A Modification on Linear Systematic Sampling for Odd Sample Size

J. Subramani

Abstract--- *The present paper deals with a modification on the selection of linear systematic sample of odd size. Consequently the proposed method is called modified linear systematic sampling. The performances of the modified linear systematic sampling are assessed with that of simple random sampling and linear systematic sampling for certain hypothetical populations. As a result, it is observed that the modified systematic sample mean performs better than the simple random sample mean and the usual systematic sample mean for estimating the population mean in the presence of linear trend among the population values.*

Keywords--- *Linear Trend, Modified Systematic Sampling, Simple Random Sampling*

I. INTRODUCTION

CONSIDER a finite population $U=\{U_1, U_2, ..., U_N\}$ of N distinct and identifiable units. Let Y be a real variable with value Y_i measured on $U_i, i=1,2,3,...,N$ giving a vector $Y=(Y_1, Y_2, ..., Y_N)$. The problem is to estimate the population mean $\bar{Y}=\frac{1}{N}\sum_{i=1}^{N} Y_i$ on the basis of a random sample selected from the population U. Any ordered sequence $S=\{u_1, u_2, ..., u_n\}=\{U_{i1}, U_{i2}, ..., U_{in}\}$, $1 \le i_l \le N$ and $1 \le l \le n$ is called a random sample of size n. Here we restrict ourselves for the sampling schemes without replacement for further discussion. Simple random sampling is one of the simplest random sampling schemes and also widely used by applied workers in different walks of life. In the past, several other sampling schemes were suggested for selecting a random sample of size n from a finite population of size N including Linear Systematic Sampling by Yates [18], Centered Systematic Sampling by Madaow [5], Balanced Systematic Sampling by Sethi [9] Diagonal Systematic Sampling Schemes by Subramani [12, 13, 14], Star Type Systematic Sampling by Subramani [15] and Determinant Sampling by Subramani and Tracy [11]. However, if there exist a linear trend among the population values, then the systematic sampling is recommended for selecting a random sample of size n, which gives a better estimator compared to simple random sampling.

Consequently several modifications have been made on the systematic sampling to achieve further improvements. For example, Madow [5], Sethi [9], Singh, Jindal and Garg [10], etc. have suggested some modifications on the selection of a systematic sample which includes centered systematic sampling, balanced systematic sampling. Yates [18] has made certain corrections in the estimator itself, commonly known as Yates end corrections. For a detailed discussion on these modifications and their related problems one may refer to Bellhouse and Rao [1], Cochran [2], Fountain and Pathak [3], Gupta and Kabe [4], Murthy [7], Sarjinder [8] and the references cited there in. In the case of linear systematic sampling (LSS), it is assumed that the population size N is a multiple of sample size n and there is no restriction on the sample size n. Recently Subramani [16] has introduced a modification on the selection of a systematic random sample of even size in the linear systematic sampling, called as modified linear systematic sampling (MLSS) scheme. He has also derived the explicit expressions for the MLSS sample means and the variance for certain hypothetical populations with a perfect linear trend among the population values and are compared with that of simple random sampling and linear systematic sampling schemes. Further it has been shown that the MLSS performs better than the simple random and the linear systematic sampling for estimating the finite population means in the presence of linear trend when the sample size is an even number. This has motivated the present study. In this study, a modification on the selection of systematic sample of odd size is made and the explicit expressions are derived for the modified linear systematic sample means together with the variance. Further it is shown that the proposed MLSS performs better than the usual systematic sampling and the simple random sampling for the populations with a perfect linear trend.

II. PROPOSED MODIFIED SYSTEMATIC SAMPLING SCHEME

The proposed modified systematic sampling scheme is explained here, firstly with the help of examples and later the generalized case. For the sake of simplicity and for the benefit of the readers, selecting a modified linear systematic sample of size n from a population of size $N = kn = k(2m+1)$ and are explained with the help of following examples for the values of $N = 3k$, $k = 3, and\ 4$.

