Improvement in Estimating the Population Mean in Simple Random Sampling using Coefficient of Skewness of Auxiliary Attribute

Tolga ZAMAN

Çankırı Karatekin University, Faculty of Science, Department of Statistics, 18100, Çankırı
(ORCID: https://orcid.org/0000-0001-8780-3655)

Abstract: This paper suggested a new family of estimators for the population mean in the simple random sampling using the information of an auxiliary attribute. Theoretically, the mean square error (MSE) equations were obtained and it was shown that all the suggested ratio estimators are more efficient than some known estimators. These results were also supported by two original data sets.

Yardımcı Özniteliğin Çarpıklık Katsayısı Kullanılarak Basit Rastgele Örneklemede Kitle Ortalamasının Tahmin Edilmesindeki İyileşme

Özet: Bu makale, basit rasgele örneklemde yardımcı öznitelik bilgisini kullanarak popülasyon ortalaması için yeni bir tahmin edici ailesi önermiştir. Teorik olarak, önerilen tüm tahmin ediciler için hata kareler ortalaması (HKO) denklemleri elde edildi ve önerilen tüm oran tahmin edicilerinin bazı bilinen tahmin edicilerden daha etkin olduğu gösterildi. Bu sonuçlar ayrıca iki orijinal veri kümesi tarafından desteklendi.

1. Introduction

When there are positive correlation between study variable y_i and auxiliary variable x_i in the simple random sampling method, ratio-type estimators are used to estimate population mean. In the sampling literature, one way to increase the efficiency of an estimator is to use auxiliary attributes. Many authors have suggested estimators based on auxiliary attributes. Zaman [1], Kadilar and Cingi [2], Naik and Gupta [3], Kadilar and Cingi [4], Shabbir and Gupta [5], Singh et al. [6], Abd-Elfattah et al. [7], Koyuncu [8], Malik and Singh [9], Zaman [10] have considered the problem of estimating population mean $\bar{Y}$ taking into consideration the correlation between the study variable and the auxiliary attribute.

Zaman [1] proposed ratio estimators in order to estimate population mean of study variable y, using information about population proportion possessing certain attributes in simple random sampling;

$$t_{pr1} = \frac{\bar{y} + b_\varphi(P-p)}{p + \beta_1(\varphi)}[P + \beta_1(\varphi)] \quad (1)$$

$$t_{pr2} = \frac{\bar{y} + b_\varphi(P-p)}{\beta_1(\varphi)p + \beta_2(\varphi)}[\beta_1(\varphi)P + \beta_2(\varphi)] \quad (2)$$

$$t_{pr3} = \frac{\bar{y} + b_\varphi(P-p)}{\beta_2(\varphi)p + \beta_1(\varphi)}[\beta_2(\varphi)P + \beta_1(\varphi)] \quad (3)$$

$$t_{pr4} = \frac{\bar{y} + b_\varphi(P-p)}{\beta_1(\varphi)p + C_p}[\beta_1(\varphi)P + C_p] \quad (4)$$

$$t_{pr5} = \frac{\bar{y} + b_\varphi(P-p)}{C_p p + \beta_1(\varphi)}[C_p P + \beta_1(\varphi)] \quad (5)$$

where C_p, $\beta_2(\varphi)$ and $\beta_1(\varphi)$ are the population coefficient of variation, the population coefficient of kurtosis of auxiliary attribute and the population coefficient of skewness of auxiliary attribute, respectively and $b_\varphi = \frac{s_{y\varphi}}{s_\varphi^2}$ is the regression coefficient. Here, s_φ^2 is the sample variance of auxiliary attribute and $s_{y\varphi}$ is the sample covariance between the auxiliary attribute and the study variable. Expressions for the MSE's of the suggested ratio-type estimators is as follows;

$$MSE(t_{pri}) \cong \frac{1-f}{n}\left[R_i^2 S_\varphi^2 + S_y^2(1-\rho_{pb}^2)\right], i = 1,2,\ldots,5 \quad (6)$$

$$R_1 = \frac{\bar{Y}}{P + \beta_1(\varphi)} \quad (7)$$

$$R_2 = \frac{\bar{Y}\beta_1(\varphi)}{P\beta_1(\varphi) + \beta_2(\varphi)} \quad (8)$$

