

MAT02036 - Amostragem 2

Aula 16 - Amostragem por Conglomerados - Estimação de Proporções

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

Intervalos de Confiança na AC1S

- Para a **média** (para M e m suficientemente grandes (?), então segue pelo TCL $IC_{AC1S}(\bar{Y}; 1 - \alpha) = \left[\bar{y}_{AC1S} \mp z_{\alpha/2} \sqrt{\widehat{Var}_{AC1S}(\bar{y}_{AC1S})} \right]$, em que $\widehat{Var}_{AC1S}(\bar{y}_{AC1S}) = \frac{s_{ec}^2}{m}$.
- 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: $IC_{AC1S}(P; 1 - \alpha) = \left[\hat{P}_{AC1S} \pm z_{\alpha/2} \sqrt{\widehat{Var}_{AC1S}(\hat{P}_{AC1S})} \right]$

Aula passada

Tamanho de amostra na AC1S

- **Determinar o número de conglomerados** m que serão sorteados para fazer parte da amostra.
- **Erro relativo** e_r

AASc de conglomerados $CV = \frac{Var_{ecT}}{\bar{Y}_c}$,

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

AASs de conglomerados $CV = \frac{S_{ec}^2}{\bar{Y}_c}$,

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

Aula passada

Tamanho de amostra na AC1S

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

Para média e proporção:

$$\text{Na AASc, } m = \frac{z_{\alpha/2}^2 \text{Var}_{ecT}}{\overline{N} e^2};$$

$$\text{Na AASs, } m = \frac{M z_{\alpha/2}^2 S_{ec}^2}{z_{\alpha/2}^2 S_{ec}^2 + \overline{N} (M-1) e^2}.$$

Para o total:

$$\text{Na AASc, } m = \frac{M^2 z_{\alpha/2}^2 \text{Var}_{ecT}}{e^2};$$

$$\text{Na AASs, } m = \frac{M^3 z_{\alpha/2}^2 S_{ec}^2}{M^2 z_{\alpha/2}^2 S_{ec}^2 + (M-1) e^2}.$$

Uso de EPA (*def*) para cálculo de tamanho de amostra

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

Estimação de Proporções

Estimação de Proporções

Parâmetro

- Assumimos a variável indicadora

$$y_{ij} = I[(i, j) \in A] = \begin{cases} 1, & \text{se a unidade } j \text{ do conglomerado } i \text{ possui o atributo, } A \subset U; \\ 0, & \text{caso contrário.} \end{cases}$$

- O total populacional, representa o **número de unidades populacionais** com o **atributo** de interesse A ,

$$T = \sum_{i \in C} T_i = N_A.$$

- A média populacional, representa a **proporção de unidades elementares** com o atributo A ,

$$\bar{Y} = \frac{1}{N} \sum_{i \in C} T_i = \frac{T}{MN} = \frac{\bar{Y}_C}{N} = \frac{N_A}{N} = P$$

- A variância populacional,...

Estimação de Proporções

Estimador natural

- O **estimador natural** HT é dado por

$$\hat{P}^{HT} = \frac{\bar{y}_C}{\bar{N}},$$

em que $\bar{y}_C = \frac{\sum_{i \in a} T_i}{m}$.

- \hat{P}^{HT} é **não viciado** para P ???
- Esse estimador pode resultar em uma **proporção estimada** $\hat{P}^{HT} > 1$ (?).
- Veremos adiante que $\bar{N} = \frac{\sum_{i \in CN_i}}{M}$ pode ser estimado por $\bar{n} = \frac{\sum_{i \in a} N_i}{m}$.

Estimação de Proporções

Variância do estimador natural

A variância de \hat{P}^{HT} na **AC1S** é dada por:

- **COM** reposição,

$$Var_{AC1S_c}(\hat{P}^{HT}) = \frac{1}{\overline{N}^2} \frac{Var_{ec_T}}{m} = \frac{1}{\overline{N}^2} \left(1 - \frac{1}{M}\right) \frac{S_{ec}^2}{m};$$

- **SEM** reposição,

$$Var_{AC1S_s}(\hat{P}^{HT}) = \frac{1}{\overline{N}^2} \left(\frac{M-m}{M-1}\right) \frac{Var_{ec_T}^2}{m} = \frac{1}{\overline{N}^2} \left(1 - \frac{m}{M}\right) \frac{S_{ec}^2}{m}.$$