Example 1: Let $N = 3k\ and\ k = 3$.

Step 1: Arrange the 9 population units as given below:

i			j		
1	2	3	1	2	3
1	2	3			
4	5	6	7	8	9

J. Subramani, Associate Professor, Department of Statistics, Pondicherry University, Puducherry-605014, India. E-mail: drjsubramani@yahoo.co.in

Step 2: Select two random numbers $1 \le i \le 3$ and $1 \le j \le 3$; include all the elements in the columns corresponding to i *and* j. The selected samples are given below:

Sample No.	i	j	Sampled units
1	1	1	1, 4, 7
2	1	2	1, 4, 8
3	1	3	1, 4, 9
4	2	1	2, 5, 7
5	2	2	2, 5, 8
6	2	3	2, 5, 9
7	3	1	3, 6, 7
8	3	2	3, 6, 8
9	3	3	3, 6, 9

Example 2: Let $N = 3k$ *and* $k = 4$.

Step 1: Arrange the 12 population units as given below:

i				j			
1	2	3	4	1	2	3	4
1	2	3	4				
5	6	7	8	9	10	11	12

Step 2: Select two random numbers $1 \le i \le 4$ and $1 \le j \le 4$; include all the elements in the columns corresponding to i *and* j. The selected samples are given below:

Sample No.	i	j	Sampled units
1	1	1	1, 5, 9
2	1	2	1, 5, 10
3	1	3	1, 5, 11
4	1	4	1, 5, 12
5	2	1	2, 6, 9
6	2	2	2, 6, 10
7	2	3	2, 6, 11
8	2	4	2, 6, 12
9	3	1	3, 7, 9
10	3	2	3, 7, 10
11	3	3	3, 7, 11
12	3	4	3, 7, 12
13	4	1	4, 8, 9
14	4	2	4, 8, 10
15	4	3	4, 8, 11
16	4	4	4, 8, 12

The steps involved in selecting a modified linear systematic sample of odd size $n = 2m+1$ from a population of size $N = nk = (2m+1)k =$, k is any positive integer are as follows:

1. Firstly arrange the $N = (2m+1)k$ population units (labels) into a $2m \times 2k$ matrix (see equation (6)).
2. The first k columns are assumed as Set 1 and the next k columns are assumed as Set 2.
3. For the two random numbers i and j in between 1 and k, select all the $2m$ units in the ith column of Set 1 and the single unit in the jth column of Set 2, which together gives the sample of size $n = 2m+1$
4. The step 3 leads to k^2 samples of size n.

Since the modified linear systematic sampling scheme has k^2 samples of size n and each unit is included in k samples, the first order and second order inclusion probabilities are obtained as given below:

$$\pi_i = \frac{1}{k}, \text{ for } i = 1,2,3...,N \tag{1}$$

$$\pi_{ij} = \begin{cases} \frac{1}{k} & \text{if i and j are in the same column of set 1} \\ \frac{1}{k^2} & \text{if i from Set 1 and j from Set 2} \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

The first order inclusion probabilities are the same for both the systematic sampling and the modified systematic sampling schemes. However on the second order inclusion probabilities, the two units in the same column will get the same probability $\frac{1}{k}$ and 0 for the two units from different columns in the case of systematic sampling where as the two units in the same column of the first set will get $\frac{1}{k}$; two units from different sets will get $\frac{1}{k^2}$ and all other pairs will get 0 in the case of modified systematic sampling. However the differences between the systematic sampling and the modified linear systematic sampling schemes are as follows:

- In systematic sampling each population unit appears in one sample only whereas in modified systematic sampling each of the population units appears in k samples.
- The number of samples selected in systematic sampling is k samples where as in modified systematic sampling it is k^2 samples.
- The number of pairs of units appeared together in systematic sampling is $kn(n-1)/2$ and all are distinct pairs where as it is $k^2 n(n-1)/2$ out of which $k(n-1)(2k+n-2)/2$ pairs are distinct in the case of modified systematic sampling.

