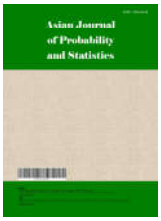

Asian Journal of Probability and Statistics

9(2): 1-8, 2020; Article no.AJPAS.61144
ISSN: 2582-0230

Development of Estimation Procedure of Population Variance in Stratified Randomized Response Technique

Nadia Mushtaq[1*] and Iram Saleem[1]

[1]Department of Statistics, Forman Christian College University, Lahore, Pakistan.

Authors' contributions

This work was carried out in collaboration between both authors. Author NM designed the study, performed the statistical analysis, wrote the protocol and wrote the first draft of the manuscript. Author IS managed the analyses of the study and managed the literature searches. Both authors read and approved the final manuscript.

Article Information

DOI: 10.9734/AJPAS/2020/v9i230221
Editor(s):
(1) Dr. Manuel Alberto M. Ferreira, Lisbon University, Portugal.
Reviewers:
(1) Avadh Pati, National Institute of Technology Silchar, India.
(2) Naveen Boiroju, National Institute of Nutrition, India.
Complete Peer review History: http://www.sdiarticle4.com/review-history/61144

Original Research Article

Received: 10 July 2020
Accepted: 17 September 2020
Published: 28 September 2020

Abstract

Singh et al. (2016) presented a ratio and regression estimators of population variance of a sensitive variable using auxiliary information based on randomized response technique (RRT). In this article, the RRT is considered in stratified random sampling for the estimation of variance. A generalized class of estimators of variance in stratified RRT is proposed and derive the procedure of variance estimation in stratified RRT. The expression of the bias and mean square error are expressed. The empirical findings support the soundness of proposed scheme of variance estimation.

Keywords: Stratified random sampling; sensitive variable; randomized response technique; variance estimation.

1 Introduction

In the sensitive issues such as abortion rates, tax evasion and use of illegal resources, RRT is used to obtain trustworthy information. Respondents who give evasive or dishonest answers introduce response bias into the study, resulting in questionable data and poor results Warner, [1]. When faced with this problem,

**Corresponding author: E-mail: nadiamushtaq@fccollege.edu.pk;*

researchers using the traditional direct questioning survey method are likely to try to gain the confidence of the respondent Warner, [1]. However, this is unreliable, because many people will not be inclined to confide certain things at all, and others would not want their confessions written down or linked to them in any way Warner, [1]. The randomized response technique (RRT) was developed over forty years ago to counter these problems with response bias by increasing the number of honest answers given to sensitive questions in a survey or interview.

It is common practice in sample survey related to market, industries and social research, and so forth that usually more than one characteristic is observed from each sampled unit of population. So stratified random sampling is more suitable than other survey designs used for obtaining information from the heterogeneous population for reasons of economy and efficiency. Variance estimation in survey sampling is of significant importance. It gives information on the accuracy of the estimators and minimum value of the variance desired.

Many survey statisticians such as Eichhorn and Hayre [2], Saha [3], Diana and Perri [4] have presented different randomized response models to estimate the population mean of a sensitive variable when there is no auxiliary information. Later, Sousa et al. [5] introduced ratio estimators to estimator population mean of sensitive study variable using auxiliary information. Sousa et al. [5]. Gupta et al. [6], Koyuncu et al. [7], Sanaullah et al. [8], Saleem et al. [9] and Sanaullah et al. [10] presented mean estimators to estimate the population mean based on randomized response technique. Mushtaq et al. [11] introduced ratio, regression and general class of estimators for the estimation of population mean using non-sensitive variable under stratified random sampling.

Gupta et al. [12] introduced a new exponential estimator for the estimation of the population mean of sensitive variable in the presence of non-sensitive auxiliary variables. Mushtaq and Noor-ul-Amin [13] presented generalized variance estimator based on additive randomized response model using non-sensitive auxiliary information

This present study focuses on the development of estimation procedure of variance in the randomized response technique. To our knowledge, no one has suggested an estimator of variance in stratified random sampling based on additive randomized response model. The proposed estimator is presented in Section 3. Section 4 represents the simulation study to examine the performance of the proposed estimators. Some concluding remarks are in Section 5.

