Applied Mathematics and Computation 215 (2010) 4198–4202

Contents lists available at ScienceDirect

Applied Mathematics and Computation

journal homepage: www.elsevier.com/locate/amc

Improvement in estimating the population mean in simple random sampling using information on auxiliary attribute

A.M. Abd-Elfattah *, E.A. El-Sherpieny, S.M. Mohamed, O.F. Abdou

Institute of Statistical Studies and Research, Cairo University, Dokki, Giza 12613, Egypt

ARTICLE INFO

Keywords:
Ratio-type estimator
Simple random sampling
Auxiliary attribute
Efficiency

ABSTRACT

This paper proposes some estimators for the population mean by adapting the estimator in Singh et al. (2008) [5] to the ratio estimators presented in Kadilar and Cingi 2006 [2]. We obtain mean square error (*MSE*) equation for all proposed estimators, and show that all proposed estimators are always more efficient than ratio estimator in Naik and Gupta (1996) [3], and Singh et al. (2008) [5]. The results have been illustrated numerically by taking some empirical population considered in the literature.

1. Introduction

Consider a sample of size n drawn by simple random sample without replacement from a population of size N. Let y_i and φ_i denoted the observation on variable y and φ, respectively, for ith unit $(i = 1, 2, 3, \ldots, N)$. Suppose there is a complete dichotomy in the population with respect to the presence or absence of an attribute, say φ, and it is assumed that attribute φ takes only the two values 0 and 1 according as

$$\begin{aligned} \varphi &= 1, \text{ if } i\text{th unit of the population possesses attribute } \varphi \\ &= 0, \text{ if otherwise.} \end{aligned}$$

Let $A = \sum_{i=1}^{N} \varphi_i$ and $a = \sum_{i=1}^{n} \varphi_i$ denoted the total number of units in the population and sample possessing attribute φ, respectively. Let $P = \frac{A}{N}$ and $\widehat{P} = \frac{a}{n}$ denoted the proportion of units in the population and sample, respectively, possessing attribute φ. Taking into consideration the point biserial correlation coefficient between auxiliary attribute and study variable, Naik and Gupta [3] defined ratio estimator of population mean when the prior information of population proportion of units, possessing the same attribute is available, as follows:

$$t_{NG} = \bar{y} \frac{P}{\widehat{P}}, \tag{1.1}$$

where $\bar{y}$ is the sample mean of study variable. The *MSE* of t_{NG} up to the first order of approximation is

$$MSE(t_{NG}) = \left(\frac{1-f}{n}\right) \left[S_y^2 + R_1^2 S_\varphi^2 - 2R_1 S_{y\varphi}\right], \tag{1.2}$$

where $f = \frac{n}{N}$; n is the sample size; N is the number of units in the population; $R_1 = \frac{\bar{Y}}{P}$, S_φ^2 is the population variance of auxiliary attribute φ, and $S_{y\varphi}$ is the population covariance between variable of interest and auxiliary attribute φ.

* Corresponding author. Address: Department of Statistics, Faculty of Science, King Abdul Aziz University Box 80203, Jedda 21589, Saudia Arabia.
E-mail address: a_afattah@hotmail.com (A.M. Abd-Elfattah).

0096-3003/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.amc.2009.12.041

Singh et al. [5] suggested the following ratio estimators for estimating the population mean $\overline{Y}$ of the study variable y in simple random sampling using known parameters of auxiliary attribute φ, such as, coefficient of variation C_P, coefficient of kurtosis $B_2(\varphi)$, and point biserial correlation coefficient ρ_{Pb} as:

$$t_1 = \frac{\bar{y} + b_\varphi(P - \widehat{P})}{\widehat{P}} P, \tag{1.3}$$

$$t_2 = \frac{\bar{y} + b_\varphi(P - \widehat{P})}{\widehat{P} + B_2(\varphi)} [P + B_2(\varphi)], \tag{1.4}$$

$$t_3 = \frac{\bar{y} + b_\varphi(P - \widehat{P})}{\widehat{P} + C_P} [P + C_P], \tag{1.5}$$

$$t_4 = \frac{\bar{y} + b_\varphi(P - \widehat{P})}{\widehat{P}B_2(\varphi) + C_P} [PB_2(\varphi) + C_P], \tag{1.6}$$

$$t_5 = \frac{\bar{y} + b_\varphi(P - \widehat{P})}{\widehat{P}C_P + B_2(\varphi)} [PC_P + B_2(\varphi)], \tag{1.7}$$

where C_P and $B_2(\varphi)$ are the population coefficient of variation and the population coefficient of kurtosis 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.