Corresponding author: tolgazaman@karatekin.edu.tr

$$R_3 = \frac{\bar{Y}\beta_2(\varphi)}{P\beta_2(\varphi)+\beta_1(\varphi)} \qquad (9)$$

$$R_4 = \frac{\bar{Y}\beta_1(\varphi)}{P\beta_1(\varphi)+C_p} \qquad (10)$$

$$R_5 = \frac{\bar{Y}C_p}{PC_p+\beta_1(\varphi)}. \qquad (11)$$

Zaman [1] deduced that all estimators, given above, were more efficient than the sample mean, the ratio estimator suggested by Naik and Gupta [3], under certain restrictions. Moreover, these results were supported by the results of the original data sets which will also be used in this article.

In the next section, the novel ratio-type estimators are proposed by improving the ratio estimators presented in Zaman [1] by combining them. Then, the MSE expressions of these novel estimators are obtained. In section 4, in addition, comparisons are done among all the proposed estimators numerically. In the last section, conclusions are summarized based on the results of the paper.

2. Suggested Estimators

In this section, new estimators are proposed following the procedure presented in Kadilar and Cingi [2] combining ratio-type estimators between (1) and (5). The general form of the proposed estimators are as follows;

$$z_{pri} = \left[\begin{array}{c} \vartheta \frac{\bar{y}+b_\varphi(P-p)}{p+\beta_1(\varphi)}[P+\beta_1(\varphi)] \\ +(1-\vartheta)\frac{\bar{y}+b_\varphi(P-p)}{(pk+l)}(Pk+l); i \end{array}\right] = 2,3,4,5 \quad (12)$$

where ϑ is a real constant to be determined such that the MSE of zt_{pri} is minimum. $k \neq 0$ and l are either real number or the functions of known parameters such as C_p, $\beta_1(\varphi)$ and $\beta_2(\varphi)$, as (1)-(5)

Expressions for the MSE's of these estimators can be computed using the first-degree approximation in the Taylor series approach as Equation 12. In general, Taylor series method for k variables can be given as;

$$h(\bar{x}_1,\bar{x}_2,\dots,\bar{x}_k) = \left[\begin{array}{c} h(\bar{X}_1,\bar{X}_2,\dots,\bar{X}_k) \\ +\sum_{j=1}^k d_j(\bar{x}_j-\bar{X}_j)+R_k(\bar{X}_k,\alpha)+O_k \end{array}\right] \quad (13)$$

where

$$d_j = \frac{\partial h(\bar{x}_1,\bar{x}_2,\dots,\bar{x}_k)}{\partial \alpha_j} \qquad (14)$$

And

$$R_k(\bar{X}_k,\alpha) = \sum_{j=1}^k \sum_{j=1}^k \left[\begin{array}{c} \frac{1}{2!}\frac{\partial^2 h(\bar{X}_1,\bar{X}_2,\dots,\bar{X}_k)}{\partial \bar{X}_i \bar{X}_j} \\ *(\bar{x}_j-\bar{X}_j)(\bar{x}_i-\bar{X}_i)+O_k \end{array}\right] \quad (15)$$

where O_k represents the terms in the expansion of the Taylor series of more than the second degree [11].

When we omit the term $R_k(\bar{X}_k,\alpha)$, we obtain Taylor series method for two variables as follows;

$$h(p,\bar{y}) - h(P,\bar{Y}) \cong \left[\begin{array}{c} \frac{\partial h(c,d)}{\partial c}\Big|_{P,\bar{Y}}(p-P) \\ +\frac{\partial h(c,d)}{\partial d}\Big|_{\bar{Y},P}(\bar{y}-\bar{Y}) \end{array}\right] \quad (16)$$

where, $h(p,\bar{y}) = z_{pri}$ and $h(P,\bar{Y}) = \bar{Y}$. MSE equations of the proposed estimators compute as follows:

