Improved Estimation of Population Mean in Stratified Random Sampling Using Information on Auxiliary Attribute

Nursel Koyuncu*
Hacettepe University, Ankara, Turkey nkoyuncu@hacettepe.edu.tr

Abstract
In the survey sampling literature, using auxiliary information increase the efficiency of estimators. Sometimes we can get auxiliary information as qualitative form. In this study, by taking the advantage of correlation between the study variable and auxiliary attribute, we have improved the estimation of population mean in stratified random sampling. Under this sampling scheme, the expressions for bias and mean square error have been obtained. To show the efficiency of proposed estimator over the existing estimator, we have made a comparative study using a real data set.

Keywords: efficiency, mean square error, ratio estimator, stratified random sampling.

1. Introduction

Let the population of size, *N*, is stratified into *L* strata with *h*-th stratum containing N_h units, where $h=1,2,...,L$ such that $\sum_{h=1}^{L} N_h = N$. A simple random sample of size n_h is drawn without replacement from the *h*-th stratum such that $\sum_{h=1}^{L} n_h = n$. Let y_{hi} and φ_{hi} denote observed values of study variable *Y* and auxiliary attribute φ, respectively, on the *i*-th unit of the *h*-th stratum, where $i=1,2,...,N_h$ and $h=1,2,...,L$. Assume that the presence or absence of an attribute φ_h introduces a complete dichotomy into the population, and assume that the attribute φ_h takes only two values, 1 or 0 as

$\varphi_{hi}=1$, if the ith unit of the population possesses attribute φ

$\varphi_{hi}=0$, otherwise.

Let $A=\sum_{i=1}^{N}\varphi_i$, $A_h=\sum_{i=1}^{N_h}\varphi_{hi}$, $a=\sum_{i=1}^{n}\varphi_i$ and $a_h=\sum_{i=1}^{n_h}\varphi_{hi}$ denote, respectively, the total number of units in the population, population stratum *h*, sample and sample stratum *h* possessing attribute φ. Let the corresponding population, *h*-th population stratum and *h*-th sample stratum proportions be $P=\frac{A}{N}$, $P_h=\frac{A_h}{N_h}$ and $p_h=\frac{a_h}{n_h}$.

Moreover, assume that $\bar{y}_h=\sum_{i=1}^{n_h}\frac{y_{hi}}{n_h}$, $\bar{y}_{st}=\sum_{h=1}^{L}W_h\bar{y}_h$, and $\bar{Y}_h=\sum_{i=1}^{N_h}\frac{Y_{hi}}{N_h}$, $\bar{Y}=\sum_{h=1}^{L}W_h\bar{Y}_h$ be the sample and population means of *Y*, respectively, where $W_h=\frac{N_h}{N}$ is the stratum weight. To obtain the MSE, let us define $e_0=\left(\bar{y}_{st}-\bar{Y}\right)/\bar{Y}$,

$e_1 = \left(p_{st} - P \right) / P$ and using these notations,

$$E\left(e_0\right) = E\left(e_1\right) = 0\,,$$

$$V_{rst} = \sum_{h=1}^{L} W_h^{r+s} \frac{E\left[\left(\bar{y}_h - \bar{Y}_h \right)^r \left(p_h - P_h \right)^s \right]}{\bar{Y}^r P^s} \qquad (1.1)$$

Using (1), we can write

$$E\left(e_0^2\right) = \frac{\sum_{h=1}^{L} W_h^2 \lambda_h S_{yh}^2}{\bar{Y}^2} = V_{20}\,, \quad E\left(e_1^2\right) = \frac{\sum_{h=1}^{L} W_h^2 \lambda_h S_{\varphi h}^2}{P^2} = V_{02}\,,$$

$$E\left(e_0 e_1\right) = \frac{\sum_{h=1}^{L} W_h^2 \lambda_h S_{y\varphi h}}{\bar{Y} P} = V_{11}\,,$$

where $S_{yh}^2 = \dfrac{\sum_{i=1}^{N_h} \left(y_{hi} - \bar{Y}_h \right)^2}{N_h - 1}$, $S_{y\varphi h} = \dfrac{\sum_{i=1}^{N_h} \left(y_{hi} - \bar{Y}_h \right) \left(p_{hi} - P_h \right)}{N_h - 1}$, $\lambda_h = \dfrac{1 - f_h}{n_h}$, and $f_h = \dfrac{n_h}{N_h}$. (see Koyuncu and Kadılar (2009))

2. Suggested Estimator

The classical combined ratio estimator of population mean using auxiliary attribute is given by

$$\bar{y}_{rst} = \frac{\bar{y}_{st}}{p_{st}} P \qquad (2.1)$$

The bias and mean square error are respectively given by

$$Bias\left(\bar{y}_{rst}\right) = \bar{Y}\left(V_{02} - V_{11}\right) \qquad (2.2)$$

$$MSE\left(\bar{y}_{rst}\right) = \bar{Y}^2\left(V_{02} + V_{20} - 2V_{11}\right) \qquad (2.3)$$

