

## Estymacja frakcji elementów z cechą wyróżnioną w populacji na podstawie próby nieprostej z niepełną realizacją

Populację ludności dorosłej kraju podzielono na warstwy z uwzględnieniem podziału na województwa i klasy miejscowości. Wybór próby z warstw był dwustopniowy. Jednostkami losowania stopnia pierwszego były niewielkie obszary gmin miejskich i wiejskich o zróżnicowanej liczbie ludności, z których następnie wylosowano osoby dorosłe.

Alokacja próby jednostek stopnia pierwszego w warstwach była proporcjonalna do liczby ludności dorosłej w warstwach.

Do losowania jednostek stopnia pierwszego zastosowano schemat Hartley'a-Rao. Z każdej warstwy wylosowano nie mniej niż dwie jednostki stopnia pierwszego. Jednostki stopnia drugiego czyli osoby dorosłe wylosowano z jednostek stopnia pierwszego według schematu bez zwracania. Z każdej jednostki stopnia pierwszego wylosowano do badania ankietowego 6 osób.

Szacowany parametr to frakcja elementów z cechą wyróżnioną w populacji, czyli iloraz wartości globalnych cech  $X$  i  $Y$  o wartościach 1 lub 0, to jest

$$R = \frac{X}{Y},$$

gdzie

$$X = \sum_{i=1}^N X_i, \quad X_i = \begin{cases} 1 & \text{dla elementów } u_i \in U \text{ posiadających cechę wyróżnioną} \\ 0 & \text{dla pozostałych } u_i \end{cases}$$

$$Y = \sum_{i=1}^N Y_i = N, \quad Y_i = \begin{cases} 1 & \text{dla elementów } u_i \in U \\ 0 & \text{dla pozostałych } u_i \end{cases}$$

$u_i$  — element skończonej populacji  $U$ ,  $i = 1, 2, \dots, N$ ,

$U$  — oznaczenie populacji,  $U = \{u_1, u_2, \dots, u_N\}$ ,

$N$  — ogółem liczebność populacji.

Estymatorem parametru  $R$  jest statystyka

$$r = \frac{x}{y},$$

gdzie

$s$  — próba zrealizowana z populacji  $U$  ( $s \subset s^* \subset U$ ) o liczebności  $n < (0.7n^*)$ ,

$s^*$  — próba wylosowana z populacji  $U$ , o liczebności  $n^*$ ,

$$x = \sum_{i=1}^n w_i x_i, \quad y = \sum_{i=1}^n w_i y_i,$$

$$x = X_i \text{ dla } u_i \in s, \quad y = Y_i \text{ dla } u_i \in s \quad (s \subset s^* \subset U),$$

$w_i$  — wagi dla  $u_i \in s$ .

Przy wyznaczaniu wag  $w_i$  uwzględniono:

- prawdopodobieństwa wyboru jednostek,
- wskaźniki realizacji w warstwach,
- dane GUS dla skorygowania proporcji cech demograficzno-społecznych z zastosowaniem stratyfikacji ex post.

Procedura estymacji jest złożona i doprowadza do skomplikowanych wzorów. Z tego powodu dla oceny wariancji estymatora stosowane są metody uproszczone bazujące na metodach resamplingu.

#### Literatura

- [1] Cz. Bracha, *Teoretyczne podstawy metody reprezentacyjnej*, Wydawnictwo Naukowe PWN, Warszawa 1996.
- [2] Cz. Bracha, *Metoda reprezentacyjna w badaniu opinii publicznej i marketingu*, Efekt, Warszawa 1998.
- [3] B. Efron, R. J. Tibshirani, *An Introduction to the Bootstrap*, Monographs on Statistics and Applied Probability 57, Chapman and Hall, New York–London 1993.
- [4] M. H. Hunsen, W. N. Hurwitz, W. G. Madow, *Sample Survey Methods and Theory*. John Wiley & Sons, Inc., New York–Chichester–Brisbane–Toronto–Singapore 1993.
- [5] Cz. Bracha, B. Lednicki, R. Wieczorkowski, *Wykorzystanie złożonych metod estymacji do dezagregacji danych z badań aktywności ekonomicznej ludności w roku 2003*, ZBSE, GUS, Warszawa 2004.
- [6] P. Tryfos, *Sampling methods for Applied Research. Text and Cases*, John Wiley & Sons, Inc., New York 1996.