International Journal of Modern Mathematical Sciences, 2018, 16(2): 87-95

International Journal of Modern Mathematical Sciences ISSN: 2166-286X

Modern Scientific Press Journal homepage: www.ModernScientificPress.com/Journals/ijmms.aspx Florida, USA

Article

Modified Ratio Estimators Using Coefficient of Skewness of Auxiliary Attribute

Tolga Zaman

Çankırı Karatekin University, Faculty of Science, Department of Statistics, Çankırı, Turkey

E-Mail: tolgazaman@karatekin.edu.tr

Article history: Received 25 June 2018, Revised 29 July 2018, Accepted 2 August 2018, Published 5 August 2018.

Abstract: This paper proposes some ratio estimators for population mean using known skewness coefficient of auxiliary attribute. Theoretically, it is obtained mean square error (MSE) equations for all proposed estimators. The efficiency between the suggested estimators and some known estimators is compared. The theoretical findings are also supported by a numerical examples using original data sets.

Keywords: Ratio estimator; Auxiliary attribute; Mean square error; Efficiency

Mathematics Subject Classification (2010): 62D05

1. Introduction

The Naik and Gupta estimator for the population mean $\bar{Y}$ of the variate of study, which make use of information regarding the population proportion possessing certain attribute, is defined by

$$\bar{y}_{NG} = \frac{\bar{y}}{p}P \tag{1.1}$$

Let y_i be ith characteristic of the population and φ_i is the case of possessing certain attributes. If ith unit has the desired characteristic, it takes the value 1, if not then the value 0. That is;

$$\varphi_i = \begin{cases} 1 \quad , & \textit{if ith unit of the population possesses attribute} \\ 0 \quad , & \textit{otherwise} \end{cases}$$

where $\bar{y}$ the sample mean of the study variable and $a = \sum_{i=1}^{n} \varphi_i$ be the total count of the units that possess certain attribute sample. $p = \frac{a}{n}$ shows the ratio of these units and it is assumed that the population proportion P of the form of attribute φ is known.

The MSE of the Naik and Gupta estimator is

$$MSE(\bar{y}_{NG}) \cong \frac{1-f}{n}\left(S_y^2 - 2R_\varphi S_{y\varphi} + R_\varphi^2 S_\varphi^2\right) \quad (1.2)$$

where, $f = \frac{n}{N}$; N is the number of units in the population; $R_\varphi = \frac{\bar{Y}}{P}$; S_φ^2 is the population variance of auxiliary attribute φ, S_y^2 is the population variance of the study variable and $S_{y\varphi}$ is the sample covariance between the auxiliary attribute and the study variable (Naik and Gupta, 1996).

2. Suggested Estimators

Following Singh et al. (2008), it is proposed the estimators using the skewness coefficient of auxiliary attribute as follows:

$$t_{pr1} = \frac{\bar{y}+b_\varphi(P-p)}{p+\beta_1(\varphi)}[P+\beta_1(\varphi)] \quad (2.1)$$

$$t_{pr2} = \frac{\bar{y}+b_\varphi(P-p)}{\beta_1(\varphi)p+\beta_2(\varphi)}[\beta_1(\varphi)P+\beta_2(\varphi)] \quad (2.2)$$

$$t_{pr3} = \frac{\bar{y}+b_\varphi(P-p)}{\beta_2(\varphi)p+\beta_1(\varphi)}[\beta_2(\varphi)P+\beta_1(\varphi)] \quad (2.3)$$

$$t_{pr4} = \frac{\bar{y}+b_\varphi(P-p)}{\beta_1(\varphi)p+C_p}[\beta_1(\varphi)P+C_p] \quad (2.4)$$

$$t_{pr5} = \frac{\bar{y}+b_\varphi(P-p)}{C_p p+\beta_1(\varphi)}[C_p P+\beta_1(\varphi)] \quad (2.5)$$

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 d 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.

