ADV MATH SCI JOURNAL
Advances in Mathematics: Scientific Journal **9** (2020), no.9, 6457–6465
ISSN: 1857-8365 (printed); 1857-8438 (electronic)
https://doi.org/10.37418/amsj.9.9.3

SOME MODIFIED ESTIMATORS OF POPULATION MEAN ON TWO OCCASIONS

VYAS DUBEY AND ARUN KUMAR SHUKLA[1]

ABSTRACT. This paper deals with some modified estimators of population mean on current occasion based on samples selected over current and previous occasions. Properties of proposed estimators have been discussed and it is seen that they are more efficient than usual estimators for optimal choice of constants included in the estimators. Numerical illustrations have been cited in support of the theoretical results.

1. INTRODUCTION

Auxiliary information is highly used for estimating population parameters like mean, total, ratio, variance etc. Estimators so defined are known as ratio, product and regression estimators which are used under different situations. The ratio (product) estimator is used for positively (negatively) correlated variables while regression estimator is useful if variables are linearly related. Amongst all such estimators, regression estimator is theoretically proved to be more efficient. Many a time, value taken by such population parameters changes over different occasions while the units of population may be same. For example, yield of a crop in a geographical area may change from year to year while the fields under the cultivation may be same. In such a situation, the information collected on

[1]*corresponding author*

2010 *Mathematics Subject Classification.* 62H12.
Key words and phrases. Auxiliary Information, Bias, Mean Squared Error (MSE), Efficiency, Order of Approximation.

the units common with earlier occasions may be used for estimating the population parameters on current occasion. Jessen [7] was the first to introduce procedure of utilizing the information obtained on first occasion in improving the estimates at the current (second) occasion. Patterson [8], Tikkiwal [17], Eckler [6], Rao and Graham [9], Sen [11–13] and others further extended the theory of successive sampling.

Avadhani [4], Artes and Garcia [10] used ratio estimator in successive sampling while Artes et al., [3], Artes and Garcia [2] used product estimator in such a situation. Sukhatme et al., [16] considered regression type estimator for estimating population mean based on the matched sample. Further, Srivastava and Srivastava [15] proposed chain type estimators. Again, Beevi [5] suggested dual to ratio estimator for mean in successive occasions.

In this present study we propose some estimators which are linear combinations of available sample means based on successive sampling on two occasions, where a fraction of previously collected sample is retained and a new sample is drawn with SRSWOR strategy from the population at the current occasion. Using combination of means of matched and unmatched portions of sample at current occasion, the optimum estimators are proposed.

2. Development of Estimators

Let y be study variable which has population mean $\overline{Y_i}$ and mean squared error ${S_i}^2$ on i^{th} occasion, i=1, 2.
Assume that n_1 units constitute the sample on the first occasion of which $n_2^{'}$ are retained on the second occasion while $n_2^{"} = n_2 - n_2^{'}$ are drawn afresh on the second occasion from $N - n_1$ units.
Again, let
S_{21} - covariance among units on first and second occasions in the population,
ρ - correlation among units on first and second occasions in the population,
β_{21} - regression coefficient of second occasion values on first occasion values,
$\overline{y_1}$ - the sample mean based on n_1 units observed on the first occasion,
$\overline{y}_2^{'}$- the sample mean based on $n_2^{'}$ units observed on the second occasion, which are common between the first and second occasions,
$\overline{y}_2^{"}$ - the sample mean based on newly drawn $n_2^{"}$ units on the second occasion, which are not common with the first occasion,

$\overline{x}_1^{'}$- the sample mean based on $n_2^{'}$ units common to both the occasions and observed on the first occasion.
Thus, $\overline{y}_1$ is an unbiased estimator of $\overline{Y}_1$ with variance

$$V(\overline{y_1}) = (\frac{1}{n_1} - \frac{1}{N})S_1^2$$

while $\overline{y}_2^{'}$ and $\overline{y}_2^{"}$ are two unbiased estimators of $\overline{Y_2}$ with variance

$$V(\overline{y}_2^{'}) = (\frac{1}{n_2^{'}} - \frac{1}{N})S_2^2$$

and

$$V(\overline{y}_2^{"}) = (\frac{1}{n_2^{"}} - \frac{1}{N})S_2^2$$

respectively. Sukhatme et al, [16] suggested following estimator of $\overline{Y_2}$ based on the matched sample $n_2^{'}$ as

$$\overline{y}_{l2}^{'} = \overline{y}_2^{'} + \beta_{21}(\overline{y}_1 - \overline{x}_1^{'}) ,$$

which is unbiased and has variance

$$V(\overline{y}_{l2}^{'}) = (\frac{1}{n_2^{'}} - \frac{1}{n_1})S_2^2(1-\rho^2) + (\frac{1}{n_1} - \frac{1}{N})S_2^2 .$$