$$zt_{pri} - \bar{Y} \cong \left(\begin{array}{c} \left[\begin{array}{c} \frac{\partial\left(\vartheta\frac{\bar{y}+b_\varphi(P-p)}{p+\beta_1(\varphi)}(P+\beta_1(\varphi))\right)}{\partial p} \\ +\frac{\partial\left((1-\vartheta)\frac{\bar{y}+b_\varphi(P-p)}{(pk+l)}(Pk+l)\right)}{\partial p} \end{array}\right]\Big|_{P,\bar{Y}}(p-P) \\ +\left[\begin{array}{c} \frac{\partial\left(\vartheta\frac{\bar{y}+b_\varphi(P-p)}{p+\beta_1(\varphi)}(P+\beta_1(\varphi))\right)}{\partial \bar{y}} \\ +\frac{\partial\left((1-\vartheta)\frac{\bar{y}+b_\varphi(P-p)}{(pk+l)}(Pk+l)\right)}{\partial \bar{y}} \end{array}\right]\Big|_{P,\bar{Y}}(\bar{y}-\bar{Y}) \end{array}\right) \quad (17)$$

$$zt_{pri} - \bar{Y} \cong \left(\begin{array}{c} \left[\begin{array}{c} \vartheta\frac{-b_\varphi(p+\beta_1(\varphi))-\bar{y}}{(p+\beta_1(\varphi))^2}(P+\beta_1(\varphi)) \\ +(1-\vartheta)\frac{-b_\varphi(pk+l)-k\bar{y}}{(pk+l)^2}(Pk+l) \end{array}\right]_{P,\bar{Y}}(p-P) \\ +\left[\begin{array}{c} \vartheta\frac{p+\beta_1(\varphi)}{(p+\beta_1(\varphi))^2}(P+\beta_1(\varphi)) \\ +(1-\vartheta)\frac{pk+l}{(pk+l)^2}(Pk+l) \end{array}\right]_{\bar{Y},P}(\bar{y}-\bar{Y}) \end{array}\right) \quad (18)$$

$$zt_{pri} - \bar{Y} \cong \left[\begin{array}{c} \vartheta(-B_\varphi - R_1) \\ +(1-\vartheta)(-B_\varphi - R_i) \end{array}\right](p-P)+(\bar{y}-\bar{Y}) \quad (19)$$

If we take the square of each side and take the expected value is passed;

$$E(z_{pri}-\bar{Y})^2 \cong \begin{array}{c} \left[\left(\begin{array}{c} \vartheta(B_\varphi+R_1) \\ +(1-\vartheta)(B_\varphi+R_i) \end{array}\right)^2\right]V(p) \\ -2\vartheta(B_\varphi+R_1) \\ -(1-\vartheta)(B_\varphi+R_i)Cov(p,\bar{y}) \\ +V(\bar{y}) \end{array} \quad (20)$$

$$\gamma_i = \vartheta(B_\varphi+R_1)+(1-\vartheta)(B_\varphi+R_i) \qquad (21)$$

$$R_i = \frac{\bar{Y}k}{Pk+l} \qquad (22)$$

$$R_1 = \frac{\bar{Y}}{P+\beta_1(\varphi)} \qquad (23)$$

$$MSE(z_{pri}) \cong \frac{1-f}{n}[\gamma_i^2 S_p^2 - 2\gamma_i S_{yp} + S_y^2]; \ i=1,2,\dots,5 \quad (24)$$

The suggested estimator by combining the estimators presented in (1) and (2) is follows;

$$z_{pr1} = \begin{array}{c} \vartheta\frac{\bar{y}+b_\varphi(P-p)}{p+\beta_1(\varphi)}(P+\beta_1(\varphi)) \\ +(1-\vartheta)\frac{\bar{y}+b_\varphi(P-p)}{(\beta_1(\varphi)p+\beta_2(\varphi))}(\beta_1(\varphi)P+\beta_2(\varphi)) \end{array} \quad (25)$$

The MSE of this estimator is found as follows;

$$MSE(z_{pr1}) \cong \frac{1-f}{n}[\gamma_1^2 S_p^2 - 2\gamma_1 S_{yp} + S_y^2] \qquad (26)$$

where

$$\gamma_1 = \vartheta\left(B_\varphi + R_1\right) + (1 - \vartheta)\left(B_\varphi + R_2\right) \qquad (27)$$

