MAT02036 - Amostragem 2

Aula 07 - Amostragem Estratificada - Tamanho Amostral

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

Comparação de alternativas de alocação da amostra

- Sob *alocação de Neyman*,

$$V_{AESne}\left(\overline{y}_{AES}\right) \leq V_{AESpr}\left(\overline{y}_{AES}\right) \leq V_{AAS}\left(\overline{y}\right)$$

- **AES** com alocação de **Neyman** é **mais eficiente** que **AES** com alocação **proporcional**.

- Ambas superam **AAS** como plano amostral para um mesmo tamanho especificado de amostra.

Efeito do Plano Amostral/Delineamento (*Design Effect*) de um *plano* amostral

$$EPA_{plano} = deff_{plano} = \frac{Var(\overline{y}_{plano})}{Var(\overline{y}_{AAS_c})}.$$

- se $deff_{plano} < 1$ então o **plano** é mais eficiente que a **AASc**.

Aula passada

Comparação de alternativas de alocação da amostra

Outro exemplo de hoje:

Sabemos mostrar $EPA_{AES_{pr}} = deff_{AES_{pr}}$ e $EPA_{AES_{pr}} = deff_{AES_{pr}}$ assumindo **AASc** dentro dos estratos

Tamanho da amostra na **AES**

Tamanho da amostra na AES

- Podemos determinar o **tamanho total da amostra** estratificada, a partir das expressões das variâncias dos estimadores, dado:
  - o tipo de **alocação**, $w_h = n_h/n$;
  - a **variância máxima**, V, desejada para a estimativa do parâmetro.
- Ex.: Estimação da **média populacional** sob **AASs** dentro dos estratos.
  - Substituindo $n_h = n \times w_h$ na expressão da variâcia do estimador, e fixando V o valor máximo para sua estimativa,

$$V \geq V_{AES}\left(\overline{y}_{AES}\right) = \sum_{h=1}^{H} W_h^2 S_{h,y}^2 \left(\frac{1}{n_h} - \frac{1}{N_h}\right) = \frac{1}{n}\sum_{h=1}^{H} \frac{W_h^2 S_{h,y}^2}{w_h} - \frac{1}{N}\sum_{h=1}^{H} W_h S_{h,y}^2$$

Logo (❓)

$$n \geq \frac{\sum_{h=1}^{H} \frac{W_h^2 S_{h,y}^2}{w_h}}{V + \frac{1}{N}\sum_{h=1}^{H} W_h S_{h,y}^2}$$

- E sob **AASc** dentro dos estratos❓

Tamanho da amostra na **AES** - média

Estimação da **média populacional** sob **AASc**

Supondo uma alocação $w_h = n_h/n$ e **AASc** dentro dos estratos,

1. A **variância do estimador** é dada por

$$V_{AES}\left(\bar{y}_{AES}\right) = \sum_{h=1}^{H} W_h^2 \frac{Var_{h,y}}{n_h} = \sum_{h=1}^{H} W_h^2 \frac{Var_{h,y}}{n w_h} = \frac{1}{n} \sum_{h=1}^{H} W_h^2 \frac{Var_{h,y}}{w_h}.$$

2. Então definindo uma **variância máxima** para a estimativa do parâmetro,

$$V \geq V_{AES}\left(\bar{y}_{AES}\right) \Leftrightarrow V \geq \frac{1}{n} \sum_{h=1}^{H} W_h^2 \frac{Var_{h,y}}{w_h} \ ,$$

ou seja,

$$n \geq \frac{\sum_{h=1}^{H} W_h^2 \frac{Var_{h,y}}{w_h}}{V}.$$

Tamanho da amostra na **AES** - total

Estimação do **total populacional**

- Sob **AASs** dentro dos estratos sabemos que

$$Var_{AES}\left(\widehat{T}_{AES}\right) = \sum_{h=1}^{H} N_h^2 S_{h,y}^2 \left(\frac{1}{n_h} - \frac{1}{N_h}\right) = \ldots = \frac{1}{N^2} V_{AES}\left(\overline{y}_{AES}\right)$$

- Fazendo $V_T \geq V_{AES}\left(\widehat{T}_{AES}\right) = \sum_{h=1}^{H} N_h^2 S_{h,y}^2 \left(\frac{1}{n_h} - \frac{1}{N_h}\right)$

$$n \geq \frac{\sum_{h=1}^{H} \frac{N_h^2 S_{h,y}^2}{w_h}}{V_T + \sum_{h=1}^{H} N_h S_{h,y}^2}$$

- E na estimação do **total populacional** sob **AASc** dentro dos estratos❓❓❓

$$n \geq \frac{\sum_{h=1}^{H} \frac{N_h^2 Var_{h,y}}{w_h}}{V_T}$$

E a margem de erro?

