Journal of Reliability and Statistical Studies; ISSN (Print): 0974-8024, (Online): 2229-5666
Vol. 9, Issue 2 (2016): 71-79

A FAMILY OF FACTOR-TYPE ESTIMATORS FOR ESTIMATION OF POPULATION VARIANCE

[1]Narendra Singh Thakur, [2]Kalpana Yadav and [3]Ranveer Singh
[1]Deptt. of Mathematics and Statistics, Banasthali University, Rajasthan, India
[2]Deptt. of Statistics, University of Delhi, Delhi, India
[3]Deptt. of Mechanical Engineering, Jaipur National University, Jaipur, India
E Mail: [1]nst_stats@yahoo.co.in, [2]kalpana22yadav@gmail.com
[1]Corresponding Author

Received April 08, 2016
Modified October 12, 2016
Accepted November 15, 2016

Abstract

In the present paper, we have proposed factor-type ratio estimator using known values of some population parameters of the auxiliary variable to estimate the population variance of the study variable. The expressions for the bias and mean squared error (m.s.e.) of the proposed estimator have been derived up to first order of approximation. The suggested estimator is bias controlled and gives choices for optimal mean squared error. A comparison has been made with some well-known estimators and it is shown that the proposed estimator is better than some other existing estimators.

Key Words: Bias, Mean Square Error, Variance, Estimation, Ratio estimator, Dual to Ratio Estimator, Auxiliary information, Factor-Type estimator.

1. Introduction

The theory and application of survey sampling is grown dramatically in the past few decades and hundreds of surveys are now carried out each year in the private sector, the academic community and various government agencies. Examples include market research and public opinion surveys, surveys allied with academic research studies and surveys on labor force contribution, health care, energy uses and economics activity. In fact it is not an over statement to say that much of the data undergoing any form of statistical analysis are collected in surveys. Survey samples impinges upon almost every field of scientific study, including agriculture, demography, education, energy, transportation, health care, economics, politics, sociology, geology, forestry and so on.

As the use of sample surveys is increased, the need for methods of analyzing and interpreting the resulting data is also increased. A basic requirement of nearly all forms of analysis, obviously, a principal requirement of good survey practice is that a measure of precision be provided for each estimate derived from the survey data. The most commonly used measure of precision is the variance of the survey estimator. In general, variance is not known but must be estimated from the survey data themselves. The problem of constructing such estimates of variance is the main problem treated in this paper.

As the preliminary to any further discussion it is important to recognize that the variance of a survey statistic is a function of both the form of the statistics and the nature of the sampling design (i.e. the procedure used in selecting the sample). An estimator of variance must take account of both the estimators and the sampling design. Subsequent sections in this paper focus specially on a factor type variance estimation methodology for simple random sampling.

To estimate the population total or mean or the variance of a variable, the use of auxiliary information (on designing and estimation stage) is most frequent phenomena in practice and widely discussed in sampling theory literature. Auxiliary variables are used in sampling theory to improve sampling designs and to achieve higher precision of the estimators for the population parameters. In survey sampling literature, a great variety of techniques using the auxiliary information by means of ratio, product and regression methods have been advocated. Particularly, in the presence of multi-auxiliary variables, a wide variety of estimators have been proposed, following different ideas, and linking together ratio, product or regression estimators, each one exploiting the variables one at a time.

Let a simple random sample without replacement (SRSWOR) of size n be drawn from a finite population $U = \{1, 2, ..., N\}$ of size N. Let (Y_i, X_i) be the observed value of study and auxiliary variable of i^{th} individual.

Isaki (1983) discussed about a ratio type variance estimator for estimating population variance and its properties. The ratio type variance estimator is

$$t_R = s_y^2 \frac{S_X^2}{s_x^2} \tag{1.1}$$

The estimator t_R is found to be biased and its m.s.e. is expressed by

$$M(t_R) = F\, S_Y^4 \left[(\lambda_{40} - 1) + (\lambda_{04} - 1) - 2(\lambda_{22} - 1)\right] \tag{1.2}$$

Dual to ratio estimator of Srivenkataramana (1980) for population variance could be written as

$$t_{st} = s_y^2 \left[\frac{N S_X^2 - n s_x^2}{S_X^2 (N - n)}\right] \tag{1.3}$$

