AMOSTRAGEM

Unidade 8

Estimação para Domínios sob AAS

Domínios de Estudo

Domínios de estudo, subpopulações ou "pequenas áreas" são quaisquer subconjuntos da população para os quais desejamos obter / produzir estimativas separadas a partir de uma amostra maior.

Por exemplo:

- ✓ homens de 40 anos ou mais;
- ✓ mulheres em idade reprodutiva (15 a 49 anos);
- ✓ fazendas com produção de café;
- ✓ crianças e adolescentes;
- ✓ municípios ou regiões específicas; etc.

Domínios de Estudo - Notação

Seja d_i a variável indicadora do domínio d, isto é:

$$d_i = \begin{cases} 1 \text{ se } i \text{ pertence ao domínio } d, \\ 0 \text{ caso contrário.} \end{cases}$$

Denotamos por $U_d = \{i : d_i = 1\}$ a população no domínio d, e por $N_d = \sum_{i \in U} d_i$ o tamanho do domínio d.

Então $y_{id} = y_i\, d_i \begin{cases} = y_i & \text{se } i \in U_d, \\ = 0 & \text{caso contrário} \end{cases}$ define a variável de estudo y relevante para estimação no domínio d.

Note que y_{id} é uma variável 'derivada', formada pelo produto de duas outras variáveis de pesquisa: y_i e d_i.

Parâmetros-Alvo Para Domínios

Total do domínio

$$Y_d = \sum_{i \in U} y_{id} = \sum_{i \in U} y_i d_i = \sum_{i \in U_d} y_i$$

Média do domínio

$$\overline{Y}_d = Y_d / N_d = \sum_{i \in U} y_{id} \Big/ \sum_{i \in U} d_i$$

➔ Caso especial de Razão de Médias das variáveis y_{id} e d_i.

Variância do domínio

$$S_d^2 = \sum_{i \in U} d_i (y_i - \overline{Y}_d)^2 \Big/ (N_d - 1)$$

Proporção de unidades populacionais no domínio

$$P_d = N_d / N = \overline{D}$$

Estimação de Parâmetros de Domínios

Passo 1: Selecionamos uma AAS de tamanho n da população U de tamanho N, e observamos $(y_i\ ;\ d_i)$ para todo $i \in s$.

Passo 2: Então construímos a variável derivada $y_{id} = y_i\, d_i$.

Passo 3: Estimamos a média $\bar{Y}_d = Y_d / N_d$ usando:

$$\bar{y}_d = \hat{Y}_d / \hat{N}_d = \sum_{i \in s} w_i y_{id} \Big/ \sum_{i \in s} w_i d_i$$

$$= (N/n) \sum_{i \in s} y_{id} \Big/ (N/n) \sum_{i \in s} d_i$$

$$= \sum_{i \in s} y_{id} \Big/ n_d$$

onde n_d é o número de observações no domínio d na amostra.

➔ Note que n_d é uma variável aleatória.

Estimação da Média

O estimador de média é uma razão de totais estimados.

Para obter sua variância, podemos recorrer aos resultados disponíveis para estimar razões.

$$V_{AAS}(\bar{y}_d) \cong \frac{1-f}{n} \frac{1}{P_d^2} \frac{1}{N-1} \sum_{i \in U} (y_{id} - \bar{Y}_d d_i)^2$$

$$= \frac{1-f}{n} \frac{1}{P_d^2} \frac{1}{N-1} \sum_{i \in U} d_i (y_i - \bar{Y}_d)^2$$

O estimador de variância correspondente é:

$$\hat{V}_{AAS}(\bar{y}_d) = \frac{1-f}{n} \frac{1}{p_d^2} \frac{1}{n-1} \sum_{i \in s} d_i (y_i - \bar{y}_d)^2$$

Estimação da Média: Propriedades Condicionais (fixando n_d)

Cochran (1977, seção 2.12) sugere analisar a distribuição da média considerando o tamanho da amostra no domínio n_d como fixo (em seu valor observado).

Nesse caso, mostra que as n_d observações na amostra s formam uma AAS da população U_d.

