International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

Some Unbiased Estimators for Estimating the Population Mean in Simple Random Sampling Using Information on Auxiliary Attribute

Monika S. Barak[1], Ashish K. Barak[2]

Department of Mathematics and Statistics, Manipal University, Jaipur-303007, India

Abstract: ***The main objective of this paper is to define some new unbiased estimators for estimating the population mean of the variable under study, which make use of information regarding the population proportion possessing certain attributes. Under simple random sampling without replacement scheme, the expressions for mean square error of the proposed estimators have been obtained, up to the first order of approximation. The conditions under which the proposed estimators are more efficient than the mean per unit estimator and ratio estimators in Naik and Gupta (1996) have been also obtained. The gains in efficiency over the existing ones have been illustrated numerically.***

Keywords: Simple random sampling, Auxiliary attributes, Percent relative efficiency, Study variable

1. Introduction

The estimation of the population mean is an unrelenting issue in sampling theory and several efforts have been made to improve the precision or accuracy of an estimator of unknown population parameter of interest when study variable y is highly correlated with the auxiliary variable x. There are many situations when auxiliary information is qualitative in nature that is auxiliary information is available in the form of an attribute, which is highly correlated with study variable. For example:

(a) Sex and height of the person.
(b) Amount of milk produced and a particular breed of the cow.
(c) Amount of yield of wheat crop and a particular variety of wheat etc.(see Jhajj et.al.[1])

In such situations, taking the advantage of point bi-serial correlation between the study variable y and the auxiliary attributes the estimators of population parameter of interest can be constructed by using prior knowledge of the population parameter of auxiliary attribute.

Now consider a finite population which consists of N identifiable units U_i ($1 \le i \le N$). Assume that a sample of size n drawn by using simple random sampling without replacement (SRSWOR) from a population of size N. Let y_i and $\emptyset_i$ denote the observations on the variable y and $\emptyset$ respectively for i^{th} unit (i=1,2,…N). Suppose there is a complete dichotomy in the population with respect to the presence or absence of an attribute, say $\emptyset$, and it is assumed that attribute $\emptyset$ takes only two values 0 and 1 according as

$\emptyset_i = 1$, if i^{th} unit of the population possesses attribute $\emptyset$
$= 0$ otherwise.

Let A= $\sum_i^N \emptyset_i$ and a = $\sum_i^n \emptyset_i$ denote the total number of units in the population and sample respectively possessing attribute $\emptyset$. Let P = $\frac{A}{N}$ and p = $\frac{a}{n}$ denote the proportion of units in the population and sample respectively possessing attribute $\emptyset$.

In order to have an estimate of population mean $\bar{Y}$of the study variable y, assuming the knowledge of the population proportion P, Naik and Gupta (1996) defined ratio estimator of population when the prior information of the population proportion of units, possessing the same attribute is available. Naik and Gupta (1996) proposed following estimator:

$$t_{NGR} = \bar{y}\left(\frac{P}{p}\right) \quad (1.1)$$

Here $\bar{y}$ is the sample mean of variable of interest. The mean square error (MSE) of $\bar{y}$ andt_{NG},up to the first order of approximation are given by-

$$\text{MSE}(\bar{y}) = \left(\frac{1-f}{n}\right) S_y^2 \quad (1.2)$$

$$\text{MSE}(t_{NGR}) = \left(\frac{1-f}{n}\right)\left\{S_y^2 + R^2 S_\emptyset^2 - 2R\rho_{y\emptyset} S_y S_\emptyset\right\} \quad (1.3)$$

where $f = \frac{n}{N}$: n is the sample size; N is the number of units in the population;$R = \frac{\bar{Y}}{P}$,$S_\emptyset^2 = \frac{1}{N-1}\sum_i^N(\emptyset_i - P)^2$ is the population variance of auxiliary attribute $\emptyset$;$S_{y\emptyset} = \frac{1}{N-1}\sum_i^N(y_i - \bar{Y})(\emptyset_i - P)$is the population covariance between variable of interest and auxiliary attribute $\emptyset$and$\rho_{y\emptyset} = \frac{S_{y\emptyset}}{S_y S_\emptyset}$ is the point bi-serial correlation coefficient.

In the present paper, some unbiased estimators for estimating the population mean of the variable under study, which make use of information regarding the population proportion certain attribute, are proposed. The expressions for MSE have been obtained. A comparison between all suggested estimators with other known estimators using real data set is considered.