It is biased and the m.s.e. of t_{st} is

$$M(t_{st}) = F\, S_Y^4 \left[(\lambda_{40} - 1) + g^2(\lambda_{04} - 1) - 2g(\lambda_{22} - 1)\right] \tag{1.4}$$

The problem of constructing efficient estimators for the population variance has been widely discussed by various authors such as Das and Tripathi (1978), Singh et al. (1988), Singh and Katyar (1991), Agrawal and Sthapit (1995), Garcia and Cebrain (1996), Arcos et al. (2005), Kadilar and Cingi (2005, 2006), Wolter (2007), Singh and Solanki (2013, 2013a), Yadav and Kadilar (2013, 2013a), Asghar et al. (2014), Kumar (2014), Singh et al. (2014), Bandyopadhyay and Singh (2015) and Misra (2016).

Remark 1.1: Define, $\alpha_1 = \dfrac{fB}{A + fB + C}, \alpha_2 = \dfrac{C}{A + fB + C}$, $f = \dfrac{n}{N}$, $\alpha = \alpha_1 - \alpha_2$.

$S_Y^2 = \dfrac{1}{N-1}\sum_{i=1}^{N}\left(Y_i - \overline{Y}\right)^2$, $S_X^2 = \dfrac{1}{N-1}\sum_{i=1}^{N}\left(X_i - \overline{X}\right)^2$, $s_y^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2$,

$s_x^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2$, $\overline{X} = \dfrac{1}{N}\sum_{i=1}^{N}X_i$, $\overline{Y} = \dfrac{1}{N}\sum_{i=1}^{N}Y_i$, $\overline{x} = \dfrac{1}{n}\sum_{i=1}^{n}x_i$, $\overline{y} = \dfrac{1}{n}\sum_{i=1}^{n}y_i$,

$\lambda_{rs} = \dfrac{\mu_{rs}}{\mu_{r0}^{\frac{r}{2}}\mu_{0s}^{\frac{s}{2}}}$, $\mu_{rs} = \dfrac{1}{N-1}\sum\left(Y_i - \overline{Y}\right)^r\left(X_i - \overline{X}\right)^s$ and $F = \dfrac{1}{n} - \dfrac{1}{N}$, $g = n/(N-n)$

2. Proposed Estimator

Singh and Shukla (1987) proposed a family of factor-type ratio estimator for population mean. This estimator is bias control and gives choice of parameter for optimum m.s.e. and the ratio, product, dual to ratio estimator etc. are special cases of this estimator at particular value of parameter as well.

Motivating form Singh and Shukla (1987), Isaki (1983) etc. we advocate the factor-type ratio estimator for population variance of the study variable as

$$T_{FT} = s_y^2\,\frac{(A+C)S_X^2 + fBs_x^2}{(A+fB)S_X^2 + Cs_x^2} \tag{2.1}$$

where, $A = (k-1)(k-2)$, $B = (k-1)(k-4)$, $C = (k-2)(k-3)(k-4)$, $0 \le k < \infty$.

2.1 Special Cases

For some specified value of *k* the estimator T_{FT} alike $\overline{y}_{FT}$ provides some existing estimators like ratio, product, dual to ratio etc. for population variance i.e. at *k=1,2,3,4* the T_{FT} is as in the following table.

Value of *k*	Estimators	
k = 1	$t_R = s_y^2\dfrac{S_X^2}{s_x^2}$	(Ratio estimator)
k = 2	$t_P = s_y^2\dfrac{s_x^2}{S_X^2}$	(Product estimator)
k = 3	$t_{st} = s_y^2\left[\dfrac{S_X^2 - s_x^2}{(1-f)S_X^2}\right]$	(Dual to ratio estimator)
k = 4	$s_y^2 = t_0$ (let)	(Unbiased estimator)

Table 2.1: Factor Type Estimator as Special Cases

3. Properties of Proposed Estimator

In order to study the properties of the proposed estimator, i.e. to discuss the bias and m.s.e. of proposed factor-type estimator, we considered for large samples

