

Amostragem por conglomerados em um único estágio (AC): parte 2

Prof. Caio Azevedo

Normalidade assintótica

- Temos, para a e A suficientemente grandes, $i = 1, 2, 3$, $j = 1, 2$, que

$$\frac{\hat{\mu}_{C_i} - \mu}{\widehat{EP}_{A_j}(\hat{\mu}_{C_i})} \xrightarrow[A \rightarrow \infty]{D} N(0, 1)$$

$$\frac{\hat{\mu}_{C_i} - \mu}{\widehat{EP}_{A_j}(\hat{\mu}_{C_i})} \xrightarrow[A \rightarrow \infty]{D} N(0, 1)$$

Intervalos de confiança

- Conglomerados de tamanhos desiguais

$$\left[\hat{\mu}_{C_i} - z_\gamma \sqrt{\hat{v}_{A_j}(\hat{\mu}_{C_i})}; \hat{\mu}_{C_i} + z_\gamma \sqrt{\hat{v}_{A_j}(\hat{\mu}_{C_i})} \right]$$

- Erro da estimativa: $z_\gamma \sqrt{\hat{v}_{A_j}(\hat{\mu}_{C_i})}$.

Intervalos de confiança

- Conglomerados de tamanhos iguais

$$\left[\hat{\mu}_C - z_\gamma \sqrt{\frac{1}{a(a-1)} \sum_{\alpha=1}^a (\hat{\mu}_\alpha - \hat{\mu}_C)^2}; \hat{\mu}_C + z_\gamma \sqrt{\frac{1}{a(a-1)} \sum_{\alpha=1}^a (\hat{\mu}_\alpha - \hat{\mu}_C)^2} \right]$$

em que $\hat{\mu}_C = \frac{1}{a} \sum_{\alpha=1}^a \hat{\mu}_\alpha$

- Erro da estimativa: $z_\gamma \sqrt{\frac{1}{a(a-1)} \sum_{\alpha=1}^a (\hat{\mu}_\alpha - \hat{\mu}_C)^2}$.

Cont.

- Os testes de hipóteses seguem, essencialmente, os mesmo procedimentos descritos nos slides 10 e 11 de http://www.ime.unicamp.br/~cnaber/aula_AE%20P2%20Amost%20S%202018.pdf, em que a estatística do teste passa a ser

$$Z_t = \frac{\hat{\mu}_{C_i} - \mu_0}{\widehat{EP}_{A_j}(\hat{\mu}_{C_i})}.$$

Determinação do tamanho da amostra

- O “tamanho da amostra” a ser determinado, nesse caso, é o número “a” de conglomerados a ser sorteado.
- Se os conglomerados tiverem tamanhos iguais o tamanho da amostra real ($n = \sum_{\alpha=1}^a n_{\alpha}$) será constante, dado a . Caso contrário, será uma variável aleatória.

Determinação do tamanho da amostra

- Supondo conglomerados de mesmo tamanho, e fixando:

$$P(|\hat{\mu}_C - \mu| \leq \delta) \geq \gamma \rightarrow P\left(|Z| \leq \frac{\delta}{\sqrt{\sigma_{ec}^2/a}}\right) = \gamma \rightarrow$$
$$z_\gamma = \frac{\delta}{\sqrt{\sigma_{ec}^2/a}} \rightarrow a = \frac{\sigma_{ec}^2 z_\gamma^2}{\delta^2}$$

em que $\sigma_{ec}^2 = \frac{1}{A} \sum_{\alpha=1}^A (\mu_\alpha - \mu)^2$

- Se se fixar o erro da estimativa, obter-se-á a mesma fórmula para o tamanho da amostra.
- Exercício: obter o tamanho da amostra considerando conglomerados de tamanhos diferentes.

Determinação do tamanho da amostra

- Exercício: obter todas as expressões para as esperanças, variâncias, EQM, testes de hipótese, IC e tamanhos de amostra considerando que os conglomerados foram selecionados segundo um plano AAS_s .
- Exercício: repetir os procedimentos para a estimação do total populacional (τ).

