ISSN 1684-8403
Journal of Statistics
Volume 17, 2010, pp. 54-65

Estimation of Current Population Variance in Two Successive Occasions

Muhammad Azam[1], Qamruz Zaman[2], Salahuddin[3]
and Javed Shabbir[4]

Abstract

The problem of estimation of the population variance for the current occasion based on the samples selected over two successive occasions has been considered. Expressions for the Optimum estimator and its variance have been derived for both the occasions. Also the expression has been derived when no information has been collected on the first occasion. Expression for the Optimum fraction of unmatched observations has been derived. The values of Optimum matched and unmatched fractions have been tabulated. The gain in efficiency of the proposed estimator is compared to the estimator without using the information collected on the first occasion. An empirical study has also been conducted to study the performance of the proposed strategy.

Keywords

Variance estimator, Sampling on successive occasions, Optimum matching fraction, Gain in efficiency

1. Introduction

The theory and practice of surveying the same population at different points of time, technically called Repetitive sampling, Rotational sampling or sampling over successive occasions have been given considerable attention by some survey statisticians. The main objective of sampling on successive occasions is to

[1] Department of Medical Statistics, Informatics and Health Economics, Medical University of Innsbruck, Austria.
[2] Department of Statistics, University of Peshawar, Pakistan
[3] Department of Statistics, Quaid-e-Azam University, Islamabad, Pakistan
Email:salahuddin_90@yahoo.com
[4] Department of Statistics, Quaid-e-Azam University, Islamabad, Pakistan

estimate some population parameters (like total, mean, ratio, variance etc.) for the most recent occasions as well as changes in these parameters from one occasion to the next (Mukhopadhyay, 1998).

Jessen (1942) was the first author who considered the problem of sampling on two successive occasions by using the information collected on the previous occasion to improve the current estimate. This technique was further extended by Eckler (1955), Patterson (1950), Rueda et al. (2008), Tikkiwal (1951), Yates (1960) and many others.

The problem of estimating the current population variance in successive sampling was considered by Sud et al. (2001). The theory of estimation of the population ratio of two characters over two occasions has been considered by Rao and Mudholkar (1967), Okafor and Arnab (1987), Artes et al. (2001) and many others.

The aim of present work is to propose an estimator, which is best linear combination of available sample variances, based on successive sampling scheme on two occasions, where a fraction of the previously collected sample is retained and a new sample is drawn with SRSWOR strategy from the population at the second occasion. The gain in efficiency of the proposed estimator has been obtained over sample variance estimator without using the information available on the first occasion and the Optimum variance estimator of the current occasion, which is the combination of the sample variances of the matched and unmatched portion of the sample at the second occasion.

2. Sampling Strategy

Let a character under study be denoted by x and y on the first and the second occasions respectively. Suppose a population of size N, which is sampled over two occasions. Assume that the size of the population remains unchanged but values of units change over occasions. Let a sample of size n be selected using SRSWOR scheme at the first occasion. Out of this sample, let m units be retained on the second occasion, while a fresh sample of the same size $u = (n\text{-}m)$ drawn on the second occasion from the remaining $(N\text{-}n)$ units of the population, so that the total sample size at the second occasion is also n i.e. $n = m + u$. Assume that the size of the population is large, so we ignore fpc. Briefly, the sampling strategy may be illustrated as follows (Table 1):

Table 1: Sampling Strategy

Occasion	Sample Units	Total Sample Size
1	$x\ x\ x\ x\ x\ x\ x\ x\ (n)\bar{y}$	n
	$x\ x\ x\ x\ x\ (m)\ \bar{x}_m$	m
2	$x\ x\ x\ x\ x\ (m)\bar{y}_m \quad x\ x\ x\ (u)\bar{y}_u$	$n = m + u$

3. The Proposed Estimator and its Properties

We are interested to estimate the current population variance S_2^2 (for the second occasion) by a linear combination of available sample variances based on:

$$\hat{\sigma}_2^2 = \alpha\, s_{1u}^2 + \beta\, s_{1m}^2 + \gamma\, s_{2m}^2 + \delta\, s_{2u}^2 \tag{3.1}$$

where the constants $\alpha, \beta, \gamma, \delta$ and the matching fraction λ are to be determined, so as to minimize the variance. Also S_1^2 and S_2^2 are the population variance of X on the first occasion and the population variance of the study variable Y on the second occasion respectively. Similarly s_{1m}^2 and s_{1u}^2 are the sample variances of matched and unmatched units on the first occasion while s_{2m}^2 and s_{2u}^2 are the sample variances of matched and unmatched units on the second occasion respectively.