$s_y^2 = S_y^2(1+e_0), s_x^2 = S_x^2(1+e_1)$ and $|e_i| < 1$, i=0, 1

where, $E(e_0) = E(e_1) = 0$, $\mathrm{E}(e_0^2) = F(\lambda_{40} - 1)$, $\mathrm{E}(e_1^2) = F(\lambda_{04} - 1)$, $E(e_0 e_1) = F(\lambda_{22} - 1)$ (3.1)

Theorem 3.1: The estimator T_{FT} could be written in terms of $e_i; i = 0,1$ up to first order of approximation as

$$T_{FT} = S_y^2\left[1 + e_0 + \alpha e_1 - \alpha\alpha_2 e_1^2 + \alpha e_0 e_1\right] \tag{3.2}$$

Proof: The estimator is

$$T_{FT} = s_y^2\left[\frac{(A+C)S_x^2 + fBs_x^2}{(A+fB)S_x^2 + Cs_x^2}\right]$$

Using large sample approximations of (3.1) and after simplification, we get,

$$= S_y^2(1+e_0)(1+\alpha_1 e_1)(1+\alpha_2 e_1)^{-1}$$

On expanding the expression and ignoring the terms of $O(n^{-1})$, we get

$$= S_y^2\left[1 + e_0 + \alpha e_1 - \alpha\alpha_2 e_1^2 + \alpha e_0 e_1\right]$$

Theorem 3.2: The bias of the proposed estimator is

$$B(T_{FT}) = S_y^2 F\alpha\left[(\lambda_{22} - 1) - \alpha_2(\lambda_{04} - 1)\right] \tag{3.3}$$

Proof: Since,

$$B(T_{FT}) = E(T_{FT} - S_y^2)$$

Using (3.2) and taking expectations we get,

$$B(T_{FT}) = S_y^2 E\left[e_0 + \alpha e_1 - \alpha\alpha_2 e_1^2 + \alpha e_0 e_1\right]$$

Using equation (3.1) we get,

$$B(T_{FT}) = S_y^2 F\alpha\left[(\lambda_{22} - 1) - \alpha_2(\lambda_{04} - 1)\right]$$

Theorem 3.3: The mean squared error of the proposed estimator is

$$M(T_{FT}) = S_y^2 F\left[(\lambda_{40} - 1) + \alpha^2(\lambda_{04} - 1) + 2\alpha(\lambda_{22} - 1)\right] \tag{3.4}$$

Proof: Since we know that the m.s.e of T_{FT} is

$$M(T_{FT}) = E(T_{FT} - S_y^2)^2$$

$$= S_y^4 E\left[e_0 + \alpha e_1 - \alpha\alpha_2 e_1^2 + \alpha e_0 e_1\right]^2$$

Squaring and ignoring higher order terms,

$$= S_y^4 E\left[e_0 + \alpha e_1\right]^2$$

$$= S_y^4 E\left[e_0^2 + \alpha^2 e_1^2 + 2\alpha e_0 e_1\right]^2$$

By using (3.1) we get the m.s.e. of the estimator as follows:

$$M(T_{FT}) = S_y^4 F\left[(\lambda_{40} - 1) + \alpha^2(\lambda_{04} - 1) + 2\alpha(\lambda_{22} - 1)\right]$$

Theorem 3.4: The minimum mean square error of T_{FT} at $\alpha = -\frac{(\lambda_{22}-1)}{(\lambda_{04}-1)} = -P$ is

$$\left[M(T_{FT})\right]_{\min} = S_y^4 F\left[\frac{(\lambda_{40}-1)(\lambda_{04}-1)-(\lambda_{22}-1)^2}{(\lambda_{04}-1)}\right] \tag{3.5}$$

Proof: On differentiating equation (3.4) with respect to α and then equate to zero we get,

$$\frac{d}{d\alpha}\left[M(T_{FT})\right]_{\min} = S_y^4 F\left[2\alpha(\lambda_{04}-1)+2(\lambda_{22}-1)\right] = 0$$

$$\Rightarrow \alpha = \frac{-(\lambda_{22}-1)}{(\lambda_{04}-1)} = -P \qquad \text{(let)}$$

On putting the value of α in equation (3.4) we get the result as (3.5).