Estimação de proporções

- Todas as quantidades são como definidas em [http:](http://www.ime.unicamp.br/~cnaber/aula_AC_Amost_2S_2018.pdf)

[//www.ime.unicamp.br/~cnaber/aula_AC_Amost_2S_2018.pdf](http://www.ime.unicamp.br/~cnaber/aula_AC_Amost_2S_2018.pdf)

- Parâmetro de interesse: $p = \frac{\sum_{\alpha=1}^A \tau_{\alpha}}{\sum_{\alpha=1}^A B_{\alpha}}$.

- Como a proporção é uma média de variáveis binárias, os desenvolvimentos seguem àqueles apresentados para a estimação da média sob AC e para a proporção sob A_1 (veja http://www.ime.unicamp.br/~cnaber/aula_AC_Amost_2S_2018.pdf) e sob A_2 (exercício).

Estimação de proporções

- Estimador (razão): $\hat{p}_{C_2} = \frac{\sum_{\alpha=1}^a \hat{\tau}_\alpha}{\sum_{\alpha=1}^a b_\alpha} = \frac{\hat{\tau}}{\hat{B}}$.
- Variância: $\mathcal{V}(\hat{p}_{C_2}) \approx \frac{1}{aAB^2} \sum_{\alpha=1}^A (\tau_\alpha - pB_\alpha)^2$
- Estimador para a variância: $\hat{\mathcal{V}}(\hat{p}_{C_2}) \approx \frac{1}{a(a-1)b^2} \sum_{\alpha=1}^a (\hat{\tau}_\alpha - \hat{p}_{C_2} b_\alpha)^2$
- Os dois outros estimadores são:

$$\hat{p}_{C_1} = \frac{\hat{\tau}}{B}$$

$$\hat{p}_{C_3} = \frac{1}{a} \sum_{\alpha=1}^a \hat{p}_\alpha$$

Cont.

- Suas variâncias são dadas, respectivamente, por:

$$\mathcal{V}(\hat{p}_{C_1}) = \frac{1}{aA\bar{B}^2} \sum_{\alpha=1}^A (\tau_{\alpha} - p)^2$$

$$\mathcal{V}(\hat{p}_{C_3}) = \frac{1}{aA} \sum_{\alpha=1}^A (p_{\alpha} - \bar{p})^2$$

- Estimadores são dados, respectivamente, por:

$$\hat{\mathcal{V}}(\hat{p}_{C_1}) = \frac{1}{a(a-1)\bar{B}^2} \sum_{\alpha=1}^A (\hat{\tau}_{\alpha} - \hat{p}_{C_1})^2$$

$$\hat{\mathcal{V}}(\hat{p}_{C_3}) = \frac{1}{a(a-1)} \sum_{\alpha=1}^A (\hat{p}_{\alpha} - \hat{p}_{C_3})^2$$

- Quando os conglomerados tiverem o mesmo tamanho, as expressões das variâncias (e de seus respectivos estimadores), simplificam-se. Veja o exercício 7.21 (livro-texto).

Normalidade assintótica

- Temos, para a e A suficientemente grandes, $i = 1, 2, 3$, $j = 1, 2$, que

$$\frac{\widehat{p}_{C_i} - p}{EP_{A_j}(\widehat{p}_{C_i})} \xrightarrow[\substack{a \rightarrow \infty, \\ A - a \rightarrow \infty}]{D} N(0, 1)$$

$$\frac{\widehat{p}_{C_i} - p}{\widehat{EP}_{A_j}(\widehat{p}_{C_i})} \xrightarrow[\substack{a \rightarrow \infty, \\ A - a \rightarrow \infty}]{D} N(0, 1)$$

Intervalos de confiança

- Conglomerados de tamanhos desiguais $j = 1, 2, i = 1, 2, 3$

$$\left[\hat{p}_{C_i} - z_\gamma \sqrt{\hat{v}_{A_j}(\hat{p}_{C_i})}; \hat{p}_{C_i} + z_\gamma \sqrt{\hat{v}_{A_j}(\hat{p}_{C_i})} \right]$$

- Erro da estimativa: $z_\gamma \sqrt{\hat{v}_{A_j}(\hat{p}_{C_i})}$.
- Exercício: repetir os procedimentos considerando que os conglomerados são sorteados segundo um plano AAS_s e/ou quando estes tiverem o mesmo tamanho.

