Journal of Statistics Applications & Probability Letters
An International Journal

http://dx.doi.org/10.12785/jsapl/020205

On The Estimation of Population Mean in Rotation Sampling

Housila P. Singh and *Surya K. Pal* *.

School of Studies in Statistics, Vikram University Ujjain, M.P., 456010 India.

Received: 14 Sep. 2014, Revised: 26 Oct. 2014, Accepted: 9 Nov. 2014,
Published online: 1 May 2015

Abstract: This paper deals with the problem of estimating the current population mean in two-occasion rotation sampling using the known population standard deviation (S_z) along with known population mean ($\bar{Z}$) of the auxiliary character z on both the occasions and the information on the study character from the previous occasion. The proposed estimator has been compared with simple mean estimator when there is no matching and the optimum estimator, which is a combination of the means of the matched and unmatched portion of the sample at the second occasion. Optimum replacement policy and the efficiency of the suggested estimator have been discussed. To examine the merits of the proposed estimator an empirical study is carried out.

Keywords Auxiliary variable, Study variable, Rotation sampling, Mean square error.

1 Introduction

In rotation sampling, it is common practice to utilize the entire information collected on the previous occasions to improve the precision of the estimates on the current occasion. The previous information may be in the form of an auxiliary variable, the character under study itself or both. There are many problems of practical interest in different fields of applied environmental sciences in which the various characters opt to change over time with respect to different parameters. Hence, one is often concerned with the measuring the characteristics of a population on several occasions to estimate the trend in time of different population parameters as time series or the current value of the parameters over several points of time. Rotation sampling provides a strong tool for generating the reliable estimate at different occasions. Jessen (1942) was the first who introduced the idea of sampling on two occasions by using the information gathered on the previous occasions to improve the precision of the current estimate. Latter various authors including Patterson (1950), Eckler (1955), Rao and Graham (1964) and Singh et al. (1992), Singh et al. (2014) among others have developed the theory of rotation sampling. Feng and Zou (1997) and Biradar and Singh (2001) used the auxiliary information on both the occasions for estimating the current population mean in the rotation sampling. Singh (2003, 2005) and Singh and Vishwakarma (2007a,b, 2009), have used the auxiliary information on both the occasion and envisaged several estimators for the estimating the population mean on current occasion in two- occasion rotation sampling. In this paper we have suggested an estimator utilizing the known population mean $\bar{Z}$ alongwith known standard deviation (S_z) of the auxiliary variable z on both the occasion in rotation sampling for estimating the current population mean in two occasion rotation sampling. The proposed estimator is compared empirically with the estimators where no such auxiliary information is used and suitable recommendations are given.

2 Proposed Estimator

Let us consider a finite population $U= (U_1, U_2 \ldots U_N)$ of size N (large) units, which is assumed to remain unchanged over two occasions. Let x (y) be the variable under study on the first (second) occasion respectively. It is assumed that information on an auxiliary variable z (stable over occasion) is readily available for the both the occasions; whose population standard deviation (S_z) alongwith population mean $\bar{Z}$ is known and it is highly positively correlated to x and y on the first and second occasion. A simple random sample of n units is drawn without replacement (WOR) on the first occasion. A random sub sample of m (= $n\lambda$) units is retained (matched) from the sample drawn on the first occasion for its use on the current (second) occasion, while a fresh sample of size u= (n-m) = $n\mu$ units is drawn on the current (second) occasion, from the entire population by simple random sampling without replacement (SRSWOR) method so that the sample size on the current

*Corresponding author e-mail: suryakantpal6676@gmail.com

(second) is also n. The fractions of the matched and fresh samples are respectively designated by λ and μ such that $\lambda+\mu=1$.

In what follows we shall use the following notations throughout this paper.

$\bar{X}$, $\bar{Y}$, $\bar{Z}$: Population means of the variables x, y and z respectively.

$\bar{x}_m$, $\bar{x}_n$, $\bar{y}_u$, $\bar{y}_m$, $\bar{z}_u$, $\bar{z}_n$: Sample means of the respective variables based on the sample sizes indicated in suffices.

C_x, C_y, C_z : Coefficients of variation of the variables x, y and z respectively,

ρ_{yx}, ρ_{yz}, ρ_{xz} : The correlation coefficients between the variables shown in suffices .

$S_x^2=(N-1)^{-1}\sum_{i=1}^{N}(x_i-\bar{X})^2$, $S_y^2=(N-1)^{-1}\sum_{i=1}^{N}(y_i-\bar{Y})^2$, $S_z^2=(N-1)^{-1}\sum_{i=1}^{N}(z_i-\bar{Z})^2$ are the population mean squares of x, y and z respectively,

$f=n/N$: Sampling fraction.

