Statistics and Applications {ISSN 2452-7395 (online)}
Volume 16 Nos. 1, 2018 (New Series), pp 365-373

Estimation of Finite Population Variance Under Systematic Sampling Using Auxiliary Information

Uzma Yasmeen[1], Muhammad Noor-ul-Amin[2] and Muhammad Hanif[3]
[1,3]*National College of Business Administration and Economics.*
[2]*Comsats Institute of Information Technology, Lahore.*

Received: April 17, 2018; Revised: May 26, 2018; Accepted: May 28, 2018

Abstract

This paper is based on the proposal of generalized estimator for finite population variance using an auxiliary variable under systematic sampling design. The expressions of bias and mean square error are obtained for the proposed estimator. The conditions are obtained for which the proposed estimators are better than the usual estimator.An empirical and simulation study is conducted to prove the superiority of the proposed estimator.

Key words: Auxiliary variables, Relative efficiency, Exponential estimator

1. Introduction

In the survey sampling, it is well recognized that the efficiency of an estimator of the population parameter can be increased by the use of auxiliary information associated to auxiliary variable, which is related with the response variable. At planning stage or at design stage, auxiliary variables may be efficiently employed to arrive at an enhanced estimator compared to those estimators not utilizing information of auxiliary variable.

We shall use information of the auxiliary variable under the context of systematic sampling technique. Systematic sampling is the simplest type of sampling design that needs only one random start. It offer good results in some situations like forest regions for assessing the volume of the timber, field operations of any survey, agriculture and meteorology, etc. Due to its simplicity, systematic sampling technique is the frequently used sampling technique in surveys of finite population. Apart from its simplicity, systematic sampling provides estimators which are more efficient than simple random sampling or stratified random sampling for certain types of populations. Various authors including Cochran (1946), Gautschi (1957), Hajeck (1959), Singh and Singh (1998), Kadilar and Cingi (2006), Koyuncu and Kadilar (2009), Singh et al. (2011), Tailor et al. (2013), Yasmeen et al (2015, 2016, 2018), Noor-ul-Amin et al(2017) have discussed the estimation of population parameters using auxiliary information under different sampling designs.

Corresponding Author: Muhammad Noor-ul-Amin
Email: nooramin.stats@gmail.com

In this paper, variance estimators have been proposed under systematic sampling technique. Section 2 presents some notations and sampling process to follow the systematic sampling technique. In section 3, recommended estimators for the estimation of population variance have been presented. Further, the bias and mean square error (MSE) of proposed estimator is obtained. The simulation study for evaluating the proposed estimator is arranged in section 4. The conclusion is given in section 5.

2. Notations

Consider a finite population $P=\{P_1,P_2,\ldots,P_N\}$ having N distinct and identifiable units serially numbered from 1 to N. It is assumed that $N = nm$, where n and m are positive integers. A systematic sample of size n units is selected from the population by selecting the first unit at random from 1 to m and then selecting every subsequent m_{th} unit. The response variable y and the auxiliary variable x are observed for each and every selected unit in the sample. Suppose (y_{ji},x_{ji}) for $(j=1,2,\ldots,m)$ and $(i=1,2,\ldots,n)$ define the value of i^{th} unit in the j^{th} sample. Then the respective systematic sample variances for the variables x and y are $s_x^{*2}=\dfrac{\sum_{i=1}^{n}(x_{ji}-\bar{x})^2}{(n-1)}$, $s_y^{*2}=\dfrac{\sum_{i=1}^{n}(y_{ji}-\bar{y})^2}{(n-1)}$ and are the estimators of the population variance of x and y respectively. Suppose

$$e_{sysy}=\frac{s_{sysy}^2-S_y^2}{S_y^2},\ e_{sysx}=\frac{s_{sysx}^2-S_x^2}{S_x^2}$$

so that

$$E(e_{sysy})=E(e_{sysx})=0,\quad E(e_{sys}^2)=A_y=\gamma^2A^2[\alpha_{2y}-1]\ ,\quad E(e_{sysx}^2)=B_x=\gamma^2B^2[\alpha_{2x}-1]\ ,$$

$$E(e_{sysy}e_{sysx})=C_{yx}=\gamma BA[\vartheta_{yx}-1],\ \gamma=\frac{N(N-1)}{N^2n}.$$

where

$$\alpha_{2y}=\frac{\sum_{i=1}^{N}\left[(y_{ji}-\bar{Y})^4\right]/(N-1)}{\left[\sum_{i=1}^{N}\left\{(y_{ji}-\bar{Y})^2/(N-1)\right\}\right]^2},\ \alpha_{2x}=\frac{\sum_{i=1}^{N}\left[(x_{ji}-\bar{X})^4\right]/(N-1)}{\left[\sum_{i=1}^{N}\left\{(x_{ji}-\bar{X})^2/(N-1)\right\}\right]^2},$$