It is obtained the MSE equations for these proposed estimators using Taylor series approach as (for details, please see the Appendix A)

$$MSE(t_{pri}) \cong \frac{1-f}{n}\left[R_i^2 S_\varphi^2 + S_y^2\left(1-\rho_{pb}^2\right)\right], \quad i = 1,2,\ldots,5 \quad (2.6)$$

where $R_1 = \frac{\bar{Y}}{P+\beta_1(\varphi)}$, $R_2 = \frac{\bar{Y}\beta_1(\varphi)}{P\beta_1(\varphi)+\beta_2(\varphi)}$, $R_3 = \frac{\bar{Y}\beta_2(\varphi)}{P\beta_2(\varphi)+\beta_1(\varphi)}$, $R_4 = \frac{\bar{Y}\beta_1(\varphi)}{P\beta_1(\varphi)+C_p}$ and $R_5 = \frac{\bar{Y}C_p}{PC_p+\beta_1(\varphi)}$.

3. Efficiency Comparisons

In this section, the MSE of some known traditional estimators are compared with the MSE of the proposed estimators.

It is well known that under simple random sampling without replacement (SRSWOR) the variance of the sample mean is;

$$V(\bar{y}) = \frac{1-f}{n}\bar{Y}^2 C_y^2 \tag{3.1}$$

In this section, firstly, the MSE of the proposed estimators given in (2.6) with the variance of sample mean it is compared, so I have the following conditions;

$$MSE(t_{pri}) < V(\bar{y})\,, i = 1,2,\dots,5 \text{ if,}$$

$$R_i^2 S_\varphi^2 < S_y^2 \rho_{pb}^2$$

$$\rho_{pb}^2 > \frac{R_i^2 S_\varphi^2}{S_y^2}, \quad i = 1,2,\dots,5 \tag{3.2}$$

When the restrictions (3.2) is satisfied, the proposed estimators are more efficient than the sample mean.

$$MSE(t_{pri}) < MSE(\bar{y}_{NG}), i = 1,2,\dots,5$$

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

$$R_i^2 S_\varphi^2 + 2R_\varphi S_{y\varphi} - R_\varphi^2 S_\varphi^2 < S_y^2 \rho_{pb}^2$$

$$\rho_{pb}^2 > \frac{S_\varphi^2(R_i^2 - R_\varphi^2) + 2R_\varphi S_{y\varphi}}{S_y^2}, i = 1,2,\dots,5 \tag{3.3}$$

When the conditions (3.3) is satisfied, the proposed estimators are more efficient than the ratio estimator suggested by Naik-Gupta (1996).

$$MSE(t_{pri}) < MSE(t_{prj}), \qquad i \neq j = 1,2,\dots,5$$

$$R_i^2 S_\varphi^2 < R_j^2 S_\varphi^2$$

$$S_\varphi^2(R_i^2 - R_j^2) < 0$$

$$R_i^2 - R_j^2 < 0$$

If $\quad R_i < R_j, R_i > -R_j, \qquad i \neq j = 1,2,\dots,5$ (3.4)

If $\quad R_i > R_j, R_i < -R_j, \qquad i \neq j = 1,2,\dots,5$ (3.5)

When either of the condition (3.4) or (3.5) is satisfied, the ith proposed estimator is more efficient than the jth proposed estimator.

4. Numerical Illustrations

The performance of various estimators considered here using the two data sets is compared in this section.

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

$y = Number \ \ of \ villages \ in \ the \ circles$

$$\phi_i = \begin{cases} 1 \quad , \quad if \ \ A \ circle \ consisting \ more \ than \ five \ vilages \\ 0 \quad , \quad \quad \quad \quad \quad \quad \quad otherwise \end{cases}$$

The statistics about the populations I and II are presented in tables 1 and 2 respectively. Note that the sample sizes as $n = 20, n = 30$ and simple random sampling (Cochran, 1977) was used respectively for populations I and II. I would like to recall that sample size has no effect on efficiency comparisons of estimators, as shown in Section 3.