By following Cochran (1977), Sukhatame et al. (1984) and Sukhatame and Sukhatame (1970) and we have:

$$E\left(s_{1u}^2\right) = E\left(s_{1m}^2\right) = S_1^2 \text{ and } E\left(s_{2u}^2\right) = E\left(s_{2m}^2\right) = S_2^2$$

So, we get:

$$E\left(\hat{\sigma}_2^2\right) = S_1^2(\alpha + \beta) + S_2^2(\gamma + \delta) \tag{3.2}$$

Here, $\hat{\sigma}_2^2$ is an unbiased estimator of σ_2^2 subject to $\alpha + \beta = 0$ and $\gamma + \delta = 1$.

Hence,

$$\hat{\sigma}_2^2 = \alpha\left(s_{1u}^2 - s_{1m}^2\right) + \gamma s_{2m}^2 + (1-\gamma)s_{2u}^2 \tag{3.3}$$

Define:

Let

$$e_{1u} = \frac{s_{1u}^2 - S_1^2}{S_1^2},\ e_{1m} = \frac{s_{1m}^2 - S_1^2}{S_1^2},\ e_{2u} = \frac{s_{2u}^2 - S_2^2}{S_2^2},\ e_{2m} = \frac{s_{2m}^2 - S_2^2}{S_2^2}.$$

Therefore,

$$E(e_{iu}) = E(e_{im}) = 0,\ (i = 1,\ 2),$$
$$E(e_{1m}e_{1u}) = E(e_{1m}e_{2u}) = E(e_{1u}e_{2u}) = 0,$$
$$E\left(e_{1u}^2\right) = \frac{\lambda_{04}-1}{u}, E\left(e_{1m}^2\right) = \frac{\lambda_{04}-1}{m}, E\left(e_{2u}^2\right) = \frac{\lambda_{40}-1}{u},$$

where

$$\lambda_{rs} = \frac{\mu_{rs}}{\mu_{20}^{r/2}\mu_{02}^{s/2}} \text{ and } \mu_{rs} = \frac{\sum_{i=1}^{N}\left(X_i - \overline{X}\right)^r\left(Y_i - \overline{Y}\right)^s}{N-1}.$$

Solving (3.1), we get:

$$\hat{\sigma}_2^2 = \alpha\left\{S_1^2(1+e_{1u}) - S_1^2(1+e_{1m})\right\} + \gamma S_2^2(1+e_{2m}) + (1-\gamma)S_2^2(1+e_{2u})$$

or

$$\hat{\sigma}_2^2 - S_2^2 = \alpha S_1^2(e_{1u} - e_{1m}) + \gamma S_2^2 e_{2m} + (1-\gamma)S_2^2 e_{2u} \tag{3.4}$$

So (3.4) becomes:

$$V\left(\hat{\sigma}_2^2\right) = \frac{\alpha^2 A}{n\mu\lambda} + \frac{B\gamma^2}{n\lambda} + \frac{B(1-\gamma)^2}{n\mu} - \frac{2\alpha\gamma C}{n\lambda} \tag{3.5}$$

where

$$A = S_1^2\left(\lambda_{04} - 1\right),\ B = S_2^2\left(\lambda_{40} - 1\right) \text{ and } C = S_1^2 S_2^2\left(\lambda_{22} - 1\right).$$

Also, $\lambda = m/n$ and $\mu = u/n$.

We obtain the Optimum/best values of α and γ by minimizing $V\left(\hat{\sigma}_2^2\right)$. The necessary condition for minimum is that the first derivatives of the function be equal to zero.

$$\frac{\partial V\left(\hat{\sigma}_2^2\right)}{\partial \alpha} = 0 \quad \text{and} \quad \frac{\partial V\left(\hat{\sigma}_2^2\right)}{\partial \gamma} = 0$$

$$A\alpha - C\mu\gamma = 0 \tag{3.6}$$

$$B\gamma - B\lambda - C\mu\alpha = 0 \tag{3.7}$$

Solving (3.6) and (3.7), we obtain the Optimum values of α and γ, which are obtained through necessary conditions stated above. These equations always satisfy such relationships.