$$\vartheta_{yx}=\frac{\sum_{i=1}^{N}\left[(y_{ji}-\bar{Y})^2(x_{ji}-\bar{X})^2/(N-1)\right]}{\left[\sum_{i=1}^{N}\left\{(y_{ji}-\bar{Y})^2/(N-1)\right\}\right]\left[\sum_{i=1}^{N}\left\{(x_{ji}-\bar{X})^2/(N-1)\right\}\right]},$$

$$A=\{1+(n-1)\rho_y\},\ B=\{1+(n-1)\rho_x\},\ \rho_y=\frac{(y_{ji}-\bar{Y})(y_{ji'}-\bar{Y})}{E(y_{ji}-\bar{Y})^2},$$

$$\rho_x=\frac{(x_{ji}-\bar{X})(x_{ji'}-\bar{X})}{E(x_{ji}-\bar{X})^2}$$

are the corresponding intra-class correlation coefficients for the response variable y and auxiliary variable x, respectively.

3. Proposed Estimators

In this section, under the framework of systematic sampling technique, the two estimators are developed using information of auxiliary variable for the estimation of finite population variance. The proposed-I variance estimator is given by

$$Q_{sysG1} = \left(s_{sys}^2 \frac{\omega S_x^2 + \eta}{\omega s_{sysx}^2 + \eta} \right). \tag{3.1}$$

where ω and η are the parameters of auxiliary variable such as the coefficient of correlation, the coefficient of variation, the coefficient of Skewness, kurtosis, Median, the Tri-mean and the Quartile Deviation.

The proposed-II exponential type estimator is given by

$$Q_{sysG2} = s_{sysy}^2 \exp\left(\frac{S_x^2 - s_{sysx}^2}{S_x^2} \right). \tag{3.2}$$

3.1. Bias and mean square error of proposed estimators

$$Q_{sysG1} = S_y^2(1+e_{sysy})\left(1 - \frac{\omega S_x^2 e_{sysx}}{\omega S_x^2 + \eta} \right). \tag{3.3}$$

After simplification, we have

$$Q_{sysG1} = S_y^2(1+\overline{e}_{sysy})\left(1+\Omega_i e_{sysx}\right)^{-1}. \tag{3.4}$$

where

$$\Omega_i = \frac{\omega S_x^2}{\omega S_x^2 + \eta}. \; i=0,\, 1\ldots.35$$

Expending and neglecting the higher order terms, we obtained

$$Q_{sysG1} - S_y^2 \approx \left[S_y^2 e_{sysy} - S_y^2 \Omega_i e_{sysx} e_{sysy} - S_y^2 \Omega_i e_{sysx} + S_y^2 \Omega_i^2 e_{sysx}^2 \right]. \tag{3.5}$$

After taking expectations, we obtain expression of bias

$$Bias\left(Q_{sysG1}\right) \approx S_y^2 \left(\gamma^2 B^2 \Omega_i^2 \left[\alpha_{2x} - 1\right] - \Omega_i \gamma BA\left[\vartheta_{yx} - 1\right]\right). \tag{3.6}$$

In order to obtain mean square error of Q_{sysG1}, again using (3.5), ignoring the terms of power two and greater, we obtained

$$E(Q_{sysG1}-S_y^2)\approx S_y^4 E\left[e_{sysy}^2+\Omega_i^2 e_{sysx}^2-2\Omega_i e_{sysy}e_{sysx}\right] \tag{3.7}$$

After applying expectations on both sides, the expression of mean square error is given by

$$MSE(Q_{sysG1})=\gamma^2 S_y^4\left[A^2\left(\alpha_{2y}-1\right)+\Omega_i^2 B^2\left(\alpha_{2x}-1\right)-\frac{2}{\gamma}\Omega_i BA\left(\vartheta_{yx}-1\right)\right]. \tag{3.8}$$

3.2. Bias and Mean Square Error of Proposed Estimator-II

By using the notations given in section 2, it can be written (3.2) as

$$Q_{sysG2}=S_y^2(1+\bar{e}_{sysy})\exp\left(\frac{S_x^2-S_x^2\left(1+e_{sysx}\right)}{S_x^2}\right). \tag{3.9}$$

