Ultra Scientist Vol. 25(2)B, 302-306 (2013).

A Family of Estimators of Population Mean with Two Phase Sampling Subject to Auxiliary Information is Attribute

MONIKA SAINI

(Acceptance Date 7th August, 2013)

Abstract

In this paper, a problem of estimating the finite population mean in two phase sampling when information regarding the population proportion possessing certain attributes is considered. 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 have also been obtained. The gains in efficiency over the existing ones have been illustrated numerically.

Key words: Simple random sampling, Auxiliary attributes, Percent relative efficiency, Study variable, Two phase sampling.

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.[1,2]

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 $U_i\ (1 \leq i \leq N)$. 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 $\varnothing_i$ denote the observations on the variable y and $\varnothing$ 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 $\varnothing$, and it is assumed that attribute $\varnothing$ takes only two values 0 and 1 according as

$\emptyset_i = 1$, if i^{th} unit of the population possesses attribute $\varnothing$

= 0 otherwise.

Let A= $\sum_i^N \varnothing_i$ and a = $\sum_i^N \varnothing_i$ denote the total number of units in the population and sample respectively possessing attribute $\varnothing$. Let P =$\frac{A}{N}$ and p = $\frac{a}{n}$ denote the proportion of units in the population and sample respectively possessing attribute $\varnothing$. Let $s_y^2 = \frac{1}{n-1}\sum_i^n (y_i - \bar{y})^2$ and $s_\emptyset^2 = \frac{1}{n-1}\sum_i^n (\emptyset_i - p)^2$ be sampling variance corresponding to the population variances $S_y^2 = \frac{1}{N-1}\sum_i^N (y_i - \bar{Y})^2$ and $S_\emptyset^2 = \frac{1}{N-1}\sum_i^N (\emptyset_i - P)^2$ respectively, where $\bar{y}$= $\frac{1}{n}\sum_i^n y_i$ and $\bar{Y}$= $\frac{1}{N}\sum_i^N y_i$

Let $s_{y\emptyset} = \frac{1}{n-1}\sum_i^n (y_i - \bar{y})(\emptyset_i - p)$

and $\hat{\rho}_{y\emptyset} = \frac{S_{y\emptyset}}{S_y S_\emptyset}$ be the sample point bi-serial covariance and bi- serial correlation between variable of interest y and auxiliary attribute $\varnothing$ corresponding to the population point bi-serial covariance and bi- serial correlation $S_{y\emptyset} = \frac{1}{N-1}$ $\sum_i^N (y_i - \bar{Y})(\emptyset_i - P)$ is the population covariance and $\rho_{y\emptyset} = \frac{S_{y\emptyset}}{S_y S_\emptyset}$, respectively.

It is assumed that population proportion P and population variance $S_\emptyset^2$ for the auxiliary attribute $\varnothing$ are unknown. In such situation, we can estimate P and $S_\emptyset^2$ from the sample by using a two phase sampling procedure as per jhajj[1,2], Kiregyra[5] and swain[6]. We use simple random sampling without replacement at both phases as described below:

(i) We draw a sample *s** of fixed size *n** from the population and observe $\varnothing$ and estimate P as well as $S_\emptyset^2$.

(ii) Given *s**, we draw a sample s (s⊆s*)of fixed size n and observe y.

Let $p^* = \frac{1}{n^*}\sum_i^{n^*} \varnothing_i$, $s_\emptyset^2 = \frac{1}{n-1}\sum_i^n (\emptyset_i - p)^2$ and $s_\emptyset^{*2} = \frac{1}{n^*-1}\sum_i^{n^*} (\emptyset_i - p^*)^2$.

In the present paper, some unbiased estimators for estimating the population mean of the variable under study, which make use estimated auxiliary 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.

2. Suggested Estimators :

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

(i) $\hat{\bar{Y}}_{d1} = \bar{y} + \left(\frac{p^*}{p}\right) - 1$ (2.1)

To obtain the characteristics of proposed estimators to the first degree of approximation, we define $\delta_y = \frac{\bar{y}}{\bar{Y}} - 1$, $\delta_1 = \frac{p}{P} - 1$ and $\delta_2 = \frac{p^*}{P} - 1$

So we have $E(\delta_i) = 0$, $i = (y, 1,2)$

and $E(\delta_y^2) = \left(\frac{1-f}{n}\right)\frac{S_y^2}{\bar{Y}^2}$, $E(\delta_1^2) = \left(\frac{1}{n} - \frac{1}{N}\right)\frac{S_\emptyset^2}{P^2}$,

