Capítulo 9 Amostragem Binomial ou de Bernoulli (AB)

A Amostragem Binomial ou de Bernoulli (AB) oferece outra alternativa (pouco usada) à AAS, para selecionar unidades com *equiprobabilidade*. Trata-se de método de amostragem que também dispensa existência de cadastro prévio para seleção da amostra, podendo a amostra ser selecionada ao mesmo tempo que o cadastro vai sendo construído, como na amostragem sistemática simples.

9.1 Método de seleção da amostra

As unidades aparecem na população ou no cadastro numa certa ordem, digamos igual à dos rótulos $i = 1, 2, \ldots, N$. Seja π a *fração amostral* desejada, tal que $0 < \pi < 1$. Sejam também $A_1, A_2, \ldots, A_N$ um conjunto de N variáveis aleatórias *independentes* e distribuídas segundo uma distribuição Uniforme no intervalo [0;1], denotada $U[0;1]$. Associamos A_i com a unidade i, para todo $i \in U$.

Então processamos sequencialmente as unidades em U, testando para cada $i = 1, \ldots, N$ a condição $A_i < \pi$. Quando $A_i < \pi$ ocorre, incluímos a unidade i na amostra s. Quando $A_i \geq \pi$, a unidade não é incluída na amostra e passamos à próxima unidade.

9.2 Probabilidades de inclusão na Amostragem Binomial ou de Bernoulli

As probabilidades de inclusão de primeira ordem são:

$$\pi_i = P(i \in s) = P(A_i < \pi) = P(U[0;1] < \pi) = \pi$$

As probabilidades de inclusão de segunda ordem são:

$$\pi_{ij} = P(i, j \in s) = P(A_i < \pi \text{ e } A_j < \pi) = P(A_i < \pi) \times P(A_j < \pi) = \pi^2 \quad \forall \; i \neq j \in U$$

Uma dificuldade associada com Amostragem Binomial ou de Bernoulli é que, antes de ser feita a seleção da amostra, o *tamanho efetivo da amostra* obtida é uma variável aleatória. Isso pode causar dificuldades para quem está planejando a pesquisa.

Para verificar isso, note que $n = \sum_{i \in U} I(A_i < \pi)$. Consequentemente, n~Binomial(N;π). Logo: $E(n) = N \times \pi$ e $V(n) = N \times \pi \times (1 - \pi)$.

9.3 Estimação de totais sob Amostragem Binomial ou de Bernoulli

O estimador HT (não viciado) do total Y é dado por:

$$\widehat{Y}_{AB} = \frac{1}{\pi} \sum_{i \in s} y_i \tag{9.1}$$

A variância desse estimador do total é dada por:

$$\begin{aligned} V_{AB}(\widehat{Y}_{AB}) &= \left(\frac{1}{\pi}\right)^2 V(\sum_{i \in U} \delta_i y_i) = \left(\frac{1}{\pi}\right)^2 \sum_{i \in U} V(\delta_i) y_i^2) \\ &= \left(\frac{1}{\pi}\right)^2 \pi(1-\pi) \sum_{i \in U} y_i^2 = \left(\frac{1}{\pi} - 1\right) \sum_{i \in U} y_i^2 \end{aligned} \tag{9.2}$$

Um estimador não viciado da variância do estimador de total é dado por:

$$\widehat{V}_{AB}(\widehat{Y}_{AB}) = \frac{1}{\pi} \left(\frac{1}{\pi} - 1\right) \sum_{i \in s} y_i^2 \tag{9.3}$$

Devido à variabilidade do tamanho efetivo da amostra, o estimador HT do total é pouco eficiente. Um estimador alternativo para o total, definido sempre que $n > 0$, é o estimador tipo razão dado por:

$$\widehat{Y}_{AB}^R = \frac{N\pi}{n} \widehat{Y}_{AB} = N \frac{1}{n} \sum_{i \in s} y_i = N\overline{y} \tag{9.4}$$

Conforme Särndal, Swensson, e Wretman (1992) (p. 65), uma aproximação da variância do estimador alternativo de total é dada por:

$$V_{AB}(\widehat{Y}_{AB}^R) \cong N^2 \left(\frac{1}{N\pi} - \frac{1}{N}\right) S_y^2 \tag{9.5}$$

