International Journal of Applied Mathematics
& Statistical Sciences (IJAMSS)
ISSN(P): 2319-3972; ISSN(E): 2319-3980
Vol. 6, Issue 1, Dec – Jan 2017; 19-32

International Academy of Science, Engineering and Technology
Connecting Researchers; Nurturing Innovations

ESTIMATION OF POPULATION PROPORTION USING TWO PHASE SAMPLING SCHEME IN THE PRESENCE OF NON-RESPONSE

B. B. KHARE[1], P. S. JHA[2] & S. KHARE[3]

[1,3]Department of Statistics, Banaras Hindu University, Varanasi, India

[2]Department of Molecular & Human Genetics, Banaras Hindu University, Varanasi, India

ABSTRACT

In this paper we have proposed two classes of two phase sampling estimators for population proportion using auxiliary character in the presence of non-response on the study character. Some members of the proposed classes of estimators are given. The expressions for bias and mean square error of the proposed classes of estimators are obtained in case of fixed sample size. Also we obtained the optimum value of sample size and sampling fraction for fixed cost and fixed precision. An empirical study has been made with the support of the proposed class of estimators.

KEYWORDS: Non-Response, Mean Square Error, Relative Bias, Auxiliary Characters, Attribute, Two Phase Sampling

1. INTRODUCTION

Sometimes, it may not be possible to collect the complete information for all the units selected in the sample due to lack of interest, person not present at home, lack of knowledge regarding the survey, ethical problems and due to refusal of the respondent to respond for the given questionnaire. In these situations, Hansen and Hurwitz (1946) have suggested a technique of sub sampling units from non-respondents. Using Hansen and Hurwitz (1946) technique, Cochran (1977) and Rao (1986, 1990) have proposed the estimators for population mean in case of known population mean of auxiliary character.

In sample survey, the use of auxiliary character has a significant role in the estimation of population parameters such as population mean, ratio and product of two population means, coefficient of variation etc. Several research works have been done for the estimation of population parameter using auxiliary character in the presence of non-response when the population mean of the auxiliary character is known / unknown. Further several research works on the estimation of population mean have been done by Khare and Srivastava (1993, 1995, 2000, 2010) and Singh and Kumar (2009, 2010). Sometimes auxiliary information is available in the form of auxiliary attribute which may also be used in the estimation of population mean of the study character. With the help of auxiliary attribute, several authors have proposed different type of estimators for the estimation of population mean of the study character. In this situation Naik and Gupta (1996), Jhajj et al. (2006), Shabbir and Gupta (2007, 2010), Singh et.al (2008), Koyuncu (2012), Abd-Elfattah et al. (2010), Singh and Solanki (2012), Malik and Singh (2013) and Sinha and Kumar (2013, 2014) have proposed different type of estimators for population mean using the information on auxiliary character as a qualitative character.

In sample surveys it generally happens that the study variable is also available in the form of qualitative characteristic such as smoking habit, educational status, socio economic group etc. For example if we want to estimates the proportion of woman smoking in a city, we may take their educational status, socio economic group and monthly expenditure etc as auxiliary character. In this situation Singh et al. (2010) have suggested a family of estimators for the population proportion. Further, Khare et al. (2015) have proposed two classes of estimators for population proportion P using auxiliary character x in the presence of non-response which are given as follow:

$$T_{c1} = H(w, v_1) \text{ and } T_{c2} = H(w, v_2) \tag{1.1}$$

Such that $H_i(P,1) = P$ and $H_{1i}(P,1) = P$, where $w = p^*, v_1 = \frac{\bar{x}^*}{\bar{X}}$, $v_2 = \frac{\bar{x}}{\bar{X}}$ and $H_{1i}(P,1) = \left(\frac{\partial H(w, v_i)}{\partial w}\right)_{(P,1)}$, $\forall i = 1, 2$. Here p^* and x^* denote the estimators for P and $\bar{X}$ using Hansen and Hurwitz (1946) technique and $\bar{x}$ and $\bar{X}$ denote the sample mean based on n units and population mean of the auxiliary character x.

In this paper, we have proposed two classes of two phase sampling estimators for population proportion using auxiliary character in the presence of non-response on the study character. The expressions for the bias and mean square error of the proposed classes of estimators are obtained and the properties of the proposed classes of estimators are studied for fixed sample sizes (n', n), for a fixed cost ($C \le C_0$) and also for a specified precision ($V = V_0$). Some members of the proposed classes of estimators are also obtained and their properties are studied. An empirical study is also given in the support of proposed classes of the estimators for P.

2. THE ESTIMATORS

Let ϕ_i and x_i denote the value of study attribute and auxiliary character for i[th] unit of the population (i=1, 2, ..., ..., N). In this case ϕ_i will take value 1 if it possessing the attribute otherwise zero. The population is supposed to be divided into N_1 responding and N_2 non-responding units such that $N_1 + N_2 = N$.