Remark 3.1: Multiple choices of *k*

The optimality condition $\alpha = \alpha_1 - \alpha_2 = -P$ provides the equation

$$\Rightarrow \frac{fB}{A+fB+C} - \frac{C}{A+fB+C} = -P$$

$$\Rightarrow fB - C + P(A+fB+C) = 0$$

$$\Rightarrow fB(1+P) + C(1-P) + PA = 0$$

The above equation provides

$$(P-1)k^3 + \left[P(f-8)+(f+9)\right]k^2 + \left[P(23-5f)-(26+5f)\right]k + \left[(4f-24)P+(4f-24)\right] = 0 \tag{3.6}$$

which is a cubic equation in the form of k. One can get three values of k like k_1, k_2, k_3 for which mean squared error is optimum. The best choice criterion for k is

Step I: Compute $\left|B(T_{FT})_{k_j}\right| \qquad for \quad j = 1,2,3$

Step II: Choose k_j as $\left|B(T_{FT})_{k_j}\right| = \min_{j=1,2,3}\left[\left|B(T_{FT})_{k_j}\right|\right]$.

This ultimately gives bias control at the optimal level of m.s.e.

4. Comparisons

(i) The variance of unbiased estimator (t_0) in SRSWOR is

$$V(t_0) = FS_y^4(\lambda_{40}-1)$$

Let, $D_1 = V(t_0) - \left[M(T_{FT})\right]_{\min} = \frac{FS_y^4(\lambda_{22}-1)^2}{(\lambda_{04}-1)}$

T_{FT} is better than t_0 if $D_1 > 0$

$$\Rightarrow (\lambda_{22}-1)^2 > 0 \Rightarrow \lambda_{22} > 1$$

(ii) Let, $D_2 = M(t_R) - \left[M(T_{FT})\right]_{\min}$

$$= FS_y^4\left[(\lambda_{04}-1)-2(\lambda_{22}-1)+\frac{(\lambda_{22}-1)^2}{(\lambda_{04}-1)}\right]$$

T_{FT} is better than t_R if $D_2 > 0$

$$\Rightarrow (\lambda_{22}-1)^2(\lambda_{04}-1)^{-1} > 2(\lambda_{22}-1)-(\lambda_{04}-1)$$

(iii) Let, $D_3 = M(t_{ST}) - [M(T_{FT})]_{\min}$

$$= FS_y^4\left[g^2(\lambda_{04}-1)-2g(\lambda_{22}-1)+\frac{(\lambda_{22}-1)^2}{(\lambda_{04}-1)}\right]$$

T_{FT} is better than t_{ST} if $D_3 > 0$

$$\Rightarrow \frac{(\lambda_{22}-1)^2}{(\lambda_{04}-1)} > 2g(\lambda_{22}-1)-g^2(\lambda_{04}-1)$$

5. Empirical study

We have considered a population of size N=200 containing values of study variable Y and auxiliary variable X [See Shukla and Thakur (2008)] with parameters of the population as

$\overline{Y}$ = 42.485; $\overline{X} = 18.515$; $S_y^2 = 199.0598$; $S_x^2 = 48.5375$; $\rho = 0.8652$;

$f = 0.3$; $\lambda_{22} = 2.47$; $\lambda_{04} = 3.74$; $\lambda_{40} = 2.56$

For known f and P, the cubic equation (3.6) in k provides three k-values as

k_1 = 1.5424; $k_2 = 2.9457$; $k_3 = 6.6783$

The bias and m.s.e. of the existing and proposed estimator are based on 10,000 repeated samples drawn by SRSWOR from population N = 200. These computations, with respect to unbiased estimator are displayed in table (5.1) where efficiency measurement is considered as

$$e(\hat{t}) = \frac{M(t_0)}{M(\hat{t})}$$

where $M(\hat{t})$ is the m.s.e. of any estimator $\hat{t}$.

The simulation process contains following steps:

STEP 1: Draw a random sample of size 20 from the population of N = 200 by SRSWOR.

STEP 2: Compute the value of different estimators and also for proposed estimator.

STEP 3: Repeat the above steps 10,000 times, which provides multiple sample based estimates $\hat{t}_1, \hat{t}_2........\hat{t}_{10,000}$.