2 Sampling Structure & Notations

Consider a finite population with N units $\Omega=\left(U_1,U_2,...,U_N\right)$ of size N. Let the population is divided into l strata with strata size N_h, such that $\sum_{h=1}^{l}N_h=N\ \left(h=1,...,l\right)$. The sample of size n_h is drawn from h^{th} stratum is n_h such that $\sum_{h=1}^{l}n_h=n$. Let (y_{hi} , x_{hi}) be the values of the sensitive study variable and the auxiliary variable on i^{th} unit of the h^{th} stratum, where $i=1,2,\ldots,N_h$ and $h=1,2,...,l$.

Let us assume that $\left(S_{yh}^2,S_{xh}^2\right)$ be the population variance of $\left(Y,X\right)$ in the stratum h, where $w=\frac{N_h}{N}$ is the stratum weight. Noor-ul-Amin et al. (2018) reported scrambled response for Y using the model, given by $Z=Y+kS$. To estimate S_{yh}^2, it is assumed that S_{xh}^2 is known and S_{zh}^2 be the population variance of the scrambled variable Z in the stratum $h.$, by using $Z=Y+kS$.

Let us define

$$e_{0st} = \left(s_{zst}^2 - S_{zst}^2\right)\Big/ S_{zst}^2 \text{ and } e_{1st} = \left(s_{xst}^2 - S_{xst}^2\right)\Big/ S_{xst}^2 ,$$

Such that $E\left(e_{ist}\right) = 0,\ i = 0,1.$ to first degree approximations, we have

$$E\left(e_{0st}^2\right) = \theta\lambda_{40h}^*,\quad E\left(e_{1st}^2\right) = \theta\lambda_{04h}^*,\quad E\left(e_{0st}e_{1st}\right) = \theta\lambda_{22h}^*, \tag{1}$$

where

$$\gamma_h = \left(\frac{1}{n_h} - \frac{1}{N_h}\right),\ \lambda_{40h}^* = \left(\lambda_{40h} - 1\right),\ \lambda_{04h}^* = \left(\lambda_{04h} - 1\right),\ \lambda_{22h}^* = \left(\lambda_{22h} - 1\right),$$

$$\lambda_{rsh} = \frac{\mu_{rsh}}{\mu_{20h}^{r/2}\mu_{02h}^{s/2}} \text{ and } \mu_{rsh} = \frac{1}{N_h - 1}\sum_{i=1}^{N_h}\left(z_{hi} - \bar{Z}_h\right)^r\left(x_{hi} - \bar{X}_h\right)^s .$$

The usual variance estimator for sensitive variable in stratified random sampling is given by

$$t_{0st} = s_{st}^2 \tag{2}$$

$$\text{var}(t_{0st}) = \gamma_h\lambda_{40h}^* \tag{3}$$

The modified ratio estimator for estimating the population variance in stratified sampling for randomized response technique is given as:

$$t_{Rst} = s_{zst}^2\left(\frac{S_{xst}^2}{s_{xst}^2}\right) \tag{4}$$

$$Bias\left(t_{Rst}\right) \approx \gamma_h S_{z_h}^2\left(\lambda_{04h}^* - \lambda_{22h}^*\right), \tag{5}$$

$$MSE\left(t_{Rst}\right) \approx \gamma_h S_{zh}^4\left(\lambda_{40h}^* + \lambda_{04h}^* - 2\lambda_{22h}^*\right). \tag{6}$$