Tamanho da amostra na **AES** - e a margem de erro?

Margem de erro absoluta - estimação da média

- Podemos pensar na **variância máxima** V como uma **função** da **margem de erro** e, tal que

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

sendo α uma 'probabilidade pequena', ou valor de significância.

- Usando a aproximação da distribuição de $\overline{y}$ para a distribuição **normal**,

$$e = z_{\frac{\alpha}{2}}\sqrt{Var(\overline{y})} \Leftrightarrow V = \frac{e^2}{z^2_{\frac{\alpha}{2}}},$$

onde $z_{\frac{\alpha}{2}}$ é o quantil da normal padrão que deixa área $\alpha/2$ à sua direita.

- Quais suposições são necessárias nesse caso❓

- A distância máxima assumida e é absoluta portando denominada **margem de erro absoluta**.

Tamanho da amostra na **AES** - e a margem de erro?

Margem de erro relativa - estimação da média

- Muitas vezes queremos controlar o **erro relativo** r na estimativa, tal que

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

- Novamente pela aproximação de $\overline{y}$ para a distribuição **normal**,

$$r\overline{Y}=z_{\frac{\alpha}{2}}\sqrt{Var(\overline{y}_{AES})}\Leftrightarrow V_T=\frac{r^2\overline{Y}^2}{z^2_{\frac{\alpha}{2}}}.$$

- Expressões de $\sqrt{Var(\overline{y})}$ na **AASc** ou **AASs**...(❓)
  - Pode ser útil utilizar a relação entre variâncias que já conhecemos

$$Var(\overline{y})=\frac{N-n}{n}\frac{S^2}{n}\Leftrightarrow\sqrt{Var(\overline{y})}=\sqrt{\frac{N-n}{n}}\frac{S}{\sqrt{n}}$$

Tamanho da amostra na **AES** - e a margem de erro?

Margem de erro relativa - estimação do total

- No caso da estimação do total o **erro relativo** r se refere ao total, tal que

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

- Novamente pela aproximação de $\overline{y}$ para a distribuição **normal**,

$$rN\overline{Y} = z_{\frac{\alpha}{2}}\sqrt{Var(\overline{T}_{AES})} \Leftrightarrow V = \frac{r^2N^2\overline{Y}^2}{z^2_{\frac{\alpha}{2}}}.$$

- O **tamanho da amostra** n é o **mesmo** na estimação da **média** e **total**,
  - se usar erro relativo $r\overline{Y}$ para definir V e usar $V \geq Var_{AES}(\overline{y})$;
  - ou com erro relativo $rN\overline{Y}$ para definir V_T e usar $V_T \geq Var_{AES}(\widehat{T})$.

- Expressões para **AASs** e **AASc** dentro dos estratos...(❓)

Tamanho da amostra e Alocações

- Para obter a expressão para uma alocação particular basta substituir w_h de forma adequada.
    - Ex. no caso da **AESpr** basta substituir w_h por W_h e obter a expressão para o cálculo do tamanho da amostra.
    - Mostrar(❓)
- Sabemos mostrar os resultados para diferentes w_h❓

Alocação	Uniforme	Proporcional	de Neyman (AASs)
w_h	$\frac{1}{H}$	$\frac{N_h}{N} = W_h$	$\frac{S_h N_h}{\sum_{h=1}^{H} S_h N_h}$

- E no caso de **AASc** dentro dos estratos?

Exemplo 🏋️

Exemplo 7 da Apostila da Profa Vanessa:

Suponha que os restaurantes em uma cidade foram divididos em 3 estratos, de acordo com a zona de localização: A $(N_1 = 600)$, B $(N_2 = 300)$ e C $(N_3 = 100)$. Queremos estimar o número médio de clientes por dia. Os valores do desvio padrão dos estratos são: 20, 30 e 50 clientes, respectivamente. Determinar o tamanho de amostra pra estimar a média de clientes por dia com um erro máximo absoluto de 3 clientes e com 99,73% de confiança (isto é, z = 3). Considere que será feita uma **AASs** em cada estrato.

Exemplo 🏃‍♀️

Ainda não falamos sobre estimação de proporções, mas será nosso próximo tópico

Exemplo 8 da Apostila da Profa Vanessa::

Seja uma população de tamanho $N = 1000$, estratificada em $H = 2$ estratos de tamanhos iguais. Uma amostra piloto de tamanho $n_1 = 40$ foi retirada do primeiro estrato, e outra de $n_2 = 60$ do segundo, com **AASs** em ambos. Na primeira amostra, houve 20 sucessos, e na segunda, 40. Calcular o tamanho da amostra necessário para estimar a proporção com 99% de confiança e erro máximo absoluto de 10%.

