Spis treści

Spis treści

1 Literatura

Literatura

- Bracha Cz. Teoretyczne podstawy metody reprezentacyjnej, PWN, 1996
- Zasępa R. Badania statystyczne metodą reprezentacyjną, PWN, 1962
- Steczkowski J. Metoda reprezentacyjna w badaniach zjawisk ekonomiczno-społecznych, PWN 1995

2 Populacja generalna

2.1 Podstawowe pojęcia

Populacja generalna

Podstawowe pojęcia

- jednostki badania $\mathcal{U} = \{u_1, u_2, \ldots, u_N\}$
- cecha statystyczna $\mathcal{Y} : \mathcal{U} \to R$
- parametr populacji $\mathcal{Y} = \{Y_1, Y_2, \ldots, Y_N\}$
- przestrzeń parametrów $\Omega \subset R^N$
- funkcja parametryczna $T : \Omega \to R$

Populacja generalna

Funkcje parametryczne (przykłady)

- wartość globalna $Y = \sum_{i=1}^{N} Y_i$
- wartość średnia $\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i$
- wariancja $S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \bar{Y})^2$
- wartość maksymalna $\max\{Y_i : i =, \ldots, N\}$

2.2 Plan losowania

Plan losowania

Definicja
rozkład prawdopodobieństwa π na zbiorze wszystkich prób $\mathcal{S}$

Prawdopodobieństwo k-tego rzędu
prawdopodobieństwo $\pi_{i_1,\ldots,i_k}$ wylosowania obiektów $u_{i_1}, \ldots, u_{i_k}$

Plan losowania

Prawdopodobieństwo pierwszego rzędu
prawdopodobieństwo π_{i_1} wylosowania do badania obiektu u_{i_1}

Prawdopodobieństwo pierwszego rzędu

$$\pi_{i_1} = \sum_{s \in \mathcal{S}(i_1)} \pi(s)$$

$\mathcal{S}(i_1) \subset \mathcal{S}$ jest zbiorem tych prób, w których występuje obiekt u_{i_1}

Plan losowania

Prawdopodobieństwo drugiego rzędu
prawdopodobieństwo π_{i_1,i_2} wylosowania do badania pary obiektów u_{i_1}, u_{i_2}

Prawdopodobieństwo drugiego rzędu

$$\pi_{i_1,i_2} = \sum_{s \in \mathcal{S}(i_1,i_2)} \pi(s)$$

$\mathcal{S}(i_1, i_2) \subset \mathcal{S}$ jest zbiorem tych prób, w których występują obiekty u_{i_1} i u_{i_2}

2.3 Schemat losowania

Schemat losowania

Schemat losowania
proces wyboru jednostek z populacji $\mathcal{U}$ ze z góry określonym prawdopodobieństwem wyboru poszczególnych jednostek

Operat losowania
wykaz jednoznacznie określonych jednostek badania

Schemat losowania

Przykłady

- losowanie proste bez zwracania (lpbz)
- losowanie proste ze zwracaniem (lpzz)
- losowanie systematyczne

Schemat losowania

Przykłady

- losowanie proporcjonalne do wartości cechy X bez zwracania (lppxbz)
- losowanie proporcjonalne do wartości cechy X ze zwracaniem (lppxzz)

Schemat losowania

Twierdzenie
Dla dowolnego planu losowania istnieje co najmniej jeden schemat losowania

2.4 Schemat lpbz

Schemat lpbz

Własności

- prawdopodobieństwo wylosowania próbki n elementowej: $\frac{1}{\binom{N}{n}}$
- $\pi_j = \frac{n}{N}$
- $\pi_{ij} = \frac{n(n-1)}{N(N-1)}$

2.5 Schemat lpzz

Schemat lpzz

Własności

- prawdopodobieństwo wylosowania próbki n elementowej: $\left(\frac{1}{N}\right)^n$
- $\pi_j = 1 - \left(\frac{N-1}{N}\right)^n$
- $\pi_{ij} = 1 - \left(\frac{N-2}{N}\right)^n$

