

A General Procedure of Estimating Population Mean Using Information on Auxiliary Attribute

¹Sachin Malik, ^{†1}Rajesh Singh and ²Florentin Smarandache

¹Department of Statistics, Banaras Hindu University
Varanasi-221005, India

²Chair of Department of Mathematics, University of New Mexico, Gallup, USA

[†] Corresponding author, rsinghstat@gmail.com

Abstract

This paper deals with the problem of estimating the finite population mean when some information on auxiliary attribute is available. It is shown that the proposed estimator is more efficient than the usual mean estimator and other existing estimators. The results have been illustrated numerically by taking empirical population considered in the literature.

Keywords Simple random sampling, auxiliary attribute, point bi-serial correlation, ratio estimator, efficiency.

1. Introduction

The use of auxiliary information can increase the precision of an estimator when study variable y is highly correlated with auxiliary variable x . There are many situations when auxiliary information is available in the form of attributes, e.g. sex and height of the persons, amount of milk produced and a particular breed of cow, amount of yield of wheat crop and a particular variety of wheat (see Jhajj et. al. (2006)).

Consider a sample of size n drawn by simple random sampling without replacement (SRSWOR) from a population of size N . Let y_i and ϕ_i denote the observations on variable y and ϕ respectively for i^{th} unit ($i=1, 2, \dots, N$).

Let $\phi_i=1$; if the i^{th} unit of the population possesses attribute $\phi=0$; otherwise.

Let $A=\sum_{i=1}^N \phi_i$ and $a=\sum_{i=1}^n \phi_i$, denote the total number of units in the population and sample respectively possessing attribute ϕ . Let $P=A/N$ and $p=a/n$ denote the proportion of units in the population and sample respectively possessing attribute ϕ . Naik and Gupta (1996) introduced a ratio estimator t_{NG} when the study variable and the auxiliary attribute are positively correlated. The estimator t_{NG} is given by

$$t_{NG} = \bar{y} \frac{-P}{p} \quad (1.1)$$

with MSE

$$MSE(t_{NG}) = f_1 (S_y^2 + R^2 S_\phi^2 - 2RS_{y\phi}) \quad (1.2)$$

$$\text{where } f_1 = \frac{N-n}{Nn}, \quad R = \frac{\bar{Y}}{P}, \quad S_y^2 = \frac{1}{N-n} \sum_{i=1}^N (y_i - \bar{Y})^2,$$

$$S_\phi^2 = \frac{1}{N-1} \sum_{i=1}^N (\phi_i - P)^2, \quad S_{y\phi} = \frac{1}{N-1} \sum_{i=1}^N (\phi_i - P)(y_i - \bar{Y})$$

(for details see Singh et al. (2008))

Jhaji et. al. (2006) suggested a family of estimators for the population mean in single and two phase sampling when the study variable and auxiliary attribute are positively correlated. Shabbir and Gupta (2007), Singh et. al. (2008) and Abd-Elfattah et. al. (2010) have considered the problem of estimating population mean \bar{Y} taking into consideration the point biserial correlation coefficient between auxiliary attribute and study variable.

The objective of this article is to suggest a generalised class of estimators for population mean \bar{Y} and analyse its properties. A numerical illustration is given in support of the present study.

2. Proposed Estimator

Let $\phi_i^* = \phi_i + mA$, m being a suitably chosen scalar, that takes values 0 and 1. Then

$$q = p + mA = p + NmP, \text{ and}$$

$$Q = (Nm + 1)P,$$

$$\text{where } q = \frac{b}{n}, Q = \frac{B}{N}, B = \sum_{i=1}^N \phi_i \text{ and } b = \sum_{i=1}^n \phi_i.$$

Motivated by Bedi (1996), we define a family of estimators for population mean \bar{Y} as

$$t = \left[w_1 \bar{y} + w_2 b(P-p) \right] \left(\frac{q}{Q} \right)^\alpha \quad (2.1)$$

where w_1, w_2 and α are suitably chosen scalars.