Licensed Under Creative Commons Attribution CC BY

2. Suggested Estimators

Under the given sampling design, we propose the following estimators of population mean as

$$\text{(i)}\hat{\hat{Y}}_{ms1} = \bar{y} + \left(\frac{P}{p}\right) - 1 \qquad (2.1)$$

To obtain the characteristics of proposed estimators to the first degree of approximation,

we define $\delta_{\emptyset} = \frac{p}{P} - 1 \delta_y = \frac{\bar{y}}{\bar{Y}} - 1$

So we have $E(\delta_i) = 0, i = (y, \emptyset)$

$E(\delta_y^2) = \left(\frac{1-f}{n}\right)\frac{S_y^2}{\bar{Y}^2}$, $E(\delta_{\emptyset}^2) = \left(\frac{1-f}{n}\right)\frac{S_{\emptyset}^2}{P^2}$, $E(\delta_y \delta_{\emptyset}) = \left(\frac{1-f}{n}\right)\rho_{y\emptyset}\frac{S_y S_{\emptyset}}{\bar{Y}P}$

Expressing (2.1) in terms of $\delta' s$, we have

$$\hat{\hat{Y}}_{ms1} = \bar{Y}(1 + \delta_y) + (1 + \delta_{\emptyset})^{-1} - 1 \qquad (2.2)$$

Expanding the right hand side of (2.2) and retaining terms up to first powers of $\delta' s$, we have

$$\hat{\hat{Y}}_{ms1} = \bar{Y}(1 + \delta_y) + (1 - \delta_{\emptyset} + \cdots.) - 1 \qquad (2.3)$$

Taking expectation on (2.3), we obtain
$\hat{\hat{Y}}_{ms1} = \bar{Y}$
Now Var$(\hat{\hat{Y}}_{ms1}) = E(\hat{\hat{Y}}_{ms1} - E(\hat{\hat{Y}}_{ms1}))^2$
$= E[(\bar{Y}(1 + \delta_y) - (1 + \delta_{\emptyset})^{-1} - 1 - \bar{Y})]^2$
$= E[\bar{Y}(\delta_y) - \delta_{\emptyset}]^2$
$= \bar{Y}^2 E(\delta_y^2) + E(\delta_{\emptyset}^2) - 2\bar{Y} E(\delta_y \delta_{\emptyset})$

$$\text{Var}(\hat{\hat{Y}}_{ms1}) = \left(\frac{1-f}{n}\right)\left\{S_y^2 + \frac{S_{\emptyset}^2}{P^2} - 2\frac{\rho_{y\emptyset}S_y S_{\emptyset}}{P}\right\} \qquad (2.4)$$

$$\text{(ii)}\ \hat{\hat{Y}}_{ms2} = \bar{y} - e^{(p-P)} + 1 \qquad (2.5)$$

Expending the right hand side of (2.5), it is clear that $E(\hat{\hat{Y}}_{ms2}) = \bar{Y}$ so that $\hat{\hat{Y}}_{ms2}$ is an unbiased estimator of the population mean.
Now $V(\hat{\hat{Y}}_{ms2}) = E(\hat{\hat{Y}}_{ms2} - E(\hat{\hat{Y}}_{ms2}))^2$
$= E(\bar{y} - e^{(p-P)} + 1 - \bar{Y})^2$
$= E[(\bar{y} - \bar{Y}) - (1 + (p - P) + \cdots..) + 1]^2$(Neglecting the higher terms)
$= E[(\bar{y} - \bar{Y}) - (p - P)]^2$
$= E(\bar{y} - \bar{Y})^2 + E(p - P)^2 - 2E(\bar{y} - \bar{Y})(p - P)$

$$V(\hat{\hat{Y}}_{ms2}) = \left(\frac{1-f}{n}\right)\{S_y^2 + S_{\emptyset}^2 - 2\rho_{y\emptyset}S_y S_{\emptyset}\} \qquad (2.6)$$

$$\text{(iii)}\ \hat{\hat{Y}}_{ms3} = \bar{y} - e^{(P-p)} + 1 \qquad (2.7)$$

Expending the right hand side of (2.7), it is clear that $E(\hat{\hat{Y}}_{ms3}) = \bar{Y}$ so that $\hat{\hat{Y}}_{ms3}$ is an unbiased estimator of the population mean.