2.6 Estymatory

Estymator

Definicja
funkcja $\hat{T} : \mathcal{S} \to R$

Charakterystyki

- wartość oczekiwana $E_\pi \hat{T}$
- wariancja $D^2_\pi \hat{T}$
- błąd średniokwadratowy $E_\pi(\hat{T} - T)^2 = D^2_\pi \hat{T} + \left(E_\pi \hat{T} - T\right)^2$

3 Estymacja wartości średniej

3.1 Estymacja wartości średniej cechy

Średnia wartość cechy

Średnia wartość cechy

$$\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i$$

Wartość globalna cechy

$$Y = \sum_{i=1}^{N} Y_i$$

Średnia próbkowa

Definicja

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

Własności

- $E\bar{y} = \frac{1}{n}\sum_{i=1}^{N} \pi_i Y_i$
- $D^2\bar{y} = \frac{1}{n^2}\left[\sum_{i=1}^{N} \pi_i(1-\pi_i)Y_i^2 + 2\sum_{i\neq j}(\pi_{ij} - \pi_i\pi_j)Y_iY_j\right]$

Estymator Horvitza-Thompsona

Definicja

$$\bar{y}_{HT} = \frac{1}{N}\sum_{i=1}^{n} \frac{y_i}{\pi_i}$$

Własności

- $E\bar{y}_{HT} = \bar{Y}$
- $D^2\bar{y}_{HT} = \frac{1}{2N^2}\sum_{i\neq j}(\pi_i\pi_j - \pi_{ij})\left(\frac{Y_i}{\pi_i} - \frac{Y_j}{\pi_j}\right)^2$

Estymator Hansena-Horvitza

Definicja

$$\bar{y}_{HH} = \frac{\bar{X}}{n} \sum_{i=1}^{n} \frac{y_i}{x_i}$$

Własności

- $E\bar{y}_{HH} = \bar{Y}$
- $D^2\bar{y}_{HH} = \frac{1}{n} \sum_{i=1}^{n} \frac{X_i}{\bar{X}} \left(\frac{\bar{X}}{X_i} \frac{Y_i}{N} - \bar{Y} \right)^2$

3.2 Losowanie proste bez zwracania

Losowanie proste bez zwracania

Średnia próbkowa

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$$

Własności

- $E\bar{y} = \bar{Y}$
- $D^2\bar{y} = \left(1 - \frac{n}{N}\right) \frac{S^2}{n}$

Schemat lpbz - minimalna liczność próby

Kryterium

$$P\left\{|t_n - T| < d\right\} \geq 1 - \alpha$$

CTG (N „duże”, n „duże”)

$$\bar{y} \text{ ma w przybliżeniu rozkład } N\left(\bar{Y}, \left(1 - \frac{n}{N}\right) \frac{S^2}{n}\right)$$

$$\frac{\bar{y} - \bar{Y}}{\sqrt{\left(1 - \frac{n}{N}\right) \frac{S^2}{n}}} \text{ ma w przybliżeniu rozkład } N(0,1)$$

Schemat lpbz - minimalna liczność próby

Rozwiązanie

$$P\left\{\left|\bar{y}-\bar{Y}\right|<d\right\}=$$

$$P\left\{-\frac{d}{\sqrt{\left(1-\frac{n}{N}\right)\frac{S^2}{n}}}<\frac{\bar{y}-\bar{Y}}{\sqrt{\left(1-\frac{n}{N}\right)\frac{S^2}{n}}}<\frac{d}{\sqrt{\left(1-\frac{n}{N}\right)\frac{S^2}{n}}}\right\}=$$

$$2\Phi\left(\frac{d}{\sqrt{\left(1-\frac{n}{N}\right)\frac{S^2}{n}}}\right)-1\geq 1-\alpha$$

