Ratio Estimators in Simple Random Sampling Using Information on Auxiliary Attribute

Rajesh Singh
Department of Statistics, Banaras Hindu University
Varanasi (U.P.), India
Email: rsinghstat@yahoo.co.in

Pankaj Chauhan
School of Statistics, DAVV, Indore (M.P.), India

Nirmala Sawan
School of Statistics, DAVV, Indore (M.P.), India

Florentin Smarandache
Department of Mathematics, University of New Mexico, Gallup, USA
Email: smarand@unm.edu

Abstract

Some ratio estimators for estimating the population mean of the variable under study, which make use of information regarding the population proportion possessing certain attribute, are proposed. Under simple random sampling without replacement (SRSWOR) scheme, the expressions of bias and mean-squared error (MSE) up to the first order of approximation are derived. The results obtained have been illustrated numerically by taking some empirical population considered in the literature.

Key words: Proportion, bias, MSE, ratio estimator.

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 exist situations when information is available in the form of attribute ϕ, which is highly correlated with y. For example

a)   Sex and height of the persons,
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]).

Consider a sample of size n drawn by 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 $(i=1,2,....N)$. Suppose there is a complete dichotomy in the population with respect to the presence or absence of an attribute, say ϕ, and it is assumed that attribute ϕ takes only the two values 0 and 1 according as

$\qquad \phi_i$ = 1, if ith unit of the population possesses attribute ϕ
$\qquad\quad$ = 0, otherwise.

Rajesh Singh, Pankaj Chauhan, Nirmala Sawan, Florentin Smarandache

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 $P = \dfrac{A}{N}$ and $p = \dfrac{a}{n}$ denote the proportion of units in the population and sample respectively possessing attribute ϕ.

Taking into consideration the point biserial correlation between a variable and an attribute, Naik and Gupta [2] defined ratio estimator of population mean when the prior information of population proportion of units, possessing the same attribute is available, as follows:

$$t_{NG} = \overline{y}\left(\frac{P}{p}\right) \tag{1.1}$$

Here $\overline{y}$ is the sample mean of variable of interest. The MSE of t_{NG} up to the first order of approximation is –

$$MSE(t_{NG}) = \left(\frac{1-f}{n}\right)\left[S_y^2 + R_1^2 S_\phi^2 - 2R_1 S_{y\phi}\right] \tag{1.2}$$

Where $f = \dfrac{n}{N}$, $\qquad R_1 = \dfrac{\overline{Y}}{P}$, $\qquad S_y^2 = \dfrac{1}{N-1}\sum_{i=1}^{N}(y_i - \overline{Y})^2$, $\qquad S_\phi^2 = \dfrac{1}{N-1}\sum_{i=1}^{N}(\phi_i - P)^2$,

$S_{y\phi} = \dfrac{1}{N-1}\sum_{i=1}^{N}(\phi_i - P)(y_i - \overline{Y})$.

In the present paper, some ratio estimators for estimating the population mean of the variable under study, which make use of information regarding the population proportion possessing certain attribute, are proposed. The expressions of bias and MSE have been obtained. The numerical illustrations have also been done by taking some empirical populations considered in the literature.

2. The suggested estimator

Following Ray and Singh [3], we propose the following estimator –

$$t_1 = \frac{\overline{y} + b_\phi(P-p)}{p}P = R^*P \tag{2.1}$$

Where $b_\phi = \dfrac{s_{y\phi}}{s_\phi^2}$, $R^* = \dfrac{\overline{y} + b_\phi(P-p)}{p}$, $s_\phi^2 = \left(\dfrac{1}{n-1}\right)\sum_{i=1}^{n}(\phi_i - p)^2$ and

$s_{y\phi} = \left(\dfrac{1}{n-1}\right)\sum_{i=1}^{n}(\phi_i - p)(y_i - \overline{Y})$.

Remark 1: When we put $b_\phi = 0$ in (2.1) the proposed estimator turns to the Naik and Gupta [2] ratio estimator t_{NG} given in (1.1).