STEP 4: Bias of $\hat{t}$ is obtained of by $B(\hat{t}) = \frac{1}{10,000}\sum_{i=1}^{10,000}\left[\hat{t}_i - S_y^2\right]$

STEP 5: Mean squared error of $\hat{t}$ is computed by $M(\hat{t}) = \frac{1}{10,000}\sum_{i=1}^{10,000}\left[\hat{t}_i - S_y^2\right]^2$

Estimator	Bias	M.S.E.	Efficiency
t_0	–13.8701	2380.7893	1
t_R	–25.9081	2356.1327	1.0105
t_{ST}	–23.8329	2391.9807	0.9953
$(T_{FT})_{k1}$	–21.1927	2115.9618	1.1252
$(T_{FT})_{k2}$	–25.1326	2491.5132	0.9556
$(T_{FT})_{k3}$	–22.8591	2129.7459	1.1179

Table 5.1: Bias, Mean Squared Error and Efficiency of Different Estimators

6. Values of k for Unbiased Estimator T_{FT}

For unbiased estimator as we know that

$$B(T_{FT})=0$$

$$\Rightarrow S_y^2\,F(\alpha_1-\alpha_2)[(\lambda_{22}-1)-\alpha_2(\lambda_{04}-1)]=0$$

$$\Rightarrow(\alpha_1-\alpha_2)[(\lambda_{22}-1)-\alpha_2(\lambda_{04}-1)]=0 \tag{6.1}$$

Case 1: $\alpha_1-\alpha_2=0\Rightarrow\dfrac{fB-C}{A+fB+C}=0\Rightarrow fB-C=0$

$$\Rightarrow f(k-1)(k-4)-(k-2)(k-3)(k-4)=0$$

$$\Rightarrow(k-4)[f(k-1)-(k-2)(k-3)]=0 \tag{6.2}$$

From (6.2) either

$$(k-4)=0\Rightarrow k=k_1^{'}=4$$

or $$k^2-(f+5)k+(6+f)=0 \tag{6.3}$$

for the equation (6.3), the two values of k are

$$k_2^{'}=\frac{(f+5)+\sqrt{(f+5)^2-4(f+6)}}{2} \tag{6.4}$$

$$k_3^{'}=\frac{(f+5)-\sqrt{(f+5)^2-4(f+6)}}{2} \tag{6.5}$$

On putting the value of f we get, $k_2^{'}=3.4971,\ k_3^{'}=1.8091$

Case 2:

$$[(\lambda_{22}-1)-\alpha_2(\lambda_{04}-1)]=0$$

$$\Rightarrow\alpha_2=\frac{(\lambda_{22}-1)}{(\lambda_{04}-1)}$$

$$\Rightarrow k^3(\lambda_{22}-\lambda_{04})-k^2[9\lambda_{04}+f(\lambda_{22}-1)-8\lambda_{22}-1]$$
$$+k[23\lambda_{22}-26\lambda_{04}-5f(\lambda_{22}-1)+3]+[24\lambda_{04}-22\lambda_{22}+4f(\lambda_{22}-1)-2]=0 \tag{6.6}$$

On putting the value of λ_{22}, λ_{04} and f in (5.6), we can obtain the three values of k as

$$k_4^{'} = 1.7205,\ k_5^{'} = 2.6017,\ k_6^{'} = 6.1887$$

7. Conclusion

The factor-type estimator for the estimation of population variance is found effective and efficient over other existing estimators. In fact, procedures of other estimators are somewhat special cases of F-T variance estimator when

$k = 1$: Ratio estimator, $k = 2$: Product estimator,

$k = 3$: Dual to ratio estimator, $k = 4$: Unbiased estimator

This shows that the factor-type estimator may be looked upon as a general class estimator consisting of other existing estimators as special cases. A specific property with variance of F-T estimator is that there are three values of constant k for which m.s.e. attains the optimum choice. The choice of k among these has the lowest bias. In this way, the factor-type variance estimator reduces the bias along with maintaining the optimum level of m.s.e. Table 5.1 shows better efficiency and bias of F-T estimator over unbiased, ratio and dual to ratio estimators. It is clear that the estimator $(T_{FT})_{k1}$ is best among all.

