Journal of Reliability and Statistical Studies; ISSN (Print): 0974-8024, (Online):2229-5666
Vol. 8, Issue 2 (2015): 107-117

ESTIMATION OF POPULATION PROPORTION USING AUXILIARY CHARACTER IN THE PRESENCE OF NON-RESPONSE

B.B. Khare*, P.S. Jha and K. Kumar
Department of statistics, Banaras Hindu University, Varanasi, India
E Mail: *bbkhare56@yahoo.com; premshankarjhabhu@gmail.com; kamalbhu03@gmail.com

Received January 13, 2015
Modified November 18, 2015
Accepted December 04, 2015

Abstract

Two classes of estimators for population proportion using auxiliary character with known population mean in the presence of non-response on the study character have been proposed. 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. The performance of the proposed classes of estimators with the relevant estimator is supported with the help of an empirical study.

Key Words: Non-response; Mean Square Error; Relative Bias; Auxiliary Characters; Attribute.

1. Introduction

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 as a qualitative variable. Sometimes auxiliary information is available in the form of auxiliary attribute which may also be used in the estimation of population mean of study character. Using auxiliary attribute, several authors have proposed different types 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, 10), Singh et.al (2008), Koyuncu (2010), Abd-Elfattah et al. (2010), Singh and Solanki (2012), Malik and Singh (2013) and Sinha and Kumar (2013, 14) have proposed different types of estimators for population mean using the information on auxiliary attribute.

In sample surveys, it generally happens that the study variable is also available in the form of qualitative characteristic such as land ownership, labour force in the population, smoking habit etc. For example if we want to estimate the proportion of student participation in politics, we may take their education, socio economic group and monthly expenditure etc as auxiliary character. Singh et al. (2010) have suggested a family of estimators for the population mean when study character itself is in the qualitative form.

Sometimes, the information on all the units selected in the sample is not available for some units due to non-response. In such situation, Hansen and Hurwitz (1946) have suggested a technique of sub-sampling from non-responding units. 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. Further several research works on the estimation of population mean have been done by Khare and Srivastava (1993, 95, 2000, 10) and Singh and Kumar (2009, 2010).

In this paper, we have proposed two classes of estimators for population proportion using auxiliary character in the presence of non-response. The expressions for the bias and mean square error of the proposed classes of estimators are also obtained and their properties are studied. Some members of the proposed classes of estimators are obtained. An empirical study is also given in the support of proposed estimators.

2. The Estimators

Suppose the population of size N is divided into N_1 responding and N_2 non-responding units. Let ϕ_i and x_i denote the value of study attribute and auxiliary character for ith unit of the population (i=1,.,.,N). In this case ϕ_i will take value 1 if it possessing the attribute otherwise zero.

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.

Now, we draw a sample of size n from population of size N by using SRSWOR method of sampling and observe that n_1 units respond and n_2 units do not respond 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. 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 estimator p^* is unbiased and its variance 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 squares of ϕ for the whole population and for the non-responding part of the population.

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

The estimator $\bar{x}^*$ is unbiased and 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 x for the whole population and for the non-responding parts of the population.

Now we propose the class of estimators for population proportion P using auxiliary character x in two different situations.

(i) When incomplete information on study attribute ϕ and corresponding $n_1 + r$ units of auxiliary character x available from the sample of size n is used and we propose the class of estimators T_{c1} for population proportion P using auxiliary character x which is given as follows:

$$T_{c1} = H(w, u_1), \tag{2.5}$$

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

(ii) When incomplete information on the study attributes ϕ and complete information on auxiliary character x from the selected sample of size n is used and we propose the class of estimators T_{c2} for population proportion P using auxiliary character x which is given as follows:

$$T_{c2} = H(w, u_2), \tag{2.6}$$

Such that $H(P, 1) = P$ and $H_1(P, 1) = 1$, where $u_2 = \frac{\bar{x}}{\bar{X}}$, $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ 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:

(i) Whenever the sample is 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 $(P,1)$.