Então segue-se que:

$$V_{AAS}\left(\bar{y}_d \middle| n_d > 0\right) = \left(1 - \frac{n_d}{N_d}\right)\frac{S_d^2}{n_d}$$

IBGE

Estimação da Média: Propriedades Condicionais (fixando n_d)

A variância pode então ser estimada usando:

$$\hat{V}_{AAS}\left(\bar{y}_d \middle| n_d > 0\right) = \left(1 - \frac{n_d}{N_d}\right)\frac{s_d^2}{n_d}$$

onde $s_d^2 = \sum_{i \in s} d_i (y_i - \bar{y}_d)^2 / (n_d - 1)$ é um estimador não viciado para S_d^2 (quando n_d é fixado).

Se N_d for desconhecido, $\hat{V}_{AAS}\left(\bar{y}_d \middle| n_d > 0\right)$ não é calculável. Cochran (1977, p. 35) sugere substituir n_d/N_d por n/N:

$$\hat{V}_{AAS}^{*}\left(\bar{y}_d \middle| n_d > 0\right) = \left(1 - \frac{n}{N}\right)\frac{s_d^2}{n_d}$$

Estimação de Parâmetros de Domínios – Total Y_d

Para obter estimadores do total populacional Y_d note que:

$$Y_d = N_d \overline{Y}_d \, .$$

Portanto, há duas situações importantes a distinguir:

- quando o tamanho do domínio (N_d) é **conhecido**,
- quando o tamanho do domínio (N_d) é **desconhecido**.

Estimação do Total com N_d Conhecido

$$\hat{Y}_d^R = N_d \, \bar{y}_d = N_d \sum_{i \in s} y_{id} \Big/ n_d$$

$$V_{AAS}\left(\hat{Y}_d^R \,\middle|\, n_d > 0\right) = N_d^2 \left(1 - \frac{n_d}{N_d}\right) \frac{S_d^2}{n_d}$$

$$\hat{V}_{AAS}\left(\hat{Y}_d^R \,\middle|\, n_d > 0\right) = N_d^2 \left(1 - \frac{n_d}{N_d}\right) \frac{s_d^2}{n_d}$$

Estimação do Total Y_d Supondo que N_d É Desconhecido

Para estimação não viciada do total supondo que N_d é desconhecido, usar:

$$\hat{Y}_d = N \sum_{i \in s} y_{id} / n .$$

Prova: basta aplicar os resultados usuais de estimação de totais sob AAS para a variável derivada y_{id}.

$$V_{AAS}(\hat{Y}_d) = N^2 \left(1 - \frac{n}{N}\right) \frac{1}{n} \frac{1}{N-1} \sum_{i \in U} \left(y_{id} - \frac{Y_d}{N} \right)^2$$

$$\hat{V}_{AAS}(\hat{Y}_d) = N^2 \left(1 - \frac{n}{N}\right) \frac{1}{n} \frac{1}{n-1} \sum_{i \in s} \left(y_{id} - \frac{t_d}{n} \right)^2$$

Intervalos de Confiança

A obtenção de intervalos de confiança para os parâmetros populacionais requer amostras grandes nos domínios.

Nesse caso, valem as seguintes aproximações:

$$\left(\hat{Y}_d^R - Y_d \right) / \sqrt{\hat{V}_{AAS}\left(\hat{Y}_d^R \right)} \approx N(0;1) \quad \text{para } n_d \text{ grande.}$$

$$\left(\hat{Y}_d - Y_d \right) / \sqrt{\hat{V}_{AAS}\left(\hat{Y}_d \right)} \approx N(0;1) \quad \text{para } n_d \text{ grande.}$$

$$\left(\bar{y}_d - \bar{Y}_d \right) \Big/ \sqrt{\hat{V}_{AAS}(\bar{y}_d)} \approx N(0;1) \quad \text{para } n_d \text{ grande}$$

Comparação da Eficiência dos Estimadores de Total

Foram propostos dois estimadores para o total populacional Yd. Portanto, é importante saber quando usar um ou outro.

Comparando as respectivas variâncias (Cochran, p. 38):

$$\frac{V_{AAS}\left(\hat{Y}_d^R \middle| n_d > 0 \text{ e } N_d \text{ conhecido}\right)}{V_{AAS}\left(\hat{Y}_d \middle| n_d > 0 \text{ e } N_d \text{ desconhecido}\right)} \cong \frac{S_d^2}{S_d^2 + (1 - P_d)\bar{Y}_d^2}$$

$$= \frac{C_d^2}{C_d^2 + (1 - P_d)}$$

➔ Conhecer o valor de N_d sempre melhora eficiência do estimador de total. ➔ Melhoria é maior quando P_d é pequena.