- Lembrando, que $Var_{ec_T} = \frac{\sum_{i \in C} (T_i - \overline{Y}_C)^2}{M} = \frac{M-1}{M} S_{ec}^2.$

Estimação de Proporções

Estimador da variância do estimador natural

- O estimador não viciado da variância de \hat{P}^{HT} na **AC1S** é dada por:
 - **COM** reposição, $\hat{Var}_{AC1S_c}(\hat{P}^{HT}) = \frac{1}{N^2} \left(1 - \frac{1}{M}\right) \frac{s_{ec}^2}{m} \approx \frac{1}{N^2} \frac{s_{ec}^2}{m}$ (?);
 - **SEM** reposição, $\hat{Var}_{AC1S_s}(\hat{P}^{HT}) = \frac{1}{N^2} \left(1 - \frac{m}{M}\right) \frac{s_{ec}^2}{m}$,

em que $s_{ec}^2 = \frac{\sum_{i \in a} (T_i - \bar{y}_C)^2}{m-1}$.

- Lembrando que o estimador $s_{ec}^2 = \hat{S}_{ec}^2 = \hat{Var}_{ecT}$ é não-viciado para:
 - Var_{ecT} se a seleção dos conglomerados for COM reposição;
 - e de S_{ec}^2 se a seleção for SEM reposição.

Estimação de Proporções

Estimador de Razão

- Se estimamos \bar{N} por \bar{n} então o estimador do tipo razão é dado por

$$\hat{P}^R = \frac{\sum_{i \in a} T_i}{\sum_{i \in a} N_i} = \frac{\bar{y}_C}{\bar{n}},$$

- É fácil ver que este estimador sempre produz estimativas menores ou iguais a 1. (?)

Variância do estimador de razão

A variância de \hat{P}^R na **AC1S** é dada por:

- COM** reposição, $Var_{AC1S_c}(\hat{P}^{HT}) = \frac{1}{\bar{N}^2} \frac{Var_{ecT}}{m} = \frac{1}{\bar{N}^2} \left(1 - \frac{1}{M}\right) \frac{S_{ec}^2}{m};$
- SEM** reposição, $Var_{AC1S_s}(\hat{P}^{HT}) = \frac{1}{\bar{N}^2} \left(\frac{M-m}{M-1}\right) \frac{Var_{ecT}^2}{m} = \frac{1}{\bar{N}^2} \left(1 - \frac{m}{M}\right) \frac{S_{ec}^2}{m}.$

Estimação de Proporções

Estimador de Razão

Estimador da variância do estimador de razão

- O estimador não viciado da variância de \hat{P}^R na **AC1S** é dada por:
 - **COM** reposição, $\hat{Var}_{AC1S_c}(\hat{P}^{HT}) = \frac{1}{N^2} \left(1 - \frac{1}{M}\right) \frac{s_{ec}^2}{m} \approx \frac{1}{N^2} \frac{s_{ec}^2}{m}$ (?);
 - **SEM** reposição, $\hat{Var}_{AC1S_s}(\hat{P}^{HT}) = \frac{1}{N^2} \left(1 - \frac{m}{M}\right) \frac{s_{ec}^2}{m}$,
- Rao (2000) menciona que, se os T_i 's forem altamente correlacionados com os N_i 's, então \hat{P}^R terá menor variância que \hat{P}^{HT} .

Estimação de Proporções

Exemplo (Apostila pg. 27)

Uma população universitária foi avaliada quanto à posse de bicicleta. Os conglomerados foram os campi da universidade. Os dados da população estão abaixo:

Campus (i)	No. pessoas com bicicleta (T_i)	Número total de pessoas (N_i)
1	2226	2950
2	1512	1726
3	315	948
Total	4053	5624

Estimação de Proporções

Exemplo (Apostila pg. 28)

Considere os dados da população universitária, construa o IC 95% para a proporção.

Estimação de Proporções

Exemplo (Apostila pg. 36)

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.
- Continuar exercícios.
- Rever os slides.

Próxima aula

- Acompanhar o material no moodle.

Amostragem por Conglomerados

- Exercícios.
- 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