Combining $\overline{y_2}^{"}$, the sample mean based on unmatched portion, Sukhatme et al, [16], further suggested estimator of $\overline{Y_2}$ as:

$$\widehat{\overline{Y_2}} = \varphi\overline{y}_{l2}^{'} + (1-\varphi)\overline{y}_2^{"} ,$$

where φ is a constant. For optimum choice of φ, the estimator $\widehat{\overline{Y_2}}$ has minimum variance

$$V(\widehat{\overline{Y_2}}) = \frac{[1-(1-\frac{n_2^{'}}{n_1})\rho^2]}{[1-\frac{n_2^{"}}{n_2}(1-\frac{n_2^{'}}{n_1})\rho^2]}\frac{S_2^2}{n_2} - \frac{S_2^2}{N} .$$

3. Proposed Estimators

We consider the following estimators of $\overline{Y_2}$ due to matched portion of sample as:

$$\begin{aligned} T_1 &= K\overline{y}_{l2}^{'} \\ T_2 &= K_1\overline{y}_2^{'} + K_2(\overline{x}_1^{'} - \overline{y}_1); K_1 + K_2 \neq 1 \\ T_3 &= W_1\overline{y}_2^{'} + W_2\overline{x}_1^{'} + (1-W_1-W_2)\overline{y}_1 , \end{aligned}$$

where K , K_1 , K_2 , W_1 and W_2 are suitable constants. The estimators T_1 and T_2 are generalizations of $\overline{y}'_{l2}$ while T_3 is a new type of estimator.
It can be seen that the above estimators are biased. Their biases are given by

$$\begin{aligned} B(T_1) &= (K-1)\overline{Y}_2\,, \\ B(T_2) &= (K_1-1)\overline{Y}_2\,, \\ B(T_3) &= (1-W_1)(\overline{Y}_1-\overline{Y}_2)\,. \end{aligned}$$

The mean squared errors of above estimators are

$$\begin{aligned} M(T_1) &= K^2\{\theta_1 S_2^2 + (\theta_2' - \theta_1)(1-\rho^2)S_2^2\} + (K-1)^2\overline{Y}_2^2 \\ M(T_2) &= K_1^2\theta_2' S_2^2 + K_2^2\,(\theta_2'-\theta_1)\,S_1^2 + 2K_1K_2(\theta_2'-\theta_1)S_{21} + (K_1-1)^2\overline{Y}_2^2 \\ M(T_3) &= W_1^2\{\theta_2' S_2^2 + \theta_1 S_1^2 - 2\theta_1 S_{21} + (\overline{Y}_2-\overline{Y}_1)^2\} + W_2^2(\theta_2'-\theta_1)S_1^2 \\ &\quad +2W_1W_2(\theta_2'-\theta_1)S_{21} - 2W_1\{\theta_1 S_1^2 - \theta_1 S_{21} + (\overline{Y}_2-\overline{Y}_1)^2\} \\ &\quad +\theta_1 S_1^2 + (\overline{Y}_2-\overline{Y}_1)^2\,, \end{aligned}$$

where $\theta_1 = \frac{1}{n_1} - \frac{1}{N}$, $\theta_2' = \frac{1}{n_2'} - \frac{1}{N}$.

Let $C_1 = S_1/\overline{Y}_1$, $C_2 = S_2/\overline{Y}_2$ be coefficients of variation of y on first and second occasions respectively,

$$\begin{aligned} C_{21} &= \frac{S_{21}}{(\overline{Y}_1\overline{Y}_2)} \\ R &= \frac{\overline{Y}_1}{\overline{Y}_2} \\ C_{\overline{y}_{l2}}^2 &= \frac{V(\overline{y}'_{l2})}{\overline{Y}_2^2} = \{(\theta_2'-\theta_1)(1-\rho^2)+\theta_1\}C_2^2\,. \end{aligned}$$

Then, the optimum values of constants K , K_1 , K_2 , W_1 and W_2 , are correspondingly given by

$$\begin{aligned} K_0 &= K_{10} = \frac{1}{1+C_{\overline{y}_{l2}}^2} \\ K_{20} &= -\beta_{21}K_{10} \\ W_{10} &= \frac{\theta_1(R^2C_1^2 - RC_{21}) + (1-R)^2}{C_{\overline{y}_{l2}}^2 + \theta_1(R^2C_1^2 - RC_{21}) + (1-R)^2} \\ W_{20} &= -\beta_{21}W_{10}\,. \end{aligned}$$