Expending and neglecting the higher order terms, we obtained

$$Q_{sysG2}-S_y^2\approx S_y^2\left[e_{sysy}-e_{sysx}e_{sysy}-e_{sysx}+\frac{e_{sysx}^2}{2}\right] \tag{3.10}$$

After taking expectations we obtain expression of bias

$$Bias\left(Q_{sysG2}\right)\approx S_y^2\left(\frac{\gamma^2}{2}B^2\left[\alpha_{2x}-1\right]-\gamma BA\left[\vartheta_{yx}-1\right]\right). \tag{3.11}$$

After simplification, the expression of mean square error is given by

$$MSE(Q_{sysG2})\approx\gamma^2 S_y^4\left[A^2\left(\alpha_{2y}-1\right)+B^2\left(\alpha_{2x}-1\right)-\frac{2}{\gamma}BA\left(\vartheta_{yx}-1\right)\right] \tag{3.12}$$

4. Some Special Cases of Suggested Estimator

The following special cases of the proposed-I estimator are summarized as

i. If $\omega=1,\rho,C_{wx},\beta_1,\beta_2$ and $\eta=0$ then the form of suggested estimator (3.1) is

$$Q_{sysG1}^{1}=\left(s_{sysy}^2\frac{S_x^2}{s_{sysx}^2}\right). \tag{4.1}$$

The respective approximate bias and expression of mean square error of estimator Q_{sysG1}^{1} is given by

$$Bias\left(Q^{1}_{sysG1}\right) \approx \gamma^{2} S_{y}^{2}\left(B^{2}\Omega_{1}^{2}\left[\alpha_{2x}-1\right]-\Omega_{1}\frac{1}{\gamma}AB\left[\vartheta_{yx}-1\right]\right).$$

and

$$MSE(Q^{1}_{sysG1}) \approx \gamma^{2} S_{y}^{4}\left[A^{2}\left(\alpha_{2y}-1\right)+\Omega_{1}^{2}B^{2}\left(\alpha_{2x}-1\right)-\frac{2}{\gamma}\Omega_{1}BA\left(\vartheta_{yx}-1\right)\right],$$

where $\Omega_1 = 1$.

ii. If $\omega = C_x$ and $\eta = 0$ then the form of suggested estimator (3.1) is

$$Q^{4}_{sysG1} = \left(s^{2}_{sysy}\frac{C_x S_x^2+1}{C_x s^{2}_{sysx}+1}\right). \tag{4.2}$$

The respective approximate bias and expression of mean square error of estimator Q^{4}_{sysG1} is given by

$$Bias\left(Q^{4}_{sysG1}\right) \approx \gamma^{2} S_{y}^{2}\left(B^{2}\Omega_{4}^{2}\left[\alpha_{2x}-1\right]-\Omega_{4}\frac{1}{\gamma}AB\left[\vartheta_{yx}-1\right]\right).$$

and

$$MSE(Q^{4}_{sysG1}) \approx \gamma^{2} S_{y}^{4}\left[A^{2}\left(\alpha_{2y}-1\right)+\Omega_{4}^{2}B^{2}\left(\alpha_{2x}-1\right)-\frac{2}{\gamma}\Omega_{4}BA\left(\vartheta_{yx}-1\right)\right],$$

where $\Omega_4 = \dfrac{C_x S_x^2}{C_x S_x^2+1}$.

iii. If $\omega = \beta_2$ and $\eta = 1$ then the form of suggested estimator (3.1) is

$$Q^{6}_{sysG1} = \left(s^{2}_{sysy}\frac{\beta_2 S_x^2+1}{\beta_2 s^{2}_{sysx}+1}\right). \tag{4.3}$$

The respective approximate bias and expression of mean square error of estimator Q^{6}_{sysG1} is given by

$$Bias\left(Q^{6}_{sysG1}\right) \approx \gamma^{2} S_{y}^{2}\left(B^{2}\Omega_{6}^{2}\left[\alpha_{2x}-1\right]-\Omega_{6}\frac{1}{\gamma}AB\left[\vartheta_{yx}-1\right]\right).$$

and

$$MSE(Q^{6}_{sysG1}) \approx \gamma^{2} S_{y}^{4}\left[A^{2}\left(\alpha_{2y}-1\right)+\Omega_{6}^{2}B^{2}\left(\alpha_{2x}-1\right)-\frac{2}{\gamma}\Omega_{6}BA\left(\vartheta_{yx}-1\right)\right],$$

where $\Omega_6 = \dfrac{\beta_2 S_x^2}{\beta_2 S_x^2+1}$.