Let $P = \sum_{i=1}^{N} \phi_i / N$ be the proportion of the units in the population possessing the attribute ϕ, $P_1 = \sum_{i=1}^{N_1} \phi_i / N_1$ be the proportion of the units in the responding part of the population possessing the attribute ϕ and $P_2 = \sum_{i=1}^{N_2} \phi_i / N_2$ is the proportion of the units in the non responding part of the population possessing the attribute ϕ.

When population mean $(\bar{X})$ of auxiliary character x is unknown, then it is suggested to select a first phase sample of size $n'(< N)$ from the population of size N by using SRSWOR method of sampling and estimate the population mean $\bar{X}$ by using $\bar{x}'$, where $\bar{x}'$ denotes the sample mean based on n' observations. Again a second phase sample of size $n(< n')$ is selected from the first phase sample of size n' by using SRSWOR method of sampling. We observed that for the study attribute ϕ, only n_1 units are responding and n_2 units are not responding in the sample of size n. Again we draw a subsample of size $r(= n_2 / k, k > 1)$ from n_2 non-responding units and collect information on r

units on the study character using personal interview method and applying extra efforts. Now using Hansen and Hurwitz (1946) technique, we propose the estimator for population proportion P which is given as follows:

$$p^{*} = \frac{n_1}{n} p_1 + \frac{n_2}{n} p_2', \tag{2.1}$$

where $p_1 = \sum_{i=1}^{n_1} \phi_i / n_1$ is the proportion of the units possessing the attribute ϕ for n_1 responding units in the sample of size n and $p_2' = \sum_{i=1}^{r} \phi_i / r$ is the proportion of the units possessing the attribute ϕ for r sub-sampled units from n_2 non respondents.

The variance of p^{*} is given as follows:

$$V(p^{*}) = \frac{f}{n} S_{\phi}^{2} + \frac{W_2(k-1)}{n} S_{\phi(2)}^{2}, \tag{2.2}$$

Where $W_2 = N_2 / N$, $f = 1 - n/N$, $Q = 1 - P$, $Q_2 = 1 - P_2$, $S_{\phi}^{2} = \frac{N-n}{n(n-1)} PQ$ and $S_{\phi(2)}^{2} = \frac{N_2 - n_2}{n_2(n_2 - 1)} P_2 Q_2$ are the population mean square of the attribute ϕ for the whole population and for the non-responding part of the population.

Similarly, for the corresponding $n_1 + r$ units related to the study attribute ϕ, the estimator $\bar{x}^{*}$ for the population mean $\bar{X}$ is given by

$$\bar{x}^{*} = \frac{n_1}{n} \bar{x}_1 + \frac{n_2}{n} \bar{x}_2', \tag{2.3}$$

Where $\bar{x}_1$ and $\bar{x}_2'$ are the sample means of the auxiliary character based on n_1 responding units and r sub sample units from n_2 non responding units.

The variance of $\bar{x}^{*}$ is given as follows:

$$V(\bar{x}^{*}) = \frac{f}{n} S_{x}^{2} + \frac{W_2(k-1)}{n} S_{x(2)}^{2}, \tag{2.4}$$

Where S_x^2 and $S_{x(2)}^2$ are the population mean squares of auxiliary character x for the whole population and for the non-responding part of the population.

The estimators for the population mean $\bar{X}$ based on sample size n' and n are given by

$$\bar{x}' = \frac{1}{n'}\sum_{i=1}^{n'} x_i \text{ And } \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{2.5}$$

And their variances are given by

$$V(\bar{x}') = \frac{f'}{n'} S_x^2 \text{ And } V(\bar{x}) = \frac{f}{n} S_x^2 \tag{2.6}$$

Where $f' = 1 - n'/N$

In case when population mean $\bar{X}$ of auxiliary character x is unknown then we estimate it by $\bar{x}'$. Further in case of non-response in the sample of size n drown from n', two situations arises

- When incomplete information on study attribute ϕ and the information on the corresponding $n_1 + r$ units of auxiliary character x is used from the sub-sample of size n selected from n' first phase sample. Now we propose the class of estimators T_{u1} for population proportion P using auxiliary character x which is given as follows:

$$T_{u1} = H(w, u_1), \text{ such that } H(P, 1) = P \text{ and } H_1(P, 1) = 1, \tag{2.7}$$

where $w = p^*$, $u_1 = \frac{\bar{x}^*}{\bar{x}'}$ and $H_1(P,1) = \left(\frac{\partial H(w, u_1)}{\partial w}\right)_{(P,1)}$.