The suggested estimator by combining the estimators presented in (1) and (3) is also as follows;

$$z_{pr2} = \begin{aligned} &\vartheta \frac{\bar{y} + b_\varphi(P-p)}{p + \beta_1(\varphi)}\left(P + \beta_1(\varphi)\right) \\ &+ (1 - \vartheta)\frac{\bar{y} + b_\varphi(P-p)}{(\beta_2(\varphi)p + \beta_1(\varphi))}\left(\beta_2(\varphi)P + \beta_1(\varphi)\right) \end{aligned} \qquad (28)$$

The mean square error of the estimator is the same as (26) but R_2 in (27) is replaced with R_3.

Moreover, the following estimator is suggested by combining ratio estimator given in (1) and (4),

$$z_{pr3} = \begin{aligned} &\vartheta \frac{\bar{y} + b_\varphi(P-p)}{p + \beta_1(\varphi)}\left(P + \beta_1(\varphi)\right) \\ &+ (1 - \vartheta)\frac{\bar{y} + b_\varphi(P-p)}{(\beta_1(\varphi)p + C_p)}\left(\beta_1(\varphi)P + C_p\right) \end{aligned} \qquad (29)$$

The mean square error of this estimator is again the same as (26) but R_2 in (27) is replaced with R_4

Finally, it is suggested the estimator by combining ratio estimators given in (1) and (5) is follows,

$$z_{pr4} = \begin{aligned} &\vartheta \frac{\bar{y} + b_\varphi(P-p)}{p + \beta_1(\varphi)}\left(P + \beta_1(\varphi)\right) \\ &+ (1 - \vartheta)\frac{\bar{y} + b_\varphi(P-p)}{(C_pP + \beta_1(\varphi))}\left(C_pP + \beta_1(\varphi)\right) \end{aligned} \qquad (30)$$

The mean square error of the estimator is also the same as (26) but R_2 in (27) is replaced with R_5.

The optimum value of ϑ to minimize (26) can easily be computed as follows;

$$\frac{\partial MSE(z_{pri})}{\partial \vartheta} = \frac{1-f}{n}\left(2\gamma_i\gamma_i'S_p^2 - 2\gamma_i'S_{yp}\right) = 0 \qquad (31)$$

$$\delta_i'\left(\delta_i S_p^2 - S_{yp}\right) = 0 \qquad (32)$$

$$(R_1 - R_i)\left[\left(\vartheta(R_1 - R_i) + B_\varphi + R_i\right)S_p^2 - S_{yp}\right] = 0 \quad (33)$$

$$\vartheta(R_1 - R_i) + B_\varphi + R_i = B_\varphi \qquad (34)$$

$$\vartheta^* = \frac{R_i}{R_i - R_1}; i = 2, \dots, 5 \qquad (35)$$

When it is used ϑ^* instead of ϑ in (16), we get $\gamma_1 = B_\varphi$. As γ_1 is independent of R_2, all suggested ratio estimators have the same minimum MSE as follows

$$MSE_{min}\left(z_{pri}\right) \cong \frac{1-f}{n}\left(S_y^2 - 2B_\varphi S_{yp} + B_\varphi^2 S_p^2\right) \qquad (36)$$

It can also write this expression by

$$MSE_{min}\left(z_{pri}\right) \cong \frac{1-f}{n}S_y^2\left(1 - \rho_{pb}^2\right) \qquad (37)$$

3. Efficiency Comparisons

In this section, it is compared the mean square error of suggested estimators, given in (36), with the MSE of ratio estimators given in Zaman [1], presented in (6). As it is obtained the following condition by these comparison:

$$MSE_{min}\left(z_{pri}\right) < MSE\left(t_{pri}\right) \qquad (38)$$

$$\frac{1-f}{n}S_y^2\left(1 - \rho_{pb}^2\right) < \frac{1-f}{n}\left[R_i^2 S_\varphi^2 + S_y^2\left(1 - \rho_{pb}^2\right)\right] \quad (39)$$

$$R_i^2 S_\varphi^2 > 0 \qquad (40)$$

We can conclude that all suggested estimators are more efficient than all ratio-type estimators presented in Zaman [1] in all restrictions, because the restriction given in (40) is always satisfied.