More, importantly, the factor-type variance estimator could be made almost unbiased also by an appropriate choice of multiple available k values. We can choose k values for different pair of (f, P) values.

Acknowledgment

The authors are thankful to the referees and editor of JRSS for their valuable comments and suggestions regarding improvement of the paper.

References

1. Agrawal, M.C. and Sthapit, A.B. (1995). Unbiased ratio-type variance estimation, Statistics and Probability Letters, 25, p. 361–364.
2. Arcos, A., Rueda, M., Martinez, M. D., Gonzalez, S. and Roman, Y. (2005). Incorporating the auxiliary information available in variance estimation, Applied Mathematics and Computation, 160, p. 387–399.
3. Asghar, A., Sanaullah, A. and Hanif, M. (2014). Generalized exponential type estimator for population variance in survey sampling, Colombian Journal of Statistics, 37 (1), p. 211–222.
4. Bandyopadhyay, A. and Singh, G. N. (2015). Estimation of ratio of population variances in absence and presence of non-response, Journal of Reliability and Statistical Studies, 8 (1), 77–93.
5. Das, A. K. and Tripathi, T. P. (1978). Use of auxiliary information in estimating the finite population variance, Sankhya, C, 40, p. 139 – 148.
6. Garcia, M. R. and Cebrain, A. A. (1996). Repeated substitution method: The ratio estimator for the population variance, Metrika, 43, p. 101–105.
7. Isaki, C. T. (1983). Variance estimation using auxiliary information, Journal of the American Statistical Association, 78, p. 117–123.

8. Kadilar, C. and Cingi, H. (2006). Improvement in variance using auxiliary information, Hecettepe Journal of Mathematics and Statistics, 35 (1), p. 111–115.
9. Kadilar, C. and Cingi, H. (2006). Ratio estimators for the population variance in simple and stratified random sampling, Applied Mathematics and Computation, 173, p. 1047–1059.
10. Kumar, S. (2014). Variance estimation in presence of random non-response, Journal of Reliability and Statistical Studies, 7 (2), p. 65–70.
11. Misra, P. (2016). On improved double sampling estimators of population variance using auxiliary information, Journal of Reliability and Statistical Studies, 9 (1), p. 93–100.
12. Shukla, D. and Thakur, N. S. (2008). Estimation of mean with imputation of missing data using factor-type estimator, Statistics in Transition, 9 (1), p. 33–48.
13. Singh, H. P. and Katyar, N. P.(1991). Variance estimation through the mean square successive differences and sample variance using apriori information, Journal of the Indian Society of Agricultural Statistics, 43 (1), p. 16–29.
14. Singh, H. P. and Solanki, R. S. (2013). A new procedure for variance estimation in simple random sampling using auxiliary information, Statistical papers, 54 (2), p. 479–497.
15. Singh, H. P. and Solanki, R. S. (2013a). Improved estimation of finite population variance using auxiliary information, Communication in statistics-Theory and Methods , 42, p. 2718–2730.
16. Singh, H. P. Upadhyaya, L. N. and Namjoshi, U. D. (1988). Estimation of finite population variance, Current Science, 57, p. 1331–1334.
17. Singh, H. P., Pal, S. K. and Solanki, R. S. (2014). A new procedure for estimation of finite population variance using auxiliary information; Journal of Reliability and Statistical Studies, 7, (2), p. 149–160.
18. Singh, V. K. and Shukla, D. (1987). One parameter family of factor-type ratio estimator, Metron, 45 (1- 2), p. 273–283.
19. Srivenkataramana, T. (1980). A dual to ratio estimator in sample surveys. Biometrika, 67, p. 199-204.
20. Wolter, K. M. (2007). Introduction to Variance Estimation; 2nd edition, Springer.
21. Yadav, S. K. and Kadilar, C. (2013). A class of ratio-cum-dual to ratio estimator of population variance, Journal of reliability and statistical studies, 6 (1), p. 29–34.
22. Yadav, S.K. and Kadilar, C. (2013a). Improved exponential type ratio estimator of population variance, Colombian Journal of Statistics, 36(1), p. 145–152.