To obtain the Bias and MSE of the estimator t , we write

$$\bar{y} = \bar{Y}(1 + e_0), \quad p = P(1 + e_1), \quad s_{\phi}^2 = S_{\phi}^2(1 + e_2),$$

$$s_{\phi y} = S_{\phi y}(1 + e_3), \quad b = \beta(1 + e_3)(1 + e_2)^{-1}$$

such that $E(e_i) = 0$, $i=0,1,2,3$ and

$$E(e_0^2) = \left(\frac{1}{n} - \frac{1}{N}\right) C_y^2, \quad E(e_1^2) = \left(\frac{1}{n} - \frac{1}{N}\right) C_p^2,$$

$$E(e_0 e_1) = \left(\frac{1}{n} - \frac{1}{N}\right) \rho_{pb} C_y C_p, \quad E(e_1 e_2) = \left(\frac{1}{n} - \frac{1}{N}\right) C_p \lambda_{03},$$

$$E(e_1 e_3) = \left(\frac{1}{n} - \frac{1}{N}\right) C_p \frac{\lambda_{12}}{\rho_{pb}},$$

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

$$t = \bar{Y} \left[w_1(1 + e_0) - w_2 \frac{\beta}{R} e_1(1 + e_3)(1 + e_2)^{-1} \right] \left(1 + \frac{e_1}{Nm + 1} \right)^{\alpha} \quad (2.2)$$

We assume that $|e_2| < 1$ and $\left| \frac{e_1}{Na + 1} \right| < 1$, so that $(1 + e_2)^{-1}$ and $\left(1 + \frac{e_1}{Nm + 1} \right)^{\alpha}$ are expandable.

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

$$t - \bar{Y} = \bar{Y} \left[w_1 \left\{ 1 + e_0 + \frac{\alpha e_1}{Nm + 1} + \frac{\alpha(\alpha - 1)}{2} \frac{e_1^2}{(Nm + 1)^2} + \frac{\alpha e_0 e_1}{Nm + 1} \right\} - w_2 \frac{\beta}{R} \left\{ e_1 + e_1 e_3 - e_1 e_2 + \frac{\alpha e_1^2}{Nm + 1} \right\} - 1 \right] \quad (2.3)$$

Taking expectation of both sides of (2.3), we get the bias of t to the first degree of approximation as :

$$B(t) = \bar{Y} \left[(w_1 - 1) + w_1 \left\{ \frac{\alpha}{Nm + 1} f_1 C_{py} + \frac{\alpha(\alpha - 1)}{2(Nm + 1)^2} f_1 C_p^2 \right\} - w_2 \frac{\beta}{R} f_1 \left\{ C_p \frac{\lambda_{12}}{\rho_{pb}} - C_p \lambda_{03} + \frac{\alpha}{Nm + 1} C_p^2 \right\} \right] \quad (2.4)$$

Squaring both sides of (2.3) and neglecting terms of e's having power greater than two, we have

$$\begin{aligned}
(t - \bar{Y})^2 = \bar{Y}^2 & \left[w_1^2 \left\{ 1 + 2e_0 + \frac{2\alpha e_1}{Nm+1} + e_0^2 + \frac{\alpha(2\alpha-1)e_1^2}{(Nm+1)^2} + \frac{4\alpha e_0 e_1}{Nm+1} \right\} \right. \\
& + w_2^2 \left(\frac{\beta}{R} \right)^2 e_1^2 + 1 - 2w_1 w_2 \frac{\beta}{R} \left\{ e_1 + e_1 e_3 + e_0 e_1 - e_1 e_2 + \frac{2\alpha e_1^2}{Nm+1} \right\} \\
& - 2w_1 \left\{ 1 + e_0 + \frac{\alpha e_1}{Nm+1} + \frac{\alpha e_0 e_1}{Nm+1} + \frac{\alpha(\alpha-1)e_1^2}{2(Nm+1)^2} \right\} \\
& \left. + 2w_2 \frac{\beta}{R} \left\{ e_1 + e_1 e_3 - e_1 e_2 + \frac{\alpha e_1^2}{Nm+1} \right\} \right] \tag{2.5}
\end{aligned}$$