Secondly, it is compared the mean square error of the suggested estimators presented in (37) with the variance of sample mean, so we have the following restriction:

$$MSE_{min}\left(z_{pri}\right) < Var(\bar{y}) \qquad (41)$$

$$\frac{1-f}{n}S_y^2\left(1 - \rho_{pb}^2\right) < \frac{1-f}{n}S_y^2 \qquad (42)$$

$$\rho_{pb}^2 > 0 \qquad (43)$$

Because this restriction is always satisfied, suggested estimators are more efficient than the sample mean.

Finally, it is compared the mean square error of the suggested estimators presented in (37) with the the ratio estimator suggested by Naik-Gupta [3], so we have the following restriction:

$$MSE_{min}\left(z_{pri}\right) < MSE\left(t_{NG}\right) \qquad (44)$$

$$\frac{1-f}{n}S_y^2\left(1 - \rho_{pb}^2\right) < \frac{1-f}{n}\left(S_y^2 - 2R_\varphi S_{y\varphi} + R_\varphi^2 S_\varphi^2\right) \quad (45)$$

$$\rho_{pb}^2 > \frac{2R_\varphi S_{y\varphi} - R_\varphi^2 S_\varphi^2}{S_y^2} \qquad (46)$$

When the restriction (46) is satisfied, the suggested estimators are more efficient than the ratio estimator suggested by Naik-Gupta [3].

4. Empirical study

We have used the same data sets as in Zaman [1] to compare the efficiencies of the suggested estimators with the ratio-type estimators numerically.

The statistics about the populations I and II are presented in Tables 1 and 2 respectively. Note that the sample sizes as $n = 20, n = 30$ [12].

Population I (Source: see Sukhatme (1957), p. 279) [13]

$$y = Number \ of \ villages \ in \ the \ circles \quad (47)$$

$$\phi_i = \begin{cases} 1 & , \quad if \ y > 5 \\ 0 & , \quad otherwise \end{cases} \quad (48)$$

Table 1. Population I Data Statistics

N = 89	$\bar{Y} = 3.3596$	$\beta_1(\varphi) = 2.3267$	$R_3 = 4.2529$
n = 20	P = 0.1236	$S_{y\varphi} = 0.5116$	$R_4 = 2.6359$
$\beta_2(\varphi) = 3.491$	$S_y = 2.0184$	$R_\varphi = 27.1812$	$R_5 = 3.3852$
$\rho_{pb} = 0.766$	$S_\varphi = 0.3309$	$R_1 = 1.3711$	
$C_y = 0.6008$	$C_p = 2.6779$	$R_2 = 2.0683$	

Population II (Source: see Zaman et al. (2014)) [14]

$$y = the \ number \ of \ teachers \quad (49)$$

$$\phi_i = \begin{cases} 1 & , \quad if \ y > 60 \\ 0 & , \quad otherwise \end{cases} \quad (50)$$

Table 2. Population II Data Statistics

N = 111	$\bar{Y} = 29.2793$	$\beta_1(\varphi) = 2.4142$	$R_3 = 39.7586$
n = 30	P = 0.1171	$S_{y\varphi} = 6.5698$	$R_4 = 23.0721$
$\beta_2(\varphi) = 3.898$	$S_y = 25.5208$	$R_\varphi = 250.0367$	$R_5 = 29.7190$
$\rho_{pb} = 0.797$	$S_\varphi = 0.3230$	$R_1 = 11.5669$	
$C_y = 0.8716$	$C_p = 2.7810$	$R_2 = 16.9073$	

When examining the conditions determined in Section 3 for these data sets, they are satisfied for the proposed estimators as follows;

For population I;

$R_i^2 S_\varphi^2 > 0, \ i = 1,2,\dots,5 \ \rightarrow$ Conditions (40) is always satisfied. $\rho_{pb}^2 > 0 \ \rightarrow$ Conditions (43) is always satisfied. $\rho_{pb}^2 = 0.587 > \frac{2R_\varphi S_{y\varphi} - R_\varphi^2 S_\varphi^2}{S_y^2} = -13.03 \ \rightarrow$ Conditions (46) is satisfied.