Intervalos de confiança

- Os testes de hipóteses seguem, essencialmente, os mesmo procedimentos descritos no slide 10 de http://www.ime.unicamp.br/~cnaber/aula_AAS%20com%20reposicao%20parte%202%20Amost%20S%202018.pdf, em que a estatística do teste passa a ser

$$Z_t = \frac{\hat{p}_{C_i} - p_0}{\widehat{EP}_{A_j}(\hat{p}_{C_i})}$$

- Exercício: repetir os procedimentos considerando que os conglomerados são sorteados segundo um plano AAS_s e/ou quando os conglomerados tiverem o mesmo tamanho.

Amostragem sistemática (AS)

- Considere uma população com N elementos em que $N = kn$, e $k \in \mathbb{Z}^+$ e que seus elementos estão ordenados de 1 a N (no sistema de referência).
- Uma unidade é selecionada aleatoriamente (segundo AAS) entre as k primeiras unidades do sistema de referência.
- As unidades seguintes, que farão parte da amostra, serão obtidas a partir da primeira unidade selecionada, em intervalos de comprimento k .

Exemplo

- Suponha uma população com $N = 1000$ e $n = 200$. Portanto, $k = 5$.
- Ou seja a população está dividida em 200 grupos de 5 unidades populacionais.
- Suponha que uma unidade seja aleatoriamente selecionada entre as 5 primeiras e que esta tenha sido a de número 3.
- Portanto, em cada um dos 199 grupos restantes, será selecionada sempre a terceira unidade, levando, assim, a uma AS de $n = 200$ unidades.

Características

- Facilidade em sua execução (mecânica do sorteio).
- Bem menos sujeita a erros do entrevistador do que os outros métodos de amostragem vistos.
- Pode ou não levar à inferências mais precisas.
- Pode ser bastante prejudicada, dependendo de como a população esteja organizada no sistema de referência.

Características

- As estimativas das variâncias dos estimadores dependem de como os elementos da população estão dispostos no sistema de referência (aleatório ou de forma tendenciosa).
- Essencialmente, a AS pode ser vista como um caso particular da AC (embora, se a população esteja disposta de forma aleatória, algumas expressões possam ser próximas daquelas obtidas sob o plano AAS).
- Outros estimadores para as variâncias dos estimadores podem ser encontrados em Bolfarine e Bussab (2005) e Cochran (1977) .

Relação com a AC

- Considere a população ordenada da seguinte forma:

$$\mathbf{d} = (y_1, \dots, y_k, y_{k+1}, \dots, y_{2k}, \dots, y_{(n-1)k+1}, \dots, y_{nk})^t$$

que pode também ser representado através de uma matriz onde na linha α tem-se a α -ésima amostra sistemática, enquanto que na coluna i tem-se o i -ésimo grupo(zona).

Relação com a AC

Amostra	grupo				média
	1	2	...	n	
1	y_1	y_{k+1}	...	$y_{(n-1)k+1}$	μ_1
2	y_2	y_{k+2}	...	$y_{(n-1)k+2}$	μ_2
\vdots	\vdots	\vdots	\ddots	\vdots	\vdots
k	y_k	y_{2k}	...	y_{nk}	μ_k
média	$\mu_{.1}$	$\mu_{.2}$...	$\mu_{.n}$	μ

Relação com a AC

- Cada linha da matriz anterior corresponde a uma amostra sistemática de média μ_{α} , $\alpha = 1, 2, \dots, k$.
- Na última coluna temos todas as médias das k amostras sistemáticas.
- Cada uma dessas amostras pode ser vista como um conglomerado (todos de tamanho n).
- Portanto a AS pode ser vista como uma AC em que $A = k$ conglomerados e destes, $a = 1$ é selecionado.

Estimador para a média (μ)

- $\hat{\mu}_{sis} = \hat{\mu}_\alpha$, a média da amostra sistemática (conglomerado) sorteada(o).
- Distribuição exata do estimador $\hat{\mu}_{sis}$,

$\hat{\mu}_{sis}$	μ_1	μ_2	\dots	μ_α	\dots	μ_k
$P(\hat{\mu}_{sis} = r)$	$\frac{1}{k}$	$\frac{1}{k}$	\dots	$\frac{1}{k}$	\dots	$\frac{1}{k}$

- Portanto $\mathcal{E}(\hat{\mu}_{sis}) = \frac{1}{k} \sum_{\alpha=1}^k \mu_\alpha = \mu$ e
 $V_k = \mathcal{V}(\hat{\mu}_{sis}) = \frac{1}{k} \sum_{\alpha=1}^k (\mu_\alpha - \mu)^2$.
- Como selecionamos apenas um único conglomerado $a = 1$, não é possível obter um estimador não viciado de $\mathcal{V}(\hat{\mu}_{sis})$.