3 Formulation of Proposed Estimation Strategy

In this section, a generalized class of estimators for the estimation of population variance of a sensitive study variable using non-sensitive auxiliary information is presented as:

$$t_{kist} = \left[k_1 s_{zst}^2 + k_2\left(S_{xst}^2 - s_{xst}^2\right)\right]\left[\alpha\left(\frac{a_{st}S_{xst}^2 + b_{st}}{a_{st}s_{xst}^2 + b_{st}}\right) + (1-\alpha)\exp\left(\frac{a_{st}\left(S_{xst}^2 - s_{xst}^2\right)}{a_{st}\left(S_{xst}^2 + s_{xst}^2\right) + 2b_{st}}\right)\right] \tag{7}$$

Where k_1 and k_2 are weights whose values are to be determined, α =0 or 1, a_{st} and b_{st} are the parameters of the auxiliary variables.

To find the bias and the mean square error for this estimator use the notations given in section 2, and the proposed estimator in (7) is given by,

$$t_{kist} \cong \left[k_1 S_z^2 \left(1+e_{0st}\right) - k_2 S_x^2 e_{1st} \right] \left[\alpha \left(1+g e_{1st}\right)^{-1} + \left(1-\alpha\right) \exp\left\{ \frac{-1}{2} g e_{1st} \left(1+\frac{1}{2} e_{1st} \right)^{-1} \right\} \right], \tag{8}$$

Where

$$g = \frac{a_{st} S_x^2}{a_{st} S_x^2 + b_{st}}$$

$$t_{kist} - S_z^2 = \left(k_1 - 1\right) S_z^2 + k_1 S_z^2 \left[e_{0st} - \frac{1}{2} g \left(1+\alpha\right) e_{1st} + \frac{1}{8} g^2 \left(3+5\alpha\right) e_{1st}^2 - \frac{1}{2} g \left(1+\alpha\right) e_{0st} e_{1st} \right]$$

$$-k_2 S_x^2 \left[e_{1st} - \frac{1}{2} g \left(1+\alpha\right) e_{1st}^2 \right]. \tag{9}$$

The respective $Bias$ and MSE of t_{kist} is given by

$$Bias\left(t_{kist}\right) \cong \left(k_1 - 1\right) S_z^2 + k_1 S_z^2 \gamma_h \left(\frac{1}{8} g^2 \left(3+5\alpha\right) \lambda_{04h}^* - \frac{1}{2} g \left(1+\alpha\right) \lambda_{22h}^* \right)$$

$$-\frac{1}{2} k_2 S_x^2 \gamma_h g \left(1+\alpha\right) \lambda_{04h}^*. \tag{10}$$

$$MSE\left(t_{kist}\right) \cong \left(S_z^2\right)^2 \left[\left(k_1 - 1\right)^2 + k_1^2 \gamma_h \left\{ \lambda_{40h}^* + \frac{1}{4} g^2 \lambda_{04h}^* \left(\alpha^2 + 7\alpha + 4\right) - 2 g \lambda_{22h}^* \left(1+\alpha\right) \right\} \right.$$

$$-2 k_1 \gamma_h \left\{ \frac{1}{8} g^2 \left(5\alpha+3\right) \lambda_{04h}^* - \frac{1}{2} g \left(1+\alpha\right) \lambda_{22h}^* \right\} + k_2^2 \frac{S_x^4}{S_z^4} \theta \lambda_{04h}^*$$

$$\left. -2 k_2 \frac{S_x^2}{S_z^2} \frac{1}{2} g \gamma_h \left(1+\alpha\right) \lambda_{04h}^* - 2 k_1 k_2 \frac{S_x^2}{S_z^2} \gamma_h \left(\lambda_{22h}^* - g \left(1+\alpha\right) \lambda_{04h}^* \right) \right]. \tag{11}$$