Table 1: Population I Data Statistics

$N = 89$	$\bar{Y} = 3.3596$	$\beta_1(\varphi) = 2.3267$	$R_3 = 4.2529$
$n = 20$	$P = 0.1236$	$S_{y\varphi} = 0.5116$	$R_4 = 2.6359$
$\beta_2(\varphi) = 3.4917$	$S_y = 2.0184$	$R_\varphi = 27.1812$	$R_5 = 3.3852$
$\rho_{pb} = 0.766$	$S_\varphi = 0.3309$	$R_1 = 1.3711$	
$C_y = 0.6008$	$C_p = 2.6779$	$R_2 = 2.0683$	

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

$y = the \ number \ of \ teachers$

$$\phi_i = \begin{cases} 1 \quad , \quad if \ the \ number \ of \ teachers \ is \ more \ than \ 60 \\ 0 \quad , \quad \quad \quad \quad \quad \quad \quad otherwise \end{cases}$$

Table 2: Population II Data Statistics

$N = 111$	$\bar{Y} = 29.2793$	$\beta_1(\varphi) = 2.4142$	$R_3 = 39.7586$
$n = 30$	$P = 0.1171$	$S_{y\varphi} = 6.5698$	$R_4 = 23.0721$
$\beta_2(\varphi) = 3.8981$	$S_y = 25,5208$	$R_\varphi = 250.0367$	$R_5 = 29.7190$
$\rho_{pb} = 0.797$	$S_\varphi = 0.3230$	$R_1 = 11.5669$	
$C_y = 0.8716$	$C_p = 2.7810$	$R_2 = 16.9073$	

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

For population I;

$$\rho_{pb}^2 = 0.587 > \frac{R_1^2 S_\varphi^2}{S_y^2} = 0.051$$

$$\rho_{pb}^2 = 0.587 > \frac{R_2^2 S_\varphi^2}{S_y^2} = 0.115$$

$$\rho_{pb}^2 = 0.587 > \frac{R_3^2 S_\varphi^2}{S_y^2} = 0.486 \quad \to \quad \text{Conditions (3.2) is satisfied.}$$

$$\rho_{pb}^2 = 0.587 > \frac{R_4^2 S_\varphi^2}{S_y^2} = 0.187$$

$$\rho_{pb}^2 = 0.587 > \frac{R_5^2 S_\varphi^2}{S_y^2} = 0.308$$

$$\rho_{pb}^2 = 0.587 > \frac{S_\varphi^2 (R_1^2 - R_\varphi^2) + 2R_\varphi S_{y\varphi}}{S_y^2} = -12.9798$$

$$\rho_{pb}^2 = 0.587 > \frac{S_\varphi^2 (R_2^2 - R_\varphi^2) + 2R_\varphi S_{y\varphi}}{S_y^2} = -19.7422$$

$$\rho_{pb}^2 = 0.587 > \frac{S_\varphi^2 (R_3^2 - R_\varphi^2) + 2R_\varphi S_{y\varphi}}{S_y^2} = -19.371 \quad \to \quad \text{Conditions (3.3) is satisfied.}$$

$$\rho_{pb}^2 = 0.587 > \frac{S_\varphi^2 (R_4^2 - R_\varphi^2) + 2R_\varphi S_{y\varphi}}{S_y^2} = -19.6704$$

$$\rho_{pb}^2 = 0.587 > \frac{S_\varphi^2 (R_5^2 - R_\varphi^2) + 2R_\varphi S_{y\varphi}}{S_y^2} = -19.5492$$

For the MSE comparison between the t_{pr1} and t_{pr2} proposed estimators, the conditions (3.4) is satisfied as follows:

$$R_1 = 1.3711 < R_2 = 2.0683 \quad \to \quad \text{Conditions (3.4) is satisfied.}$$

Thus, the first proposed estimator is more efficient than the second one.