Elfattah et al. (2010) and Koyuncu (2012) have proposed some estimators using auxiliary attribute in simple random sampling. In this study we try to improve estimators in stratified random sampling. Following Koyuncu and Kadilar (2010) we propose following estimator of population mean using information of auxiliary attribute :

$$\bar{y}_{Nst} = \left[w_1 \bar{y}_{st} + w_2 \left(P - p_{st} \right) \right] \left(\frac{a_{st} P + b_{st}}{a_{st} p_{st} + b_{st}} \right) \qquad (2.4)$$

where w_1 and w_2 are suitable constants, a_{st} and b_{st} are either real numbers or the functions of the known parameters for h-th stratum of the auxiliary attribute, φ, such as standard deviation $S_{\varphi(st)} = \sum_{h=1}^{L} W_h \sigma_{\varphi h}$, coefficient of variation $C_{\varphi(st)} = \sum_{h=1}^{L} W_h C_{\varphi h}$, skewness $\beta_{1(\varphi)st} = \sum_{h=1}^{L} W_h \beta_{1h}(\varphi)$, kurtosis $\beta_{2(\varphi)st} = \sum_{h=1}^{L} W_h \beta_{2h}(\varphi)$ and correlation coefficient $\rho_{(\varphi y)st} = \sum_{h=1}^{L} W_h \rho_{(\varphi y)h}$. We can generate many new estimators using suitable values of a_{st} and b_{st} as given in Table 1. Using e terms we can rewrite as

$$\bar{y}_{Nst} = \left[w_1 \bar{Y}(1+e_0) - w_2 P e_1 \right](1+\tau_{st} e_1)^{-1} \tag{2.5}$$

where $\tau_{st} = \dfrac{a_{st} P}{a_{st} P + b_{st}}$. Up to the first order of approximation we have

$$\begin{aligned} \bar{y}_{Nst} - \bar{Y} = \Big[&(w_1 - 1)\bar{Y} + w_1 \bar{Y}\left(e_0 - \tau_{st} e_1 - \tau_{st} e_0 e_1 + \tau_{st}^2 e_1^2\right) \\ &- w_2 P\left(e_1 - \tau_{st} e_1^2\right) \Big] \end{aligned} \tag{2.6}$$

Taking expectation of (2.6) we get the bias of proposed estimator as

$$Bias(\bar{y}_{Nst}) = \left[(w_1 - 1)\bar{Y} + w_1 \bar{Y}\left(\tau_{st}^2 V_{02} - \tau_{st} V_{11}\right) + \tau_{st} w_2 P V_{02} \right] \tag{2.7}$$

Squaring and applying expectation, the mean square error of (2.4) is given by

$$\begin{aligned} &(\bar{y}_{Nst} - \bar{Y})^2 \\ &= \Big[(w_1 - 1)^2 \bar{Y}^2 + w_1^2 \bar{Y}^2 \left(e_0^2 + 3\tau^2 e_1^2 - 4\tau e_0 e_1 + 2e_0 - 2\tau e_1\right) \\ &+ w_2^2 P^2 e_1^2 - 2 w_1 \bar{Y}^2 \left(e_0 - \tau e_1 - \tau e_0 e_1 + \tau^2 e_1^2\right) \\ &- 2\bar{Y} P w_2 w_1 \left(e_1 - 2\tau e_1^2 + e_0 e_1\right) + 2\bar{Y} P w_2 \left(e_1 - \tau e_1^2\right) \Big] \end{aligned} \tag{2.8}$$

$$\begin{aligned} MSE(\bar{y}_{Nst}) = \Big[&(w_1 - 1)^2 \bar{Y}^2 + w_1^2 \bar{Y}^2 \left(V_{20} + 3\tau^2 V_{02} - 4\tau V_{11}\right) \\ &+ w_2^2 P^2 V_{02} - 2 w_1 \bar{Y}^2 \left(\tau^2 V_{02} - \tau V_{11}\right) \\ &- 2\bar{Y} P w_2 w_1 \left(-2\tau V_{02} + V_{11}\right) - 2 w_2 \tau \bar{Y} P V_{02} \Big] \end{aligned} \tag{2.9}$$

The optimum values of w_1 and w_2 are obtained by minimizing (2.9)

$$w_1^* = \frac{V_{02}\left(1-\tau^2 V_{02}\right)}{\left(V_{02}+V_{02}V_{20}-\tau^2 V_{02}^2 - V_{11}^2\right)} \tag{2.12}$$

$$w_2^* = \frac{\bar{Y}}{P}\left(\tau + \frac{1-\tau^2 V_{02}\left(V_{11}-2\tau V_{02}\right)}{\left(V_{02}+V_{02}V_{20}-\tau^2 V_{02}^2 - V_{11}^2\right)}\right) \tag{2.13}$$

Using optimum values in (2.9) the minimum mean square error of proposed estimator is obtained as:

$$MSE_{\min}\left(\bar{y}_{Nst}\right) = \bar{Y}^2\left(1-\frac{\left(V_{02}-\tau^2 V_{02}V_{11}^2-\tau^2 V_{02}^2+\tau^2 V_{02}^2 V_{20}\right)}{\left(V_{02}+V_{02}V_{20}-\tau^2 V_{02}^2 - V_{11}^2\right)}\right) \tag{2.14}$$

