Resumo Cap.6 e 7

Usando o estimador Regressão:

Domicílio	REIDENTES(X)	MASCULINO (Y)	Domicílio	REIDENTES	MASCULINO
1	5	1	16	5	3
2	6	3	17	4	3
3	3	1	18	4	3
4	3	1	19	3	2
5	2	1	20	3	1
6	3	1	21	4	1
7	3	1	22	3	2
8	3	1	23	3	2
9	4	2	24	1	0
10	4	3	25	2	1
11	3	2	26	4	3
12	2	1	27	3	1
13	7	3	28	4	2
14	4	3	29	2	1
15	3	2	30	4	2

Estimador:

$$\bar{y}_{Reg} = \bar{y} + \hat{B}(\bar{X} - \bar{x}),$$

com

$$\hat{B} = \frac{S_{xy}}{s_x^2}.$$

Calcular: $V_{Reg} = Var[\bar{y}_{Reg}]$
Estimador $\hat{V}_{Rg} = s_y^2(1 - \hat{\rho}_{xy}^2)$, com

$$\hat{\rho} = \frac{s_{xy}}{s_x s_y}.$$

AMOSTRAGEM POR CONGLOMERADOS - CAP. 7

População dividida em A subpopulações (COGLOMERADOS) disjuntos de tamanhos $B_1, \ldots, B_A$ tais que

$$B_1 + \ldots + B_A = N,$$

e

$$\bar{B} = \frac{N}{A} \to N = A\bar{B}.$$

Total Populacional:

$$\tau = \sum_{\alpha=1}^{A} \tau_\alpha = \sum_{\alpha=1}^{A} B_\alpha \mu_\alpha \longrightarrow \mu = \frac{\tau}{N} = \frac{1}{A\bar{B}} \sum_{\alpha=1}^{A} B_\alpha \mu_\alpha = \frac{1}{A} \sum_{\alpha=1}^{A} \frac{B_\alpha}{\bar{B}} \mu_\alpha,$$

Sendo selecionada uma AASc $\mathbf{s}$ de "a" conglomerados, sendo observadas TODAS as unidades nos conglomerados selecionados (e,portanto, μ_α nos conglomerados selecionados) temos o estimador de μ:

$$\hat{\mu}_c = \frac{1}{a} \sum_{\alpha \in \mathbf{s}} \frac{B_\alpha}{\bar{B}} \mu_\alpha.$$

Para calcular o estimador acima é necessario conhecer $\bar{B}$, ou seja, N.

Também

$$E[\hat{\mu}_c] = E[\frac{1}{a} \sum_{\alpha \in \mathbf{s}} \frac{B_\alpha}{\bar{B}} \mu_\alpha] = \frac{1}{N} \sum_{\alpha=1}^{A} \tau_\alpha = \mu,$$

e

$$Var[\hat{\mu}_c] = \frac{\sigma_{ec}^2}{a},$$

onde

$$\sigma_{ec}^2 = \frac{1}{A} \sum_{i=1}^{A} (\frac{B_\alpha}{\bar{B}} \mu_\alpha - \mu)^2$$

que é estimado por

$$\hat{\sigma}^2_{ec} = \frac{1}{a-1}\sum_{i\in \mathbf{s}}(\frac{B_\alpha}{\bar{B}}\mu_\alpha - \hat{\mu}_c)^2$$

Para o total populacional τ temos o estimador

$$\hat{\tau}_c = N\hat{\mu}_c = A\bar{B}\frac{1}{a}\sum_{\alpha\in \mathbf{s}}\frac{B_\alpha}{\bar{B}}\mu_\alpha = A\frac{1}{a}\sum_{\alpha\in \mathbf{s}}B_\alpha\mu_\alpha = A\frac{1}{a}\sum_{\alpha\in \mathbf{s}}\tau_\alpha = A\hat{\bar{\tau}}_c,$$

com

$$\hat{\bar{\tau}}_c = \frac{1}{a}\sum_{\alpha\in \mathbf{s}}\tau_\alpha$$

Temos

$$Var[\hat{\tau}_c] = \frac{A^2}{Aa}\sum_{\alpha=1}^{A}(\tau_\alpha - \bar{\tau})^2.$$

Como estimador não viciado temos

$$\widehat{Var}[\hat{\tau}_c] = \frac{A}{a-1}\sum_{\alpha\in \mathbf{s}}(\tau_\alpha - \hat{\bar{\tau}}_c)^2.$$

