

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, \dots, B_A tais que

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

e

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

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 s de " a " conglomerados, sendo observadas TODAS as unidades nos conglomerados selecionados (e, portanto, μ_{α} nos conglomerados selecionados) temos o estimador de μ :

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

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

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

e

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

onde

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

que é estimado por

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

Para o total populacional τ temos o estimador

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

com

$$\hat{\tau}_c = \frac{1}{a} \sum_{\alpha \in S} \tau_\alpha$$

Temos

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

Como estimador não viciado temos

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

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

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

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

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

onde

$$\sigma_R^2 = \frac{1}{A} \sum_{\alpha=1}^A (\tau_\alpha - B_\alpha \mu)^2 = \frac{1}{A} \sum_{\alpha=1}^A \left(\frac{B_\alpha}{\bar{B}} \right)^2 (\mu_\alpha - \mu)^2 = \sigma_{eq}^2,$$

pois

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

Usar o estimador

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

Coefficiente 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

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

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

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

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

Amostragem Sistemática (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 necessário dois ($a=2$) conglomerados para estimar a variância não podemos estimar a variância da AS. Podemos usar a variância da AASs (s^2/n) que pode não ser conveniente em muitas situações. Uma alternativa é usar vários valores iniciais, selecionar ao acaso mais de uma unidade na primeira zona

Revisão Cap. 8

Temos uma Pop. dividida em A conglomerados e uma AASc, s , é selecionada (1o. estágio). De cada conglomerado selecionado no 1o. estágio uma AASc, s_α , com b_α unidades é selecionada. Temos então amostragem em dois estágios.

No conglomerado α selecionado no 1o. estágio, estimamos o total populacional por

$$\tau_\alpha = B_\alpha * \hat{\mu}_\alpha,$$

onde

$$\hat{\mu}_\alpha = \frac{1}{b_\alpha} \sum_{j \in s_\alpha} Y_{\alpha j}.$$

Temos então o estimador da M.P.

$$\hat{\mu}_{2c} = \frac{1}{a} \sum_{\alpha \in s} \frac{B_\alpha}{B} \hat{\mu}_\alpha = \frac{\hat{\tau}_\alpha}{B}.$$

Temos, $E[\hat{\mu}_{2c}]$ é não viciado

$$E[\hat{\mu}_{2c}] = E_1\{E_2[\frac{1}{a} \sum_{\alpha \in s} \frac{B_\alpha}{\bar{B}} \hat{\mu}_\alpha]\} = E_1[\frac{1}{a} \sum_{\alpha \in s} \frac{B_\alpha}{\bar{B}} \mu_\alpha] = \frac{1}{A} \sum_{\alpha=1}^A \frac{B_\alpha}{\bar{B}} \mu_\alpha = \mu.$$

$$\begin{aligned} V_{2c} &= Var[\hat{\mu}_{2c}] = E_1[Var_2] + Var_1[E_2] \\ &= \frac{1}{aA} \sum_{\alpha=1}^A \left(\frac{B_\alpha}{\bar{B}} \mu_\alpha - \mu\right)^2 + \frac{1}{aA} \sum_{\alpha=1}^A \left(\frac{B_\alpha}{\bar{B}}\right)^2 \frac{\sigma_\alpha^2}{b_\alpha}. \end{aligned}$$

Estimador para variância para AASc:

$$\hat{V}_{2c} = \frac{s_{2ec}^2}{a},$$

onde

$$s_{2ec}^2 = \frac{1}{a-1} \sum_{\alpha \in s} \left(\frac{B_\alpha}{\bar{B}} \hat{\mu}_\alpha - \hat{\mu}_{2c}\right)^2.$$

Sendo N desconhecido substituir \bar{B} por $\hat{\bar{B}}$ obtendo o estimador

$$\hat{\mu}_{2c2} = \frac{1}{a} \sum_{\alpha \in s} \frac{B_\alpha}{\hat{\bar{B}}} \hat{\mu}_\alpha,$$

com variância estimada

$$\hat{V}_{2c2} = \frac{s_{2eq}^a}{a}$$

onde

$$s_{2eq}^2 = \frac{1}{a-1} \sum_{\alpha \in s} \left(\frac{B_\alpha}{\hat{\bar{B}}}\right)^2 (\hat{\mu}_\alpha - \hat{\mu}_{2c2})^2.$$