- When incomplete information on $n_1 + r$ units of the study attributes ϕ and the complete information on auxiliary character x from the selected sample of size n from n' is used and we propose the class of estimators T_{u2} for the population proportion P using auxiliary character x which is given as follows:

$$T_{u2} = H(w, u_2), \text{ such that } H(P, 1) = P \text{ and } H_1(P, 1) = 1, \tag{2.8}$$

Where $u_2 = \frac{\bar{x}}{\bar{x}'}$, and $H_1(P,1) = \left(\frac{\partial H(w, u_2)}{\partial w}\right)_{(P,1)}$.

The function $H(w, u_i)\ i = 1, 2$ also satisfies the following regularity conditions:

- Whenever be the sample chosen, w and u_i assume positive values in a bounded closed convex subset D_i of the two dimensional real space containing the point $(P, 1)$.
- The first and second partial derivatives of $H(w, u_i)$ with respect to w and u_i exit and are assumed to be continuous and bounded in two dimensional real spaces D_i. (2.9)

Now, expanding $H(w, u_i)$ about the point $(\bar{Y}, 1) = M$ by using Taylor's series up to the second order derivatives, we have

$$T_{ui} = H(P,1) + (w-P)H_1(M) + (u_i-1)H_{2(i)}(M)$$

$$+\frac{1}{2}\left\{(w-P)^2H_{11}(m^*)+(u_i-1)^2H_{22(i)}(m^*)+2(w-P)(u_i-1)H_{12(i)}(m^*)\right\} \quad (2.10)$$

After putting $H(P,1)=P$, $H_1(M)=1$ and $H_{11}(M)=0$, we have

$$T_{ui} = w+(u_i-1)H_{2(i)}(M)+\frac{1}{2}\{(u_i-1)^2H_{22(i)}(m^*)+2(w-P)(u_i-1)H_{12(i)}(m^*)\} \quad (2.11)$$

Where

$$H_1(M)=\left(\frac{\partial H(m)}{\partial w}\right)_M, \quad H_{2(i)}(M)=\left(\frac{\partial H(m)}{\partial u_i}\right)_M, \quad H_{11}(m^*)=\frac{\partial^2 H(m^*)}{\partial w^2},$$

$$H_{22(i)}(m^*)=\frac{\partial^2 H(m^*)}{\partial u_i^2}, \quad H_{12(i)}(m^*)=\frac{\partial^2 H(m^*)}{\partial w \partial u_i}, \quad m=(w,u_i), \quad m^*=(w^*,u_i^*), \quad w^*=P+\theta_i(w-P),$$

$u_i^*=1+\theta_i(u_i-1)$, $0<\theta_i<1$.

3. BIAS AND MEAN SQUARE ERROR (MSE) OF THE PROPOSED CLASSES OF ESTIMATORS

Under the regularity conditions specified for the function $H(w,u_i)$, bias and mean square error of the estimator T_{ui} always exists. In order to derive the expressions for bias and mean square error of the estimators, we use the large sample approximations.

Let $p^*=P(1+\varepsilon_0)$, $\bar{x}^*=\bar{X}(1+\varepsilon_1)$, $\bar{x}=\bar{X}(1+\varepsilon_2)$, $\bar{x}'=\bar{X}(1+\varepsilon_3)$, such that $E(\varepsilon_\ell)=0$ and $|\varepsilon_\ell|<1$ $\forall\ \ell=0,1,2,3$.

Now, using SRSWOR method of sampling, we have

$$E(\varepsilon_0^2)=\frac{f}{n}C_\phi^2+\frac{W_2(k-1)}{n}C_{\phi(2)}^2, \quad E(\varepsilon_1^2)=\frac{f}{n}C_x^2+\frac{W_2(k-1)}{n}C_{x(2)}^2, \quad E(\varepsilon_2^2)=\frac{f}{n}C_x^2,$$

$$E(\varepsilon_3^2)=\frac{f'}{n'}C_x^2,\ E(\varepsilon_0\varepsilon_1)=\frac{f}{n}C_{\phi x}+\frac{W_2(k-1)}{n}C_{\phi x(2)},\ E(\varepsilon_0\varepsilon_2)=\frac{f}{n}C_{\phi x},\ E(\varepsilon_0\varepsilon_3)=\frac{f'}{n'}C_{\phi x},$$

where $C_\phi^2=\frac{S_\phi^2}{P^2}$, $C_{\phi(2)}^2=\frac{S_{\phi(2)}^2}{P^2}$, $C_x^2=\frac{S_x^2}{\bar{X}^2}$, $C_{x(2)}^2=\frac{S_{x(2)}^2}{\bar{X}^2}$, $C_{\phi x}=\rho_{\phi x}\frac{S_\phi S_x}{P\bar{X}}$,

$C_{\phi x(2)}=\rho_{\phi x(2)}\frac{S_{\phi(2)}S_{x(2)}}{P\bar{X}}$ and $\rho_{\phi x}$ and $\rho_{\phi x(2)}$ are the point-by serial correlation coefficient between ϕ and x for the whole population and for the non-responding part of the population.