iv. If $\omega = 1$ and $\eta = QD$ then the form of suggested estimator (3.1) is

$$Q_{sysG1}^{17} = \left(s_{sysy}^{2} \frac{S_x^2 + QD}{s_{sysx}^{2} + QD} \right). \tag{4.4}$$

The respective approximate bias and expression of mean square error of estimator Q_{sysG1}^{17} is given by

$$Bias\left(Q_{sysG1}^{17}\right) \approx S_y^2 \left(\gamma^2 B^2 \Omega_{17}^2 \left[\alpha_{2x} - 1\right] - \Omega_{17} \gamma AB \left[\vartheta_{yx} - 1\right]\right).$$

and

$$MSE(Q_{sysG1}^{17}) \approx \gamma^2 S_y^4 \left[A^2 \left(\alpha_{2y} - 1\right) + \Omega_{17}^2 B^2 \left(\alpha_{2x} - 1\right) - \frac{2}{\gamma} \Omega_{17} BA \left(\vartheta_{yx} - 1\right) \right],$$

where $\Omega_{17} = \dfrac{S_x^2}{S_x^2 + QD}$.

v. If $\omega = \beta_1$ and $\eta = QD$ then the form of suggested estimator (3.1) is

$$Q_{sysG1}^{20} = \left(s_{sysy}^{2} \frac{\beta_1 S_x^2 + QD}{\beta_1 s_{sysx}^{2} + QD} \right). \tag{4.5}$$

The respective approximate bias and expression of mean square error of estimator Q_{sysG1}^{20} is given by

$$Bias\left(Q_{sysG1}^{20}\right) \approx S_y^2 \left(\gamma^2 B^2 \Omega_{20}^2 \left[\alpha_{2x} - 1\right] - \Omega_{20} \gamma BA \left[\vartheta_{yx} - 1\right]\right).$$

and

$$MSE(Q_{sysG1}^{20}) \approx \gamma^2 S_y^4 \left[A^2 \left(\alpha_{2y} - 1\right) + \Omega_{20}^2 B^2 \left(\alpha_{2x} - 1\right) - \frac{2}{\gamma} \Omega_{20} BA \left(\vartheta_{yx} - 1\right) \right],$$

where $\Omega_{20} = \dfrac{\beta_1 S_x^2}{\beta_1 S_x^2 + QD}$.

5. Numerical Illustration

A simulation study is conducted to demonstrate the efficiency of developed estimators using the information on single auxiliary variable under systematic sampling technique. In this section, the performance of the proposed estimators is evaluated for real population. The comparison is presented in the form of percentage relative efficiency.

The percentage relative efficiencies (PRE), given in Table 1, are computed using the following formulae

$$PRE = \frac{\mathrm{var}(Q_0)}{\mathrm{var}(Q_*)} \times 100 \tag{5.1}$$

$Q_0 = s_y^2, Q_* = Q_0, Q_{sysG1}^1, Q_{sysG1}^4, Q_{sysG1}^6, Q_{sysG1}^{17}, Q_{sysG1}^{20}, Q_{sysG2},$

where s_{sysy}^2 is denoted by usual sample variance and Q_{sysG1}^1, Q_{sysG1}^4, Q_{sysG1}^6, Q_{sysG1}^{17}, Q_{sysG1}^{20}, Q_{sysG2} are denoted by developed estimators used in the formulae of the percent relative efficiencies.

For this study, consider the population given in the Table (See Appendix) for summarizing the simulation procedures. The procedure is used to find the efficiency of the developed estimators over usual sample variance.

The following steps are used in R-Language to perform simulation:
Step 1: From the described population, the sample size is considered as $n = 3$. The procedure is repeated 100,000 times to calculate the several values of estimators under systematic sampling technique.
Step 2: Using the sample obtained in step 1, the 100,000 values of s_{sysy}^2 and Q_{sysG1}, Q_{sysG2}, separately are obtained using (3.1) and (3.2), respectively.
Step 3: Using the values found in step 2, the values of percent relative efficiency (PRE) is computed by (5.1) and reported in Table 1.

This simulation study involves the evaluation of the percent relative efficiencies of the proposed estimators and usual sample variance for sample size $n = 3$.The results obtained from the simulation study indicate that the developed variance estimators are more efficient than the usual estimator.