In order to obtain the optimum values of k_1 and k_2, partially differentiating (11) and equating to zero, the expression obtained is as

$$k_{1(opt)} = \frac{1 - \frac{1}{8} \gamma_h g^2 \left(4\alpha^2 + 3\alpha + 1\right) \lambda_{04h}^*}{1 + \gamma_h \left\{ \lambda_{40h}^* \left(1 - \rho_{zxh}^2\right) - g^2 \frac{1}{4} \left(\alpha + 3\alpha^2\right) \lambda_{04h}^* \right\}}$$

$$k_{2(opt)} = \frac{S_z^2}{S_x^2} \left\{ \frac{1}{2} g \left(1+\alpha\right) + k_{1(opt)} \left(\frac{\lambda_{22h}^*}{\lambda_{04h}^*} - g \left(1+\alpha\right) \right) \right\}$$

Substituting these optimum values of k_1 and k_2 in (11), the minimum MSE of t_{kist} is given by

$$MSE\left(t_{kist}\right)_{\min} \cong \left(S_z^2\right)^2 \left[1-\frac{1}{4}g^2\gamma_h\left(1+\alpha\right)^2\lambda_{04h}^* - \frac{\left\{1-\frac{1}{8}\gamma_h g^2\left(4\alpha^2+3\alpha+1\right)\lambda_{04h}^*\right\}^2}{\left\{1+\gamma_h\left\{\lambda_{40h}^*\left(1-\rho_{zxh}^2\right)-g^2\frac{1}{4}\left(\alpha+3\alpha^2\right)\lambda_{04h}^*\right\}\right\}}\right] \quad (12)$$

By using (12), for different values of a, b and $\alpha=0$ or $\alpha=1$, we can get the minimum MSE_s of t_{kist} $\left(i=0,1,2,3,4,5\right)$.

3.1 Some members of proposed class of estimators

Different estimators can be generated from the proposed estimator given in (7) by substituting the suitable choices of a, b and $\alpha=0$ or $\alpha=1$, Some generated estimators are listed in Tables 1 & 2.

Table 1. Some members of proposed class of estimator t_{kist} $\left(i=0,1,2\right)$, When $\alpha=0$

a	b	***Estimator***
1	**0**	$t_{k0st}=\left[k_1 s_{zst}^2+k_2\left(S_{xst}^2-s_{xst}^2\right)\right]\left[\exp\left(\frac{\left(S_{xst}^2-s_{xst}^2\right)}{\left(S_{xst}^2+s_{xst}^2\right)}\right)\right]$
1	ρ_x	$t_{k1st}=\left[k_1 s_{zst}^2+k_2\left(S_{xst}^2-s_{xst}^2\right)\right]\left[\exp\left(\frac{\left(S_{xst}^2-s_{xst}^2\right)}{\left(S_{xst}^2+s_{xst}^2+2\rho_{xst}\right)}\right)\right]$
C_x	$\beta_2\left(x\right)$	$t_{k2st}=\left[k_1 s_{zst}^2+k_{2st}\left(S_{xst}^2-s_{xst}^2\right)\right]\left[\exp\left(\frac{C_x\left(S_{xst}^2-s_{xst}^2\right)}{C_x\left(\left(\left(S_{xst}^2+s_{xst}^2\right)\right)+2\beta_2\left(x\right)\right)}\right)\right]$

Table 2. Some members of proposed class of estimator t_{kist} $\left(i=3,4,5\right)$, When $\alpha=1$

a	b	***Estimator***
1	**0**	$t_{k3st}=\left[k_1 s_{zst}^2+k_2\left(S_{xst}^2-s_{xst}^2\right)\right]\left[\left(\frac{S_{xst}^2}{s_{xst}^2}\right)\right]$
1	ρ_x	$t_{k4st}=\left[k_1 s_{zst}^2+k_2\left(S_{xst}^2-s_{xst}^2\right)\right]\left[\left(\frac{S_{xst}^2+\rho_x}{s_{xst}^2+\rho_x}\right)\right]$
C_x	$\beta_2\left(x\right)$	$t_{k5st}=\left[k_1 s_{zst}^2+k_2\left(S_{xst}^2-s_{xst}^2\right)\right]\left[\left(\frac{C_x S_{xst}^2+\beta_2(x)}{C_x s_{xst}^2+\beta_2(x)}\right)\right]$

4 Empirical Investigation and Simulation Study

In this section, a simulation study is presented by comparing the performance of estimators discussed in this paper. Consider that the sensitive study variable *Y* and auxiliary variable *X* are related to each other and is defined as:

$$Y_i = RX_i + e_i \tag{13}$$

where $e_i \sim N(0,1)$ and R=1.5.