Now retaining the terms of ε_0, ε_1 and ε_2, upto second order only in the expression of bias and mean square error, the expressions for bias and mean square error of the estimator T_{ui} up to terms of order (n^{-1}) are given as follows:

$$Bias(T_{u1}) = \frac{1}{2}\left[\left\{\left(\frac{1}{n}-\frac{1}{n'}\right)C_x^2 + \frac{W_2(k-1)}{n}C_{x(2)}^2\right\}H_{22(1)}(M) + 2P\left\{\left(\frac{1}{n}-\frac{1}{n'}\right)C_{\phi x} + \frac{W_2(k-1)}{n}C_{\phi x(2)}\right\}H_{12(1)}(M)\right], \tag{3.1}$$

$$Bias(T_{u2}) = \frac{1}{2}\left[\left\{\left(\frac{1}{n}-\frac{1}{n'}\right)C_x^2\right\}H_{22(1)}(M) + 2P\left\{\left(\frac{1}{n}-\frac{1}{n'}\right)C_{\phi x}\right\}H_{12(1)}(M)\right], \tag{3.2}$$

$$MSE(T_{u1}) = V(p^*) + \left\{\left(\frac{1}{n}-\frac{1}{n'}\right)C_x^2 + \frac{W_2(k-1)}{n}C_{x(2)}^2\right\}H_{2(1)}^2(M) + 2P\left\{\left(\frac{1}{n}-\frac{1}{n'}\right)C_{\phi x} + \frac{W_2(k-1)}{n}C_{\phi x(2)}\right\}H_{2(1)}(M), \tag{3.3}$$

and

$$MSE(T_{u2}) = V(p^*) + \left(\frac{1}{n}-\frac{1}{n'}\right)\left[H_{2(2)}^2(M)C_x^2 + 2PH_{2(2)}(M)C_{\phi x}\right]. \tag{3.4}$$

The optimum values of $H_{2(1)}(M)$ and $H_{2(2)}(M)$ which minimize the MSE of the estimator T_{ui}, $i = 1, 2$ are given as follows:

$$H_{2(1)}(M)_{opt} = -P\frac{\left(\frac{1}{n}-\frac{1}{n'}\right)C_{\phi x} + \frac{W_2(k-1)}{n}C_{\phi x(2)}}{\left(\frac{1}{n}-\frac{1}{n'}\right)C_x^2 + \frac{W_2(k-1)}{n}C^2{}_{x(2)}} \tag{3.5}$$

and

$$H_{2(2)}(M)_{opt} = -P\frac{C_{\phi x}}{C_x^2}. \tag{3.6}$$

Now putting the optimum value of $H_{2(1)}(M)_{opt}$ and $H_{2(2)}(M)_{opt}$ in the expression of minimum value of $MSE(T_{u1})$ and $MSE(T_{u2})$ given by (3.3) and (3.4) respectively, we get

$$MSE(T_{u1})_{\min} = V(p^*) - P^2\frac{\left\{\left(\frac{1}{n}-\frac{1}{n'}\right)C_{\phi x} + \frac{W_2(k-1)}{n}C_{\phi x(2)}\right\}^2}{\left(\frac{1}{n}-\frac{1}{n'}\right)C_x^2 + \frac{W_2(k-1)}{n}C_{x(2)}^2} \tag{3.7}$$

and

$$MSE(T_{u2})_{\min} = V(p^*) - P^2\left(\frac{1}{n} - \frac{1}{n'}\right)\frac{(C_{\phi x})^2}{C_x^2} \quad (3.8)$$

The $MSE(T_{u1})_{\min}$ given by (3.7) may be approximated to

$$MSE(T_{u1})_{\min} = V(p^*) - P^2\frac{C_{\phi x}}{C_x^2}\left\{\left(\frac{1}{n} - \frac{1}{n'}\right)C_{\phi x} + \frac{W_2(k-1)}{n}C_{\phi x(2)}\right\}, \text{ for } \frac{C_{\phi x}}{C_x^2} = \frac{C_{\phi x(2)}}{C_{x(2)}^2} \quad (3.9)$$

Also, the expression of $MSE(T_{u1})_{\min}$ and $MSE(T_{u2})_{\min}$ from equation (3.9) and (3.8) is expressed in the form of the coefficient of $\frac{1}{n}$ and $\frac{1}{n'}$ which are given by

$$MSE(T_{u1})_{\min} = \frac{1}{n}P^2\{(C_\phi^2 + W_2(k-1)C_{\phi_2}^2) - \frac{(C_{\phi x})^2}{C_x^2} - W_2(k-1))\frac{(C_{\phi x}C_{\phi x_2})}{C_x^2}\} + \frac{1}{n'}P^2\frac{(C_{\phi x})^2}{C_x^2} - \frac{1}{N}C_\phi^2 \quad (3.10)$$

$$MSE(T_{u2})_{\min} = \frac{1}{n}P^2\{(C_\phi^2 + W_2(k-1)C_{\phi_2}^2) - \frac{(C_{\phi x})^2}{C_x^2})\} + \frac{1}{n'}P^2\frac{(C_{\phi x})^2}{C_x^2} - \frac{1}{N}C_\phi^2 \quad (3.11)$$