Taking expectation of both sides of (2.5), we get the MSE of t to the first degree of approximation as:

$$\text{MSE}(t) = \bar{Y}^2 \left[1 + w_1^2 A_{1(\alpha)}^{(m)} + w_2^2 A_2 - 2w_1 w_2 A_{3(\alpha)}^{(m)} - 2w_1 A_{4(\alpha)}^{(m)} + 2w_2 A_{5(\alpha)}^{(m)} \right] \tag{2.6}$$

where,

$$A_{1(\alpha)}^{(m)} = \left[1 + f_1 \left\{ C_y^2 + \frac{\alpha C_p^2}{Nm+1} \left(\frac{2\alpha-1}{Nm+1} + 4k \right) \right\} \right]$$

$$A_2 = \left(\frac{\beta}{R} \right)^2 f_1 C_p^2$$

$$A_{3(\alpha)}^{(m)} = \frac{\beta}{R} f_1 \left[C_p^2 \left\{ \frac{2\alpha}{Nm+1} + k \right\} + \frac{C_p}{\rho_{pb}} \lambda_{12} - C_p \lambda_{03} \right]$$

$$A_{4(\alpha)}^{(m)} = \left[1 + \frac{\alpha}{Nm+1} f_1 \left\{ \frac{\alpha-1}{2(Nm+1)} + k \right\} \right]$$

$$A_{5(\alpha)}^{(m)} = \frac{\beta}{R} f_1 \left[\frac{\alpha C_p^2}{Nm+1} + \frac{C_p}{\rho_{pb}} \lambda_{12} - C_p \lambda_{03} \right]$$

where , $k = \rho_{pb} \frac{C_y}{C_p}$.

The MSE(t) is minimised for

$$w_1 = \frac{(A_2 A_{4(\alpha)}^{(m)} - A_{3(\alpha)}^{(m)} A_{5(\alpha)}^{(m)})}{(A_{1(\alpha)}^{(m)} A_2 - (A_{3(\alpha)}^{(m)})^2)} = w_{10} \tag{2.7}$$

$$w_2 = \frac{(A_{3(\alpha)}^{(m)} A_{4(\alpha)}^{(m)} - A_{1(\alpha)}^{(m)} A_{5(\alpha)}^{(m)})}{(A_{1(\alpha)}^{(m)} A_2 - (A_{3(\alpha)}^{(m)})^2)} = w_{20} \tag{2.8}$$

3. Members of the family of estimator of t and their Biases and MSE

Table 3.1: Different members of the family of estimators of t

Choice of scalars				Estimator
w_1	w_2	α	m	
1	0	0	0	$t_1 = \bar{y}$
w_1	0	0	0	$t_2 = w_1 \bar{y}$ Searls (1964) type estimator
w_1	0	α	m	$t_3 = w_1 \bar{y} \left(\frac{q}{Q} \right)^\alpha$
w_1	0	α	0	$t_4 = w \bar{y} \left(\frac{p}{P} \right)^\alpha$
1	0	-1	0	$t_5 = \bar{y} \left(\frac{P}{p} \right)$, Naik and Gupta (1996) estimator
1	1	-1	0	$t_6 = \left[\bar{y} + b(P-p) \right] \frac{P}{p}$ Singh et. al. (2008) estimator
w_1	w_2	0	0	$t_7 = [w_1 \bar{y} + w_2 b(P-p)]$
w_1	1	0	0	$t_8 = [w_1 \bar{y} + b(P-p)]$
w	w	0	0	$t_9 = w [\bar{y} + b(P-p)]$
1	1	0	0	$t_{10} = \bar{y} + b(P-p)$ Regression estimator

