^2}$.

Tamanho de amostra na AC1S

- O coeficiente de variação CV, ou variâncias entre totais de conglomerados, Var_{ec_T} ou S^2_{ec}, são necessários para os cálculos de m;
    - geralmente devemos **estimar**, por estudos prévios, similares, ou piloto.

Exemplo: pg. 30 da Apostila da Profa. Vanessa

Suponha que se deseja estimar a média de renda familiar em certa cidade com um erro não maior que 10% e 99,57% de confiança $(z = 3)$. Para tanto, vai se fazer **AC1S** nos bairros da cidade $(M = 5)$. Num estudo anterior, obteve-se $CV = 0,077$.

Tamanho de amostra na AC1S

Uso de $EPA\ (deff)$ para cálculo de tamanho de amostra

Um método aproximado para obter-se o tamanho de amostra necessário na **AC1S** é multiplicar o tamanho de amostra necessário para uma **AASc**, n_{AASc}, pelo EPA_{AC1S} $(deff_{AC1S})$ e então dividir pelo tamanho médio dos conglomerados $\overline{N}$.

$$m \approx \frac{n_{AASc} \times EPA_{AC1S}}{\overline{N}} (\text{❓}).$$

Exemplo: pg. 36 da Apostila da Profa. Vanessa

Em uma certa região, deseja-se fazer uma **AC1S** de fazendas criadores de gado. Em média, as fazendas têm 50 animais. O interesse é estimar a prevalência de uma doença, isto é, a proporção de animais doentes. Numa região vizinha, um estudo mostrou que 10% dos animais estavam doentes e $r_{int} = 0,1225$. Quantas fazendas devem pertencer à amostra, considerando que se deseja uma margem de erro de 1% para mais ou para menos e 95% de confiança?

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 R

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