The optimum value of $H_{2(1)}(M)_{opt}$ and $H_{2(2)}(M)_{opt}$ in equation (3.5) and (3.6) are in the form of the value of unknown parameters. Sometimes the unknown constant used in the estimators are in the form of some unknown parameters. In these situations, the optimum value of the constant may be obtained from the past data (Reddy 1978) and if the information on these parameters are not available from past data, then one may estimate it on the basis of sample observations without having any loss in efficiency of the proposed estimators. It has been shown that up to the terms of order (n^{-1}), the minimum values of the mean square error of the estimator is unchanged if we estimate the optimum values of constants by using the sample values (Srivastava and Jhajj 1983). Any parametric function $H(w, u_i)$ satisfying the regularity condition (2.7) can generate a class of asymptotic estimators. Some members of the proposed class of estimators are given as follows:

$T_{i1} = w(a + (1-a)u_i)$, $T_{i2} = w(u_i^\alpha)$,

$T_{i3} = (w + a_1(1-u_i))u_i^\beta$, $T_{i4} = w(2 - u_i^\gamma)$,

$$T_{i5} = w\left[1 + \alpha_0\left(\frac{u_i - 1}{u_i + 1}\right)\right] \text{ and } T_{i6} = w\exp\left(\frac{u_i - 1}{u_i + 1}\right), \; i = 1,2 \quad (3.12)$$

Where $a, a_1, \alpha, \alpha, \beta$ and γ are constants

4. DETERMINATION OF n', n AND k FOR A FIXED COST ($C \le C_0$)

Let C_0 be the total cost (fixed) of the survey apart from the overhead cost. The cost function C' for the cost incurred in the survey apart from overhead expenses, can be expressed by

$$C' = C_1'n' + C_1 n + C_2 n_1 + C_3 \frac{n_2}{k} \tag{4.1}$$

Since C' will vary from sample to sample, so we consider the expected cost C to be incurred in the survey apart from overhead expenses, which is given by

$$C = E(C') = C_1'n' + n\left[C_1 + C_2 W_1 + C_3 \frac{W_2}{k}\right], \tag{4.2}$$

Where

C_1': The cost per unit of identifying and observing auxiliary variable,

C_1: The cost per unit of mailing questionnaire/visiting the unit at the second phase,

C_2: The cost per unit of collecting and processing data for the study variable ϕ for responding units and

C_3: The cost per unit of obtaining and processing data for the study variable ϕ (after extra efforts) from sub-sampled units from n_2 non-responding units.

From equation (3.9) and (3.8) the $MSE(T_{ui})_{\min}$, $i = 1,2$; can be expressed in terms of the notation Ψ_{0i}, Ψ_{1i} and Ψ_{2i} which is given as

$$MSE(T_{ui}) = \frac{1}{n'}\Psi_{1i} + \frac{1}{n}\Psi_{0i} + \frac{k}{n}\Psi_{2i} + \text{terms of independent of } n' \text{ and } n, \tag{4.3}$$

Where $\Psi_{1i} =$ coefficient of $\frac{1}{n'}$ terms, $\Psi_{0i} =$ coefficient of $\frac{1}{n}$ terms and $\Psi_{2i} =$ coefficient of $\frac{k}{n}$ terms.

Now we minimize MSE for the fixed cost $C \le C_0$ and to obtain the optimum values of n', n and k.

Let us consider a function

$$\Omega = MSE(T_{ui}) + \lambda_i\left\{C_1'n' + n\left(C_1 + C_2 W_1 + C_3 \frac{W_2}{k}\right) - C_0\right\},\ i = 1,2 \tag{4.4}$$

Where λ_i are Lagrange's multipliers

Now differentiating Ω with respect to n', n and k and equating to zero, we get

$$n'_{opt} = \sqrt{\frac{\Psi_{1i}}{\lambda_i C'_1}} \text{ ,(4.5)}$$

$$n_{opt} = \sqrt{\frac{\Psi_{0i} + k_{opt}\Psi_{2i}}{\lambda_i\left(C_1 + C_2W_1 + C_3\frac{W_2}{k_{opt}}\right)}}, \tag{4.6}$$

And $\frac{n}{k} = \sqrt{\frac{\Psi_{2i}}{\lambda_i C_3 W_2}}$ (4.7)