The estimator $t_1 = \bar{y}$ is an unbiased estimator of the population mean \bar{Y} and has the variance

$$\text{Var}(t_1) = f_1 S_y^2 \tag{3.1}$$

To, the first degree of approximation the biases and MSE's of t_i 's, $i=1,2,\dots,10$ are respectively given by

$$B(t_2) = \bar{Y}(w_1 - 1) \quad (3.2)$$

$$B(t_3) = \bar{Y} \left[(w_1 - 1) + f_1 w_1 \left\{ \frac{\alpha}{Nm + 1} \rho_{pb} C_y C_p + \frac{\alpha(\alpha - 1)}{2(Nm + 1)} C_p^2 \right\} \right] \quad (3.3)$$

$$B(t_4) = \bar{Y} \left[(w_1 - 1) + w_1 f_1 \left\{ \alpha \rho_{pb} C_y C_p + \frac{\alpha(\alpha - 1)}{2} C_p^2 \right\} \right] \quad (3.4)$$

$$B(t_5) = \bar{Y} f_1 [C_p^2 - \rho_{pb} C_y C_p] \quad (3.5)$$

$$B(t_6) = \bar{Y} f_1 \left[(C_p^2 - \rho_{pb} C_y C_p) - \frac{\beta}{R} \left(C_p \frac{\lambda_{12}}{\rho_{pb}} - C_p \lambda_{03} \right) \right] \quad (3.6)$$

$$B(t_7) = \bar{Y} \left[(w_1 - 1) - \frac{\lambda_{12} \beta}{R} f_1 \left\{ C_p \frac{\lambda_{12}}{\rho_{pb}} - C_p \lambda_{03} \right\} \right] \quad (3.7)$$

$$B(t_8) = \bar{Y} \left[(w_1 - 1) - \frac{\beta}{R} f_1 \left\{ C_p \frac{\lambda_{12}}{\rho_{pb}} - C_p \lambda_{03} \right\} \right] \quad (3.8)$$

$$B(t_9) = \bar{Y} \left[(w - 1) - w f_1 \frac{\beta}{R} \left\{ C_p \frac{\lambda_{12}}{\rho_{pb}} - C_p \lambda_{03} \right\} \right] \quad (3.9)$$

$$B(t_{10}) = -\bar{Y} \frac{\beta}{R} f_1 \left[C_p \frac{\lambda_{12}}{\rho_{pb}} - C_p \lambda_{03} \right] \quad (3.10)$$

The corresponding MSE's will be

$$MSE(t_2) = \bar{Y}^2 \left[1 + w_1^2 A_{1(0)}^{(0)} - 2w_1 A_{4(0)}^{(0)} \right] \quad (3.11)$$

$$MSE(t_3) = \bar{Y}^2 \left[1 + w_1^2 A_{1(\alpha)}^{(m)} - 2w_1 A_{4(\alpha)}^{(m)} \right] \quad (3.12)$$

$$MSE(t_4) = \bar{Y}^2 \left[1 + w_1^2 A_{1(\alpha)}^{(0)} - 2w_1 A_{4(\alpha)}^{(0)} \right] \quad (3.13)$$

$$MSE(t_5) = \bar{Y}^2 \left[1 + A_{1(-1)}^{(0)} - 2A_{4(-1)}^{(0)} \right] \quad (3.14)$$

$$\text{MSE}(t_6) = \bar{Y}^2 \left[1 + A_{1(-1)}^{(0)} + A_2 - 2A_{3(-1)}^{(0)} - 2A_{4(-1)}^{(0)} + 2A_{5(-1)}^{(0)} \right] \quad (3.15)$$

$$\text{MSE}(t_7) = \bar{Y}^2 \left[1 + w_1^2 A_{1(0)}^{(0)} + w_2^2 A_2 - 2w_1 w_2 A_{3(0)}^{(0)} - 2w_1 A_{4(0)}^{(0)} + 2w_2 A_{5(0)}^{(0)} \right] \quad (3.16)$$