In Singh et al. [5], mean square error *MSE* equation of these ratio estimators were given by

$$MSE(t_i) = \frac{1-f}{n}\left[R_i^2 S_\varphi^2 + S_y^2(1 - \rho_{Pb}^2)\right], \quad (i = 1, 2, 3, \ldots, 5), \tag{1.8}$$

where $R_1 = \frac{\overline{Y}}{P},\ R_2 = \frac{\overline{Y}}{P+B_2(\varphi)},\ R_3 = \frac{\overline{Y}}{P+C_P},\ R_4 = \frac{\overline{Y}B_2(\varphi)}{PB_2(\varphi)+C_P}$ and $R_5 = \frac{\overline{Y}C_P}{PC_P+B_2(\varphi)}$.

Singh et al. [5] concluded that the ratio estimators t_i $(i = 1, 2, \ldots, 5)$ which uses some known value of population proportion were more efficient than the sample mean $\bar{y}$ and ratio estimator in Naik and Gupta [3].

In the next section, we develop new estimators combining ratio estimators in Singh et al. [5] and obtain the *MSE* equations of these new estimators. In Section 3, we compare the efficiencies, theoretically, based on *MSE* equations, between the proposed estimators and the ratio estimators presented in Singh et al. [5]. In Section 4, we also discuss the comparison among all the suggested estimators numerically. In Section 5, we give a hint to obtain different estimators by a similar method presented in this study.

2. Suggested estimators

We propose the estimator using the procedure presented in Kadilar and Cingi 2006 [2] combining ratio estimators (1.3) and (1.4) as follows:

$$t_{pro1} = m_1 \frac{\bar{y} + b_\varphi(P - \widehat{P})}{\widehat{P}} P + m_2 \frac{\bar{y} + b_\varphi(P - \widehat{P})}{\widehat{P} + B_2(\varphi)} (P + B_2(\varphi)), \tag{2.1}$$

where m_1 and m_2 are weights that satisfy the condition $m_1 + m_2 = 1$.

The *MSE* of this estimator can be found using the first degree approximation in the Taylor series method defined by

$$MSE(t_{pro1}) \cong d \sum d', \tag{2.2}$$

where d is a vector defined as $d = \left[\frac{\partial h(a,b)}{\partial a}\Big|_{\overline{Y},P} \quad \frac{\partial h(a,b)}{\partial b}\Big|_{\overline{Y},P}\right]$, $\sum$ is the variance covariance matrix as $\sum = \frac{1-f}{n}\begin{bmatrix} S_y^2 & S_{y\varphi} \\ S_{\varphi y} & S_\varphi^2 \end{bmatrix}$ (see Wolter [7]). Here $h(a,b) = h(\bar{y}, \widehat{P}) = t_{pro1}$. According to this definition, we obtain d for the proposed estimator as follows:

$$d = \begin{bmatrix} 1 & -m_1(R_1 + B_\varphi) - m_2(R_2 + B_\varphi) \end{bmatrix},$$

where $B_\varphi = \frac{S_{y\varphi}}{S_\varphi^2} = \frac{\rho_{Pb} S_y}{S_\varphi}$. Note that we omit the difference : $b - B$ (Cochran [1]).

We obtain the *MSE* of the proposed estimator using (2.2) as

$$MSE(t_{pro1}) = \frac{1-f}{n}\left(S_y^2 - 2\eta S_{y\varphi} + \eta^2 S_\varphi^2\right), \tag{2.3}$$

where

$$\eta = m_1(R_1 + B_\varphi) + m_2(R_2 + B_\varphi). \tag{2.4}$$

We also propose the estimator combining ratio estimators (1.3) and (1.5) as

$$t_{pro2} = m_1 \frac{\bar{y} + b_\varphi(P - \widehat{P})}{\widehat{P}} P + m_2 \frac{\bar{y} + b_\varphi(P - \widehat{P})}{\widehat{P} + C_P} (P + C_P). \tag{2.5}$$

The *MSE* of this estimator is the same as (2.3) but R_2 in (2.4) is replaced with R_3.

In addition, we propose the following estimator combining ratio estimators (1.3) and (1.6) as

$$t_{pro3} = m_1 \frac{\bar{y} + b_\varphi(P - \widehat{P})}{\widehat{P}} P + m_2 \frac{\bar{y} + b_\varphi(P - \widehat{P})}{\widehat{P} B_2(\varphi) + C_P} [P B_2(\varphi) + C_P]. \tag{2.6}$$

The mean square error of this estimator is again the same as (2.3) but R_2 in (2.4) is replaced with R_4.