Solving (4.6) and (4.7), we have

$$k_{opt.} = \sqrt{\frac{\Psi_{0i}C_3W_2}{\Psi_{2i}(C_1 + C_2W_1)}} \tag{4.8}$$

Putting the values of n', n and k in (4.4), we have

$$\sqrt{\lambda_i} = \frac{1}{C_0}\left[\sqrt{\Psi_{1i}C'_1} + \sqrt{(\Psi_{0i} + k_{opt}\Psi_{2i})(C_1 + C_2W_1 + C_3\frac{W_2}{k_{opt.}})}\right] \tag{4.9}$$

The minimum value of $MSE(T_i), i = 1,2$ is given by

$$MSE(T_{ui})_{\min.} = \frac{1}{C_0}\left[\sqrt{\Psi_{1i}C'_1} + \sqrt{(\Psi_{0i} + k_{opt.}\Psi_{2i})(C_1 + C_2W_1 + C_3\frac{W_2}{k_{opt.}})}\right]^2 - \frac{S_\phi^2}{N} \tag{4.10}$$

In case of P^*, the expected total cost is given by

$$C = E(C') = n\left(C_1 + C_2W_1 + C_3\frac{W_2}{k}\right) \tag{4.11}$$

and $MSE(P^*)_{\min.} = \frac{1}{C_0}\left[\sqrt{(S_\phi^2 + W_2(k-1)S_{\phi 2}^2)(C_1 + C_2W_1 + C_3\frac{W_2}{k})}\right]^2 - \frac{S_\phi^2}{N}$ (4.12)

5. DETERMINATION OF n', n AND k FOR A SPECIFIED VARIANCE $V = V'_0$

Let V'_0 be the variance of the estimator $MSE(T_{ui})_{\min}$, $i = 1,2$ fixed in advance and we have

$$V'_0 = \frac{1}{n}\Psi_{0i} + \frac{1}{n'}\Psi_{1i} + \frac{k}{n}\Psi_{2i} - \frac{S_\phi^2}{N} \tag{5.1}$$

For minimizing the average total cost C for the specified variance of the Estimator $MSE(T_{ui})_{\min}$, $i = 1,2$, we define a function

$$\Omega_1 = \left\{ C_1' n' + n\left(C_1 + C_2 W_1 + C_3 \frac{W_2}{k} \right) \right\} + \mu_i \{ MSE(T_{ui}) - V_0' \}, \tag{5.2}$$

where $\mu_i \ (i = 1,2)$ are Lagrange multiplier.

Now differentiating with respect to n', n and k and equating to zero, we have

$$n'_{opt} = \sqrt{\frac{\mu_i \Psi_{1i}}{C_1'}}, \tag{5.3}$$

$$n_{opt} = \sqrt{\frac{\mu_i (\Psi_{0i} + k_{opt} \Psi_{2i})}{\left(C_1 + C_2 W_1 + C_3 \frac{W_2}{k_{opt.}} \right)}}, \tag{5.4}$$

and $$\frac{n}{k} = \sqrt{\frac{\mu_i \Psi_{2i}}{C_3 W_2}} \tag{5.5}$$

From (5.2) , (5.4) and (5.5), we have

$$k_{opt.} = \sqrt{\frac{\Psi_{0i} C_3 W_2}{\Psi_{2i} (C_1 + C_2 W_1)}}, \tag{5.6}$$

and $$\sqrt{\mu_i} = \frac{1}{\left[V_0' + \frac{S_\phi^2}{N} \right]} \left[\sqrt{C_1' \Psi_{1i}} + \sqrt{(\Psi_{0i} + k_{opt.} \Psi_{2i}) \left(C_1 + C_2 W_1 + C_3 \frac{W_2}{k_{opt.}} \right)} \right] \tag{5.7}$$

The minimum expected total cost for the specified variance V_0' for the estimator $MSE(T_{ui})_{\min}$, $i = 1,2$ is given by

$$C(T_{ui})_{\min.} = \frac{\left[\sqrt{C_1' \Psi_{1i}} + \sqrt{(\Psi_{0i} + k_{opt.} \Psi_{2i}) \left(C_1 + C_2 W_1 + C_3 \frac{W_2}{k_{opt.}} \right)} \right]^2}{\left[V_0' + \frac{S_\phi^2}{N} \right]}, \ i = 1,2 \tag{5.8}$$

In case of P^*, the fixed precision is given by

$$V_0' = \frac{1}{n} \Psi_0 + \frac{k}{n} \Psi_1 - \frac{S_\phi^2}{N} \tag{5.9}$$

$$C(P^*)_{\min.} = \frac{\left[\sqrt{(\Psi_0 + k_{opt.}\Psi_1)\left(C_1 + C_2 W_1 + C_3 \frac{W_2}{k_{opt.}}\right)}\right]^2}{\left[V_0' + \frac{S_\phi^2}{N}\right]}, \tag{5.10}$$