Para casa 🏠

- Continuar os Exemplos.
- Mostrar tamanho de amostra n para **AASc** dentro dos estratos nas alocações **AESun**, **AESpr**, **AESne**.
- Fazer exercícios 11.7 e 11.10 do livro 'Amostragem: Teoria e Prática Usando R' https://amostragemcomr.github.io/livro/estrat.html#exerc11
- Fazer exercício 5 da lista 1.
- Rever os slides.

Próxima aula 📊

- Amostragem Estratificada
  - Estimação de proporções
  - Exercícios e Intervalos de confiança

Muito obrigado!

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

Resumo da notação

- **Tamanho da amostra** na **AES** dado um tipo de **alocação**, w_h, e **fixando a variância** máxima que se deseja para a estimativa do parâmetro, V

Parâmetro	Sob AASc dentro dos estratos	Sob AASs dentro dos estratos
Média	$n \geq \frac{\sum_{h=1}^{H} W_h^2 \frac{Var_{h,y}}{w_h}}{V}$	$n \geq \frac{\sum_{h=1}^{H} \frac{W_h^2 S_{h,y}^2}{w_h}}{V+\frac{1}{N}\sum_{h=1}^{H} W_h S_{h,y}^2}$
Total	$n \geq \frac{\sum_{h=1}^{H} N_h^2 \frac{Var_{h,y}}{w_h}}{V}$	$n \geq \frac{\sum_{h=1}^{H} \frac{N_h^2 S_{h,y}^2}{w_h}}{V+\sum_{h=1}^{H} N_h S_{h,y}^2}$

- **Margem de erro** para o estimador $\hat{\theta}$ (approx. normal para dist. de $\hat{\theta}$)
  - **Absoluta**: $e = z_{\frac{\alpha}{2}} \sqrt{Var(\hat{\theta})} \Leftrightarrow V = \frac{e^2}{z_{\frac{\alpha}{2}}^2}$
  - **Relativa**: $r\overline{Y} = z_{\frac{\alpha}{2}} \sqrt{Var(\hat{\theta})} \Leftrightarrow V = \frac{r^2 \overline{Y}^2}{z_{\frac{\alpha}{2}}^2}$
- Pode ser útil utilizar a relação entre variâncias que já conhecemos $Var(\overline{y}) = \frac{N-n}{n} \frac{S^2}{n} \Leftrightarrow \sqrt{Var(\overline{y})} = \sqrt{\frac{N-n}{n}} \frac{S}{\sqrt{n}}$

Resumo da notação

- Tamanho mínimo de amostra para **estimação da média** populacional

Alocação	AASc dentro dos estratos	AASs dentro dos estratos
AES_{un}	$n \geq \frac{H\sum_{h=1}^{H} W_h^2 Var_{h,y}}{V}$	$n \geq \frac{H\sum_{h=1}^{H} W_h^2 S_h^2}{V+\frac{1}{N}\sum_{h=1}^{H} W_h S_h^2}$
AES_{pr}	$n \geq \frac{\sum_{h=1}^{H} W_h Var_{h,y}}{V}$	$n \geq \frac{\sum_{h=1}^{H} W_h S_h^2}{V+\frac{1}{N}\sum_{h=1}^{H} W_h S_h^2}$
AES_{ne}	$n \geq \frac{\left(\sum_{h=1}^{H} W_h DP_{h,y}\right)^2}{V}$	$n \geq \frac{\left(\sum_{h=1}^{H} W_h S_{h,y}\right)^2}{V+\frac{1}{N}\sum_{h=1}^{H} W_h S_h^2}$
AES_{ot}	$n \geq \frac{\left(\sum_{h=1}^{H} W_h DP_{h,y}\sqrt{C_h}\right)\left(\sum_{h=1}^{H} W_h DP_{h,y}/\sqrt{C_h}\right)}{V}$	$n \geq \frac{\left(\sum_{h=1}^{H} W_h S_{h,y}\sqrt{C_h}\right)\left(\sum_{h=1}^{H} W_h S_{h,y}/\sqrt{C_h}\right)}{V+\frac{1}{N}\sum_{h=1}^{H} W_h S_h^2}$

Referências

Slides baseados no Capítulo 11 do livro

- Amostragem: Teoria e Prática Usando o R

Citações do Capítulo

- Neyman(1934)
- Cochran(1977)