Thus, the minimum mean squared errors of corresponding estimators are obtained as

$$M_0(T_1) = M_0(T_2) = \frac{V(\overline{y}'_{l2})}{1 + C^2_{\overline{y}_{l2}}} \,,$$

$$M_0(T_3) = \frac{V(\overline{y}'_{l2})\{\theta_1 S_1^2 + (\overline{Y}_2 - \overline{Y}_1)^2\} - \theta_1^2 S_{21}^2}{V(\overline{y}'_{12}) + \theta_1(S_1^2 - 2S_{21}) + (\overline{Y}_2 - \overline{Y}_1)^2} \qquad (3.1)$$

Simplifying (3.1), $M_0(T_3)$ can be expressed as

$$M_0(T_3) = \frac{V(\overline{y}'_{l2})(1-R)^2 + \theta_1\theta'_2 C_2^2(1-\rho^2)S_1^2}{1 + C^2_{\overline{y}_{l2}} + \theta_1(R^2C_1^2 - 2RC_{21})} \,.$$

It is clear that $M_0(T_3) < V(\overline{y}'_{l2})$. If $0 < R < 2$, i.e. the population has increasing trend, the first term in numerator of $M_0(T_3)$ will be smaller than $V(\overline{y}'_{l2})$. The second term in numerator is of $O(n^{-2})$ which may be neglected. A sufficient condition that the denominator in $M_0(T_3)$ will be positive is that $\rho C_2/ C_1 > 1/2$ which seems to always happen in practice in the situation of successive occasions.
Again, T_3 is more efficient than $\overline{y}'_{l2}$ if $M_0(T_3) < V(\overline{y}'_{l2})$, or if

$$\frac{V(\overline{y}'_{l2})\{\theta_1 S_1^2 + (\overline{Y}_2 - \overline{Y}_1)^2\} - \theta_1^2 S_{21}^2}{V(\overline{y}'_{l2}) + \theta_1(S_1^2 - 2S_{21}) + (\overline{Y}_2 - \overline{Y}_1)^2} < V(\overline{y}'_{l2})$$

or if

$$\{\sqrt{V(\overline{y}'_{l2})} - \frac{\theta_1 S_{21}}{\sqrt{V(\overline{y}'_{l2})}}\}^2 > 0 \,,$$

which is always true.
If n'_2 = n/2 and N is large, then $M_0(T_1)$ and $M_0(T_3)$ simplify to:

$$M_0(T_1) = \frac{V(\overline{y}'_{l2})}{1 + \frac{(2-\rho^2)C_2^2}{n}}$$

$$\begin{aligned} M_0(T_3) &= \frac{V(\overline{y}'_{l2})[(1-R)^2 + \frac{2}{n}\frac{(1-\rho^2)}{(2-\rho^2)}R^2C_1^2]}{(1-R)^2 + \frac{1}{n}\{2(1-\rho^2)C_2^2 + (RC_1 - \rho C_2)^2\}} \\ &= \frac{V(\overline{y}'_{l2})[1 + \frac{2}{n}\frac{(1-\rho^2)}{(2-\rho^2)}\eta C_1^2]}{1 + \frac{1}{n(1-R)^2}\{2(1-\rho^2)C_2^2 + (RC_1 - \rho C_2)^2\}} \\ &= \frac{V(\overline{y}'_{l2})[1 + \frac{A_1}{n}]}{1 + \frac{A_2}{n}} \,, \end{aligned}$$

where

$$\eta = \frac{R^2}{(1-R)^2}, A_1 = \frac{2(1-\rho^2)\eta C_1^2}{(2-\rho^2)}, A_2 = \frac{1}{(1-R)^2}\{2(1-\rho^2)C_2^2+(RC_1-\rho C_2)^2\}.$$

Now, combining the above estimators with $\overline{y}_2^{"}$, we have following estimators of $\overline{Y}_2$:

$$\begin{aligned} T_1^* &= \lambda T_{10} + (1-\lambda)\overline{y}_2^{"}, \\ T_2^* &= \phi T_{30} + (1-\phi)\overline{y}_2^{"}. \end{aligned}$$

4. Bias and MSEs of proposed estimators

It can easily be seen that the biases of the proposed estimators are:

$$\begin{aligned} B(T_1^*) &= (1-\lambda)(K_0-1)\overline{Y}_2, \\ B(T_2^*) &= (1-\phi)(1-W_{10})(\overline{Y}_2-\overline{Y}_1). \end{aligned}$$