Schemat lpbz - minimalna liczność próby

Rozwiązanie

$$\Phi\left(\frac{d}{\sqrt{\left(1-\frac{n}{N}\right)\frac{S^2}{n}}}\right)\geq 1-\frac{\alpha}{2}$$

$$n^*\geq\frac{Nu_{1-\alpha/2}^2S^2}{Nd^2+u_{1-\alpha/2}^2S^2}$$

$$n\geq\lfloor n^*\rfloor+1$$

Schemat lpbz - minimalna liczność próby

Populacja $N=1000000$**; liczność próby?**

d/S	$\alpha=0.10$	$\alpha=0.05$	$\alpha=0.01$
0.01	26343	36994	62221
0.05	1082	1535	2647
0.10	271	384	664
0.20	68	97	166

Schemat lpbz - zadana liczność próby

Rozwiązanie

$$\frac{d}{S}=u_{1-\alpha/2}\sqrt{\frac{1}{n}-\frac{1}{N}}$$

Populacja $N = 1000000$; dokładność estymacji?

n	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
100	164.49%	196.00%	257.58%
500	73.56%	87.65%	115.19%
1000	52.01%	61.98%	81.45%
1500	42.47%	50.61%	66.51%

Schemat lpbz - estymatory złożone

Estymator iloczynowy

$$\bar{y}_p = \frac{\bar{x}\bar{y}}{\bar{X}}$$

Schemat lpbz - estymatory złożone

Estymator iloczynowy: własności

- $E(\bar{y}_p) = \bar{Y} + \left(1 - \frac{n}{N}\right) \frac{S_{xy}}{n\bar{X}} + O(n^{-2})$
- $MSE(\bar{y}_p) = \left(1 - \frac{n}{N}\right) \frac{S_y^2 + 2RS_{xy} + R^2 S_{xy}^2}{n} + O(n^{-2})$
- $S_{xy} = \frac{1}{N-1} \sum_{j=1}^{N} (X_j - \bar{X})(Y_j - \bar{Y}) \quad R = \frac{\bar{Y}}{\bar{X}}$

Schemat lpbz - estymatory złożone

Estymator ilorazowy

$$\bar{y}_q = \frac{\bar{y}}{\bar{x}} \bar{X}$$

Schemat lpbz - estymatory złożone

Estymator ilorazowy: własności

- $E(\bar{y}_q) = \bar{Y} + \left(1 - \frac{n}{N}\right) \frac{RS_x^2 - S_{xy}}{n\bar{X}} + O(n^{-2})$
- $MSE(\bar{y}_p) = \left(1 - \frac{n}{N}\right) \frac{S_y^2 - 2RS_{xy} + R^2 S_{xy}^2}{n} + O(n^{-2})$

- $S_{xy} = \dfrac{1}{N-1}\sum_{j=1}^{N}(X_j - \bar{X})(Y_j - \bar{Y}) \quad R = \dfrac{\bar{Y}}{\bar{X}}$

Schemat lpbz - estymatory złożone

Estymator regresyjny

$$\bar{y}_{lr} = \bar{y} + b(\bar{X} - \bar{x}); \qquad b = \frac{s_{xy}}{s_x^2}$$

Schemat lpbz - estymatory złożone

Estymator regresyjny: własności

- $E(\bar{y}_{lr}) = \bar{Y} + \left(1 - \dfrac{n}{N}\right)\dfrac{B\sum_{j=1}^{N}(Y_j - \bar{Y})^3 - \sum_{j=1}^{N}(X_j - \bar{X})^2(Y_j - \bar{Y})}{(n-1)(N-1)S_x^2} + O(n^{-2})$
- $MSE(\bar{y}_{lr}) = \left(1 - \dfrac{n}{N}\right)\dfrac{S_y^2(1 - \varrho_{xy}^2)}{n} + O(n^{-2})$
- $\varrho_{xy} = \dfrac{S_{xy}}{S_x S_y} = B\dfrac{S_x}{S_y}$