MSE of this estimator can be found by using Taylor series expansion given by –

$$f(p,\overline{y}) \cong f(P,\overline{y}) + \frac{\partial f(c,d)}{\partial c}\bigg|_{P,\overline{Y}}(p-P) + \frac{\partial f(c,d)}{\partial c}\bigg|_{P,\overline{Y}}(\overline{y}-\overline{Y}) \qquad (2.2)$$

Where $f(p,\overline{y}) = R^*$ and $f(P,\overline{Y}) = R_1$.

Expression (2.2) can be applied to the proposed estimator in order to obtain MSE equation as follows:

$$R^* - R_1 \cong \frac{\partial\left((\overline{y}+b_\phi(P-p))\right)/p}{\partial p}\bigg|_{P,\overline{Y}}(p-P) + \frac{\partial\left((\overline{y}+b_\phi(P-p))\right)/p}{\partial\overline{y}}\bigg|_{P,\overline{Y}}(\overline{y}-\overline{Y})$$

$$\cong -\left(\frac{\overline{y}}{p^2} + \frac{b_\phi P}{p^2}\right)\bigg|_{P,\overline{Y}}(p-P) + \frac{1}{p}\bigg|_{P,\overline{Y}}(\overline{y}-\overline{Y})$$

$$E(R^* - R_1)^2 \cong \frac{(\overline{Y}+B_\phi P)^2}{P^4}V(p) - \frac{2(\overline{Y}+B_\phi P)}{P^3}Cov(p,\overline{y}) + \frac{1}{P^2}V(\overline{y})$$

$$\cong \frac{1}{P^2}\left\{\frac{(\overline{Y}+B_\phi P)^2}{P^2}V(p) - \frac{2(\overline{Y}+B_\phi P)}{P}Cov(p,\overline{y}) + V(\overline{y})\right\} \qquad (2.3)$$

Where $B_\phi = \dfrac{S_{\phi y}}{S_\phi^2} = \dfrac{\rho_{pb}S_y}{S_\phi}$.

$\rho_{pb} = \dfrac{S_{y\phi}}{S_y S_\phi}$, is the point biserial correlation coefficient.

Now,

$$MSE(t_1) = P^2 E(R_1 - R_\phi)^2$$

$$\cong \frac{(\overline{Y}+B_\phi P)^2}{P^2}V(p) - \frac{2(\overline{Y}+B_\phi P)}{P}Cov(p,\overline{y}) + V(\overline{y}) \qquad (2.4)$$

Simplifying (2.4), we get MSE of t_1 as

$$MSE(t_1) \cong \left(\frac{1-f}{n}\right)\left[R_1^2 S_\phi^2 + S_y^2\left(1 - \rho_{pb}^2\right)\right] \qquad (2.5)$$

Several authors have used prior value of certain population parameters (s) to find more precise estimates. Searls (1964) used Coefficient of Variation (CV) of study character at estimation stage. In practice this CV is seldom known. Motivated by Searls (1964) work, Sen (1978), Sisodiya and Dwivedi (1981), and Upadhyaya and Singh (1984) used the known CV of the auxiliary character for estimating population mean of a study character in ratio method of estimation. The use of prior value of Coefficient of Kurtosis in estimating the population variance of study character y was first made by Singh et. al. (1973). Later, used by and Searls and Intarapanich (1990), Upadhyaya and Singh (1999), Singh (2003) and Singh et. al. (2004) in the estimation of population mean of study character. Recently Singh and Tailor (2003) proposed a modified ratio estimator by using the known value of correlation coefficient.

In next section, we propose some ratio estimators for estimating the population mean of the variable under study using known parameters of the attribute ϕ such as coefficient of variation C_p, Kurtosis $(\beta_2(\phi))$ and point biserial correlation coefficient ρ_{pb}.

3. Suggested Estimators

We suggest following estimator –

$$t = \frac{\bar{y} + b_\phi(P - p)}{(m_1 p + m_2)}(m_1 P + m_2) \tag{3.1}$$

Where $m_1(\neq 0)$, m_2 are either real number or the functions of the known parameters of the attribute such as C_p, $(\beta_2(\phi))$ and ρ_{pb}.