3. Emprical Study

To illustrate the efficiency of suggested estimators in the application, we consider the data concerning the number of teachers as the study variable (*y*) and for auxiliary attribute we use number of students classifying more or less than 750, in both primary and secondary schools as auxiliary variable for 923 districts at 6 regions (as 1:Marmara 2:Agean 3:Mediterranean 4:Central Anatolia 5:Black Sea 6:East and Southeast Anatolia) in Turkey in 2007 (source: The Turkish Republic Ministry of Education). The summary statistics of the data are given in Table 2. We used Neyman allocation for allocating the samples to different strata (Cochran,1977).

The *MSE* values of the stratified suggested estimators $\bar{y}_{Nst(1)}$ - $\bar{y}_{Nst(4)}$ and classical combined ratio estimator $\bar{y}_{rst}$ have been obtained. These values are given in Table 3. From Table 3, we observe that $\bar{y}_{Nst(4)}$ is the most efficient estimator for this data set.

4. Conclusion

In this study we proposed a family of estimator of population mean using information of auxiliary attribute such as coefficient of variation, skewness, kurtosis etc. We showed that using different auxiliary information affects the efficiency. It was found that the suggested estimators are more efficient than existing estimator.

References

Abd-Elfattah, A.M., El-Sherpieny, E.A., Mohamed, S.M., Abdou, O.F. (2010)"Improvement in estimating the population mean in simple random sampling using information on auxiliary attribute," *Applied Mathematics and Computation*, 215, 4198-4202.

Koyuncu N. (2012)"Efficient estimators of population mean using auxiliary attributes," *Applied Mathematics Computation*, 218, 10900-10905.

Koyuncu, N., Kadılar, C. (2010)"On Improvement in Estimating Population Mean in Stratified Random Sampling," *Journal of Applied Statistics*, 37, 6, 999-1013.

Koyuncu, N., Kadılar, C. (2009) "Family of estimators of population mean using two auxiliary variables in stratified random sampling," *Communications in Statistics: Theory and Methods*, 38, 14, 2398-2417.

Table 1:Some members of the family of estimators of $\bar{y}_{Nst}$

Estimator	a_{st}	b_{st}
$\bar{y}_{Nst(1)}$	1	$\beta_{2(\varphi)st}=\sum_{h=1}^{L}W_h\beta_{2h}(\varphi)$
$\bar{y}_{Nst(2)}$	1	$C_{\varphi(st)}=\sum_{h=1}^{L}W_hC_{\varphi h}$
$\bar{y}_{Nst(3)}$	$C_{\varphi(st)}=\sum_{h=1}^{L}W_hC_{\varphi h}$	$\beta_{2(\varphi)st}=\sum_{h=1}^{L}W_h\beta_{2h}(\varphi)$
$\bar{y}_{Nst(4)}$	$\beta_{2(\varphi)st}=\sum_{h=1}^{L}W_h\beta_{2h}(\varphi)$	$C_{\varphi(st)}=\sum_{h=1}^{L}W_hC_{\varphi h}$

Table 2:Data Statistics

N_1 =127	N_2 =117	N_3 =103
N_4 =170	N_5 =205	N_6 =201
n_1 =31	n_2 =21	n_3 =29
n_4 =38	n_5 =22	n_6 =39
S_{y1} =883.835	S_{y2} =644.922	S_{y3} =1033.467
S_{y4} =810.585	S_{y5} =403.654	S_{y6} =711.723
$\overline{Y}_1$ =703.74	$\overline{Y}_2$ =413	$\overline{Y}_3$ =573.17
$\overline{Y}_4$ =424.66	$\overline{Y}_5$ =267.03	$\overline{Y}_6$ =393.84
$S_{\varphi1}$ =0.213	$S_{\varphi2}$ =0.159	$S_{\varphi3}$ =0.253
$S_{\varphi4}$ =0.316	$S_{\varphi5}$ =0.284	$S_{\varphi6}$ =0.218
p_1 =0.952	p_2 =0.974	p_3 =0.932
p_4 =0.888	p_5 =0.912	p_6 =0.950
S_{xy1} =25.267	S_{xy2} =9.982	S_{xy3} =37.453
S_{xy4} =44.625	S_{xy5} =21.04	S_{xy6} =18.66
$\beta_2(\varphi_1)$ =16.922	$\beta_2(\varphi_2)$ =35.579	$\beta_2(\varphi_3)$ =10.34
$\beta_2(\varphi_4)$ =4.231	$\beta_2(\varphi_5)$ =6.675	$\beta_2(\varphi_6)$ =15.56

Table 3: MSE Values of Estimators

Estimator	MSE
$\overline{y}_{rst}$	2189.222
$\overline{y}_{Nst(1)}$	2160.101
$\overline{y}_{Nst(2)}$	2074.195
$\overline{y}_{Nst(3)}$	2160.64
$\overline{y}_{Nst(4)}$	2023.49