Table 1: Percent Relative Efficiencies and Relative Biases of Different Estimators with respect to s_{sysy}^2

Estimator	PRE	RB
Q_0	100.00	0.0191
Q_{sysG1}^1 **(proposed)**	215.00	–0.2365
Q_{sysG1}^4 **(proposed)**	272.00	–0.0703
Q_{sysG1}^6 **(proposed)**	236.00	–0.0348
Q_{sysG1}^{17} **(proposed)**	295.87	–0.0201
Q_{sysG1}^{20} **(proposed)**	245.76	–0.0864
Q_{sysG2} **(proposed)**	496.97	–0.0976

It is concluded From Table 1 that proposed estimators are more efficient than the usual sample variance for the sample size used in the Example. It is clear that proposed estimators are more useful than usual variance estimator as the performance of suggested estimators are better than the usual variance estimator. We have suggested ratio estimators for estimating variance of a finite population. From the simulation results, given in Table 1, we infer that the suggested estimators are more efficient estimators than the usual sample variance for the finite population. Hence, the suggested estimators are recommended for its practical use for estimating variance of a finite population when single auxiliary variable is available. The simulated efficiencies for the estimators are obtained in Table 1 for the sample size 3. The superiority of proposed estimators may be concluded from the information of Table 1. However, the simulation studies may be extended with different population characteristics and different sample sizes.

References

Cochran, W.G. (1946). Relative accuracy of systematic and stratified random samples for a certain class of populations. *The Annals of Mathematical Statistics*, **17**, 164–177.

Gautschi, W. (1957). Some remarks on systematic sampling. *The Annals of Mathematical Statistics*, **28**, 385–394.

Hajeck, J. (1959). Optimum strategy and other problems in probability sampling. *Casopis pro Pestovani Matematiky*, **84**, 387–423.

Singh, H.P. and Singh, R. (1998). Almost unbiased ratio and product type estimators in systematic sampling. *Questiio*, **22**(**3**), 403–416.

Kadilar, C. and Cingi, H. (2006). An improvement in estimating the population mean by using the correlation coefficient. *Hacettepe Journal of Mathematics and Statistics*, **35**(**1**), 103–109.

Koyuncu, N. and Kadilar, C. (2009). Family of estimators of population mean using two auxiliary variables in stratified random sampling. *Communications in Statistics: Theory and Methods*, **38**(**13–15**), 2398–2417.

Singh, H. P., Tailor, R. and Jatwa, N.K. (2011). Modified ratio and product estimators for population mean in systematic sampling. *Journal of Modern Applied Statistical Methods*, **10**(**2**), 424–435.

Tailor, T., Jatwa, N. and Singh, H. P. (2013). A ratio-cum-product estimator of finite population mean in systematic sampling. *Statistics in Transition*, **14**(**3**), 391–398.

Khan, M. and Singh, R. (2015). Estimation of population mean in chain ratio-type estimator under systematic sampling. *Journal of Probability and Statistics*, Article ID 248374, 5 pages http://dx.doi.org/10.1155/2015/248374.

Khan, M. and Abdullah, H. (2016). A note on a difference type estimator for population mean under two phase sampling design. *Springer Plus* **5**:723 https://doi.org/10.1186/s40064-016-2368-1.

Noor-ul-Amin, M., Javaid, A. and Hanif, M. (2017). Estimation of population mean in systematic random sampling using auxiliary information. *Journal of Statistics and Management Systems*, **20**(**6**), 1095–1106.

Yasmeen, U., Noor ul Amin, M.N. and Hanif, M. (2015). Generalized exponential estimators of finite population mean using transformed auxiliary variables. *International Journal of Applied and Computational Mathematics*, **1**(**2**), 1–10.

Yasmeen U, Noor ul Amin M. and Hanif M. (2016). Exponential ratio and product type estimators of finite population mean. *Journal of Statistics and Management Systems*, **19**, 55–71.

Yasmeen, U., Noor-ul-Amin, M. and Hanif, M. (2018). Exponential estimators of finite population variance using transformed auxiliary variables. *Proceedings of the National Academy of Sciences*, India, Sect. A Physical Sciences: https://doi.org/10.1007/s40010-017-0410-5.

Appendix

We consider the data given by Cochran (1977) to examine the performances of different estimators of finite population variance.

Description of population parameters

Population	**Y**	**X**	**z**
	Food cost	Size	Income

Value of population parameters

Parameter	**Population**
N	33
n_1	9
n_2	4
$\bar{Y}$	27.4909
$\bar{X}$	3.7272
$\bar{Z}$	72.5454
C_y	0.3629
C_x	0.4025
C_z	0.1436
ρ_{yx}	0.4237
ρ_{yz}	0.2522
ρ_{xz}	–0.0660
λ_{400}	5.7200
λ_{040}	2.3800
λ_{004}	2.1400
λ_{220}	1.4300
λ_{202}	2.2870
λ_{022}	1.4920
λ_{210}	0.6365
λ_{201}	0.5506