Lastly, we propose the following estimator combining ratio estimators (1.3) and (1.7) as

$$t_{pro4} = m_1 \frac{\bar{y} + b_\varphi(P - \widehat{P})}{\widehat{P}} P + m_2 \frac{\bar{y} + b_\varphi(P - \widehat{P})}{\widehat{P} C_P + B_2(\varphi)} [P C_P + B_2(\varphi)]. \tag{2.7}$$

The mean square error of this estimator is again the same as (2.3) but R_2 in (2.4) is replaced with R_5.

The optimal values of m_1 and m_2 to minimize (2.3) can easily be found as follows:

$$m_1^* = \frac{R_2}{R_2 - R_1} \quad \text{and} \quad m_2^* = \frac{R_1}{R_1 - R_2}, \tag{2.8}$$

when we use m_1^* and m_2^* instead of m_1 and m_2 in (2.4), we get $\eta = B_\varphi$. As η is independent of R_2, all proposed estimators have the same minimum *MSE* as follows:

$$MSE_{\min}(t_{proi}) = \frac{1-f}{n}\left(S_y^2 - 2B_\varphi S_{y\varphi} + B_\varphi^2 S_\varphi^2\right), \quad i = 1,2,3,4.$$

We can also write this expression as

$$MSE_{\min}(t_{proi}) = \frac{1-f}{n} S_y^2 (1 - \rho_{Pb}^2). \tag{2.9}$$

2.1. New ratio estimators

We suggest following estimator:

$$t_{proi} = \frac{\bar{y}}{m_1 \widehat{P} + m_2} (m_1 P + m_2), \tag{2.10}$$

where m_1 and m_2 are either real number or the function of the known parameter of auxiliary attribute such as C_P, $B_2(\varphi)$ and ρ_{Pb}, note that the sum of m_1 and m_2 not necessarily equal to one.

The following scheme presents some of the important estimators of the population mean, which can be obtained by suitable choice of constants m_1 and m_2:

Estimator	Values of	
	m_1	m_2
$t_{pro1SD} = \bar{y}\frac{P+C_P}{\widehat{P}+C_P}$	1	C_P
$t_{pro2SK} = \bar{y}\frac{P+B_2(\varphi)}{\widehat{P}+B_2(\varphi)}$	1	$B_2(\varphi)$
$t_{pro3US1} = \bar{y}\frac{PB_2(\varphi)+C_P}{\widehat{P}B_2(\varphi)+C_P}$	$B_2(\varphi)$	C_P
$t_{pro3US2} = \bar{y}\frac{PC_P+B_2(\varphi)}{\widehat{P}C_P+B_2(\varphi)}$	C_P	$B_2(\varphi)$
$t_{pro4ST} = \bar{y}\frac{P+\rho_{Pb}}{\widehat{P}+\rho_{Pb}}$	1	ρ_{Pb}

We obtain the *MSE* equation for these proposed estimators as

$$MSE(t_{proi}) = \frac{1-f}{n} \overline{Y}^2 \left[C_y^2 + C_P^2 \psi_i(\psi_i - 2K_{Pb})\right] \quad i = 1,2,\ldots,5, \tag{2.11}$$

where $\psi_{1SD} = \frac{P}{P+C_P}$; $\psi_{2SK} = \frac{P}{P+B_2(\varphi)}$; $\psi_{3US1} = \frac{PB_2(\varphi)}{PB_2(\varphi)+C_P}$; $\psi_{3US2} = \frac{PC_P}{PC_P+B_2(\varphi)}$ and $\psi_{4ST} = \frac{P}{P+\rho_{Pb}}$.

3. Efficiency comparison

In this section, firstly, we compare *MSE* of proposed estimators, given in (2.9), with the *MSE* of ratio estimator presented in Singh et al. [5], given in (1.8). As we obtain the following condition by these comparison:

$$R_i^2 S_\varphi^2 > zero. \tag{3.1}$$

Table 1
Percent relative efficiencies of $\bar{y},\ t_{NG},\ t_i\ (i = 1,2,\ldots,5)$ and t_{proi} with respect to $\bar{y}$.

Estimator	PREs $(.,\bar{y})$ Population	
	I	II
$\bar{y}$	100	100
t_{NG}	7.124	7.804
t_1	4.931	5.695
t_2	237.766	161.873
t_3	221.45	155.231
t_4	72.098	69.341
t_5	215.271	152.612
t_{proi}	242.19	163.61

We can infer that all proposed estimators are more efficient than all ratio estimators presented in Singh et al. [5] in all conditions, because the condition given in (3.1) is always satisfied.