The following scheme presents some of the important estimators of the population mean, which can be obtained by suitable choice of constants m_1 and m_2:

Estimator	Values of	
	m_1	m_2
$t_1 = \dfrac{\bar{y} + b_\phi(P - p)}{p}P$	1	1
$t_2 = \dfrac{\bar{y} + b_\phi(P - p)}{(p + \beta_2(\phi))}[P + \beta_2(\phi)]$	1	$\beta_2(\phi)$
$t_3 = \dfrac{\bar{y} + b_\phi(P - p)}{(p + C_p)}[P + C_p]$	1	C_p
$t_4 = \dfrac{\bar{y} + b_\phi(P - p)}{(p + \rho_{pb})}[P + \rho_{pb}]$	1	ρ_{pb}
$t_5 = \dfrac{\bar{y} + b_\phi(P - p)}{(p\beta_2(\phi) + C_p)}[P\beta_2(\phi) + C_p]$	$\beta_2(\phi)$	C_p
$t_6 = \dfrac{\bar{y} + b_\phi(P - p)}{(pC_p + \beta_2(\phi))}[PC_p + \beta_2(\phi)]$	C_p	$\beta_2(\phi)$
$t_7 = \dfrac{\bar{y} + b_\phi(P - p)}{(pC_p + \rho_{pb})}[PC_p + \rho_{pb}]$	C_p	ρ_{pb}
$t_8 = \dfrac{\bar{y} + b_\phi(P - p)}{(p\rho_{pb} + C_p)}[P\rho_{pb} + C_p]$	ρ_{pb}	C_p
$t_9 = \dfrac{\bar{y} + b_\phi(P - p)}{(p\beta_2(\phi) + \rho_{pb})}[P\beta_2(\phi) + \rho_{pb}]$	$\beta_2(\phi)$	ρ_{pb}
$t_{10} = \dfrac{\bar{y} + b_\phi(P - p)}{(p\rho_{pb} + \beta_2(\phi))}[P\rho_{pb} + \beta_2(\phi)]$	ρ_{pb}	$\beta_2(\phi)$

Following the approach of section 2, we obtain the MSE expression for these proposed estimators as –

$$\text{MSE}(t_i) \cong \left(\frac{1-f}{n}\right)\left[R_i S_\phi^2 + S_y^2(1-\rho_{pb}^2)\right], \qquad (i=1,2,3,....,10) \qquad (3.2)$$

Where $R_1 = \dfrac{\overline{Y}}{P}$, $R_2 = \dfrac{\overline{Y}}{P+\beta_2(\phi)}$, $R_3 = \dfrac{\overline{Y}}{P+C_p}$, $R_4 = \dfrac{\overline{Y}}{P+\rho_{pb}}$,

$$R_5 = \frac{\overline{Y}\beta_2(\phi)}{P\beta_2(\phi)+C_p}, \quad R_6 = \frac{\overline{Y}C_p}{PC_p+\beta_2(\phi)}, \quad R_7 = \frac{\overline{Y}C_p}{PC_p+\rho_{pb}}, \quad R_8 = \frac{\overline{Y}\rho_{pb}}{P\rho_{pb}+C_p},$$

$$R_9 = \frac{\overline{Y}\beta_2(\phi)}{P\beta_2(\phi)+\rho_{pb}}, \quad R_{10} = \frac{\overline{Y}\rho_{pb}}{P\rho_{pb}+\beta_2(\phi)}.$$

4. Efficiency comparisons

It is well known that under simple random sampling without replacement (SRSWOR) the variance of the sample mean is

$$V(\overline{y}) = \left(\frac{1-f}{n}\right)S_y^2 \qquad (4.1)$$

From (4.1) and (3.2), we have

$$V(\overline{y}) - \text{MSE}(t_i) \geq 0, \qquad i=1,2,.....,10$$

$$\Rightarrow \rho_{pb}^2 > \frac{S_\phi^2}{S_y^2}R_i^2 \qquad (4.2)$$

When this condition is satisfied, proposed estimators are more efficient than the sample mean.