Now $V(\hat{\hat{Y}}_{ms3}) = E(\hat{\hat{Y}}_{ms3} - E(\hat{\hat{Y}}_{ms3}))^2$
$= E(\bar{y} - e^{(P-p)} + 1 - \bar{Y})^2$
$= E[(\bar{y} - \bar{Y}) - (1 + (P - p) + \cdots..) + 1]^2$(Neglecting the higher terms)
$= E[(\bar{y} - \bar{Y}) + (p - P)]^2$
$= E(\bar{y} - \bar{Y})^2 + E(p - P)^2 + 2E(\bar{y} - \bar{Y})(p - P)$

$$V(\hat{\hat{Y}}_{ms3}) = \left(\frac{1-f}{n}\right)\{S_y^2 + S_{\emptyset}^2 + 2\rho_{y\emptyset}S_y S_{\emptyset}\} \qquad (2.8)$$

3. Efficiency Comparisons

In this section, the conditions for which the proposed estimators $\hat{\hat{Y}}_{ms1}$, $\hat{\hat{Y}}_{ms2}$ and $\hat{\hat{Y}}_{ms3}$ are better than the mean square error and ratio estimator proposed by Naik and Gupta (1996).

1.1 Suggested Estimators vs. Mean per unit

From equation (1.2) and (2.4)

MSE $(\bar{y}) - V(\hat{\hat{Y}}_{ms1}) \geq 0$ if $\rho_{y\emptyset} > \frac{1}{2P}\frac{S_{\emptyset}}{S_y}$

Thus we arrive at the following theorem:

Theorem 3.1.1: The estimator$(\hat{\hat{Y}}_{ms1})$is more efficient than $\bar{y}$ if $\rho_{y\emptyset} > \frac{1}{2P}\frac{S_{\emptyset}}{S_y}$

Corollary 3.1.1: If $S_{\emptyset}$ and S_y are approximately the same then $(\hat{\hat{Y}}_{ms1})$ is more efficient than $(\bar{y})$ whenever

$$\rho_{y\emptyset} > \frac{1}{2P}$$

From equation (1.2) and (2.6), we obtain

MSE $(\bar{y}) - V(\hat{\hat{Y}}_{ms2}) \geq 0$ if $\rho_{y\emptyset} > \frac{1}{2}\frac{S_{\emptyset}}{S_y}$

Thus we arrive at the following theorem:

Theorem 3.1.2: The estimator$(\hat{\hat{Y}}_{ms2})$is more efficient than $\bar{y}$ if$\rho_{y\emptyset} > \frac{1}{2}\frac{S_{\emptyset}}{S_y}$.

Corollary 3.1.2: If $S_{\emptyset}$ and S_y are approximately the same then $(\hat{\hat{Y}}_{ms2})$ is more efficient than $(\bar{y})$ whenever

$$\rho_{y\emptyset} > \frac{1}{2}$$

From equation (1.2) and (2.8), we obtain

MSE $(\bar{y}) - V(\hat{\hat{Y}}_{ms3}) \geq 0$ if $\rho_{y\emptyset} < -\frac{1}{2}\frac{S_{\emptyset}}{S_y}$

Thus we arrive at the following theorem:

Theorem 3.1.3: The estimator$(\hat{\hat{Y}}_{ms3})$is more efficient than $\bar{y}$ if.

Licensed Under Creative Commons Attribution CC BY

$$\rho_{y\emptyset} < -\frac{1}{2}\frac{S_\emptyset}{S_y}$$

Corollary 3.1.3: If $S_\emptyset$ and S_y are approximately the same then $\left(\hat{\hat{\bar{Y}}}_{ms3}\right)$ is more efficient than $(\bar{y}\ )$ whenever

$$\rho_{y\emptyset} < -\frac{1}{2}$$

3.2Suggested Estimators vs. Ratio Estimator (Naik and Gupta (1996))

From equation (1.3) and (2.4)

MSE $(t_{NGR}) - V\left(\hat{\hat{\bar{Y}}}_{ms1}\right) \geq 0$ whenever $\rho_{y\emptyset} > \frac{1}{2P}\frac{S_\emptyset}{S_y}(1+\ \bar{Y})$

Thus we arrive at the following theorem:

Theorem 3.2.1: The estimator$\left(\hat{\hat{\bar{Y}}}_{ms1}\right)$is more efficient than t_{NGR}

whenever $\rho_{y\emptyset} > \frac{1}{2P}\frac{S_\emptyset}{S_y}(1+\ \bar{Y})$

Corollary 3.2.1: If $S_\emptyset$ and S_y are approximately the same then $\left(\hat{\hat{\bar{Y}}}_{ms1}\right)$ is more efficient than (t_{NGR}) whenever$\rho_{y\emptyset} > \frac{1}{2P}(1+\ \bar{Y})$.

From equation (1.3) and (2.6), we obtain

MSE $(t_{NGR}) - V\left(\hat{\hat{\bar{Y}}}_{ms2}\right) \geq 0$ if $\rho_{y\emptyset} > -\frac{1(1-R)}{2}\frac{S_\emptyset}{S_y}$.