(ii) 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. (2.7)

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

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

$$+ \frac{1}{2}\left\{\begin{matrix}(w-P)^2 H_{11}(m^*) + (u_i - 1)^2 H_{22(i)}(m^*) \\ + 2(w-P)(u_i - 1) H_{12(i)}(m^*)\end{matrix}\right\} \tag{2.8}$$

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

$$T_{ci} = w + (u_i - 1) H_{2(i)}(M)$$
$$+ \frac{1}{2}\{(u_i - 1)^2 H_{22(i)}(m^*) + 2(w - P)(u_i - 1) H_{12(i)}(m^*)\} \tag{2.9}$$

where

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

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

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

3. Bias and Mean Square Error (MSE) Of the Estimators

Under the regularity conditions specified for the function $H(w, u_i)$, the bias and mean square error of the class of estimators T_{ci} $(i = 1,2)$ always exists. In order to derive the expressions for bias and mean square error of the proposed classes of estimators, we use the large sample approximations.

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

Now, using SRSWOR method of sampling, we have

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

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

$$E(\varepsilon_0 \varepsilon_2) = \frac{f}{n} C_{\phi x} \quad ,$$

where

$C_\phi^2 = \dfrac{S_\phi^2}{P^2}, C_{\phi(2)}^2 = \dfrac{S_{\phi(2)}^2}{P^2}, C_x^2 = \dfrac{S_x^2}{\overline{X}^2}, C_{x(2)}^2 = \dfrac{S_{x(2)}^2}{\overline{X}^2}, C_{\phi x} = \rho_{\phi x} \dfrac{S_\phi S_x}{P\overline{X}},$

$C_{\phi x(2)} = \rho_{\phi x(2)} \dfrac{S_{\phi(2)} S_{x(2)}}{P\overline{X}}$ and $\rho_{\phi x}$ and $\rho_{\phi x(2)}$ are the Point-Byserial Correlation coefficient between ϕ and x for the whole population and for the non-responding part of the population.

We are retaining the terms involving the powers in $\varepsilon_0, \varepsilon_1$ 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 class of estimators T_{ci} $(i = 1,2)$ up to terms of order (n^{-1}) are given as follows:

$$Bias(T_{c1}) = \frac{1}{2}\left[\left\{\frac{f}{n} C_x^2 + \frac{W_2(k-1)}{n} C_{x(2)}^2\right\} H_{22(1)}(M) + 2P\left\{\frac{f}{n} C_{\varphi x} + \frac{W_2(k-1)}{n} C_{\varphi x(2)}\right\} H_{12(1)}(M)\right] \tag{3.1}$$

$$Bias(T_{c2}) = \frac{1}{2}\left[\left\{\frac{f}{n} C_x^2\right\} H_{22(1)}(M) + 2P\left\{\frac{f}{n} C_{\varphi x}\right\} H_{12(1)}(M)\right] \tag{3.2}$$

$$MSE(T_{c1}) = V(p^*) + \left\{\frac{f}{n} C_x^2 + \frac{W_2(k-1)}{n} C_{x(2)}^2\right\} H_{2(1)}^2(M) + 2P\left\{\frac{f}{n} C_{\varphi x} + \frac{W_2(k-1)}{n} C_{\varphi x(2)}\right\} H_{2(1)}(M) \tag{3.3}$$

and

$$MSE(T_{c2}) = V(p^*) + \frac{f}{n}\left[H_{2(2)}^2(M) C_x^2 + 2PH_{2(2)}(M) C_{\phi x}\right] \tag{3.4}$$

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

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

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 $MSE(T_{c1})$ and $MSE(T_{c2})$ respectively, we get

$$MSE(T_{c1})_{opt} = V(p^*) - P^2\frac{(\frac{f}{n}C_{\phi x} + \frac{W_2(k-1)}{n}C_{\phi x(2)})^2}{\frac{f}{n}C_x^2 + \frac{W_2(k-1)}{n}C_{x(2)}^2} \tag{3.7}$$

and

$$MSE(T_{c2})_{opt} = V(p^*) - P^2\frac{f}{n}\frac{(C_{\phi x})^2}{C_x^2} \tag{3.8}$$

The optimum value of $H_{2(1)}(M)_{opt}$ and $H_{2(2)}(M)_{opt}$ in the equation (3.5) and (3.6) are in the form of the values 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 value (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:

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

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

$$T_{i5} = w\left[1+\alpha_0\left(\frac{u_i-1}{u_i+1}\right)\right] \quad \text{and} \quad T_{i6} = w\exp\left(\frac{u_i-1}{u_i+1}\right) \tag{3.9}$$

where $a, a_1, \alpha_0, \alpha, \beta$ and γ are constants.