For estimating the population mean $\bar{Y}$ on the second (current) occasion two estimators are given. One is based on a sample of size $u=n\mu$ drawn, afresh on the second occasion and is defined by

$$T_u=\bar{y}_u\left(\frac{\bar{Z}+S_z}{\bar{z}_u+S_z}\right) \tag{2.1}$$

Which is on the line of Singh (2003).

The second estimator is chain-type ratio estimator based on the sample of size m (= $n\lambda$) common to both the occasions and is defined as

$$T_m=\bar{y}_m\left(\frac{\bar{x}_m}{\bar{x}_n}\right)\left(\frac{\bar{Z}+S_z}{\bar{z}_n+S_z}\right). \tag{2.2}$$

Combining the estimator T_u and T_m, we have the final estimator of the population mean $\bar{Y}$ as

$$T=\varphi T_u+(1-\varphi)T_m, \tag{2.3}$$

Where ϕ is an unknown scalar (constant) to be determined such that the *MSE* of the estimator T is least

3 Bias and Mean Square Error (*MSE*) of Suggested Estimator 'T'

Since T_u and T_m both are biased estimators of the population mean $\bar{Y}$, therefore resulting estimator T is also a biased estimator of the population mean $\bar{Y}$.The bias *B* (.) and *MSE* (.) up to the first order of approximation are respectively given by

$$B(T)=\phi B(T_u)+(1-\phi)B(T_m) \tag{3.1}$$

And

$$MSE(T)=[\phi^2 MSE(T_u)+(1-\phi)^2 MSE(T_m)+2\phi(1-\phi)Cov(T_u,T_m)], \tag{3.2}$$

Where

$$B(T_u)=\bar{Y}\left(\frac{1}{u}-\frac{1}{N}\right)\gamma(\gamma C_z^2-\rho_{yz}C_yC_z),$$

$$B(T_m)=\bar{Y}\left[\left(\frac{1}{m}-\frac{1}{N}\right)(C_x^2-\rho_{yx}C_yC_x)+\left(\frac{1}{n}-\frac{1}{N}\right)\gamma(\gamma C_z^2-\rho_{yz}C_yC_z)\right],$$

$$MSE(T_u)=\bar{Y}^2\left(\frac{1}{u}-\frac{1}{N}\right)\delta_3,\quad MSE(T_m)=\bar{Y}^2\left[\frac{1}{m}\delta_1+\frac{1}{n}\delta_2-\frac{1}{N}\delta_3\right],$$

$$Cov(T_u, T_m) = -\left(\frac{\bar{Y}^2}{N}\right)\delta_3, \delta_1 = (C_y^2 + C_x^2 - 2\rho_{yx}C_yC_x),$$

$$\delta_2 = [\gamma(\gamma C_z^2 - 2\rho_{yz}C_yC_z) + (2\rho_{yx}C_yC_x - C_x^2)],$$

$$\delta_3 = [C_y^2 + \gamma(\gamma C_z^2 - 2\rho_{yz}C_yC_z)] = \delta_1 + \delta_2 \text{ And } \gamma = \bar{Z} / (\bar{Z} + S_z).$$

4 Minimum Mean Square Error (*MSE*) of '*T*'

Minimizing the $MSE(T)$ with respect to ϕ we get the optimum value of ϕ as

$$\phi_{opt} = \frac{[MSE(T_m) - Cov(T_u, T_m)]}{[MSE(T_u) + MSE(T_m) - 2Cov(T_u, T_m)]} = \frac{\mu(\delta_3 - \mu\delta_2)}{(\delta_3 - \mu^2\delta_2)}, \quad (4.1)$$

Thus the resulting minimum *MSE* of *T* is given by

$$Min.MSE(T) = \frac{[MSE(T_u)MSE(T_m) - \{Cov(T_u, T_m)\}^2]}{[MSE(T_u) + MSE(T_m) - 2Cov(T_u, T_m)]}$$

$$= \frac{\bar{Y}^2[\delta_4 - \mu\delta_2\delta_3 + \mu^2\delta_2\delta_3 f]}{n(\delta_3 - \mu^2\delta_2)}$$

$$= \frac{\bar{Y}^2\delta_3[(1-f)\delta_3 - \mu\delta_2 + \mu^2\delta_2 f]}{n(\delta_3 - \mu^2\delta_2)}, \quad (4.2)$$

Where

$$\delta_4 = (\delta_1\delta_3 + \delta_2\delta_3 - f\delta_3^2) = (1-f)\delta_3^2.$$

5 Optimum Replacement Policy

To determine the optimum value of μ so that the population mean $\bar{Y}$ may be estimated with maximum efficiency, we minimize $Min.MSE(T)$ in equation (4.2) with respect to μ, which results in a quadratic equation in μ, which is shown as

$$\delta_2\mu^2 - 2\delta_3\mu + \delta_3 = 0. \quad (5.1)$$

Solving equation (5.1) for $\hat{\mu}$, the solutions are given as

$$\hat{\mu} = \frac{\alpha_3 \pm \sqrt{\alpha_1\alpha_3}}{\alpha_2}. \quad (5.2)$$