Thus we arrive at the following theorem:

Theorem 3.2.2: The estimator$\left(\hat{\hat{\bar{Y}}}_{ms2}\right)$is more efficient than t_{NGR} if $\rho_{y\emptyset} > -\frac{1(1-R)}{2}\frac{S_\emptyset}{S_y}$.

Corollary 3.2.2: If $S_\emptyset$ and S_y are approximately the same then $\left(\hat{\hat{\bar{Y}}}_{ms2}\right)$ is more efficient than (t_{NGR}) whenever

$$\rho_{y\emptyset} > -\frac{1(1-R)}{2}$$

From equation (1.3) and (2.8), we obtain

MSE $(t_{NGR}) - V\left(\hat{\hat{\bar{Y}}}_{ms3}\right) \geq 0$ if $\rho_{y\emptyset} > \frac{1(1+R)}{2}\frac{S_\emptyset}{S_y}$

Thus we arrive at the following theorem

Theorem 3.2.3: The estimator$\left(\hat{\hat{\bar{Y}}}_{ms3}\right)$is more efficient than t_{NGR} if.

$\rho_{y\emptyset} > \frac{1(1+R)}{2}\frac{S_\emptyset}{S_y}$.

Corollary 3.2.3: If $S_\emptyset$ and S_y are approximately the same then $\left(\hat{\hat{\bar{Y}}}_{ms3}\right)$ is more efficient than (t_{NGR}) whenever

$$\rho_{y\emptyset} > \frac{1(1+R)}{2}$$

4. Numerical Illustration

To get a rough idea about the gain in efficiency for the proposed estimators. We compare the performance of various estimators considered here using the two data sets as previously used by Shabbir and Gupta(2010).

Population I (Source: Sukhatme and Sukhatme(1970), pp. 256).

y = Number of villages in the circles.
$\emptyset$ = A circle consisting more than five villages.
$N = 89, \bar{Y} = 3.36, P = 0.124, \rho_{y\emptyset} = 0.766,$
$n = 23, S_y = 2.019, S_\emptyset = 0.332, S_{y\emptyset} = 0.513$

Population II (Source: Sukhatme and Sukhatme(1970), pp. 256).

y = Area (in acres) under wheat crop in the circles.
$\emptyset$ = A circle consisting more than five villages.
$N = 89, \bar{Y} = 1102, P = 0.124, \rho_{y\emptyset} = 0.624$
$n = 23, S_y = 2.019, S_\emptyset = 0.332, S_{y\emptyset} = 0.418$

The percent relative efficiency (PRE's) of the suggested estimators with respect to the mean per unit estimator and ratio estimators in Naik and Gupta (1996) have been commuted and complied in table I and table II.

Table 1: Suggested Estimators vs. Mean per unit

Estimator	PREs Population - I Population – II
$\bar{y}$ $\hat{\hat{\bar{Y}}}_{ms1}$ $\hat{\hat{\bar{Y}}}_{ms2}$ $\hat{\hat{\bar{Y}}}_{ms3}$	100 100 130 124 132.65 134.13 114.04 111

Table 2: Suggested Estimators vs. Ratio Estimator (Naik and Gupta(1996))

Estimator	PREs Population - I Population – II
t_{NGR} $\hat{\hat{\bar{Y}}}_{ms1}$ $\hat{\hat{\bar{Y}}}_{ms2}$ $\hat{\hat{\bar{Y}}}_{ms3}$	100 100 120115.14 122.49118.69 119.49118.74

5. Conclusion

We have developed new estimators and obtained the minimum MSE equations for the proposed estimators. Theoretically, we have demonstrated that all proposed estimators are always more efficient than the mean square error and ratio estimators proposed by Naik and Gupta (1996).In addition, we support this theoretical result

Licensed Under Creative Commons Attribution CC BY

numerically using the data used by Shabbir and Gupta (2010).

References

[1] Jhajj, H.S., Sharma, M.K. and Grover, L.K. (2006), A family of estimators of population mean using information on auxiliary attribute. Pakistan journal of Statistics, Vol.22 (1), pp. 43-50.
[2] Naik, V.D. and Gupta, P.C. (1996), A note on estimating of mean with known population proportion of an auxiliary character, Journal of Indian Society of agricultural statistics, Vol. 48(2), pp. 151-158.
[3] P.V. Sukhatme, B.V. Sukhatme, (1970), Sampling Theory of Surveys with Applications, Lowa State University Press, Ames, USA.
[4] Shabbir, J. and Gupta, S. (2010). Estimation of the finite population mean in two phase sampling when auxiliary variables are attributes, Hecettepe Journal of Mathematics and Statistics, Vol. 39(1) 2010, pp. 121-129

Licensed Under Creative Commons Attribution CC BY