Journal of Reliability and Statistical Studies; ISSN (Print): 0974-8024, (Online): 2229-5666
Vol. 11, Issue 1 (2018): 01-07

ESTIMATION OF POPULATION MEAN THROUGH IMPROVED CLASS OF RATIO TYPE ESTIMATORS USING AUXILIARY INFORMATION

Banti Kumar*, Manish Kumar, S.E.H. Rizvi and M. Iqbal Jeelani Bhat
Division of Statistics and Computer Science, FBSc, Sher-e- Kashmir
University of Agricultural Sciences and Technology of Jammu, India
E Mail: bantikumar1573@gmail.com*; manshstat@gmail.com;
sehrizvi_stats@yahoo.co.in; jeelani.miqbal@gmail.com
*Corresponding author

Received March 27, 2017
Modified February 04, 2018
Accepted March 05, 2018

Abstract

In survey sampling we are often concerned with the estimation of population parameters with the use of auxiliary information at pre-selection stage, selection stage and estimation stage. If used properly, this information may provide better estimates than those where such information is not used. In this paper, an attempt has been made to develop a general class of improved ratio type estimators for estimation of population mean by modifying conventional ratio estimator whose large sample properties are compared with the conventional ratio estimator and estimators proposed by Sharma *et al*. (2010). It is observed that the proposed class of estimators performed better than conventional ratio estimator and estimators proposed by Sharma et al. (2010) on the basis of unbiasedness, mean squared error and efficiency criterion. An empirical study has also been presented in support of the present investigation.

Key Words: Auxiliary Information, Ratio Estimator, Relative Bias, Relative Mean Squared Error, Population Mean.

1. Introduction

Survey sampling is a method of drawing an inference about the characteristic of a population or universe by observing only a part of the population. It was outstanding contribution of Neyman (1934) by providing the inferential basis of the representative method, sampling design and assessment of purposive selection, marked a turning point in the history of sampling and upon numerous new avenues for fruitful research in the theory and philosophy of sample surveys**.** In survey sampling, we are often concerned with the estimation of population mean ($\bar{Y}$) using auxiliary information, which may be available (or may be made available by diverting a part of the resources) in one form or the other. Since, sample mean ($\bar{y}$) has been found minimum variance unbiased estimator under simple random sampling while estimating the population mean $\bar{Y}$. One can use ratio type estimator for estimating the $\bar{Y}$ population mean when there is high correlation between the study variate and the auxiliary variate or it can be used when the auxiliary variate satisfy the condition (i) if $C_x/2C_y < \rho \leq 1$ and both Y and X are positive or negative (ii) if $-C_x/2C_y < \rho \leq 1$ and either Y or X is negative (Singh and Chaudhary, 1995), where, $C_x = \frac{s_x}{\bar{X}}$ and $C_y = \frac{s_y}{\bar{Y}}$ are the coefficient of variation of x and y respectively. Tin (1965) proposed some modifications in the ratio estimators for making it unbiased. Thereafter, several modification of ratio estimators in sampling theory have been proposed, an interesting done by Sharma et al. (2010). Singh and Naqvi (2015) proposed a class of estimators of finite population mean of study variable using the knowledge of population mean of an auxiliary variable In the presence of non-response. Mishra et al. (2017) discussed the problem of estimation of population mean in stratified sampling using information on two auxiliary variables. In this paper, an attempt has been made to propose a general class of ratio estimators which does not involve either any additional financial burden or any reduction in sample size but at same time utilize the auxiliary information.

2. Results and Discussion

Consider a random sample of size n is drawn from a population of size N and observations on study variable Y and auxiliary variables X are obtained. Further, the sample means $\bar{y}$ and $\bar{x}$ are unbiased estimators of population means and $\bar{Y}$ and $\bar{X}$respectively, while s_y^2 and s_x^2 are unbiased estimators of population variances σ_y^2 and σ_x^2 respectively. Similarly, $s_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$ is an unbiased estimator of population covariance σ_{xy}. Also, $\theta = \left(\frac{C_{02}}{C_{20}}\right)^{1/2}$and $\rho = \frac{C_{11}}{(C_{02})^{1/2}(C_{20})^{1/2}}$.