$$\text{MSE}(t_8) = \bar{Y}^2 \left[1 + w_1^2 A_{1(0)}^{(0)} + A_2 - 2w_1 \left(A_{3(0)}^{(0)} + A_{4(0)}^{(0)} \right) + 2A_{5(0)}^{(0)} \right] \quad (3.17)$$

$$\text{MSE}(t_9) = \bar{Y}^2 \left[1 + w^2 \left(A_{1(0)}^{(0)} + A_2 - 2A_{3(0)}^{(0)} \right) - 2w \left(A_{4(0)}^{(0)} - A_{5(0)}^{(0)} \right) \right] \quad (3.18)$$

$$\text{MSE}(t_{10}) = \bar{Y}^2 \left[1 + A_{1(0)}^{(0)} + A_2 - 2 \left(A_{3(0)}^{(0)} + A_{4(0)}^{(0)} - A_{5(0)}^{(0)} \right) \right] \quad (3.19)$$

The MSE's of the estimators of t_i , $i=2,3,4,7,8,9$ will be minimised respectively, for

$$w_1 = \frac{A_{4(0)}^{(0)}}{A_{1(0)}^{(0)}} \quad (3.20)$$

$$w_1 = \frac{A_{4(\alpha)}^{(m)}}{A_{1(\alpha)}^{(m)}} \quad (3.21)$$

$$w_1 = \frac{A_{4(\alpha)}^{(0)}}{A_{1(\alpha)}^{(0)}} \quad (3.22)$$

$$\left. \begin{aligned} w_1 &= \frac{\left(A_2 A_{4(0)}^{(0)} - A_{3(0)}^{(0)} A_{5(0)}^{(0)} \right)}{\left(A_2 A_{1(0)}^{(0)} - \left(A_{3(0)}^{(0)} \right)^2 \right)} \\ w_2 &= \frac{\left(A_{3(0)}^{(0)} A_{4(0)}^{(0)} - A_{1(0)}^{(0)} A_{5(0)}^{(0)} \right)}{A_2 A_{1(0)}^{(0)} - \left(A_{3(0)}^{(0)} \right)^2} \end{aligned} \right\} \quad (3.23)$$

$$w_1 = \frac{\left(A_{3(0)}^{(0)} + A_{4(0)}^{(0)} \right)}{A_{1(0)}^{(0)}} \quad (3.24)$$

$$w = \frac{A_{4(0)}^{(0)} - A_{5(0)}^{(0)}}{A_{1(0)}^{(0)} + A_2 - 2A_{3(0)}^{(0)}} \quad (3.25)$$

Thus the resulting minimum MSE of t_i , $i= 2,3,4,7,8,9$ are, respectively given by

$$\min. \text{MSE}(t_2) = \bar{Y}^2 \left[1 - \frac{(A_{4(0)}^{(0)})^2}{A_{1(0)}^{(0)}} \right] \quad (3.26)$$

$$\min. \text{MSE}(t_3) = \bar{Y}^2 \left[1 - \frac{(A_{4(\alpha)}^{(m)})^2}{A_{1(\alpha)}^{(m)}} \right] \quad (3.27)$$

$$\min. \text{MSE}(t_4) = \bar{Y}^2 \left[1 - \frac{(A_{4(\alpha)}^{(0)})^2}{A_{1(\alpha)}^{(0)}} \right] \quad (3.28)$$

$$\min. \text{MSE}(t_7) = \bar{Y}^2 \left[1 - \frac{\left\{ A_2 (A_{4(0)}^{(0)})^2 - 2A_{3(0)}^{(0)} A_{5(0)}^{(0)} + A_{1(0)}^{(0)} (A_{5(0)}^{(0)})^2 \right\}}{\left\{ A_2 A_{1(0)}^{(0)} - A_{3(0)}^{(0)} \right\}^2} \right] \quad (3.29)$$

$$\min. \text{MSE}(t_8) = \bar{Y}^2 \left[1 + A_2 + 2A_{5(0)}^{(0)} - \frac{(A_{3(0)}^{(0)} - A_{4(0)}^{(0)})^2}{A_{1(0)}^{(0)}} \right] \quad (3.30)$$

$$\min. \text{MSE}(t_9) = \bar{Y}^2 \left[1 - \frac{(A_{4(0)}^{(0)} - A_{5(0)}^{(0)})^2}{A_{1(0)}^{(0)} + A_2 - 2A_{3(0)}^{(0)}} \right] \quad (3.31)$$