Schemat lpbz - porównanie strategii

$(lpbz, \bar{y})$ **efektywniejsza niż** $(lpbz, \bar{y}_p)$

$$D^2(\bar{y}) < MSE(\bar{y}_p)$$

$$\Updownarrow$$

$$S_y^2 < S_y^2 + 2RS_{xy} + R^2 S_x^2$$

$$\Updownarrow$$

$$\varrho_{xy} > -\frac{V_x}{2V_y}$$

Schemat lpbz - porównanie strategii

$(lpbz, \bar{y})$ **efektywniejsza niż** $(lpbz, \bar{y}_q)$

$$D^2(\bar{y}) < MSE(\bar{y}_q)$$
$$\Updownarrow$$
$$S_y^2 < S_y^2 - 2RS_{xy} + R^2 S_x^2$$
$$\Updownarrow$$
$$\varrho_{xy} < \frac{V_x}{2V_y}$$

Schemat lpbz - porównanie strategii

$(lpbz, \bar{y}_{lr})$ **efektywniejsza niż** $(lpbz, \bar{y})$

$$MSE(\bar{y}_{lr}) < D^2(\bar{y})$$
$$\Updownarrow$$
$$S_y^2(1 - \varrho_{xy}^2) < S_y^2$$
$$\Updownarrow$$
$$\varrho_{xy} \neq 0$$

Schemat lpbz - porównanie strategii

$(lpbz, \bar{y}_{lr})$ **efektywniejsza niż** $(lpbz, \bar{y}_p)$

$$MSE(\bar{y}_{lr}) < MSE(\bar{y}_p)$$
$$\Updownarrow$$
$$S_y^2(1 - \varrho_{xy}^2) < S_y^2 + 2RS_{xy} + R^2 S_x^2$$
$$\Updownarrow$$
$$R \neq -\frac{S_{xy}}{S_x^2}$$

Schemat lpbz - porównanie strategii

$(lpbz, \bar{y}_{lr})$ **efektywniejsza niż** $(lpbz, \bar{y}_q)$

$$MSE(\bar{y}_{lr}) < MSE(\bar{y}_q)$$

$$\Updownarrow$$

$$S_y^2(1 - \varrho_{xy}^2) < S_y^2 - 2RS_{xy} + R^2 S_x^2$$

$$\Updownarrow$$

$$R \neq \frac{S_{xy}}{S_x^2}$$

Schemat lpbz - porównanie strategii

$(lpbz, \bar{y}_q)$ **efektywniejsza niż** $(lpbz, \bar{y}_p)$

$$MSE(\bar{y}_q) < MSE(\bar{y}_p)$$

$$\Updownarrow$$

$$S_y^2 + 2Rs_{xy} + R^2 S_x^2 < S_y^2 - 2Rs_{xy} + R^2 S_x^2$$

$$\Updownarrow$$

$$\varrho_{xy} > 0$$

3.3 Losowanie warstwowe

Wprowadzenie

Warstwy

$$\mathcal{U} = \bigcup_{i=1}^{M} \mathcal{U}_h, \qquad \mathcal{U}_g \cap \mathcal{U}_h = \emptyset \text{ dla } g \neq h$$

$$\mathcal{U}_h = \{u_{h1}, \ldots, u_{hN_h}\}$$

$$N_1 + \cdots + N_M = N$$

Wprowadzenie

Badana cecha

$$\bar{Y}_h = \frac{1}{N_h} \sum_{j=1}^{N_h} Y_{hj}$$

$$S_h^2 = \frac{1}{N_h - 1} \sum_{j=1}^{N_h} (Y_{hj} - \bar{Y}_h)^2$$

Wprowadzenie

Badana cecha

$$\bar{Y} = \frac{1}{N} \sum_{h=1}^{M} \sum_{j=1}^{N_h} Y_{hj} = \sum_{h=1}^{M} W_h \bar{Y}_h$$