In general, for the given population size $N = kn = k(2m+1)$, the selected modified systematic samples (labels of the population units) for the random starts i and j are given below:

$$S_{ij} = (i, i+k, i+3k, ..., i+(n-2)k, j+(n-1)k) \quad i, j = 1,2,3,...,k$$

The modified linear systematic sample based on the random starts i and j and the variance are obtained respectively as

$$\bar{y}_{mlss} = \bar{y}_{ij} = \frac{1}{n}\left[\sum_{l=0}^{n-2} y_{i+kl} + y_{j+(n-1)k}\right] \quad i, j = 1,2,3,...,k \quad (3)$$

and

$$V(\bar{y}_{mlss}) = \frac{1}{k^2}\sum_{i=1}^{k}\sum_{j=1}^{k}\left(\bar{y}_{ij} - \bar{Y}\right)^2 \quad (4)$$

III. Relative Performance of Modified Linear Systematic Sampling in the Presence of Perfect Linear Trend

It is well known that the linear systematic sampling (LSS) is preferred over the simple random sampling if the population consists solely of a linear trend among the population values. For a detailed discussion on estimation of finite population means, one may refer to Cochran [2], Fountain and Pathak [3], Sukhatme et.al. [17] and the references cited therein. In this section, the relative efficiencies of the modified systematic sampling with that of simple random sampling and systematic sampling for estimating the mean of finite populations with linear trend among the population values are assessed for certain hypothetical populations.

A. Population with a Linear Trend

In this hypothetical population, the values of $N = kn = k(2m+1)$ population units are in arithmetic progression. That is,

$$Y_i = a + ib, \; i = 1,2,3,...,N \quad (5)$$

The method of obtaining the modified systematic sample means for the above hypothetical population with the random starts i and j are as given below:

Step1: Arrange the $N = kn = k(2m+1)$ population units in a $2m \times 2k = (n-1) \times 2k$ matrix array as given below:

$$\begin{array}{cccccccccccc}
1 & 2 & & i & & k & k+1 & & k+2 & & k+j & 2k \\
\hline
1 & 2 & ... & i & ... & k & & & & & & \\
k+1 & k+2 & ... & k+i & ... & 2k & & & & & & \\
. & . & . & . & . & . & & & & & & \\
(n-2)k+1 & (n-2)k+2 & ,,, & (n-2)k+i & ... & (n-1)k & (n-1)k+1 & & (n-1)k+2 & ... & (n-1)k+j & ... \; nk
\end{array} \quad (6)$$

Step 2: As stated earlier, the first k columns are assumed as Set 1 and the next k columns are assumed as Set 2.

Step 3: For the two random numbers i and j in between 1 and k, then select all the $2m = n-1$ units in the ith column of Set 1 and the single unit in the jth column of Set 2, which give the sample totals of size n together with the modified systematic sample mean as given below.

$$\sum_{l=1}^{n-1} y_{il} = \left[a + ib + a + (k+i)b + ... + a + [(n-2)k + i]b\right]$$

$$= \left[(n-1)a + [(n-1)i + \frac{k(n-2)(n-1)}{2}]b\right]$$

$$= (n-1)\left\{a + [i + \frac{k(n-2)}{2}]b\right\}$$

$$y_j = \left\{a + [j + k(n-1)]b\right\}$$

$$\bar{y}_{mlss} = \bar{y}_{ij} = \frac{1}{n}\left[\sum_{l=1}^{n-1} y_{il} + y_j\right] = a + \frac{1}{n}\left[(n-1)i + j + \frac{kn(n-1)}{2}\right]b$$

$$i, j = 1,2,3,...,k \quad (7)$$

For the above population with a linear trend, the variances of the simple random sample mean $V(\bar{y}_r)$, systematic sample mean $V(\bar{y}_{sy})$ and modified linear systematic sample mean $V(\bar{y}_{mlss})$ are obtained in terms of k as given below:

$$V(\bar{y}_r) = \frac{(k-1)(N+1)b^2}{12} \quad (8)$$

$$V(\bar{y}_{sy}) = \frac{(k-1)(k+1)b^2}{12} \quad (9)$$

$$V(\bar{y}_{mlss}) = \left(\frac{(n-1)^2+1}{n^2}\right)\frac{(k-1)(k+1)b^2}{12} \quad (10)$$