Cochran (1940) proposed ratio type estimator for estimating the population mean is as

$$\bar{y}_r = \frac{\bar{y}}{\bar{x}}\bar{X} \tag{1}$$

Where, $\bar{y}$ and $\bar{x}$ are unbiased estimators of $\bar{Y}$ and $\bar{X}$, the population means of the characteristics under study and auxiliary characteristics respectively.

The relative bias upto order O (n^{-1}) and relative mean squared error of the ratio type estimator ($\bar{y}_r$) upto O (n^{-2}) , as

$$RB(\bar{y}_r) = \frac{1}{n}(C_{20} - C_{11}) \tag{2}$$

$$RM(\bar{y}_r) = \frac{1}{n}(C_{02} - 2C_{11} + C_{20}) + \frac{1}{n^2}[(2(2C_{21} - C_{12} - C_{30}) + 3(3C_{20}^2 - 6C_{20}C_{11} + 2C_{11}^2 + C_{20}C_{02})]. \tag{3}$$

Sharma *et al.,* (2010) proposed the following general class of ratio type estimators as

$$t_s = \bar{y}_r + \frac{1}{n}\left[p\frac{\bar{x}}{\bar{X}}\frac{s_y^2}{\bar{y}^2} + q\frac{s_{xy}}{\bar{X}}\right] \tag{4}$$

where , p and q are the scalars specifying the estimator.

The relative bias and relative mean squared error of t_s upto order O (n^{-1}) and O (n^{-2}) are as

$$RB(t_s)=RB(\bar{y}_r) + \frac{1}{n}(pC_{02} + qC_{11}) \tag{5}$$

$$RM(t_s) = RM(\bar{y}_r) + \frac{p}{n^2}[2C_{03} - 2C_{12} + 2C_{02}C_{11} - 2C_{02}^2] + \frac{q}{n^2}[2C_{20}C_{11} - 2C_{11}^2 + C_{12} - C_{21}] + \frac{1}{n^2}[pC_{02} + qC_{11}]^2 \tag{6}$$

By keeping the above, the following improved class of ratio type estimators for $\bar{Y}$ have been proposed as

$\boldsymbol{t}^{*} = \bar{\boldsymbol{y}}_r + \frac{\boldsymbol{p}}{\boldsymbol{n}}\bar{\boldsymbol{y}}\left[\frac{s_x^2}{\bar{x}\bar{X}} + \boldsymbol{q}\frac{s_{xy}}{\bar{y}\bar{X}}\right]$ where, p and q are unknown scalars specifying the estimator. (7)

The relative bias and relative mean squared error of the estimator $\boldsymbol{t}^{*}$ upto order O (n^{-1}) and O (n^{-2}) respectively are as

$$RB(\boldsymbol{t}^{*}) = RB(\bar{y}_r) + \frac{p}{n}(C_{20} + qC_{11}) \quad (8)$$

$$RM(\boldsymbol{t}^{*}) = RM(\bar{y}_r) + \frac{p}{n^2}[2C_{20}(C_{20} - 3C_{11} + 2C_{02}) + 2C_{21} - C_{30}] + \frac{pq}{n^2}[C_{11}(C_{20} - C_{11}) + 2C_{12} - 2C_{21}] + \frac{p^2}{n^2}[C_{20} + qC_{11}]^2. \quad (9)$$

If $p = 0$ in (7), the general class of improved ratio type estimator t^{*} will reduces to conventional ratio type estimator. Thus, $t^{*}_{(0,q)} = \bar{y}_r$, is the special case of proposed ratio type estimator t^{*}.