Now, we compare the MSE of the proposed estimators with the MSE of Naik and Gupta (1996) estimator t_{NG}. From (3.2) and (1.1) we have

$$\text{MSE}(t_{NG}) - \text{MSE}(t_i) \geq 0, \qquad (i=1,2,.....,10)$$

$$\Rightarrow \rho_{pb}^2 \geq \frac{S_\phi^2}{S_y^2}\left[R_i^2 - R_\phi^2 + 2R_\phi K_{yp}\right] \qquad (4.3)$$

where $K_{yp} = \rho_{yp}\dfrac{C_y}{C_p}$.

Pak.j.stat.oper.res. Vol.IV No.1 2008 pp47-53

51

5. 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, $\overline{Y}$ = 3.36, P = 0.1236, ρ_{pb} = 0.766, C_y = 0.604, C_p = 2.19, $\beta_2(\phi)$ = 6.23181.

In table 5.1 percent relative efficiencies (PRE) of various estimators are computed with respect to $\overline{y}$.

Table 5.1: PRE of different estimators of $\overline{Y}$ with respect to $\overline{y}$

Estimator	PRE $(.,\overline{y})$
$\overline{y}$	100
t_{NG}	11.61
t_1	7.36
t_2	236.55
t_3	227.69
t_4	208.09
t_5	185.42
t_6	230.72
t_7	185.27
t_8	230.77
t_9	152.37
t_{10}	237.81

From table 5.1, we observe that the proposed estimators $t_i (i = 2,.....,10)$ which uses some known values of population proportion performs better than the usual sample mean $\overline{y}$ and Naik and Gupta (1996) estimator t_{NG}.

Conclusion

We have suggested some ratio estimators for estimating $\overline{Y}$ which uses some known value of population proportion. For practical purposes the choice of the estimator depends upon the availability of the population parameters.

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. Pak. J. Statist., 22 (1), 43-50.

2. Naik, 6 V.D. and Gupta, P.C. (1996): A note on estimation of mean with known population proportion of an auxiliary character. Jour. Ind. Soc. Agr. Stat., 48 (2), 151-158.

3. Ray, S. K. and Singh, R. K. (1981): Difference-cum-ratio type estimators. J. Ind. Statist. Assoc., 19, 147-151.

4. Searls, D.T. (1964): The utilization of known coefficient of variation in the estimation procedure. Journal of American Statistical Association, 59, 1125-1126.

5. Searls, D.T. and Intarapanich, P. (1990): A note on an estimator for the variance that utilizes the kurtosis. The American Statistician, 44, 4, 295-296.

6. Sen, A.R. (1978): Estimation of the population mean when the coefficient of variation is known. Commun. Statist., Theory – Meth. A (7), 657-672.

7. Singh, G.N. (2003): On the improvement of product method of estimation in sample surveys. Jour. Ind. Soc. Agri. Statistics, 56(3), 267-275.

8. Singh H.P. And Tailor, R. (2003): Use of known correlation coefficient in estimating the finite population mean. Statistics in Transition, 6, 4, 555-560.

9. Singh H.P., Tailor, R., Tailor, R. and Kakran, M.S. (2004): An improved estimator of population mean using power transformation. Jour. Ind. Soc. Agr. Stat., 58 (2), 223-230.

10. Singh, J. Pandey, B.N. and Hirano, K. (1973): On the utilization of a known coefficient of kurtosis in the estimation procedure of variance. Ann. Inst. Stat. Math., 25, 51-55.

11. Sisodia, B.V.S. And Dwivedi, V.K. (1981): A modified ratio estimator using coefficient of variation of auxiliary variable. Jour.Ind.Soc.Agril. Statist., 33, 2, 13-18.

12. Sukhatme, P.V. and Sukhatme, B.V. (1970): Sampling theory of surveys with applications. Iowa State University Press, Ames, U.S.A.

13. Upadhyaya, L.N. and Singh, H.P. (1984): On the estimation of the population mean with known coefficient of variation. Biom. Jour., 26 (8), 915-922.

14. Upadhyaya, L.N. and Singh, H.P. (1999): Use of transformed auxiliary variable in estimating the finite population mean. Biometrical Journal, 41, 5, 627-636.