Where $\Psi_0 =$ coefficient of $\frac{1}{n}$ terms, $\Psi_1 =$ coefficient of $\frac{k}{n}$ terms which are obtained by (2.2)

6. AN EMPIRICAL STUDY

To illustrate the efficiency of the proposed class of estimators, we have considered the data from Sukhatme and Sukhatme (1997), p-256, which gives the number of villages and the area under wheat in each of the 89 administrative areas (These area are known as Patwari circle in the local terminology) in Hapur Subdivision of Meerut District, published by Govt. of India (1951). The last 25% administrative areas have been considered as non-response group of the population. Here we have taken the study attribute (ϕ) which is the administrative areas having no of villages greater than 5 and auxiliary character (x) as.

$$\phi = \begin{cases} 1, \text{ if the administrative areas having greater than 5 villages} \\ 0, \text{otherwise.} \end{cases}$$

x = Area under Wheat (Acres) in each administrative area.

The values of the parameters of the population under the study are as follows:

$P = 0.124$, $\bar{X} = 1102$, $C_p = 2.672$, $C_x = 0.65$, $P_2 = 0.182$, $\bar{X}_2 = 1242.68$, $C_{p(2)} = 3.185$, $C_{x(2)} = 0.582$, $\rho_{px} = 0.621$, $\rho_{px(2)} = 0.665$.

The problem considered here is to estimate the population proportion of administrative areas having the no of villages greater than 5, by using area under wheat in the administrative areas as the auxiliary character. The value of $H_{2(1)}(M)_{opt}$ and $H_{2(2)}(M)_{opt}$ from equation (3.5) and (3.6) are given below:

$H_{2(1)}(M)_{(k=2)} =$ -0.3667 , $H_{2(1)}(M)_{(k=3)} =$ -0.3897 ,
$H_{2(1)}(M)_{(k=4)} =$ -0.4029 and $H_{2(2)}(M) =$ -0.5097

The mean square error (MSE) and relative efficiency (in %) of the proposed classes of estimators T_{u1} and T_{u2} with respect to P^* at different level of sub-sampling fractions are given in Table 1.

Table 1: Relative Efficiency (In %) of the Estimators with Respect To P^*

Estimator	$N = 89$, $n' = 60$, $n = 40$ RE (.) in %		
	$1/k$		
	1/2	**1/3**	**1/4**
p^*	100.00 (0.00247)*	100.00 (0.00344)	100.00 (0.00440)
T_{u1}	143.92 (0. 00171)	151.61 (0.00227)	156.56 (0.00281)
T_{u2}	115.23 (0.00215)	110.51 (0.00311)	108.03 (0.00408)

*Figures in parenthesis give the MSE (.)

Here we have also computed the relative efficiency and expected cost of the proposed classes of estimators with respect to P^* for the fixed cost as well as for fixed precision respectively, which is given in Table 2.

Table 2: Relative Efficiency and Expected Cost

	$C_1' =$ □5, $C_1 =$ □15, $C_2 =$ □50, $C_3 =$ □400							
Estimators	$C_0 =$ □3000(fixed)					$V_0 = 0.00250$(fixed)		
	$k_{opt.}$ (~)	$n'_{opt.}$ (~)	$n_{opt.}$ (~)	$MSE(.)$ (~)	**R.E. in %**	$n'_{opt.}$ (~)	$n_{opt.}$ (~)	**Exp. cost in □.**
p^*	1.27	---	28	0.00496	100.00	---	47	4980.44
T_{u1}	1.70	87	30	0.00212	233.85	78	27	2697.04
T_{u2}	1.19	83	23	0.00247	200.78	82	23	2978.00

7. CONCLUSIONS

From Table 1, we find that the proposed estimators T_{u1} and T_{u2} have substantial gain in efficiency as compared to P^* for different value of sub-sampling fractions and their MSE increases by increasing the value of sub-sampling fraction. It also exhibits that relative efficiency of T_{u1} increases when sampling fraction increases but, we also observed that the relative efficiency of T_{u2} decreases as sampling fraction is increases. Similarly from Table 2, we observe that the estimators T_{u1} and T_{u2} are more efficient than P^* for fixed cost and the expected cost incurred in T_{u1} and T_{u2} are less in comparison to the expected cost incurred for P^* in the case of specified precision.

Hence on the basis of the theoretical comparisons and empirical study, we may suggest the use of the proposed classes of estimators T_{u1} and T_{u2} in the case of fixed sample size, fixed cost and for specified precision.

8. REFERENCES

1. Abd-Elfattah, A. M., El-Sherpieny, E. A, Mohamed, S. M. and Abdou, O. F. (2010). Improvement in estimating the population mean in simple random sampling using information on auxiliary attribute, Applied Mathematics and Computation, 215, p. 4198–4202.