From (5.2) two values of $\hat{\mu}$ are possible, hence to choose a value of $\hat{\mu}$, it should be recommended that $0 \le \hat{\mu} \le 1$, all other values of $\hat{\mu}$ are inadmissible. Substituting the value of $\hat{\mu}$ from equation (5.2) into equation (5.1), we have

$$Min.MSE(T)_{opt} = \frac{\bar{Y}^2\delta_3[(1-f)\delta_3 - \mu_0\delta_2 + \mu_0^2\delta_2 f]}{n(\delta_3 - \mu_0^2\delta_2)}, \quad (5.3)$$

Where μ_0 is the admissible value of μ obtained from (5.2).

For the sake of convenience, we suppose that since *x* and *y* denote the same study variable over two occasions and *z* is the stable auxiliary variable correlated to *x* and *y*, therefore, looking on the stability nature [Reddy (1978)] of the coefficient of variation and following Cochran (1977) and Feng and Zou (1997) the coefficients of variation of *x*, *y*, *z* are considered to be equal (i.e. $C_y \cong C_x \cong C_z$).

Under this supposition $\hat{\mu}$ and $Min.MSE(T)$ in (5.2) and (5.3) respectively, reduce to:

$$\hat{\mu}^* = \frac{\delta_3^* \pm \sqrt{\delta_1^* \delta_3^*}}{\delta_2^*} \tag{5.4}$$

And

$$Min.MSE(T)^*_{opt} = \frac{S_y^2 \delta_3^* [(1-f)\delta_3^* - \mu_0^* \delta_2^* + \mu_0^{*2} \delta_2^* f]}{n(\delta_3^* - \mu_0^{*2} \delta_2^*)}, \tag{5.5}$$

Where

$\delta_1^* = 2(1-\rho_{yx})$, $\delta_2^* = [\gamma(\gamma - 2\rho_{yz}) + (2\rho_{yx} - 1)]$, $\delta_3^* = [1 + \gamma(\gamma - 2\rho_{yz})]$ and μ_0^* is the admissible value $\hat{\mu}$ of obtained from (5.4).

6 Efficiency Comparison and Conclusion

The percent relative efficiencies of the proposed estimator 'T' with respect to (i) usual unbiased estimator $\bar{y}_n$, when there is no matching and (ii) the estimator $\hat{\bar{Y}} = \psi \, \bar{y}_u + (1-\psi)\bar{y}'_m$, when no auxiliary information is used at any occasion, where $\bar{y}'_m = \bar{y}_m + \beta_{yx}(\bar{x}_n - \bar{x}_m)$ and β_{yx} is the known population regression coefficient; have been obtained for various choices ρ_{yx}, ρ_{yz} and γ. Following Sukhatme et al. (1984), the variance of usual unbiased estimator $\bar{y}_n$ and optimum variance of $\hat{\bar{Y}}$ are respectively given by

$$Var(\bar{y}_n) = (1-f)\frac{S_y^2}{n}, \tag{6.1}$$

And

$$Var(\hat{\bar{Y}}) = [(1/2)\left(1 + \sqrt{(1-\rho_{yx}^2)}\right) - f]\frac{S_y^2}{n}, \tag{6.2}$$

For N=1000, n=100 and the various choices ρ_{yx}, ρ_{yz}, $\bar{Z}$ and S_z, Table 1 shows the optimum values of $\hat{\mu}$ and the percent relative efficiencies E_1 and E_2 of the suggested estimator T with respect to $\bar{y}_n$ and $\hat{\bar{Y}}$ respectively by using the following formulae:

$$E_1 = \frac{V(\bar{y}_n)}{Min.MSE(T)^*_{opt}} \times 100$$

$$= \frac{(1-f)(\delta_3^* - \mu_0^{*2}\delta_2^*)}{\delta_3^*[(1-f)\delta_3^* - \mu_0^*\delta_2^* + \mu_0^{*2}\delta_2^* f]} \times 100, \tag{6.3}$$

And

$$E_2 = \frac{V(\hat{\bar{Y}})_{opt}}{Min.MSE(T)^*_{opt}} \times 100$$

$$= \frac{[\left(1 + \sqrt{(1-\rho_{yx}^2)}\right) - 2f](\delta_3^* - \mu_0^{*2}\delta_2^*)}{2\delta_3^*[(1-f)\delta_3^* - \mu_0^*\delta_2^* + \mu_0^{*2}\delta_2^* f]} \times 100, \tag{6.4}$$

Table1: Optimum values μ_0^* and percent relative efficiencies of T with respect to $\bar{y}_n$ and $\hat{\bar{Y}}$ for $\bar{Z}=125$.