$$S^2 = \frac{1}{N - 1} \sum_{h=1}^{M} \sum_{j=1}^{N_h} (Y_{hj} - \bar{Y})^2$$

Losowanie warstwowe

Schemat $lw : lpbz$

- Dla każdego h losowana jest próba o liczebności n_h według schematu $lpbz$
- losowania z warstw są niezależne od siebie
- całkowita liczebność próby

$$n = \sum_{h=1}^{M} n_h$$

Estymator warstwowy

Estymator warstwowy

$$\bar{y}_w = \sum_{h=1}^{M} W_h \bar{y}_h = \frac{1}{N} \sum_{h=1}^{M} \frac{N_h}{n_h} \sum_{i=1}^{n_h} y_{hi}$$

Własności

- $E\bar{y}_w = \bar{Y}$
- $D^2(\bar{y}_w) = \sum_{h=1}^{M} W_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{S_h^2}{n_h}$
- $S^2 = \sum_{h=1}^{M} W_h S_h^2 + \sum_{h=1}^{M} W_h (\bar{Y}_h - \bar{Y})^2$

Estymator warstwowy

Losowanie proporcjonalne $lwp : lpbz$

- $n_h = nW_h = n\frac{N_h}{N}$
- $\bar{y}_w = \bar{y}$
- $D^2_{prop}(\bar{y}_w) = \left(1 - \frac{n}{N}\right) \frac{1}{n} \sum_{h=1}^{M} W_h S_h^2$

Alokacja

Optymalna alokacja

- koszt: $K = k_0 + \sum_{h=1}^{M} k_h n_h$
- $D^2(\bar{y}_w) = \sum_{h=1}^{M} W_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{S_h^2}{n_h}$

Alokacja

Optymalna alokacja
funkcja Lagrange'a:

$$\mathcal{L}(n_1, \ldots, n_h, \lambda) = \sum_{h=1}^{M} \frac{W_h^2 S_h^2}{n_h} - \frac{1}{N} \sum_{h=1}^{M} W_h S_h^2 - \lambda \left(K - k_0 - \sum_{h=1}^{M} n_h k_h \right)$$

$$\frac{\partial \mathcal{L}}{\partial n_h} = 0 \Rightarrow n_h^* = \frac{W_h S_h}{\sqrt{\lambda k_h}}$$

Alokacja

Optymalna alokacja

$$\frac{1}{\sqrt{\lambda}} = (K - k_0) \left(\sum_{h=1}^{M} \sqrt{k_h} W_h S_h \right)^{-1}$$

$$n_h^* = \frac{W_h S_h (K - k_0)}{\sqrt{k_h} \sum_{h=1}^{M} \sqrt{k_h} W_h S_h}$$

3.4 Losowanie dwustopniowe

Wprowadzenie

Warstwy

$$\mathcal{U} = \bigcup_{i=1}^{M} \mathcal{U}_h, \qquad \mathcal{U}_g \cap \mathcal{U}_h = \emptyset \text{ dla } g \neq h$$

$$\mathcal{U}_h = \{u_{h1}, \ldots, u_{hN_h}\}$$

$$N_1 + \cdots + N_M = N$$

Wprowadzenie

Badana cecha

$$\bar{Y}_h = \frac{1}{N_h}\sum_{j=1}^{N_h} Y_{hj} \qquad \bar{Y} = \frac{1}{N}\sum_{h=1}^{M}\sum_{j=1}^{N_h} Y_{hj} = \sum_{h=1}^{M} W_h \bar{Y}_h$$

$$S_h^2 = \frac{1}{N_h - 1}\sum_{j=1}^{N_h}(Y_{hj} - \bar{Y}_h)^2$$

$$S^2 = \frac{1}{N-1}\sum_{h=1}^{M}\sum_{j=1}^{N_h}(Y_{hj} - \bar{Y})^2$$

Wprowadzenie

Schemat $l2s : I - lpbz, II - lpbz$

- Krok 1: losujemy $2 < m < M$ warstw według schematu $lpbz$
- krok 2: z wylosowanych warstw losujemy próby według schematu $lpbz$
- całkowita liczebność próby