For population II;

$R_i^2 S_\varphi^2 > 0, \ i = 1,2,\dots,5 \ \rightarrow$ Conditions (40) is always satisfied. $\rho_{pb}^2 > 0 \ \rightarrow$ Conditions (43) is always satisfied. $\rho_{pb}^2 = 0.635 > \frac{2R_\varphi S_{y\varphi} - R_\varphi^2 S_\varphi^2}{S_y^2} = -4.97 \ \rightarrow$ Conditions (46) is satisfied.

In Table 3, values of mean square error, which are computed using equations given in Sections 1 and 2, are presented. When it is examined Table 3, it is observed that the suggested estimators have the smallest mean square error value among all ratio-type estimators presented Section 1. This is an expected results, as mentioned in Section 3 since the conditions presented in (40) and (43) are always satisfied.

Table 3. MSE values of the Ratio Estimators

Estimator	MSE

	Population I	Population II
$\bar{y}$	0.1579	15.8427
t_{NG}	2.2157	94.5823
t_{pr1}	0.0732	6.1188
t_{pr2}	0.0834	6.5047
t_{pr3}	0.142	9.7908
t_{pr4}	0.0948	7.1302
t_{pr5}	0.1139	8.0206
z_{pri}	0,0653	5.7793

From the result of these numerical illustrations, it is deduced that all the suggested estimators are more efficient than all ratio-type estimators for these data sets.

5. Conclusions

New ratio-type estimators were produced by combining the ratio estimators considered in Zaman [1] the minimum MSE equations were obtained for the suggested estimators. Theoretically, it was shown that all the suggested estimators are always more efficient than the ratio-type estimators. These theoretical results are also supported numerically using the same original data sets as in Zaman [1].

References

[1] Zaman, T. 2018. Modified Ratio Estimators Using Coefficient of Skewness of Auxiliary Attribute. International Journal of Modern Mathematical Sciences. 16(2) 87-95.

[2] Kadilar C., Cingi, H. 2006. Improvement in estimating the population mean in simple random sampling. Applied Mathematics Letters. 19. 75-79

[3] Naik, V.D. and Gupta, P.C. 1996. A note on estimation of mean with known population of an auxiliary character, Journal of Indian Society Agricultural Statistics, 48 (2) 151–158.

[4] Kadilar, C. and Cingi, H. 2005. A new estimator using two auxiliary variables. Applied Mathematics and Computation, 162(2) 901-908.

[5] Shabbir, J., and Gupta, S. 2007. On estimating the finite population mean with known population proportion of an auxiliary variable. Pakistan Journal of Statistics 23(1), 1-9.

[6] Singh, R., Chauhan, P., and Sawan, N. 2008. On linear combination of ratio and product type exponential estimator for estimating the finite population mean. Statistics in Transition, 9(1), 105-115.

[7] Abd-Elfattah, A. M., El-Sherpieny, E. A., Mohamed, S. M., and Abdou, O. F. 2010. Improvement in estimating the population mean in simple random sampling using information on auxiliary attribute. Applied mathematics and computation, 215(12), 4198-4202.

[8] Koyuncu, N. 2012. Efficient estimators of population mean using auxiliary attributes. Applied Mathematics and Computation, 218(22), 10900-10905.

[9] Malik, S., and Singh, R. 2013. An improved estimator using two auxiliary attributes. Applied mathematics and computation, 219(23), 10983-10986.

[10] Zaman, T. 2018. New family of estimators using two auxiliary attributes. International Journal of Advanced Research I Engineering & Management (IJAREM), 4(11), 11-16.

[11] Wolter, K.M. 1985. Introduction to Variance Estimation, (Springer-Verlag).

[12] Cochran, W.G. 1977. Sampling techniques, John Wiley and Sons, New York.

[13] Sukhatme, P.V. 1957. Sampling theory of surveys with applications. The Indian Society of Agrcultural Statistics, New Delhi, pp. 279-280.

[14] Zaman, T., Saglam, V., Sagir, M., Yucesoy, E., and Zobu, M. 2014. Investigation of some estimators via taylor series approach and an application. American Journal of Theoretical and Applied Statistics, 3(5) 141-147.