4. Empirical study

The data for the empirical study is taken from natural population data set considered by Sukhatme and Sukhatme (1970):

y = Number of villages in the circles and

ϕ = A circle consisting more than five villages

$N = 89, \bar{Y} = 3.36, P = 0.1236, \rho_{pb} = 0.766, C_y = 0.6040, C_p = 2.190$

$\lambda_{04} = 6.1619, \lambda_{40} = 3.810, \lambda_{12} = 146.475, \lambda_{03} = 2.2744$

In the Table 4.1 percent relative efficiencies (PRE's) of various estimators are computed with respect to \bar{y} .

Table 4.1: PRE of different estimators of \bar{Y} with respect to \bar{y} .

Estimator	PRE's
$t_1 = \bar{y}$	100.00
t_2	101.41
t_3	90.35
t_4	6.92
t_5	11.64
t_6	7.38
t_7	100.44
t_8	243.39
t_9	243.42
t_{10}	241.98

Conclusion

The MSE values of the members of the family of the estimator t have been obtained using (2.6). These values are given in Table 4.1. When we examine Table 4.1, we observe the superiority of the proposed estimators t_2 , t_7 , t_8 , t_9 and t_{10} over usual unbiased estimator t_1 , t_3 , t_4 , Naik and Gupta (1996) estimator t_5 and Singh et. al. (2008) estimator t_6 . From this result we can infer that the proposed estimators t_8 and t_9 are more efficient than the rest of the estimators considered in this paper for this data set.

We would also like to remark that the value of the min. $MSE(t_{10})$, which is equal to the value of the MSE of the regression estimator is 241.98. From Table 4.1 we notice that the value of MSE of the estimators t_8 and t_9 are less than this value, as shown in Table 4.1. Finally, we can say that the proposed estimators t_8 and t_9 are more efficient than the regression estimator for this data set.

References

1. Abd-Elfattah, A.M. El-Sherpieny, E.A. Mohamed, S.M. Abdou, O. F. (2010): Improvement in estimating the population mean in simple random sampling using information on auxiliary attribute. Appl. Mathe. and Compt. doi:10.1016/j.amc.2009.12.041
2. Bedi, P. K. (1996). Efficient utilization of auxiliary information at estimation stage. Biom. Jour, 38:973–976.
3. Jhajj, H.S., Sharma, M.K. and grover, L.K. (2006) : A family of estimators of population mean using information on auxiliary attribute. Pak. Journ. of Stat., 22(1), 43-50.
4. Naik,V.D. and Gupta,P.C.(1996): A note on estimation of mean with known population proportion of an auxiliary character. Journ. of the Ind. Soc. of Agr. Stat., 48(2), 151-158.
5. Searls, D.T. (1964): The utilization of known coefficient of variation in the estimation procedure. Journ. of the Amer. Stat. Assoc., 59, 1125-1126.

6. Singh, R. Chauhan, P. Sawan, N. Smarandache, F. (2008): Ratio estimators in simple random sampling using information on auxiliary attribute. Pak. J. Stat. Oper. Res. 4(1) 47–53.
7. Shabbir, J. and Gupta, S.(2007): On estimating the finite population mean with known population proportion of an auxiliary variable. Pak. Journ. of Stat., 23 (1), 1-9.
8. Sukhatme, P.V. and Sukhatme, B.V. (1970): Sampling theory of surveys with applications. Iowa State University Press, Ames, U.S.A.