From (9) and (3), it is observed that the relative mean squared error of both estimators i.e., t^{*} and $\bar{y}_r$, are identical upto order O (n^{-1}). By comparing the relative mean squared error of both estimators upto O (n^{-2}), we find that the estimator t^{*} has smaller relative mean squared error than that of $\bar{y}_r$, if

$$\frac{p}{n^2}[2C_{20}(C_{20} - 3C_{11} + 2C_{02}) + 2C_{21} - C_{30}] + \frac{pq}{n^2}[C_{11}(C_{20} - C_{11}) + 2C_{12} - 2C_{21}] + \frac{p^2}{n^2}[C_{20} + qC_{11}]^2 < 0 \quad (10)$$

Under bivariate normal distribution, the expression (10) reduces to

$$\frac{p}{n^2}[2C_{20}(C_{20} - 3C_{11} + 2C_{02})] + \frac{pq}{n^2}[C_{11}(C_{20} - C_{11})] + \frac{p^2}{n^2}[C_{20} + qC_{11}]^2 < 0$$

Further, t^{*} has smaller bias than t_s upto order O (n^{-1}), if

$pC_{20} + pqC_{11} < pC_{02} + qC_{11}$ and vice versa.

From the eq. (12) for $p < 0$ and $q > 0$, consider $p = -1$ and $q = 1$, t^{*} have smaller bias than t_s, if it satisfy the condition $\rho > \frac{\theta^2 - 1}{2\theta}$ and the value of θ lie between $1.001 < \theta < 2.414$.

From (9) and (6), it is observed that the t^{*} has smaller relative mean squared error than t_s upto order O (n^{-2}), if

$$\frac{p}{n^2}[2C_{20}(C_{20} - 3C_{11} + 2C_{02}) + 2C_{21} - C_{30} - 2C_{03}^2 + 2C_{12} - 2C_{02}C_{11} + 2C_{02}^2] + \frac{pq}{n^2}[C_{11}(C_{20} - C_{11}) + 2C_{12} - 2C_{21}] + \frac{p^2}{n^2}[C_{20} + qC_{11}]^2 + \frac{q}{n^2}[2C_{11}^2 - 2C_{20}C_{11} + C_{12} - C_{21}] + \frac{1}{n^2}[pC_{02} + qC_{11}]^2 < 0 \quad (11)$$

Under bivariate normal distribution, expression (11) becomes

$$\frac{p}{n^2}[2C_{20}(C_{20} - 3C_{11} + 2C_{02}) + 2C_{21} - C_{30} - 2C_{03}^2 + 2C_{12} - 2C_{02}C_{11} + 2C_{02}^2] + \frac{pq}{n^2}[C_{11}(C_{20} - C_{11}) + 2C_{12} - 2C_{21}] + \frac{p^2}{n^2}[C_{20} + qC_{11}]^2 + \frac{q}{n^2}[2C_{11}^2 - 2C_{20}C_{11} + C_{12} - C_{21}] + \frac{1}{n^2}[pC_{02} + qC_{11}]^2 < 0 \quad (12)$$

From the eq. (12) for $p < 0$ and $q > 0$, consider $p = -1$ and $q = 1,$ the estimator $t_{1(-1,1)}$ will be more efficient than $t_{s(-1,1)}$, if $\rho < \frac{3\theta^4+\theta^2+1}{4\theta^3+5\theta}$ which results to $0.63 < \theta < 1.65$.

If we take p= -1 and q= -1, in the estimator t^*, the relative bias of the estimator becomes zero. i.e., the proposed estimator is unbiased.
Thus, the estimator ${t^*}_{(-1,-1)}$ is more efficient than estimator $t_{s_{(-1,-1)}}$ upto order O (n^{-2}) under bivariate normal distribution, if $\rho < \frac{2\theta^4+5\theta^2+1}{7\theta}$which results in θ lies between $0.407 < \theta < 0.891$.

3. Empirical Illustration

For numerical evaluation of the large sample properties of the proposed ratio type estimator for population mean under study, symmetrical and asymmetrical data have been generated by carrying out simulation through R and SAS softwares. Samples of sizes 30, 60, 120 and 150 have been drawn from the population of size 200. Higher sample sizes i.e., 120 and 150 have been taken in order to see the consistency of the estimator.

The parameters of the dataset I were N=200,$\bar{Y} = 11.83$,$\bar{X} = 11.995$, σ_y^2 =53.37 and σ_x^2 =55.96. As per the Normality tests i.e., Shapiro-Wilk, Kolmogorov-Smirnov, Cramer-Von Mises and Anderson Darling test having values 0.888, 0.168, 0.901 and 7.374 respectively which were significant which indicates that the data were asymmetrical in nature.