Estimador para a média (μ)

- Na maioria dos casos, essa variância é estimada por:

$$\hat{V}_k = \hat{V}_s = \frac{1-f}{n(n-1)} \sum_{i=1}^n (Y_i - \hat{\mu}_{sis})^2,$$

em que $f=n/N$ e $(Y_1, \dots, Y_n)^t$ são variáveis aleatórias que representam a amostra a ser sorteada.

- Tal estimador é adequado quando a AS é aproximadamente equivalente à AAS (elementos organizados de forma “aleatória”).
- Contudo, se se tiver algum padrão na disposição dos elementos no sistema de referência, os resultados podem ser diferentes daquelas da AAS.

Estimador para a média (μ)

- Sob AS (lembrando que AS corresponde ao plano AC, sob certas restrições), temos que

$$V_k = \{1 + \rho_{int}(n - 1)\} \frac{\sigma^2}{n}.$$

- E $\rho_{int} = \frac{\sigma_{ec}^2 - \sigma_{dc}^2 / (n - 1)}{\sigma^2}$.
- Como todos os conglomerados têm o mesmo tamanho e $B = n$ (tamanho de cada conglomerado) então

$$\sigma_{ec}^2 = \frac{1}{k} \sum_{\alpha=1}^k (\mu_{\alpha} - \mu)^2, \quad \sigma_{dc}^2 = \frac{1}{k} \sum_{\alpha=1}^k \sigma_{\alpha}^2,$$

$$\sigma_{\alpha}^2 = \frac{1}{n} \sum_{i=1}^n (y_{\alpha i} - \mu_{\alpha})^2$$

Exemplos

- (Exemplo 7.6, do livro texto) Considere uma população onde $D = (2, 6, 10, 8, 10, 12)$, $N = 6$ e dividida em $n = 2$ e $k = 3$ unidades cada, como se segue:

	Grupo		
Amostra	1	2	Média
1	2	8	5
2	6	10	8
3	10	12	11

Exemplos

- Distribuição do $\hat{\mu}_{sis}$

$\hat{\mu}_{sis}$	5	8	11
$P(\hat{\mu}_{sis} = r)$	1/3	1/3	1/3

- O que leva à $\mathcal{E}(\hat{\mu}_{sis}) = 8$, $\mathcal{V}(\hat{\mu}_{sis}) = 6$.

Exemplos

- Distribuição do $\hat{\mu}$

$\hat{\mu}$	4	5	6	7	8	9	10	11
$P(\hat{\mu} = r)$	1/15	1/15	2/15	2/15	2/15	1/5	2/15	2/15

- O que leva à $\mathcal{E}(\hat{\mu}) = 8$, $\mathcal{V}(\hat{\mu}) = 4,27$. Assim, nota-se que $\hat{\mu}$ é melhor do que $\hat{\mu}$. Com efeito, $\rho_{int} = 0,125 > 0$

Exemplos

- Exemplo 7.7 (livro-texto). Considere uma população com $N = 40$ elementos distribuídos em $k = 10$ amostras sistemáticas de tamanho $n = 4$.
- Análise link <http://www.ime.unicamp.br/~cnaber/Exemplo%207.7%20Amost%20S%202018.ods>.
- Neste caso $\mathcal{V}_{AE_2}(\hat{\mu}) = (1 - f) \frac{s^2}{n} = \left(1 - \frac{1}{10}\right) \frac{136,25}{4} = 30,7$.

Exemplos

- Além disso, $\mathcal{V}(\hat{\mu}_{sis}) = \frac{1}{k} \sum_{\alpha=1}^k (\mu_{\alpha} - \mu)^2 = 8,74$.
- Com efeito, neste caso $\rho_{int} = -0,217 < 0$
- Sob AE: cada um dos grupo é um estrato. Seleciona-se um elemento de cada estrato. $\mathcal{V}(\hat{\mu}_{es}) = \left(1 - \frac{1}{10}\right) \frac{13,49}{4} \approx 3,03$.