Similarly for population II;

$$\rho_{pb}^2 = 0.635 > \frac{R_1^2 S_\varphi^2}{S_y^2} = 0.021$$

$$\rho_{pb}^2 = 0.635 > \frac{R_2^2 S_\varphi^2}{S_y^2} = 0.046$$

$$\rho_{pb}^2 = 0.635 > \frac{R_3^2 S_\varphi^2}{S_y^2} = 0.253 \quad \to \quad \text{Conditions (3.2) is satisfied.}$$

$$\rho_{pb}^2 = 0.635 > \frac{R_4^2 S_\varphi^2}{S_y^2} = 0.085$$

$$\rho_{pb}^2 = 0.635 > \frac{R_5^2 S_\varphi^2}{S_y^2} = 0.141$$

$$\rho_{pb}^2 = 0.635 > \frac{S_\varphi^2 (R_1^2 - R_\varphi^2) + 2R_\varphi S_{y\varphi}}{S_y^2} = -62379.5$$

$$\rho_{pb}^2 = 0.635 > \frac{S_\varphi^2(R_2^2 - R_\varphi^2) + 2R_\varphi S_{y\varphi}}{S_y^2} = -62227.5$$

$$\rho_{pb}^2 = 0.635 > \frac{S_\varphi^2(R_3^2 - R_\varphi^2) + 2R_\varphi S_{y\varphi}}{S_y^2} = -60932.6 \quad \vec{\rightarrow} \quad \text{Conditions (3.3) is satisfied.}$$

$$\rho_{pb}^2 = 0.635 > \frac{S_\varphi^2(R_4^2 - R_\varphi^2) + 2R_\varphi S_{y\varphi}}{S_y^2} = -61981$$

$$\rho_{pb}^2 = 0.635 > \frac{S_\varphi^2(R_5^2 - R_\varphi^2) + 2R_\varphi S_{y\varphi}}{S_y^2} = -61630.1$$

For the MSE comparison between the t_{pr1} and t_{pr2} proposed estimators, the conditions (3.4) is satisfied as follows:

$$R_1 = 11.5669 < R_2 = 16.9073 \quad \vec{\rightarrow} \quad \text{Conditions (3.4) is satisfied.}$$

Thus, the first proposed estimator is more efficient than the second.

Using the above results I calculated the MSE and the efficiency for all the estimators in Section 2. The efficiency for each estimator with respect to the sample mean of a SSR is defined as follows:

$$e(\hat{y}) = \frac{MSE(\bar{y})}{MSE(\hat{y})}$$

Where $MSE(\hat{y})$ is the mean square error for the each estimator suggested in Section 2 and $MSE(\bar{y}) = V(\bar{y})$ for a SSR of sizes 20 and 30, respectively.

Table 3 presents the MSE and the efficiency all for the estimators given in Section 2. As we can see that the suggested estimators t_{pr1} and t_{pr2} dominate all other estimators with efficiency 2.16 and 2.59; 1.89 and 2.44, respectively.

Table 3: MSE values of the Ratio Estimators

Estimator	**MSE**		**Efficiency**	
	Population I	**Population II**	**Population I**	**Population II**
$\bar{y}$	0.1579	15.8427	1	1
t_{NG}	2.2157	94.5823	0.0713	0.1675
t_{pr1}	0.0732	6.1188	2.1562	2.5892
t_{pr2}	0.0834	6.5047	1.8931	2.4356
t_{pr3}	0.142	9.7908	1.1119	1.6181
t_{pr4}	0.0948	7.1302	1.6667	2.2219
t_{pr5}	0.1139	8.0206	1.3865	1.9752

5. Conclusions

In the study, some ratio estimators for population mean using known skewness coefficient of auxiliary attribute is developed and a theoretical argument to show that proposed estimators have a smaller MSE than all other estimators is given. Based on these results, we noticed that the suggested estimator t_{pr1} has the highest efficiency.

References

Cochran, W.G. *Sampling techniques*, John Wiley and Sons, New York, **1977.**

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

Singh, R., Chauhan, P., Sawan, N. and Smarandache, F. Ratio estimators in simple random sampling using information on auxiliary attribute, *Pakistan Journal of Statistics and Operation Research* IV (1) **(2008)** 47–53.