Consider a population of size N=1000. The auxiliary variable $X_i \sim G(a,b)$ is generated from gamma distribution with parameters a=2 and b=3. Assume the scrambling variable $S_i \sim B(\alpha,\beta)$ with, $\alpha = 6.5$ and $\beta = 0.5$. Then, the scrambled responses on the study variable as $Z_i = Y_i + kS_i$, $i = 1,2,3,\ldots,n$ and $k = 0.3, 0.5, 0.7$. Table 3 gives the results of MSE and percent relative efficiency (PRE) for the proposed estimators as compare to ratio estimator and mean estimator using the following expression:

$$PRE = \frac{MSE(t_{0st})}{MSE(t_{\omega})} \times 100$$

Where $= t_{Rst}, t_{ist}, t_{k0st}, t_{k1st}, t_{k2st}, t_{k3st}, t_{k4st}, t_{k4st}$. This process is repeated M=3000 times, for the different values of n such as 30, 150 and 300.

Table 3. The MSE and PRE of all estimators

N	N_h	n	Estimator	k=0.3		k=0.5		k=0.7	
				MSE	***PRE***	***MSE***	***PRE***	***MSE***	***PRE***
1000	N_1= 550 N_2=450	30	t_{0st}	0.04866	100.000	0.04965	100.000	0.05057	100.000
			t_{Rst}	0.03468	140.311	0.04572	108.596	0.04855	104.161
			t_{kist}	0.03158	**154.085**	0.03694	**134.407**	0.03715	**136.124**
			t_{k0st}	0.01033	**471.055**	0.01114	**445.691**	0.01108	**456.408**
			t_{k1st}	0.01504	**323.537**	0.01715	**289.504**	0.01749	**289.137**
			t_{k2st}	0.03116	**156.162**	0.03420	**145.175**	0.03483	**145.191**
			t_{k3st}	0.03788	**128.458**	0.03991	**124.405**	0.03964	**127.573**
			t_{k4st}	0.01315	**370.038**	0.01381	**359.522**	0.01411	**358.398**
			t_{k5st}	0.03309	**147.053**	0.03360	**147.768**	0.03292	**153.615**
		150	t_{0st}	0.00801	100.000	0.01003	100.000	0.01022	100.000
			t_{Rst}	0.00551	145.372	0.00707	141.867	0.00785	130.191
			t_{kist}	0.00423	**189.362**	0.00284	**353.169**	0.002923	**349.641**
			t_{k0st}	0.00280	**286.071**	0.002919	**343.611**	0.002898	**352.657**
			t_{k1st}	0.00154	**520.130**	0.00186	**539.247**	0.002066	**494.676**
			t_{k2st}	0.00606	**132.178**	0.00690	**145.362**	0.007168	**142.578**
			t_{k3st}	0.00277	**289.170**	0.00222	**451.802**	0.002281	**448.049**
			t_{k4st}	0.00265	**302.264**	0.00315	**318.413**	0.003440	**297.093**
			t_{k5st}	0.00505	**158.614**	0.00688	**145.785**	0.007149	**142.957**