By comparing the variance expressions given above one can easily show that for every odd n

$$V(\bar{y}_{mlss}) \le V(\bar{y}_{sy}) \le V(\bar{y}_r) \quad (11)$$

IV. A Modification on Modified Linear Systematic Sampling

It has been shown that in Section 3.1, the modified linear systematic sampling performs better than the other sampling schemes including the systematic sampling and simple random sampling. However, the proposed sampling scheme is not completely trend free sampling (Mukerjee and Sengupta [6]). Further improvement can be achieved by modifying the modified linear systematic sample means, as done in the case of linear systematic sample mean by introducing Yates [18] type end corrections.

A. Yates Type End Corrections

The modification involves the usual modified linear systematic sampling, but the modified sample mean is defined as

$$\bar{y}^*_{mlss} = \bar{y}_{mlss} + \alpha(y_1 - y_n) = \bar{Y} \quad (12)$$

That is, the units selected first and last are given the weights $\frac{1}{n} + \alpha$ and $\frac{1}{n} - \alpha$ respectively, where as the remaining units will get the equal weight of $\frac{1}{n}$, so as to make the proposed estimator is equal to the population mean. That is, the value of α is obtained as

$$\alpha = \frac{\bar{Y} - \bar{y}_{mlss}}{(y_1 - y_n)} \quad (13)$$

For the hypothetical population defined in Section 3.1, after a little algebra we have obtained the value of α for the two random starts i and j given below:

$$\alpha = \frac{2(n-1)i + 2j - n(k+1)}{2n[(n-1)k + j - i)]} \quad (14)$$

Remark 4.1: In the presence of a perfect linear trend, the revised modified linear systematic sample estimator $\bar{y}^*_{mlss} = \bar{Y}$

and hence the variance $V(\bar{y}^*_{mlss}) = 0$. That is, it becomes a trend free sampling (Mukerjee and Sengupta [3]).

V. Conclusion

In this paper a new systematic type sampling scheme namely, modified linear systematic sampling (MLSS) scheme is introduced for the odd sample size. That is, the proposed sampling scheme can be used as an alternate to linear systematic sampling, when $N = kn = k(2m+1)$. The explicit expression for the sample means and the variance of the MLSS are derived for certain hypothetical populations with a perfect linear trend among the population values. The performances of the proposed sampling scheme are assessed with that of the simple random sampling and linear systematic sampling for the above said hypothetical population. It is shown that the modified linear systematic sampling performs well compared to other sampling schemes considered here. In fact $V(\bar{y}_{mlss}) \le V(\bar{y}_{sy}) \le V(\bar{y}_r)$ is true for every odd integer n.

References

[1] D.R. Bellhouse, and J.N.K. RAO, Systematic sampling in the presence of linear trends, Biometrika,62, 694-697, 1975.

[2] W.G. COCHRAN, "Sampling Techniques", 3rd Edition, John Wiley and Sons, New York, 1977.

[3] R.L. Fountain and P.L PATHAK, "Systematic and Non-random Sampling in the Presence of Linear Trends", Communications in statistics- Theory and Methods 18, 2511-2526, 1989.

[4] A.K. Gupta and D.G.KABE, Theory of Sample Surveys, World Scientific Publishing Company, 2011.

[5] W.G.MADOW, On the theory of systematic sampling III. Comparison of centered and random start systematic sampling, Ann.Math.Statist., 24, 101-106, 1953.

[6] R. Mukerjee and S. SENGUPTA, "Optimal Estimation of a Finite Population Means in the Presence of Linear Trend", Biometrika 77, 625-630, 1990.

