Available online at http://scik.org

J. Math. Comput. Sci. 8 (2018), No. 4, 523-530

https://doi.org/10.28919/jmcs/3717

ISSN: 1927-5307

A NEW CHAIN RATIO ESTIMATOR USING INFORMATION ON AUXILIARY ATTRIBUTE

TOLGA ZAMAN[1,*], AND ÇAĞLAR SÖZEN[2]

[1]Department of Statistics, Çankırı Karatekin University, Çankırı 18100, Turkey

[2]Department of Banking and Finance, Giresun University, Giresun 28100, Turkey

Abstract: In this paper, we develop to ratio estimator suggested by Naik-Gupta [J. Indian Soc. Agric. Stat., 48 (2), 151-158] [1] and obtain its MSE equation. We prove that the proposed chain ratio estimator is more efficient than the Naik-Gupta estimator under certain conditions. In addition, this theoretical result is supported by an application with original data sets.

Keywords: ratio estimator; mean square error; auxiliary attribute; efficiency.

2010 AMS Subject Classification: 97K80.

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

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

---

*Corresponding author

E-mail address: tolgazaman@karatekin.edu.tr

Received April 4, 2018

where it is assumed that the population proportion P of the form of attribute ϕ is known.

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;

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

Let $A = \sum_{i=1}^{N} \phi_i$ and $a = \sum_{i=1}^{n} \phi_i$ be the the total count of the units that possess certain attribute in population and sample, respectively. And $P = \frac{A}{N}$ and $p = \frac{a}{n}$ shows the ratio of these units, respectively.

The MSE of the Naik and Gupta estimator is

$$MSE(\bar{y}_{NG}) \cong \frac{1-f}{n}\left(S_y^2 - 2RS_{y\phi} + R^2 S_\phi^2\right) \tag{1.2}$$

where, $f = \frac{n}{N}$; N is the number of units in the population; $R = \frac{\bar{Y}}{P}$ is the population ratio; S_ϕ^2 is the population variance of the form of attribute and S_y^2 is the population variance of the study variable [1].

2. The Proposed Chain Estimator

Following Kadılar and Cingi (2003) [2], We propose a chain estimator using information about population proportion possessing certain attributes. When $\bar{y}$ in (1.1) is replaced with $\bar{y}_{NG}$, the proposed chain estimator is obtained as

$$\bar{y}_{cNG} = \frac{\bar{y}_{NG}}{p} P \tag{2.1}$$

We can re-write (2.1) using (1.1) as,

$$\bar{y}_{cNG}(\alpha) = \bar{y}\left(\frac{P}{p}\right)^{\alpha} \tag{2.2}$$

where α is real numbers. MSE of this estimator can be found using Taylor series method defined as;

$$MSE(\bar{y}_{cNG}(\alpha)) \cong d \sum d' \tag{2.3}$$

where, $d = \left[\frac{\partial h(a,b)}{\partial a}\Big|_{\bar{Y},P} \quad \frac{\partial h(a,b)}{\partial b}\Big|_{\bar{Y},P} \right]$ and $\sum = \frac{1-f}{n}\begin{bmatrix} S_y^2 & S_{y\phi} \\ S_{\phi y} & S_\phi^2 \end{bmatrix}$

[3]. Where,$h(a,b) = h(\bar{y},\bar{x}) = \bar{y}_{pr1}(\alpha)$. S_y^2 and S_ϕ^2 denote the population of variances of the study variable and unit ratios possessing certain attributes, respectively. $S_{y\phi} = S_{\phi y}$ denotes the population covariance between units ratio possessing certain attributes and study variable. According to this definition, we obtain d for this estimator as follows;

$$d = \left[1 \quad -\frac{\alpha\bar{Y}}{P}\right]$$

We obtain the MSE equation of this estimator using (2.3) as follows;