$$\alpha = \frac{BC\mu\lambda}{AB - C^2\mu^2} \quad \text{and} \quad \gamma = \frac{AB - \mu AB}{AB - C^2\mu^2}$$

Substituting the Optimum values of α and γ in (3.4) and (3.5), we get:

$$\hat{\sigma}_2^2 = \left[\frac{BC\mu\lambda}{AB - C^2\mu^2}(s_{1u}^2 - s_{1m}^2) + \frac{AB - AB\mu}{AB - C^2\mu^2} s_{2m}^2 + (1 - \frac{AB - AB\mu}{AB - C^2\mu^2}) s_{2u}^2 \right] \tag{3.8}$$

and

$$V\left(\hat{\sigma}_2^2\right) = \left[\frac{A}{n}\frac{(BC)^2\mu\lambda}{\left(AB - C^2\mu^2\right)} + \frac{\lambda}{n}\frac{A^2B^3}{\left(AB - C^2\mu^2\right)^2} + \frac{B}{n}\frac{\mu\left(AB - C^2\mu\right)^2}{\left(AB - C^2\mu^2\right)^2} - \frac{2\mu\lambda}{n}\frac{AB^2C^2}{\left(AB - C^2\mu^2\right)^2} \right]$$

or

$$V\left(\hat{\sigma}_2^2\right) = \frac{B(AB - \mu AB)}{n(AB - C^2\mu^2)} \tag{3.9}$$

If $\mu = 0$, $\lambda = 1$ i.e. complete matching or $\mu = 1$, $\lambda = 0$ *i.e.* no matching. We obtained the same variance in both cases i.e.

$$V\left(\hat{\sigma}^2\right) = \frac{B}{n} = \frac{S_2^4(\lambda_{40} - 1)}{n} \tag{3.10}$$

It means, we get the same precision either by keeping the same sample or by changing it on every occasion. Also note that an estimate that the first occasion is given by (3.8) just by interchanging A and B and the occasions.

$$\hat{\sigma}_1^2 = \left[\frac{AC\mu\lambda}{AB - C^2\mu^2}(s_{2u}^2 - s_{2m}^2) + \frac{AB - AB\mu}{AB - C^2\mu^2} s_{1m}^2 + (1 - \frac{AB - AB\mu}{AB - C^2\mu^2}) s_{1u}^2\right] \tag{3.11}$$

This can be achieved if we have data on both occasions. To get optimum $V\left(\hat{\sigma}_2^2\right)$, we differentiate (3.11) w. r. t. μ , then equating zero to get optimum value of μ i.e.

$$\mu_{opt} = \frac{AB - \sqrt{A^2B^2 - C^2AB}}{C^2} \tag{3.12}$$

or

$$\mu_{opt} = \frac{1}{1 + \sqrt{1 - \frac{C^2}{AB}}} \tag{3.13}$$

Substituting Optimum value of μ in (3.9), we get

$$V\left(\hat{\sigma}_2^2\right)_{opt} = \frac{B}{2n}\frac{AB + \sqrt{A^2B^2 - C^2AB}}{AB} \tag{3.14}$$

Substituting the values of A, B and C in (3.14), we get

$$V\left(\hat{\sigma}_2^2\right)_{opt} = \frac{S_2^4(\lambda_{40} - 1)}{2n}\left[1 + \sqrt{1 - \frac{(\lambda_{22} - 1)^2}{(\lambda_{04} - 1)(\lambda_{40} - 1)}}\right] \tag{3.15}$$

However, if only the estimate using information collected on the second occasion, then we have estimator of population variance as:

$$\hat{\sigma}^2 = \lambda s_{2m}^2 + \mu s_{2u}^2 \tag{3.16}$$

Such that $\mu + \lambda = 1$.

Also, we find that:

$V\left(\hat{\sigma}_2^2\right)_{opt} \le V\left(\hat{\sigma}^2\right)$ if

$$\frac{B}{2n}\frac{AB + \sqrt{A^2B^2 - C^2AB}}{AB} \le \frac{B}{n} \quad \text{or } C^2AB \ge 0$$
(3.17)

Because, A, B and C^2 are all positive quantities.