The MSEs of the estimators are given by

$$\begin{aligned} M(T_1^*) &= \lambda^2 M(T_1) + 2\lambda(1-\lambda)Cov(\overline{y}_2^{"}, T_1) + (1-\lambda)^2 V(\overline{y}_2^{"}), \\ M(T_2^*) &= \phi^2 M(T_3) + 2\phi(1-\phi)Cov(\overline{y}_2^{"}, T_3) + (1-\phi)^2 V(\overline{y}_2^{"}). \end{aligned}$$

The optimum values of λ, ϕ for which the above MSEs will be minimum are

$$\lambda_0 = \frac{V(\overline{y}_2^{"}) - Cov(\overline{y}_2^{"}, T_1)}{V(\overline{y}_2^{"}) + M_0(T_1) - 2Cov(\overline{y}_2^{"}, T_1)},$$

and

$$\phi_0 = \frac{V(\overline{y}_2^{"}) - Cov(\overline{y}_2^{"}, T_3)}{V(\overline{y}_2^{"}) + M_0(T_3) - 2Cov(\overline{y}_2^{"}, T_3)}.$$

The minimum values of the MSEs of the above estimators are given by

$$\begin{aligned} M_0(T_1^*) &= \frac{V(\overline{y}_2^{"})M_0(T_{10}) - \{Cov(\overline{y}_2^{"}, T_{10})\}^2}{V(\overline{y}_2^{"}) + M_0(T_{10}) - 2Cov(\overline{y}_2^{"}, T_{10})}, \\ M_0(T_2^*) &= \frac{V(\overline{y}_2^{"})M_0(T_{30}) - \{Cov(\overline{y}_2^{"}, T_{30})\}^2}{V(\overline{y}_2^{"}) + M_0(T_{30}) - 2Cov(\overline{y}_2^{"}, T_{30})}. \end{aligned}$$

Following Sukhatme et al. [16], we have:

$$\begin{aligned} Cov(\overline{y}_2^{"}, T_{10}) &= -K_0\frac{S_2^2}{N}, \\ Cov(\overline{y}_2^{"}, T_{30}) &= -W_{10}\frac{S_2^2}{N}. \end{aligned}$$

Therefore, the estimators T_1^*, T_2^* have minimum mean squared errors as follows

$$M_0(T_1^*) = \frac{V(\overline{y}_2^{"})M_0(T_{10}) - \{K_0(\frac{S_2^2}{N})\}^2}{V(\overline{y}_2^{"}) + M_0(T_{10}) + 2K_0(\frac{S_2^2}{N})},$$

$$M_0(T_2^*) = \frac{V(\overline{y}_2^{"})M_0(T_{30}) - \{W_{10}(\frac{S_2^2}{N})\}^2}{V(\overline{y}_2^{"}) + M_0(T_{30}) + 2W_{10}(\frac{S_2^2}{N})}.$$

5. Efficiency Comparisons

Variance of $\widehat{\overline{Y}_2}$ may be written as Sukhatme et al., see [16]:

$$V(\widehat{\overline{Y}_2}) = \frac{V(\overline{y}_2^{"})V(\overline{y}_{l2}^{'}) - (\frac{S_2^2}{N})^2}{V(\overline{y}_2^{"}) + V(\overline{y}_{l2}^{'}) + 2(\frac{S_2^2}{N})}.$$

It is difficult to compare the minimum MSEs of the proposed estimators with $V(\widehat{\overline{Y}_2})$. If N is large, so that the terms S_2^2/N and $(S_2^2/N)^2$ can be ignored, their MSEs reduce to:

$$M_0(T_1^*) \approx \frac{V(\overline{y}_2^{"})M_0(T_{10})}{V(\overline{y}_2^{"}) + M_0(T_{10})},$$

$$M_0(T_2^*) \approx \frac{V(\overline{y}_2^{"})M_0(T_{30})}{V(\overline{y}_2^{"}) + M_0(T_{30})}.$$

The optimum values of constants involved in the estimators may not be known in advance. Therefore, we replace them by their consistent estimators. It can easily be seen that such a replacement will not affect MSEs of the estimators up to first order of approximation [10].