$$MSE(\bar{y}_{cNG}(\alpha)) \cong \left[1 \quad -\frac{\alpha\bar{Y}}{P}\right] \frac{1-f}{n} \begin{bmatrix} S_y^2 & S_{yx} \\ S_{xy} & S_\phi^2 \end{bmatrix} \begin{bmatrix} 1 \\ -\frac{\alpha\bar{Y}}{P} \end{bmatrix}$$

$$MSE(\bar{y}_{cNG}(\alpha)) \cong \frac{1-f}{n}\left(S_y^2 - \frac{2\alpha\bar{Y}S_{y\phi}}{P} + \frac{\alpha^2\bar{Y}^2S_\phi^2}{P^2}\right)$$

$$MSE(\bar{y}_{cNG}(\alpha)) \cong \frac{1-f}{n}\left(S_y^2 - 2\alpha R S_{y\phi} + \alpha^2 R^2 S_\phi^2\right) \tag{2.4}$$

where, $R = \frac{\bar{Y}}{P}$, $S_y^2 = (N-1)^{-1}\sum_{i=1}^N (y_i - \bar{Y})^2$, $S_\phi^2 = (N-1)^{-1}\sum_{i=1}^N (\phi_i - P)^2$,

$S_{y\phi} = (N-1)^{-1}\sum_{i=1}^N (y_i - \bar{Y})(\phi_i - P)$.

We can have the optimal values of α (2.4) by following equations:

$$\frac{\partial MSE(\bar{y}_{cNG}(\alpha))}{\partial \alpha} = \frac{1-f}{n}\left(2\alpha R^2 S_\phi^2 - 2RS_{y\phi}\right) = 0$$

$$\alpha R S_{\phi}^{2} - S_{y\phi} = 0$$

$$\alpha = \frac{B_{\phi}}{R} \tag{2.5}$$

where $B_{\phi} = \frac{S_{y\phi}}{S_{\phi}^{2}}$.

We can obtain minimum MSE of the proposed chain estimator using the optimal equations of α in (2.5).

$$MSE_{min}\left(\bar{y}_{cNG}(\alpha)\right) = \frac{1-f}{n}\left(S_{y}^{2} - 2B_{\phi}S_{y\phi} + B_{\phi}^{2}S_{\phi}^{2}\right) \tag{2.6}$$

3. Efficiency Comparisons

In this section, we compare the MSE of the proposed chain estimator, given in (2.2), with the MSE of the Naik-Gupta estimators, given in (1.1). We have the condition;

$$MSE\left(\bar{y}_{cNG}(\alpha)\right) < MSE\left(\bar{y}_{NG}\right)$$

$$\frac{1-f}{n}\left(S_{y}^{2} - 2B_{\phi}S_{y\phi} + B_{\phi}^{2}S_{\phi}^{2}\right) < \frac{1-f}{n}\left(S_{y}^{2} - 2RS_{y\phi} + R^{2}S_{\phi}^{2}\right)$$

$$S_{\phi}^{2}\left(B_{\phi}^{2} - R^{2}\right) - 2S_{y\phi}\left(B_{\phi} - R\right) < 0$$

$$\left(B_{\phi} - R\right)\left[S_{\phi}^{2}\left(B_{\phi} + R\right) - 2S_{y\phi}\right] < 0$$

For $B_{\phi} - R < 0$, That is, $B_{\phi} < R$

$$S_{\phi}^{2}\left(B_{\phi} + R\right) - 2S_{y\phi} > 0 \tag{3.1}$$

Similarly, $B_{\phi} - R > 0$, That is, $B_{\phi} > R$

$$S_{\phi}^{2}\left(B_{\phi} + R\right) - 2S_{y\phi} < 0 \tag{3.2}$$

When condition (3.1) or (3.2) is satisfied, the proposed chain estimator given in (2.2), are more efficient than the Naik-Gupta estimator, given in (1.1).

4. Numerical illustrations

We compare the performance of various estimators considered here using the two data sets.

Population I (Source: see Sukhatme (1957), p. 279)[4].