Esta variância é portanto aproximadamente igual à de uma AAS de tamanho igual a $N\pi$, que seria também o tamanho esperado da amostra sob AB. Ainda conforme Särndal, Swensson, e Wretman (1992) (p. 63), a variância do estimador simples do total pode ser reescrita como:

$$V_{AB}(\widehat{Y}_{AB}) = N^2 \left(\frac{1}{N\pi} - \frac{1}{N} \right) S_y^2 \left[1 - \frac{1}{N} + \frac{1}{CV_y^2} \right] \tag{9.6}$$

Logo, de acordo com Särndal, Swensson, e Wretman (1992) (p. 65), a eficiência relativa do estimador alternativo é dada por:

$$\frac{V_{AB}(\widehat{Y}_{AB})}{V_{AB}(\widehat{Y}_{AB}^R)} \cong \left[1 - \frac{1}{N} + \frac{1}{CV_y^2} \right] \cong \left[1 + \frac{1}{CV_y^2} \right] \tag{9.7}$$

Essa expressão mostra que o estimador alternativo será tanto mais eficiente que o estimador HT quanto menor for o CV da variável de interesse y. Isso faz sentido, porque a variância do estimador HT depende da variabilidade do tamanho efetivo da amostra, e esse componente domina a variância total quando o CV da variável de interesse y diminui.

9.4 Estimação de médias sob Amostragem Binomial

O estimador HT (não viciado) da média $\overline{Y}$ é dado por:

$$\widehat{\overline{Y}}_{AB} = \frac{1}{N\pi} \sum_{i \in s} y_i \tag{9.8}$$

A variância do estimador HT da média é dada por:

$$V_{AB}\left(\widehat{\overline{Y}}_{AB}\right) = \frac{1}{N^2} \left(\frac{1}{\pi} - 1 \right) \sum_{i \in U} y_i^2 \tag{9.9}$$

Um estimador não viciado da variância do estimador HT da média é dado por:

$$\widehat{V}_{AB}\left(\widehat{\overline{Y}}_{AB}\right) = \frac{1}{N^2\,\pi} \left(\frac{1}{\pi} - 1 \right) \sum_{i \in s} y_i^2 \tag{9.10}$$

Assim como no caso do estimador de total, devido à variabilidade do tamanho efetivo da amostra, o estimador HT da média é pouco eficiente. Um estimador alternativo para a média, definido sempre que $n > 0$, é o estimador tipo razão dado por:

$$\widehat{\overline{Y}}^{R}_{AB} = \frac{N\pi}{n}\widehat{\overline{Y}}_{AB} = \frac{1}{n}\sum_{i \in s} y_i \tag{9.11}$$

Uma aproximação da variância do estimador alternativo da média é dada por:

$$V_{AB}\left(\widehat{\overline{Y}}^{R}_{AB}\right) \cong \left(\frac{1}{N\pi} - \frac{1}{N}\right) S_y^2 \tag{9.12}$$

Esta variância é portanto aproximadamente igual à do estimador de média sob uma AAS de tamanho igual a $N\pi$, que seria também o tamanho esperado da amostra sob AB. As análises feitas na seção anterior sobre a eficiência relativa dos estimadores de total são válidas também para os estimadores da média.

9.5 Exemplos de aplicação da Amostragem Binomial ou de Bernoulli

Um exemplo clássico ocorre na inspeção alfandegária praticada na saída de passageiros chegando de vôos internacionais. Quando cada passageiro aperta o botão para saber se sua bagagem será ou não inspecionada pela Receita Federal, está em ação um processo de Amostragem Binomial. Não sabemos o valor da fração amostral estabelecida pela Receita Federal, mas este é o único parâmetro necessário para especificar completamente o processo de amostragem em questão.

A Amostragem Sistemática também seria viável nesse caso, mas poderia ser mais facilmente burlada por pessoas interessadas em não ter sua bagagem inspecionada, caso fossem capazes de detectar qual é o valor do pulo K sendo praticado. O emprego de Amostragem Binomial impede essa prática, ao tornar imprevisível o resultado de cada sorteio que define se o passageiro terá ou não sua bagagem inspecionada.

Aplicações de Amostragem Binomial para populações em fluxo são convenientes, pois dispensam lista ou cadastro prévios. Deve-se evitar seu uso sempre que $P(n = 0)$ for 'grande'.