Secondly, we compare the *MSE* of the new estimators given in (2.11) with the variance of sample mean, so we have the following condition:

$$\begin{aligned}
&MSE(t_{proi}) < V(\bar{y}); \quad i = 1,2,\ldots,5, \quad \text{if},\\
&\psi_i - 2\rho_{Pb}\frac{C_y}{C_P} < zero,\\
&2\rho_{Pb}\frac{C_y}{C_P} > \psi_i,\\
&\therefore\ \rho_{Pb} > \frac{1}{2}\frac{C_P}{C_y}\psi_i, \quad i = 1,2,\ldots,5. \qquad (3.2)
\end{aligned}$$

When this condition is satisfied, proposed estimators are more efficient than the sample mean.

4. Empirical study

We now compare the performance of various estimators considered here using the two data sets as previously used by Shabbir and Gupta [4].

Population I (Source: Sukhatme and Sukhatme [6], p. 256).

y = Number of villages in the circles.
φ = A circle consisting more than five villages.
$N = 89,\ \overline{Y} = 3.36,\ P = 0.124,\ \rho_{Pb} = 0.766,$
$C_y = 0.601,\ C_P = 2.678,\ n = 23,\ B_2(\varphi) = 6.162,\ R_1 = 27.18,\ R_2 = 0.534,\ R_3 = 1.199,\ R_4 = 6.019,\ R_5 = 1.386.$

Population II (Source: Sukhatme and Sukhatme [6], p. 256).

y = Area (in acres) under wheat crop in the circles.
φ = A circle consisting more than five villages.
$N = 89,\ \overline{Y} = 1102,\ P = 0.124,\ \rho_{Pb} = 0.624,$
$C_y = 0.65,\ C_P = 2.678,\ n = 23,\ B_2(\varphi) = 6.162,\ R_1 = 8915,\ R_2 = 175.31,\ R_3 = 393.31,\ R_4 = 6.019,\ R_5 = 454.468.$

We have computed the percent relative efficiencies (PREs) of $\bar{y},\ t_{NG},\ t_i\ (i = 1,2,\ldots,5)$ and t_{proi} with respect to usual unbiased estimator $\bar{y}$ and displayed in Table 1.

From Table 1 it can be concluded that all proposed estimators $t_{proi}\ (i = 1,2,3,4)$ are more efficient than the usual unbiased estimator $\bar{y}$, ratio estimators of Naik and Gupta [3], and the ratio estimators presented in Singh et al. [5].

5. Conclusion

We have developed new estimators combining ratio estimators considered in Singh et al. [5] and obtained the minimum *MSE* equation for the proposed estimators. Theoretically, we have demonstrated that all proposed estimators are always more efficient than ratio estimators. In addition, we support this theoretical result numerically using the data used by Shabbir and Gupta [4].

Some other estimators can also be derived combining ratio estimators given in (1.4)–(1.7) in the form (2.1), but all these estimators have again the same minimum *MSE* equation given in (2.9). We would like to recall that R_1 and R_2 in (2.4) and in (2.8) should be changed according to ratio estimators that are combined.

Acknowledgements

The authors are deeply grateful to the referee and the editor of the journal for their extremely helpful comments and valued suggestions that led to this improved version of the paper.

References

[1] W.G. Cochran, Sampling Techniques, John Wiley and Sons, New York, 1977.
[2] C. Kadilar, H. Cingi, Improvement in estimating the population mean in simple random sampling, Applied Mathematics Letters 19 (2006) 75–79.
[3] V.D. Naik, P.C. Gupta, A note on estimation of mean with known population of an auxiliary character, Journal of Indian Society Agricultural Statistics 48 (2) (1996) 151–158.
[4] J. Shabbir, S. Gupta, On estimating the finite population mean with known population proportion of an auxiliary variable, Pakistan Journal of Statistics 23 (1) (2007) 1–9.
[5] R. Singh, P. Chauhan, N. Sawan, F. Smarandache, Ratio estimators in simple random sampling using information on auxiliary attribute, Pakistan Journal of Statistics and Operation Research IV (1) (2008) 47–53.
[6] P.V. Sukhatme, B.V. Sukhatme, Sampling Theory of Surveys with applications, Iowa State University Press, Ames, USA, 1970.
[7] K.M. Wolter, Introduction to Variance Estimation, second ed., Springer-Verlag, 1985.