$E(\delta_2^2) = \left(\frac{1}{n} - \frac{1}{n'}\right)\frac{S_\emptyset^2}{P^2}$ $E(\delta_1\delta_2) = \left(\frac{1}{n} - \frac{1}{n'}\right)\rho_{y\emptyset}\ \frac{S_y S_\emptyset}{\bar{Y}P}$

Expressing (2.1) in terms of δ's, we have

$\hat{\bar{Y}}_{d1} = \bar{Y}(1 + \delta_y) + (1 + \delta_2)(1 + \delta_1)^{-1} - 1$ (2.2)

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

$\hat{\bar{Y}}_{d1} = \bar{Y}(1 + \delta_y) + (1 + \delta_2)(1 + \delta_\emptyset + \cdots.) - 1$ (2.3)

Taking conditional expectation on (2.3), we obtain

$$\hat{\bar{Y}}_{d1} = \bar{Y}$$

Now to find out the sampling variance, we have to the first order of approximation

Var $(\hat{\bar{Y}}_{d1}) = E(\hat{\bar{Y}}_{d1} - E(\hat{\bar{Y}}_{d1}))^2$

Var $(\hat{\bar{Y}}_{d1}) = \left(\frac{1}{n} - \frac{1}{N}\right)S_y^2 + \left(\frac{1}{n} - \frac{1}{n'}\right)\left\{\frac{S_\emptyset^2}{P^2} - 2\ \frac{\rho_{y\emptyset} S_y S_\emptyset}{P}\right\}$ (2.4)

(ii) $\hat{\bar{Y}}_{d2} = \bar{y} - e^{(p-p^*)} + 1$ (2.5)

Expending the right hand side of (2.5), it is clear that $E\left(\hat{\bar{Y}}_{d2}\right) = E_1E_2\left(\hat{\bar{Y}}_{d2}\right) = \bar{Y}$ so that $\hat{\bar{Y}}_{d2}$ is conditionally unbiased estimator of the population mean.

Now conditional variance for $\hat{\bar{Y}}_{d2}$ is

$V\left(\hat{\bar{Y}}_{d2}\right) = V_1E_2\left(\hat{\bar{Y}}_{d2}\right) + E_1V_2\left(\hat{\bar{Y}}_{d2}\right)$

$= \left(\frac{1}{n} - \frac{1}{N}\right)S_y^2 + \left(\frac{1}{n} - \frac{1}{n'}\right)\{S_y^2 + S_\emptyset^2 - 2\rho_{y\emptyset}\ S_y S_\emptyset\}$

(Neglecting the higher terms)

After simplification the above equation can be written as

$V\left(\hat{\bar{Y}}_{d2}\right) = \left(\frac{1}{n} - \frac{1}{N}\right)S_y^2 + \left(\frac{1}{n} - \frac{1}{n'}\right)\{S_\emptyset^2 - 2\rho_{y\emptyset}\ S_y S_\emptyset\}$ (2.6)

(iii) $\hat{\bar{Y}}_{d3} = \bar{y} - e^{(p^*-p)} + 1$ (2.7)

Expending the right hand side of (2.7), it ` is clear that $E\left(\hat{\bar{Y}}_{d3}\right) = E_1E_2\left(\hat{\bar{Y}}_{d3}\right) = \bar{Y}$ so that $\hat{\bar{Y}}_{d3}$ is conditionally unbiased estimator of the population mean.

Now conditional variance for $\hat{\bar{Y}}_{d3}$ is

$V\left(\hat{\bar{Y}}_{d3}\right) = V_1E_2\left(\hat{\bar{Y}}_{d3}\right) + E_1V_2\left(\hat{\bar{Y}}_{d3}\right)$

$= \left(\frac{1}{n'} - \frac{1}{N}\right)S_y^2 + \left(\frac{1}{n} - \frac{1}{n'}\right)\{S_y^2 + S_\emptyset^2 + 2\rho_{y\emptyset}\ S_y S_\emptyset\}$

(Neglecting the higher terms)

After simplification the above equation can be written as

$V\left(\hat{\bar{Y}}_{d3}\right) = \left(\frac{1}{n} - \frac{1}{N}\right)S_y^2 + \left(\frac{1}{n} - \frac{1}{n'}\right)\{S_\emptyset^2 + 2\rho_{y\emptyset}\ S_y S_\emptyset\}$ (2.8)

3. Efficiency Comparisons :

In this section, the conditions for which the proposed estimators $\hat{\bar{Y}}_{d1}$, $\hat{\bar{Y}}_{d2}$ and $\hat{\bar{Y}}_{d3}$ are better than the mean per unit.