S_z		45			65			85		
ρ_{yz}	ρ_{yx}	μ_0^*	E_1	E_2	μ_0^*	E_1	E_2	μ_0^*	E_1	E_2
0.2	0.5	0.42	101.15	100.02	0.41	103.84	102.67	0.41	105.31	104.12
	0.7	0.36	137.05	135.51	0.36	136.95	135.42	0.36	135.36	133.84
	0.9	0.27	235.88	233.23	0.28	216.97	214.54	0.30	200.21	197.96
0.4	0.5	0.45	110.61	105.48	0.45	113.62	108.35	0.44	115.26	109.92
	0.7	0.39	150.97	143.97	0.40	150.87	143.87	0.40	149.07	142.16
	0.9	0.30	263.10	250.90	0.31	241.58	230.38	0.33	222.52	212.20
0.6	0.5	0.50	124.40	110.58	0.50	127.90	113.69	0.49	129.82	115.40
	0.7	0.44	171.67	152.60	0.44	171.55	152.49	0.45	169.43	150.61
	0.9	0.34	304.84	270.97	0.36	279.13	248.11	0.37	256.40	227.91
0.8	0.5	0.59	148.53	115.52	0.58	152.97	118.98	0.58	155.40	120.87
	0.7	0.53	209.01	162.56	0.53	208.85	162.44	0.53	206.12	160.32
	0.9	**0.42**	**384.28**	**298.88**	**0.44**	**350.03**	**272.25**	**0.46**	**319.90**	**248.81**

Table 1 exhibits that the values of E_1 and E_2 are larger than 100. It follows that the proposed estimator T is more efficient than estimators $\bar{y}_n$ and $\hat{\bar{Y}}$ under optimal conditions. Thus in practice the suggested estimator T is to be preferred over estimators $\bar{y}_n$ and $\hat{\bar{Y}}$.

Acknowledgements

The authors are thankful to the editor and two anonymous referees for their valuable suggestions on the paper.

References

[1] R. S. Biradar, and H.P. Singh, Successive sampling using auxiliary information on both the occasions. *Calcutta Statist. Assoc. Bull.*, **51**, 243-251, (2001).

[2] W.G. Cochran, Sampling Techniques .Third Edition. Johan Wiley, (1977).

[3] A. R. Eckler, Rotation sampling. *Ann. Math. Statist.* **26**, 664-685, (1955).

[4] S. Feng and G. Zou, Sampling rotation method with auxiliary variable. *Commun. Statist. Theo. Meth.* **26(6)**, 1497-1509, (1997).

[5] R. J. Jessen, Statistical investigation of a sample survey for obtaining form facts. Iowa Agricultural Experiment Station Road Bulletin no. 304, Ames, USA, (1942).

[6] H. D. Patterson, Sampling on successive occasions with partial replacement of units, *Jour. Roy. Statist. Assoc*, B, **12**, 241- 255, (1950).

[7] J.N.K. Rao and J. E. Graham, Rotation design for sampling on repeated occasions. *Jour. Amer. Statist. Assoc.* **59**, 492-509, (1964).

[8] V. N. Reddy, A study on the use of prior knowledge on certain population parameters in estimation. Sankhya C, **40**, 29-37, (1978).

[9] G. N. Singh, on the improvement of product method of estimation in sample surveys. *Jour. Ind. Soc. Agril. Statist.* **56 (3)**, 267-275, (2003).

[10] Singh, G.N. (2005): On the use of chain type ratio estimator in successive Sampling. Statistics in Transition - new series,7(1), 21-26.

[11] H.P. Singh, M. K. Srivastava, N. Srivastava, T. A. Tarray, V. Singh, and S. Dixit, Chain regression-type estimator using Multiple auxiliary information in successive sampling. *Hac. Jour. Math. Statist*, **45(6)**, 1-11, (2014).

[12] H. P. Singh and G. K. Vishwakarma, Modified exponential ratio product estimators for finite population mean in double sampling. *Austrian Jour. Statist.* **36 (3)**, 217 – 225, (2007a).

[13] H. P. Singh and G. K. Vishwakarma, A general procedure for estimating population mean in successive sampling. *Commun. Statist. Theo. Meth.*, **38**, 293-308, (2009).

[14] H. P. Singh and G. K. Vishwakarma, A general class of estimators in successive sampling. *Metron*, **65(2)**, 201-227, (2007b).

[15] Singh, H.P. Singh, H.P. and Singh, V.P. A generalized efficient class of estimators of population mean in two-phase and successive Sampling. *Int. Jour. Manag. Systems*, **8(2)**, 173-183, (1992).

[16] P. V. Sukhatme, B. V. Sukhatme, and C. Ashok. Sampling theory of surveys with applications, 3rd ed. Ames, IA, Iowa State University Press, (1984).