4. Comparison of the Proposed Class of Estimators with the Relevant Estimators

When we compare p^*, T_{c1} and T_{c2}, we observed.

$$V(p^*) - MSE(T_{c1})_{\min.} = P^2 \frac{\left(\frac{f}{n}C_{\phi x} + \frac{W_2(k-1)}{n}C_{\phi x(2)}\right)^2}{\frac{f}{n}C_x^2 + \frac{W_2(k-1)}{n}C_{x(2)}^2} \geq 0, \tag{4.1}$$

$$V(p^*) - MSE(T_{c2})_{\min.} = P^2 \frac{f}{n}\frac{(\mathrm{C}_{\phi x})^2}{\mathrm{C}_x^2} \geq 0, \tag{4.2}$$

and

$$MSE(T_{c2})_{\min.} - MSE(T_{c1})_{\min.} \geq 0, \text{ if } \frac{C_{x_{(2)}}^2}{C_x^2} \leq \frac{C_{\phi x_{(2)}}}{C_{\phi x}}\left(\frac{W_2(k-1)}{f}\frac{C_{\phi x_{(2)}}}{C_{\phi x}} + 2\right) \tag{4.3}$$

Hence $MSE(T_{c1})_{\min.} < MSE(T_{c2})_{\min.}$.

5. 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 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 (3.4) and (3.5) are given below:

$H_{2(1)}(M)_{(k=2)} =$ -0.352 , $H_{2(1)}(M)_{(k=3)} =$ -0.373 ,
$H_{2(1)}(M)_{(k=4)} =$ -0.386 and $H_{2(2)}(M) =$ -0.5097

Estimator	$N = 89$, $n = 40$ RE (.) in %		
	$1/k$		
	1/2	1/3	1/4
p^*	100.00 (0.00247)*	100.00 (0.00344)	100.00 (0.00440)
T_{c1}	165.67(0. 00149)	167.93 (0.00205)	169.59 (0.00260)
T_{c2}	127.93 (0.00193)	118.64 (0.00290)	113.99 (0.00386)

*Figures in parenthesis give the MSE (.)

Table 1: Relative Efficiency (RE in %) of the proposed estimators with respect to usual estimator P^*.

6. Conclusion

The proposed class of estimators T_{c1} and T_{c2} for the estimation population proportion P using auxiliary character in the presence of non response on study character are wider classes of estimators having a number of estimators as member of the classes. The members of the class of estimators T_{c1} and T_{c2} will have same minimum value of MSE for the optimum values of the constants. The class of estimators T_{c1} and T_{c2} are more efficient than the conventional estimator P^*. However, the $MSE(T_{c1})_{\min.} < MSE(T_{c2})_{\min.}$ holds true under (4.3). So it is suggested to use T_{c1} in most of the cases rather than T_{c2}.

Acknowledgment
The authors are thankful to the referees for their useful suggestions.

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. 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.
6. 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.
7. 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.
8. 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.
9. Koyuncu, N. (2012). Efficient estimators of population mean using auxiliary attributes, Appl. Math. Comp, 218, p. 10900-10905.
10. Malik, S. and Singh, R. (2013). An improved estimator using two auxiliary attributes, Applied Mathematics and Computation, 219, p. 10983-10986.
11. 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.
12. Rao, P. S. R. S. (1986). Ratio estimation with sub sampling the no respondents, Survey Methodology, 12, p. 217-230.

13. 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.
14. Reddy, V.N. (1978). A study of use of prior knowledge on certain population parameters in estimation, Sankhya Sec. B., 40, p. 29-37.
15. 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.
16. 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.
17. 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.
18. 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.
19. 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.
20. 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.
21. 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.
22. 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.
23. 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.
24. 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.
25. Sukhatme, P.V. and Sukhatme, B.V. (1997). Sampling theory of surveys with application, Delhi, 1st edition. Piyush Publications.