The variance of is given by

Var $(\bar{y}) = \left(\frac{1}{n} - \frac{1}{N}\right)S_y^2$ (3.1)

Suggested Estimators vs. Mean per unit :

From equation (3.1) and (2.4)

$V(\bar{y}) - V\left(\hat{\bar{Y}}_{d1}\right) > 0$ if $\rho_{y\emptyset} > \frac{1}{2P}\frac{S_\emptyset}{S_y}$

Thus we arrive at the following theorem:

Theorem 3.1: The estimator $\left(\hat{\bar{Y}}_{d1}\right)$ is more efficient than $\bar{y}$ if $\rho_{y\emptyset} > \frac{1}{2P}$

Corollary 3.1: If $S_\varnothing$ and S_y are approximately the same then $\left(\hat{\bar{Y}}_{d1}\right)$ is more efficient than $(\bar{y})$ whenever

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

From equation (3.1) and (2.6), we obtain

$V(\bar{y}) - V\left(\hat{\bar{Y}}_{d2}\right) > 0$ if $\rho_{y\emptyset} > \frac{1}{2}\frac{S_\emptyset}{S_y}$

Thus we arrive at the following theorem:

Theorem 3.2: The estimator $\left(\hat{\bar{Y}}_{d2}\right)$ is more efficient than $\bar{y}$ if $\rho_{y\emptyset} > \frac{1}{2}\frac{S_\emptyset}{S_y}$.

Corollary 3.2: If $S_\varnothing$ and S_y are approximately the same then $\left(\hat{\bar{Y}}_{d2}\right)$ is more efficient than $(\bar{y})$ whenever

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

From equation (3.1) and (2.8), we obtain

$V(\bar{y}) - V\left(\hat{\bar{Y}}_{d3}\right) > 0$ if $\rho_{y\emptyset} < -\frac{1}{2}\frac{S_\emptyset}{S_y}$

Thus we arrive at the following theorem:

Theorem 3.3: The estimator $\left(\hat{\bar{Y}}_{d3}\right)$ is more efficient than $\bar{y}$ if.

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

Corollary 3.3: If $S_\varnothing$ and S_y are approximately the same then $\left(\hat{\bar{Y}}_{d3}\right)$ is more efficient than $(\bar{y})$ whenever

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

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[4].

Population[3] (Source: Sukhatme and Sukhatme(1970), pp. 256).

y = Number of villages in the circles.
∅=A circle consisting more than five villages.
$N = 89,\ \bar{Y} = 3.36,\ P = 0.124, \rho_{y\emptyset} = 0.766,$ $n' = 45,\ n = 23,\ S_y = 2.019\ S_\emptyset = 0.332,$ $S_{y\emptyset} = 0.513$

The percent relative efficiency (PRE's) of the suggested estimators with respect to the mean per unit estimator have been computed and complied in table 1.

Table 1. Suggested Estimators vs. Mean per unit

Population (Source: Sukhatme and Sukhatme(1970), pp. 256)	Estimator PREs			
	$\bar{y}$	$\hat{\bar{Y}}_{d1}$	$\hat{\bar{Y}}_{d2}$	$\hat{\bar{Y}}_{d3}$
	100	120	133	104

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. In addition, we support this theoretical result numerically using the data used by Shabbir and Gupta[4].

References

1. Jhajj, H.S., Sharma, M.K. and Grover, L.K., An efficient class of chain estimators of population variance under sub-sampling scheme, Journal of Japan Statistical Society *35(2)*, 273-286 (2005).
2. Jhajj, H.S., Sharma, M.K. and Grover, L.K., A family of estimators of population mean using information on auxiliary attribute. Pakistan journal of Statistics, Vol. *22(1)*, pp. 43-50 (2006).
3. P.V. Sukhatme, B.V. Sukhatme, Sampling Theory of Surveys with Applications, Lowa State University Press, Ames, USA (1970).
4. Shabbir, J. and Gupta, S., Estimation of the finite population mean in two phase sampling when auxiliary variables are attributes, Hecettepe Journal of Mathematics and Statistics, Vol. *39(1)*, pp. 121-129 (2010).
5. Kiregyra, B., Regression type estimators using two auxiliary variables and the models of double sampling from finite populations, Metrika, *31(3-4)*, 215-226 (1984).
6. Swain, A.K.P.C., A note on the use of multiple auxiliary variables in sample surveys, Trabajos de Estadistica, *30*, 135-141 (1970).