6. Numerical Illustration

Consider the data from Singh and Chaudhary [14] which relates to data on arecanut trees in 30 randomly selected villages from a population of 1878 villages. Let y_1 and y_2 denote no. of arecanut trees in the years 1980-81 and 1981-82 respectively. For this data we have:

N=1878

n_1=30 ; $\overline{y}_1$=1394.30;

n_2=30 ; $\overline{y}_2$=1414.57;

$n_2^{'}$=10 ; $\overline{y}_2^{'}$=801.70;
$n_2^{"}$=20 ; $\overline{y}_2^{"}$=1721.00;
s_1^2=642884.0 ; s_2^2=692589.5 ; s_{21}=625528.5;
r=0.9374
b=0.9730
K_0=0.98604591
W_{10}=0.11163827; W_{20}=-0.10862445.

Relative Efficiency of the estimators defined by $REF = \frac{V(\overline{y}_2)}{M(.)}$ % is given in Table 1. Table 1 shows that proposed estimators T_1^* and T_2^* are more efficient than existing estimator $\widehat{\overline{Y}}_2$.

Estimator	MSE	REF
$\overline{y}_2$	22717.5257	100%
$\widehat{\overline{Y}}_2$	15320.6546	148.28%
T_1^*	15292.6544	148.55%
T_2^*	13148.4813	172.77%

TABLE 1. Relative Efficiency of Estimators

REFERENCES

[1] E. ARTES, A. GARCIA: *A note on successive sampling using auxiliary information*, Proceedings of the 15^{th} International Workshop on Statistical Modelling, (2000), 376–379.

[2] E. ARTES, A. GARCIA: *Estimating the current mean in successive sampling using a product estimate*, Conference on Agricultural and Environmental Statist. Application in Rome, (2001), XLIII–1–XLIII–2.

[3] E. ARTES, M. RUEDA, A. ARCOS: *Successive sampling using a product estimate*, WIT Transactions on Ecology and the Environment, 30, ISSN 1743-3541, (1998), 85–90.

[4] M. S. AVADHANI *Contribution to the theory of sampling from finite population and its application*, Ph.D.Thesis, Delhi University, 1968.

[5] N. BEEVI: *Dual to ratio estimators for mean estimation in successive sampling using auxiliary information on two occasions*, Journal of Statistics Applications and Probability, **7**(1) (2018), 49–58.

[6] A. R. ECKLER: *Rotation Sampling*, Ann. Math. Stat., **26** (1955), 664–685.

[7] R. J. JESSEN: *Statistical investigation of a sample survey for obtaining farm facts*, Iowa Expt. Stat. Res., Bull., **304**, 1942.

[8] H. D. PATTERSON: *Sampling on successive occasions with partial replacement of units*, Jour. Roy. Stat. Soc.,B **12** (1950), 241–255.

[9] J. N. K. RAO, J. E. GRAHAM *Rotation design for sampling on repeated occasions*, Jour. Amer. Statist. Assoc., **59** (1964), 492–509.

[10] S. SAMPATH: *On the optimal choice of unknowns in ratio type estimator*. J.Ind. Soc. Agri. Stat., **41** (1989), 166–172.

[11] A. R. SEN: *Successive sampling with two auxiliary variables*, Sankhya, Ser. B, **33** (1971), 371–378.

[12] A. R. SEN: *Successive sampling with p (p=1) auxiliary variables*, Ann. Math. Statist., **43** (1972), 2031-2034.

[13] A. R. SEN: *Theory and application of sampling on repeated occasions with several auxiliary variables*, Biometrics, **29** (1973), 381–385.

[14] D. SINGH, F. S. CHAUDHARY: *Theory and analysis of sample survey designs*, New Age International (P) Ltd. Publishers, New Delhi, 1986, 299–302.

[15] M. K. SRIVASTAVA, N. SRIVASTAVA *Chain type ratio estimator in successive sampling using multi -auxiliary information*, Communications in Statistics- Theory and Methods, **45**(16) (2016), 4798–4811.

[16] P. V. SUKHATME, B. V. SUKHATME, S. SUKHATME, C. ASHOK: *Sampling theory of Survey With Application*, Iowa University Press, Ames Iowa, U.S.A. , II Edition, 1984.

[17] B. D. TIKKIWAL: *Theory of Successive Sampling*, Unpublished Diploma Dissertation, Institute of Agricultural Statistics, New Delhi, India, 1951.

SCHOOL OF STUDIES IN STATISTICS
PT. RAVISHANKAR SHUKLA UNIVERSITY, RAIPUR (C.G.) 492010 INDIA

SCHOOL OF STUDIES IN STATISTICS
PT. RAVISHANKAR SHUKLA UNIVERSITY, RAIPUR (C.G.) 492010 INDIA
Email address: arun_shukla@rediffmail.com