N	N_h	n	Estimator	k=0.3		k=0.5		k=0.7	
				MSE	***PRE***	***MSE***	***PRE***	***MSE***	***PRE***
		300	t_{0st}	0.00310	100.000	0.00315	100.000	0.00329	100.000
			t_{Rst}	0.00262	118.321	0.00237	132.911	0.00222	148.198
			t_{kist}	0.00145	**213.793**	0.00149	**211.409**	0.00158	**208.228**
			t_{k0st}	0.00134	**231.343**	0.00147	**214.286**	0.00154	**213.636**
			t_{k1st}	0.00074	**418.919**	0.00069	**456.522**	0.00066	**498.485**
			t_{k2st}	0.00286	**108.392**	0.00273	**115.385**	0.00280	**117.500**
			t_{k3st}	0.00087	**356.322**	0.00090	**350.000**	0.00103	**319.417**
			t_{k4st}	0.00138	**224.638**	0.00126	**250.000**	0.00122	**269.672**
			t_{k5st}	0.00286	**108.392**	0.00272	**115.809**	0.00279	**117.921**

From the results, it is observed that the generalized proposed variance estimator in stratified sampling using randomized response model performs efficiently as compare to ordinary mean and ratio estimator under randomized response model.

5 Conclusion

Results clearly shows that the proposed estimators for the estimation of population variance using non-sensitive auxiliary information based on stratifies sampling design performs more efficiently. The proposed generalized estimator provides lower MSE's than the MSE of usual variance and ratio estimator for different values of '*k*' under the proposed scrambled randomized response model. Also, as the sample size increase the MSE decreases for all estimators and there is an increase in the efficiency of all estimators.

Competing Interests

Authors have declared that no competing interests exist.

References

[1] Warner SL. Randomized response: A survey technique for eliminating evasive answer bias. Journal of the American Statistical Association. 1965;60(309):63–69.

[2] Eichhron BH, Hayre LS. Scrambled randomized response methods for obtaining sensitive quantitative data. Journal of Statistical Planning and Inference.1983;7:307–316.

[3] Saha A. A randomized response technique for quantitative data under unequal probability sampling. Journal of Statistical Theory and Practice. 2008;2(4):589–596.

[4] Diana G, Perri PF. New scrambled response models for estimating the mean of a sensitive quantitative character. Journal of Applied Statistics. 2010;37(11):1875–1890.

[5] Sousa R, Shabbir J, Real PC, Gupta S. Ratio estimation of the mean of a sensitive variable in the presence of auxiliary information. Journal of Statistical Theory and Practice. 2010;4(3):495–507.

[6] Gupta S, Shabbir J, Sousa R, Real PC. Estimation of the mean of a sensitive variable in the presence of auxiliary information. Communications in Statistics-Theory and Methods. 2012;41:1–12.

[7] Koyuncu N, Gupta S, Sousa R. Exponential-type estimators of the mean of a sensitive variable in the presence of non sensitive auxiliary information. Communications in Statistics-Simulation and Computation. 2014;43(7):1583–1594.

[8] Sanaullah A, Saleem I, Shabbir J. Use of scrambled response for estimating mean of the sensitivity variable. Communications in Statistics-Theory and Methods. 2020;49(11):2634–2647.

[9] Saleem I, Sanaullah A, Hanif M. Double-sampling regression-cum-exponential estimator of the mean of a sensitive variable. Mathematical Population Studies. 2009;26(3):163–182.

[10] Sanaullah A, Saleem I, Gupta S, Hanif M. Mean estimation with generalized scrambling using two-phase sampling. Communications in Statistics-Simulation and Computation. 2020;1-15.

[11] Mushtaq N, Noor-ul-Amin M, Hanif M. Estimation of a population mean of a sensitive variable in stratified two-phase sampling. Pakistan Journal of Statistics. 2016;32(5).

[12] Gupta S, Shabbir J, Sousa R, Corte-Real P. Improved exponential type estimators of the mean of a sensitive variable in the presence of non-sensitive auxiliary information. Communications in Statistics-Simulation and Computation. 2016;45 (9):3317–3328.

[13] Mushtaq N, Noor-Ul-Amin M. Generalized variance estimators using randomized device in the presence of auxiliary information. Journal of Statistics and Management Systems. 2019;22(8):1417-1424.

Peer-review history:
The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)
http://www.sdiarticle4.com/review-history/61144