Quando N é desconhecido e então $\bar{B}$ podemos definir o estimador

$$\hat{\mu}_{c2} = \frac{ET}{EN} = \frac{A\hat{\bar{\tau}}}{A\hat{\bar{B}}} = \frac{\hat{\bar{\tau}}}{\hat{\bar{B}}}$$

Temos então um estimador do tipo razao (Chap. 5) com $X_i = B_i$ e $Y_i = \tau_i$ de modo que

$$Var[\hat{\mu}_{c2}] = \frac{\sigma^2_R}{\bar{B}^2 a},$$

onde

$$\sigma^2_R = \frac{1}{A}\sum_{\alpha=1}^{A}(\tau_\alpha - B_\alpha\mu)^2 = \frac{1}{A}\sum_{\alpha=1}^{A}(\frac{B_\alpha}{\bar{B}})^2(\mu_\alpha - \mu)^2 = \sigma^2_{eq},$$

pois

$$R = \frac{\mu_Y}{\mu_X} = \frac{\bar{\tau}}{\bar{B}} = \mu.$$

Usar o estimador

$$\hat{\sigma}_R^2 = \frac{1}{a-1}\sum_{\alpha\in s}(\tau_\alpha - B_\alpha\hat{\mu}_{2c})^2 = \frac{1}{a-1}\sum_{\alpha\in s}B_\alpha^2(\mu_\alpha - \hat{\mu}_{2c})^2 = s_{eq}^2.$$

Coeficiente de correlação intraclasse.

1. Formar todos os pares de unidades distintas nos conglomerados
2. Calcular coeficiente de correlação de Pearson para os pares formados.
3. Para conglomerados de igual tamanho,

$$\rho_{int} = \frac{\sigma_e^2 - \frac{\sigma_d^2}{B}}{\sigma^2},$$

onde

$$\sigma^2 = \sigma_{ec}^2 + \sigma_{dc}^2,$$

com

$$\hat{\rho}_{int} = r_{int} = \frac{s_{ec}^2 - \frac{s_{dc}^2}{B-1}}{s_{ec}^2 + s_{dc}^2}.$$

4. Para conglomerados de tamanhos diferentes podemos usar

$$\rho_{2c} = \frac{\sigma_{ec}^2 - \frac{\sigma_{dp}^2}{B}}{\sigma_{ec}^2 + \sigma_{dp}^2}.$$

Podemos tbém usar σ_{eq}^2 (no lugar de σ_{ec}^2 que corresponde a $\hat{\mu}_{2c}$.

Com estimadores

$$\hat{\sigma}_{ec}^2 = s_{ec}^2 = \frac{1}{a-1}\sum_{\alpha\in s}(\frac{B_\alpha}{\bar{B}}\mu_\alpha - \hat{\mu}_{2c})^2 \quad \hat{\sigma}_{dp} = \frac{1}{a}\sum_{\alpha\in s}(\frac{B_\alpha}{\hat{\bar{B}}})\sigma_\alpha^2.$$

Caso $B_\alpha = B$, $\alpha = 1, \ldots, A$,

$$\hat{\mu}_c = \hat{\mu}_c = \frac{\hat{\bar{\tau}}}{B}.$$

$$Var[\hat{\mu}_c] = \frac{1}{aA}\sum_{\alpha=1}^{A}(\mu_\alpha - \mu)^2 = \sigma_e^2.$$

Amostragem Sitemática (AS)

População dividida em n subpopulações (Zonas) de k unidades cada. Na primeira zona uma unidade é selecionada de acordo com AAS. Nas outras zonas seleciona-se (sistematicamente) a unidade na posição correspondente a posição da unidade selecionada na primeira zona. O estimador da média populacional é a média da amostra sistemática selecionada.

Exemplo. Considere $N = 9$, com $n = k = 3$. Enquanto temos $\binom{9}{3} = 84$ amostras AASs distintas (não levando em conta a ordem), temos apenas k=3 amostras sistemáticas, ou seja:

$$s_1 = (1, 4, 7), \quad s_2 = (2, 5, 8), \quad s_3 = (3, 6, 9).$$

A AS pode ser vista como uma AC, onde temos n conglomerados e apenas UM (a=1) é selecionado. Portanto, como na amostragem por conglomerados é necessario dois (a=2) conglomerados para estimar a variância não podemos estimar a variância da AS. Poldemos usar a variância da AASs (s^2/n) que pode não ser conveniente em muitas situações. Uma alternativa é usar vários valores iniciais, seleconar ao acaso mais de uma unidade na primeira zona.