Pode-se estimar o CCI por

$$r_{2c} = \frac{s_{2ec}^2 - \frac{s_{2dp}^2}{\bar{B}}}{s_{2ec}^2 + s_{2dp}^2}.$$

Cap. 9. Probs desiguais

População: $\mathcal{U} = \{1, \dots, N\}$

Planejamento amostral: Unidades são selecionadas com reposição e com probabilidade proporcional a Z_i , $\sum_{i=1}^N Z_i = 1$.

$$Y : Y_1, \dots, Y_N,$$

$$Z : Z_1, \dots, Z_N,$$

com estimador não viciado para o total populacional

$$\hat{\tau}_{ppz} = \frac{1}{n} \sum_{i \in s} \frac{Y_i}{Z_i} = \frac{1}{n} \sum_{i=1}^N \frac{Y_i}{Z_i} f_i,$$

$$f_i \sim \text{Bin}(n, Z_i), \quad i = 1, \dots, N.$$

Temos tbém da pg. 228, que

$$V_{ppz} = \frac{1}{n} \sum_{i=1}^N Z_i \left(\frac{Y_i}{Z_i} - \tau \right)^2,$$

com

$$\hat{V}_{ppz} = \frac{1}{n(n-1)} \sum_{i \in s} \left(\frac{Y_i}{Z_i} - \hat{\tau}_{ppz} \right)^2.$$

Amostragem por conglomerados

Conglomerados selecionados com probabilidade proporcional ao tamanho dos conglomerados ou seja B_α , isto é

$$Z_\alpha = \frac{B_\alpha}{N}, \quad \alpha = 1, \dots, A,$$

de modo que

$$\begin{aligned} \hat{\tau}_{cppz} &= \frac{1}{a} \sum_{\alpha \in s} \frac{\tau_\alpha}{Z_\alpha} = \frac{1}{a} \sum_{\alpha \in s} \frac{\tau_\alpha}{B_\alpha/N} \\ &= \frac{N}{a} \sum_{\alpha \in s} \mu_\alpha. \end{aligned}$$

Tbém, para a média pop. temos o estimador

$$\hat{\mu}_{cppz} = \frac{\hat{\tau}_{cPPZ}}{N} = \frac{1}{a} \sum_{\alpha \in s} \mu_\alpha.$$

E,

$$V_{cppz} = Var[\hat{\tau}_{cppz}] = \frac{N}{a} \sum_{\alpha=1}^A B_{\alpha} (\mu_{\alpha} - \mu)^2,$$

com

$$\hat{V}_{cppz} = \frac{N^2}{a(a-1)} \sum_{\alpha \in s} (\mu_{\alpha} - \hat{\mu}_{cppz})^2.$$

A2E com probabilidades desiguais

Temos o estimador

$$\hat{\tau}_{2cppz} = \frac{1}{a} \sum_{\alpha \in s} \frac{B_{\alpha}}{Z_{\alpha}} \hat{\mu}_{\alpha}.$$

Com Z_{α} no 1o. estágio, e AASc no 2o. estágio,

$$V_{2cppz} = Var[\hat{\tau}_{2cppz}] = \frac{1}{a} \sum_{\alpha=1}^A Z_{\alpha} \left(\frac{\tau_{\alpha}}{Z_{\alpha}} - \tau \right)^2 + \frac{1}{a} \sum_{\alpha=1}^A \frac{B_{\alpha}^2 \sigma_{\alpha}^2}{Z_{\alpha} b_{\alpha}}.$$

Tbém,

$$V_{2cppz} = \frac{1}{a(a-1)} \sum_{\alpha \in s} \left(\frac{\hat{\tau}_{\alpha}}{Z_{\alpha}} - \hat{\tau}_{2cppz} \right)^2.$$

Exerçios: 1 Trabalhar o caso $Z_{\alpha} = B_{\alpha}/N$.

2. Trabalhar o caso $Z_{\alpha} = 1/A$, $\alpha = 1, \dots, A$.

3. Trabalhar o caso $Z_{\alpha} = B_{\alpha}/N$ para o Exemplo das aldeias.

4. Trabalhar os dois casos para a população do Ex. 7.7 da pg. 188.