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

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

Table 1: Population 1 Data Statistics

N:89	$\overline{Y}$: 3.3596
n: 20	P: 0.1236
R: 27.181	S_y: 2.018
ρ_{pb}: 0.766	S_ϕ: 0.331
$S_{\phi y}$: 0.512	B_ϕ: 4.673

Population 2: We use the teachers data which means the number of teachers working at school in Trabzon. (Directorate of National Education, Trabzon)[5]. The schools in Trabzon are taken as unit of population. The data is defined as following;

$$y = the\ number\ of\ teachers$$

$$\phi_i = \begin{cases} 1 & , \quad if\ the\ number\ of\ teachers\ is\ more\ than\ 40 \\ 0 & , \quad otherwise \end{cases}$$

The population statistics of the data are given in Table 2.

Table 2: Population 2 Data Statistics

N: 111	$\overline{Y}$: 31.837
n: 40	P: 0.279
R: 114	S_y: 36.185
ρ_{pb}: 0.878	S_ϕ: 1.010
$S_{\phi y}$: 32.100	B_ϕ: 31.431

For Populations 1 and 2, We take the sample sizes as $n = 20$ and $n = 40$ using simple random sampling [6] . The MSE of the Naik-Gupta and proposed chain estimators are computed as given in (1.2) and (2.6), respectively, and these estimators are compared to each other with respect to their MSE values.

In tables 1 and 2, There are the statistics about the population for data 1, data 2 sets. Note that the correlations between the variate are 0.766 and 0.878, respectively.

Table3: The MSE values for proposed chain and Naik-Gupta estimator

Estimator	MSE	
	Population 1	Population 2
Proposed Chain	0.065	297.725
Naik-Gupta	2.217	7197.818

In table 3, the values of the MSE are given. From table, it is seen that the proposed chain estimator has a smaller than the Naik-Gupta estimator. Therefore, it is concluded that proposed chain estimator more efficient than the Naik-Gupta estimator for both population 1 and population 2 data sets.

This results are expected because the condition (3.1) is satisfied for proposed estimator as follows:

For Population 1,

$$B_{\phi} = 4.673; \quad R = 27.181 \quad \text{and} \quad B_{\phi} < R$$

$$S_{\phi}^2(B_{\phi} + R) - 2S_{y\phi} = 2.466 > 0$$

Similarly, for Population 2;

$$B_{\phi} = 31.431; \; R = 114 \; \text{ and } \quad B_{\phi} > R$$

$$S_{\phi}^2(B_{\phi} + R) - 2S_{y\phi} = 84.327 > 0$$

Thus, the condition mentioned in section 3 is satisfied for Population 1 and 2 data sets.

5. Conclusion

We have analyzed the proposed chain estimator and obtained its MSE equation. According to the theoretical discussion in Section 3 and the results of the numerical examples, we infer that the proposed chain estimator are more efficient than the Naik-Gupta ratio estimator. In forthcoming studies, we hope to adapt the proposed chain estimators in stratified random sampling.

Conflict of Interests

The authors declare that there is no conflict of interests.

REFERENCES

[1] V.D. Naik, P.C. Gupta. A note on estimation of mean with known population proportion of an auxiliary character. J. Indian Soc. Agric. Stat., 48 (2) (1996), 151-158.

[2] C. Kadılar, H. Cingi. A study on the chain ratio-type estimator. Hacettepe J. Math. Stat., 32 (2003), 105-108.

[3] K. M. Wolter. Introduction to Variance Estimation (Springer-Verlag, 2003).

[4] P.V. Sukhatme, Sampling theory of surveys with applications. The Indian Society of Agricultural Statistics; New Delhi (1957).

[5] T. Zaman, V. Sağlam, M. Sağır, E. Yücesoy, M. Zobu, Investigation of some estimators via taylor series approach and an application. Amer. J. Theor. Appl. Stat., 3(5) (2014), 141-147.

[6] W. G. Cochran, Sampling techniques. John Wiley & Sons.New York, (1977).