$$n = \sum_{h=1}^{m} n_{(h)}$$

Estymator dwustopniowy

Estymator dwustopniowy

$$\bar{y}_{(2)} = \frac{M}{m}\sum_{h=1}^{m} W_{(h)}\bar{y}_{(h)}$$

Własności

- $E\bar{y}_{(2)} = Y$
- $D^2(\bar{y}_{(2)}) = \frac{M^2}{m}\left[\left(1 - \frac{m}{M}\right) S_1^2 + \frac{1}{M}\sum_{h=1}^{M} W_h^2 \left(1 - \frac{n_h}{N_h}\right)\frac{S_h^2}{n_h}\right]$
- $S_1^2 = \frac{1}{M-1}\sum_{h=1}^{M}(Y_h - \bar{Y})^2$

Alokacja

Optymalna alokacja

- Znaleźć liczbę m losowanych warstw oraz frakcję f_2 losowanych obiektów z warstw
- Założenie: dla wszystkich warstw $\frac{n_h}{N_h} = f_2 = const$
- $D^2(\bar{y}_{(2)}) = \frac{M^2}{m}\left[\left(1 - \frac{m}{M}\right) S_1^2 + \frac{1-f_2}{f_2}\bar{N}S_2^2\right]$
- $S_2^2 = \frac{1}{N}\sum_{h=1}^{M} N_h S_{2h}^2; \quad \bar{N} = \frac{N}{M}$

Alokacja

Optymalna alokacja

- koszt: $\hat{K} = k_0 + mk_1 + k_2\sum_{h=1}^{m} n_{(h)}$
- $E(\hat{K}) = k_0 + m(k_1 + f_2\bar{N}k_2)$

Alokacja

Optymalna alokacja

- funkcja Lagrange'a:

$$\begin{aligned}&\mathcal{L}(m, f_2, \lambda) = \\ &\quad M^2\left[\left(\frac{1}{m} - \frac{1}{M}\right) S_1^2 + \frac{\bar{N}(1-f_2)}{mf_2}S_2^2\right] \\ &\quad + \lambda\left[(K - k_0) - mk_1 - mf_2\bar{N}k_2\right]\end{aligned}$$

Alokacja

Optymalna alokacja

$$\frac{\partial \mathcal{L}}{\partial m} = -M^2 \left[\frac{S_1^2}{m^2} + \frac{\bar{N}(1-f_2)}{f_2 m^2} S_2^2 \right] - \lambda(k_1 + f_2 \bar{N} k_2)$$

$$\frac{\partial \mathcal{L}}{\partial f_2} = -\frac{m^2 \bar{N} S_2^2}{m f_2^2} - \lambda m \bar{N} k_2$$

$$\frac{\partial \mathcal{L}}{\partial m} = 0 \quad \frac{\partial \mathcal{L}}{\partial f_2} = 0$$

Alokacja

Optymalna alokacja

$$\lambda m \bar{N} k_2 = -\frac{M^2 \bar{N} S_2^2}{m f_2^2}$$

$$\lambda(k_1 + f_2 \bar{N} k_2) = -\frac{M^2}{m^2} \left[S_1^2 + \bar{N} \frac{(1-f_2)}{f_2} S_2^2 \right]$$

$$\lambda = -\frac{M^2 S_2^2}{k_2 m^2 f_2^2}$$

Alokacja

Optymalna alokacja

$$f_2 = \sqrt{\frac{k_1 S_2^2}{k_2(S_1^2 - \bar{N} S_2^2)}}$$

$$m = \frac{K - k_0}{k_1 + k_2 \bar{N} f_2}$$