(p,q)	Sample Size / Estimator	30	60	120	150
	Conventional	1.399	0.795	0.374	0.341
(-1,0)	t^*	-0.380	-0.193	-0.132	-0.041
(-1,0)	t_s	1.019	0.602	0.242	0.299
(-1,-1)	t^*	0.000	0.000	0.000	0.000
(-1,-1)	t_s	2.797	1.590	0.748	0.681
(-2,-1)	t^*	-1.399	-0.795	-0.374	-0.341
(-2,-1)	t_s	4.196	2.385	1.121	1.022
(-1,-3)	t^*	0.760	0.385	0.264	0.082
(-1,-3)	t_s	6.355	3.565	1.759	1.445

Table 1: Relative bias of the conventional ($\bar{y}_r$), Sharma et al., (2010) and proposed ratio type estimators for asymmetrical data under different values of p and q.

Table 1 reveals that for p=-1 and q=0, the proposed estimator t* has smaller bias than $\bar{y}_r$ and t_s. The bias approximately reduces to zero as sample size increases. For p=-1 and q=-1, the proposed estimator t* is unbiased, the similar type of result is also proposed by Tin (1965). In case of p= -2 and q= -1, the bias of t* is similar as that of $\bar{y}_r$ with negative sign. For p=-1 and q=-3, t* has smaller bias than that of both $\bar{y}_r$ and t_s upto order O (n^{-1}).

(p,q)	Sample size / Estimator	30	60	120	150
(-1,0)	y_r	174.30	157.06	139.91	140.24
	t_s	159.74	144.19	124.70	134.83
(-1,-1)	y_r	185.59	163.25	145.44	141.83
	t_s	274.61	235.57	197.88	195.60
(-2,-1)	y_r	494.46	285.53	206.01	192.40
	t_s	1030.41	574.22	377.27	359.37
(-1,-3)	y_r	201.35	170.58	148.33	144.04
	t_s	646.41	524.06	430.68	378.25

Table 2: Percent relative efficiency of estimator t* with respect to $\bar{y}_r$ and t_s estimators in case of asymmetrical data

It has been observed from Table 2 that the range of percent relative efficiency of the proposed estimator t^* with respect to $\bar{y}_r$ and t_s is 124.70 to 1030.41. The estimator t* at p=-1 and q=-1 is unbiased, consistent and efficient in case of asymmetrical data. In case of unbiased ratio estimator, the relative efficiency was found to lie in the range 141.83 to 274.61 per cent.

It has been observed that the range of percent relative efficiency of the proposed estimator $t^*_{(-1,0)}$ is 139.91 to 174.30 with respect to $\bar{y}_r$ whereas in case of t_s, it varies from 124.70 to 159.74 at different sample sizes i.e., 30, 60, 120 and 150. In case of $t^*_{(-1,-1)}$, the estimator is unbiased and ranges varies from 141.83 to 274.61 with respect to $\bar{y}_r$ and t_s estimators. Further, for p<-1 and q>-1, the estimator the ranges varies from 141.83 to 494.46 and 195.60 to 1030.41 respectively with respect to $\bar{y}_r$ and t_s. The parameters of the dataset II were N=200, $\bar{Y} = 0.0517$, $\bar{X} = 10.002$, σ_y^2 =3.845 and σ_x^2 =25.383. The tests of normality i.e., Shapiro-Wilk, Kolmogorov-Smirnov, Cramer-Von Mises and Anderson Darling test were applied on the data and were found to be non-significant having values 0.998, 0.023, 0.011 and 0.081. Thus, the data were symmetrical in nature.