Now, we compute the gain in the precision G of the proposed estimator over an estimate using no information collected on the first occasion i.e.

$$G = \frac{V\left(\sigma^2\right)}{V\left(\hat{\sigma}_2^2\right)} = \frac{AB - \mu^2C^2}{AB - \mu C^2} \tag{3.18}$$

and

$$G_{opt} = \frac{V\left(\hat{\sigma}^2\right)}{V\left(\hat{\sigma}_2^2\right)_{opt}} = \frac{2AB}{AB - \sqrt{A^2B^2 - C^2AB}} \tag{3.19}$$

or $\qquad G_{opt} = \dfrac{2}{1 + \sqrt{1 - \dfrac{C^2}{AB}}}$.

Let

$$K = \frac{C^2}{AB} = \frac{(\mu_{22} - \mu_{20}\mu_{02})^2}{(\mu_{04} - \mu_{02}^2)(\mu_{40} - \mu_{20}^2)}. \tag{3.20}$$

We observe that, in most of the cases, $0 \le K \le 1$. If the value of K is not real, then the Optimum value of μ as well as variance will be inadmissible. We obtained values of μ_{opt}, λ_{opt} and G_{opt} for different values of K in the Table 2 below.

Table 2: Computation of Fraction of Unmatched Units and Gain

K	μ_{opt}	λ_{opt}	$G(\%)$
0.00	0.5000	0.5000	100.000
0.10	0.5132	0.4868	102.633
0.20	0.5279	0.4721	105.573
0.30	0.5445	0.4555	108.893
0.40	0.5635	0.4365	112.702
0.50	0.5858	0.4142	117.157
0.60	0.6126	0.3874	122.515
0.70	0.6461	0.3539	129.222
0.80	0.6910	0.3090	138.197
0.90	0.7597	0.2403	151.949
1.00	1.0000	0.0000	200.000

In Table 2, $100 \leq G(\%) \leq 200$ for all values of *K*.

4. Empirical Study

In the following section the gain in the proposed estimator (3.1) is calculated relative to usual variance estimator in the following examples.

Example 1: The data has been taken from Singh and Singh (1996) containing 528 university professors. A random sample of 25 units has been taken and values have been calculated.
The performance of suggested estimator has been expressed in Tables 3 and 4.

Example 2: The data on electricity, total production of 206 countries for two years (1998 and 1999) as under has been taken from Industrial Commodity Statistics Yearbook 2000 (United Nations, 2003). A random sample of 25 units has been taken and values have been calculated.
The performance of suggested estimator has been expressed in Tables 5 and 6.

Example 3: The data about the District-Wise Production of Fruits of 107 districts for the year 2001-02 and 2002-03 has been taken from Ministry of Food,

Agriculture and Livestock (Economic Wing), Islamabad. A random sample of 30 units has been taken and values have been calculated.
The performance of suggested estimator has been expressed in Tables 7 and 8.

5. Concluding Remarks

In the present study, we proposed an estimator $\hat{\sigma}_2^2$ discussed in Section 2 for the estimation of one of important statistical measures i.e. variance using two successive occasions. The proposed estimator is unbiased under certain conditions. It is proved theoretically that the proposed estimator $\hat{\sigma}_2^2$ is more efficient than the usual unbiased variance estimator $\hat{\sigma}^2$ under certain condition provided in equation 3.17 (which is always true). An empirical study is also conducted to validate the performance of the proposed estimator and it is observed in three data sets that the proposed estimator performs better than usual variance estimator for estimating the population variance.

Table 3: Descriptive Statistics Based on Example 1

Values			
n	25	s_{1u}^2	5.7790
m	12	s_{2u}^2	3.6131
$\hat{\mu}_{02}$	3.5106	$\hat{\lambda}_{04}$	6.3550
$\hat{\mu}_{20}$	3.23272	$\hat{\lambda}_{40}$	2.9428
$\hat{\mu}_{22}$	38.7856	$\hat{\lambda}_{22}$	3.4176
$\hat{\mu}_{04}$	78.3207	$\hat{\alpha}$	-0.14075
$\hat{\mu}_{40}$	30.7539	$\hat{\gamma}$	-0.13896
$\hat{\mu}_{opt}$	0.601696	K	0.561798

Table 4: Bias, MSE and RE (%) of Estimators Based on Example 1

Estimator	Bias	MSE	RE
$\hat{\sigma}^2$	0	1.0805	100
$\hat{\sigma}_2^2$	0	0.8979	120.34