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

Wolter, K.M. (1985). *Introduction to Variance Estimation*, (Springer-Verlag), **1985.**

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

Appendix

In general, Taylor series method for k variables can be given as;

$$h(\bar{x}_1,\bar{x}_2,\dots,\bar{x}_k) = h(\bar{X}_1,\bar{X}_2,\dots,\bar{X}_k) + \sum_{j=1}^{k} d_j\left(\bar{x}_j - \bar{X}_j\right) + R_k(\bar{X}_k,\alpha) + O_k$$

where

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

and

$$R_k(\bar{X}_k,\alpha) = \sum_{j=1}^{k}\sum_{j=1}^{k} \frac{1}{2!}\frac{\partial^2 h(\bar{X}_1,\bar{X}_2,\dots,\bar{X}_k)}{\partial \bar{X}_i \bar{X}_j}\left(\bar{x}_j - \bar{X}_j\right)(\bar{x}_i - \bar{X}_i) + O_k$$

where O_k represents the terms in the expansion of the Taylor series of more than the second degree (Wolter, 1985). When we omit the term $R_k(\bar{X}_k,\alpha)$, we obtain Taylor series method for two variables as follows;

$$h(p,\bar{y}) - h(P,\bar{Y}) \cong \left.\frac{\partial h(c,d)}{\partial c}\right|_{P,\bar{Y}}(p-P) + \left.\frac{\partial h(c,d)}{\partial d}\right|_{\bar{Y},P}(\bar{y}-\bar{Y})$$

where, $h(p,\bar{y}) = t_{pri}$ and $h(P,\bar{Y}) = \bar{Y}$

MSE equations of the proposed estimators given in Table 1 compute as follows:

$$t_{pri} - \bar{Y} \cong \left.\frac{\partial\left(\frac{\bar{y}+b(P-p)}{m_1 p + m_2}[m_1P+m_2]\right)}{\partial p}\right|_{P,\bar{Y}}(p-P) + \left.\frac{\partial\left(\frac{\bar{y}+b(P-p)}{m_1 p + m_2}[m_1P+m_2]\right)}{\partial \bar{y}}\right|_{\bar{Y},P}(\bar{y}-\bar{Y})$$

$$E\left(t_{pri}-\bar{Y}\right)^2 \cong \left(B^2 + \frac{m^1\bar{Y}^2}{(m_1P+m_2)^2} + 2B\frac{m_1\bar{Y}}{m_1P+m_2}\right)V(p) - 2\left(\frac{B(m_1P+m_2)+m_1\bar{Y}}{m_1p+m_2}\right)Cov(p,\bar{y}) + V(\bar{y})$$

If $R_i = \frac{m_1\bar{Y}}{m_1p+m_2}$ taken,

$$E\left(t_{pri}-\bar{Y}\right)^2 \cong \frac{1-f}{n}\left[(B^2 + R_i^2 + 2BR_i)S_\varphi^2 - 2(B+R_i)S_{y\varphi} + S_y^2\right]$$

where, $B = \frac{S_{y\varphi}}{S_\varphi^2}$

$$\cong \frac{1-f}{n}\left[\frac{S_{y\varphi}^2}{S_\varphi^4}S_\varphi^2 + R_i^2 S_\varphi^2 + 2R_i S_{y\varphi} - 2\frac{S_{y\varphi}^2}{S_\varphi^2} - 2R_i S_{y\varphi} + S_y^2\right]$$

$$MSE(t_{pri}) \cong \frac{1-f}{n}\left[R_i^2 S_\varphi^2 + S_y^2\left(1 - \frac{S_{y\varphi}^2}{S_\varphi^2 S_y^2}\right)\right]$$

$$MSE(t_{pri}) \cong \frac{1-f}{n}\left[R_i^2 S_\varphi^2 + S_y^2(1-\rho_{pb}^2)\right]; \quad i = 1,2,3,4,5 \qquad (A)$$

where, $R_1 = \frac{\bar{Y}}{P+\beta_1(\varphi)}$, $R_2 = \frac{\bar{Y}\beta_1(\varphi)}{P\beta_1(\varphi)+\beta_2(\varphi)}$, $R_3 = \frac{\bar{Y}\beta_2(\varphi)}{P\beta_2(\varphi)+\beta_1(\varphi)}$, $R_4 = \frac{\bar{Y}\beta_1(\varphi)}{P\beta_1(\varphi)+C_p}$ and $R_5 = \frac{\bar{Y}C_p}{PC_p+\beta_1(\varphi)}$