(p,q)	Sample Size / Estimator	30	60	120	150
	$\bar{y}_r$	1.053	0.302	0.245	0.161
(-1,0)	t^*	0.039	0.012	0.005	0.005
(-1,0)	ts	0.949	0.220	0.211	0.131
(-1,1)	t^*	0.078	0.024	0.009	0.011
(-1,1)	ts	0.910	0.208	0.207	0.126
(-1,-1)	t^*	0.000	0.000	0.000	0.000
(-1,-1)	ts	0.988	0.232	0.216	0.137
(-2,1)	t^*	-0.897	-0.255	-0.227	-0.139
(-2,1)	ts	0.806	0.126	0.173	0.097

Table 3: Relative bias of the estimators $\bar{y}_r$, t_s and t^* in case of symmetrical data under different values of p and q

Table 3 reveals that the estimator t*(-1,-1) was again found to be unbiased. Further, the relative biases of the estimators t*(-1,0), t*(-1,1), t*(-1,-1) and t*(-2,1) have the lesser relative bias than the conventional as well as corresponding t_s ratio type estimators.

(p,q)	Sample Size / Estimator	30	60	120	150
(-1,0)	$\bar{y}_r$	113.212	120.989	118.217	109.867
	t_s	114.950	126.772	120.510	112.800
(-1,1)	$\bar{y}_r$	120.989	123.760	113.447	115.406
	t_s	120.698	129.485	118.093	116.527
(-1,-1)	$\bar{y}_r$	106.354	118.232	123.569	104.569
	t_s	109.913	124.066	123.302	109.137
(-2,1)	$\bar{y}_r$	119.711	136.097	113.309	119.075
	t_s	119.366	150.157	122.833	121.965

Table 4: Relative efficiency of estimator t* with respect to $\bar{y}_r$ and $t_s$ratio type estimator in case of symmetrical data

From Table 4, it has been observed that the estimator t* at p=-1 and q=-1 is unbiased, consistent and efficient in case of symmetrical data. The range of relative efficiency of the unbiased estimator t*(-1,-1) was found to be 109.137 to 124.066. Further, the range of relative efficiency of the proposed estimator for $p < 0$ and $q > 0$ with respect to $\bar{y}_r$ and t_s is 113.309 to 136.097 and 116.527 to 150.157 respectively.

4. Conclusions

In this paper, a general class of ratio type estimators has been developed. Different estimators have been obtained by setting different values of p and q. The values of p and q at which the value of relative mean squared error was found minimum have been proposed in this paper. An unbiased ratio type estimator $t^*_{(-1,-1)} = \bar{y}_r + \frac{1}{n}\bar{y}\left[\frac{s_{xy}}{\bar{y}\bar{X}} - \frac{s_x^2}{\bar{x}\bar{X}}\right]$ has been proposed that was efficient in case of symmetrical and asymmetrical data under different sample sizes. But by sacrificing the property of unbiasedness a more efficient estimator have been obtained at p= -2 and q= -1 i.e., the estimator $t^*_{(-2,-1)} = \bar{y}_r - \frac{2}{n}\bar{y}\left[\frac{s_x^2}{\bar{x}\bar{X}} - \frac{s_{xy}}{\bar{y}\bar{X}}\right]$ and proposed.

References

1. Cochran, W.G.(1940). The estimation of yield in cereal experiments by using sampling for the ratio of grains to total produce, Journal of Agricultural Science, 30, p. 262-275.
2. Mishra, M. Singh, B. P. and Singh, R. (2017). Estimation of Population Mean Using two Auxiliary Variables in Stratified Random Sampling, Journal of Reliability and Statistical Studies, 10(1), p. 59-68.
3. Neyman, J. (1934). On the two different aspects of the representative method, Journal of the Royal Statistical Society, 97(4), 558-625.

4. Sharma, M, Bhatnagar, S, Khan, I (2010). A general class of improved ratio type estimators through auxiliary information, International Journal of Agricultural and Statistical Sciences, 6, p.115-118
5. Singh R.K. and Naqvi, N. (2015). Estimation of Population Mean Using Auxiliary Information in Non-Response, Journal of Reliability and Statistical Studies, 8(1), p.135-146.
6. Singh, D. and Chaudhary, F.S. (1986). Theory and Analysis of Sample Survey Designs, New Age International Publishers, New Delhi.
7. Tin, M. (1965). Comparison of some ratio estimators, Journal of American Statistical Association. 60, p.294-307.