Table 5: Descriptive Statistics Based on Example 2

Values			
n	25	s_{1u}^2	1.19E+09
m	12	s_{2u}^2	5.46E+09
$\hat{\mu}_{02}$	6.51E+08	$\hat{\lambda}_{04}$	8.8720
$\hat{\mu}_{20}$	657318750	$\hat{\lambda}_{40}$	8.7910
$\hat{\mu}_{22}$	3.78E+18	$\hat{\lambda}_{22}$	8.8313
$\hat{\mu}_{04}$	3.76E+18	$\hat{\alpha}$	-3.299E-12
$\hat{\mu}_{40}$	3.8E+18	$\hat{\gamma}$	-9.704E-22
$\hat{\mu}_{opt}$	0.996645	K	0.99998867

Table 6: Bias, MSE and RE (%) of Estimators Based on Example 2

Estimator	Bias	MSE	RE
$\hat{\sigma}^2$	0	3.6646E+18	100
$\hat{\sigma}_2^2$	0	1.8385E+18	199.33

Table 7: Descriptive Statistics Based on Example 3

Values			
n	30	s_{1u}^2	2406895765.67
m	15	s_{2u}^2	6.40162E+09
$\hat{\mu}_{02}$	5588789823.00	$\hat{\lambda}_{04}$	3.6050
$\hat{\mu}_{20}$	5508530275021	$\hat{\lambda}_{40}$	3.9945
$\hat{\mu}_{22}$	1.16621E+20	$\hat{\lambda}_{22}$	3.7881
$\hat{\mu}_{04}$	1.2602E+20	$\hat{\alpha}$	-1.48017E-11

$\hat{\mu}_{40}$	1.21209E+20	$\hat{\gamma}$	0.5
$\hat{\mu}_{opt}$	0.944259	K	0.996515

Table 8: Bias, MSE and RE (%) of Estimators Based on Example 3

Estimator	Bias	MSE	RE
$\hat{\sigma}^2$	0	6.36878E+08	100
$\hat{\sigma}_2^2$	0	337236925.69	188.85

References

1. Artes, R., Eva, M. and Garcia, A. (2001). Estimation of current population ratio in successive sampling. *Journal of Indian Society of Agricultural Statistics*, **54 (3)**, 342-354.
2. Cochran, W. G. (1977). *Sampling Techniques*. Third Edition, John Wiley and Sons, New York.
3. Eckler, A. R. (1955). Rotation sampling. *Annals of Mathematical Statistics*, **26**, 664-685.
4. Jessen, R. J. (1942). Statistical investigation of a sample survey for obtaining farm facts. *Iowa Agricultural Experiment Station Research Buliten*, **304**, 54-59.
5. Mukhopadhyay, P. (1998). *Theory and Methods of Survey Sampling*. Prentice Hall of India, New Delhi.
6. Okafor, F. C. and Arnab, R. (1987). Some strategies of two stage sampling for estimating population ratios over two occasions. *Australian Journal of Statistics*, **29 (2)**, 128-142.
7. Patterson, H.D (1950): Sampling on successive occasions with partial replacement of units. *Journal of Royal Statistical Society*-B, **12**, 241-255.
8. Rao, P. and Mudholkar, G. (1967). Generalized multivariate estimators for the mean of finite population parameters. *Journal of American Statistical Society*, **62**, 1008-1012.
9. Rueda, M., Munoz, J. and Arcos, A. (2008). Successive sampling to estimate quantiles with P-auxiliary variables. *Quality and Quantity*, **42**, 427-443.

10. Singh, R. and Singh, M. N. (1996). *Elements of Survey Sampling*. Vol. 15, Kluwer Academic Publishers, The Netherlands.
11. Sud, U. C., Srivastava, A. K. and Sharma, D. P. (2001). On the estimation of population variance in repeat surveys. *Journal of Indian Society of Agricultural Statistics*, **54(2)**, 354-369.
12. Sukhatame, P. V. and Sukhatame, B. V. (1970). *Sampling Theory of Survey with Applications*. 2nd Edition, New Delhi, India.
13. Sukhatame, P. V., Sukhatame, B. V., Sukhatame, S. and Asok, C. (1984). *Sampling Theory of Survey with Applications*. Iowa State University Press, U.S.A.
14. Tikkiwal, B. D. (1951). *Theory of Successive Sampling*, Unpublished Diploma Dissertation. Institute of Agricultural Research Statistics, New Delhi, India.
15. UNO. (2003): *Industrial Commodity Statistics Yearbook*: United Nations, New York.
16. Yates, F. (1960): *Sampling Methods for Censuses and Surveys*. Charles Griffin and Company, London.