2. Cochran, W.G. (1977). Sampling Techniques, New York, 3rd edn. John Wiley and Sons.
3. Hansen, M. H. and Hurwitz, W. N., (1946). The problem of non-response in sample surveys, J. Amer. Stat. Assoc., 41, p. 517-529.
4. Jhajj, H. S., Sharma, M. K. and Grover,L. K. (2006). A family of estimators of population mean using auxiliary attribute, Pak. J. Stat., 22 (1), p. 43-50.
5. Khare, B. B., Jha, P. S. and Kumar, K. (2015). Estimation of population proportion using auxiliary character in the presence of non-response, J. Relia. Stat. Stu., 8(2), p. 107-117.
6. Khare, B. B. and Srivastava, S. (1993). Estimation of population mean using auxiliary character in presence of non-response, Nat. Acad. Sc. Letters India, 16, p. 111-114.
7. Khare, B. B. and Srivastava, S. (1995). Study of conventional and alternative two phase sampling ratio, product and regression estimators in presence of non-response, Proc. Nat. Acad. Sci. India, 65(A) II, p. 195-203.
8. Khare, B. B. and Srivastava, S. (2000). Generalized estimators for population mean in presence of no response, Internal. J. Math. & Statist. Sci., 9, p. 75-87.
9. Khare, B. B. and Srivastava, S. (2010). Generalized two phase estimators for the population mean in the presence of non-response, Aligarh. Jour. Stat., 30, p. 39-54.
10. Koyuncu, N. (2012). Efficient estimators of population mean using auxiliary attributes, Appl. Math. Comp, 218, p. 10900-10905.
11. Malik, S. and Singh, R. (2013). An improved estimator using two auxiliary attributes, Applied Mathematics and Computation, 219, p. 10983-10986.
12. Naik, V. D. and Gupta, P. C. (1996). A note on estimation of mean with known population proportion of an auxiliary character, J. Ind. Soc. Agri. Stat., 48(2), p. 151-158.
13. Rao, P. S. R. S. (1986). Ratio estimation with sub sampling the no respondents. Survey Methodology, 12, p. 217-230.
14. Rao, P. S. R. S. (1990). Regression estimators with sub sampling of no respondents. In-Data Quality Control, Theory and Pragmatics, (Eds.) Gunar E. Liepins and V.R.R. Uppuluri, Marcel Dekker, New York, p. 191-208.
15. Reddy, V.N. (1978). A study of use of prior knowledge on certain population parameters in estimation, Sankhya Sec. B., 40, p. 29-37.
16. Shabbir, J. and Gupta, S. (2007). On estimating the finite population mean with known population proportion of an auxiliary variable, Pak. J. Stat., 23 (1), p. 1-9.
17. Shabbir, J. and Gupta, S. (2010). Estimation of the finite population mean in two phases sampling when auxiliary variables are attributes, Hacett. J. Math. Stat., 39(1), p. 121-129.
18. Singh, H. P. and Solanki, R. S. (2012). Improved estimation of population mean in simple random sampling using information on auxiliary attribute, Appl. Math. Comp., 218, p. 7798-7812.
19. Singh, H.P. and Kumar, S. (2009). A general procedure of estimating the population mean in the presence of non response under double sampling using auxiliary information, SORT- Stat. and Opera. Res. Tran., 33, p. 71-83.

20. Singh, H.P. and Kumar, S. (2010). Estimation of mean in the presence of non response using two phase sampling scheme, Statistical Papers, 51, p. 559-582.
21. Singh, R., Chauhan, P., Sawan, N. and Smarandache, F. (2008). Ratio estimators in simple random sampling using information on auxiliary attribute, Pak. J. Stat. and Opera. Res., 4, p. 47-53.
22. Singh, R., Kumar, M. and Smarandache, F. (2010). Ratio estimator in simple random sampling when study variable is an attributes, World Appl. Sci. J., 11 (5), p. 586-589.
23. Sinha, R. R. and Kumar, V. (2013). Improved estimators for population mean using attributes and auxiliary characters under incomplete information, Int. J. of math. And Stat., 14, p. 43-54.
24. Sinha, R. R. and Kumar, V. (2014). Improved classes of estimators for population mean using information on auxiliary character under double sampling the non-respondents, Nat. Acad. Sci. Lett., 37(1), p. 71-79.
25. Srivastava, S.K. and Jhajj, H.S. (1983). A class of estimators of the population means using multi- auxiliary information, Cal. Stat. Assoc. Bull., 32, p. 47-56.
26. Sukhatme, P.V. and Sukhatme, B.V. (1997). Sampling theory of surveys with application, Delhi, 1st edition. Piyush Publications.