[7] M.N. Murthy, "Sampling Theory and Methods ", Statistical Publishing House, Calcutta, India, 1967.

[8] Sarjinder Singh, Advanced Sampling- Theory with Publications, Springer, 2003

[9] V.K. SETHI, On optimum pairing of units, Sankhya, B27, 315-320, 1965.

[10] D. Singh, K.K. Jindal and J.N.GARG, On modified systematic sampling, Biometrika, 55, 541-546, 1968.

[11] J. Subramani and D.S. TRACY, "Determinant Sampling Scheme for Finite Populations, International Journal of Mathematics and Statistical Sciences, 8, (1), 27-41, 1999.

[12] J. Subramani, "Diagonal Systematic Sampling Scheme for Finite Populations, Journal of Indian Society of Agricultural Statistics, 53, 187-195, 2000.

[13] J. Subramani, "Further Results on Diagonal Systematic Sampling Scheme for Finite Populations, Journal of Indian Society of Agricultural Statistics – Vol 63, No.3, 277-282, 2009

[14] J. Subramani, "Generalization of Diagonal Systematic Sampling Scheme for Finite Populations, Model Assisted Statistics and Applications, 5, 117-128, 2010.

[15] J. Subramani, "Star Type Systematic Sampling for Finite Populations, Communications in Statistics – Theory and Methods, Accepted for Publication, 2012.

[16] J. Subramani, "A Modification on Systematic Sampling", Proceeding of the 14th Annual Conference of the Society of Statistic, Computer and Applications, 2012.

[17] P.V. Sukhatme, B.V. Sukhatme, S. Sukhatme and C. ASOK, "Sampling Theory of Surveys with Applications", Iowa State Univ. Press, Ames, Iowa, USA, 1984.

[18] F. Yates, "Systematic Sampling", Phil. Trans. Royal. Soc., A, 241, 345-377, 1948.

Appendix

Derivation of Variance of Modified Linear Systematic Sample Mean In The Presence of Linear Trend

Let $N = kn = k(2m+1)$ be the population size, where k any positive integer. The population units are $U_1, U_2, ..., U_N$. In this hypothetical (labeled) population, the values of N population units are in arithmetic progression. That is,

$$Y_i = a + ib, \quad i = 1,2,3,...,N$$

Derivation of Variance Of Modified Linear Systematic Sample Mean

For the above population with a linear trend, the modified linear systematic sample mean with the random starts i and j and the population mean are obtained as given below:

$$\bar{y}_{mlss} = \bar{y}_{ij} = a + \frac{1}{n}\left[(n-1)i + j + \frac{kn(n-1)}{2}\right]b \qquad i, j = 1,2,3,...,k$$

$$\bar{Y} = a + \left[\frac{N+1}{2}\right]b = a + \left[\frac{kn+1}{2}\right]b$$

The variance of modified linear systematic sample mean $\bar{y}_{mlss}$ is obtained as

$$V(\bar{y}_{mlss}) = E\left(\bar{y}_{mlss} - \bar{Y}\right)^2 = \frac{1}{k^2}\sum_{i=1}^{k}\sum_{j=1}^{k}\left(\bar{y}_{ij} - \bar{Y}\right)^2$$

$$= \frac{1}{k^2}\sum_{i=1}^{k}\sum_{j=1}^{k}\left(a + \frac{1}{n}\left[(n-1)i + j + \frac{kn(n-1)}{2}\right]b - a - \left[\frac{kn+1}{2}\right]b\right)^2$$

$$= \frac{b^2}{k^2n^2}\sum_{i=1}^{k}\sum_{j=1}^{k}\left((n-1)i + j - \frac{n(k+1)}{2}\right)^2$$

$$= \frac{b^2}{k^2n^2}\sum_{i=1}^{k}\sum_{j=1}^{k}\left(\begin{array}{l}(n-1)^2i^2 + j^2 + 2(n-1)ij \\ -n[k+1][(n-1)i + j] + \left[\frac{n(k+1)}{2}\right]^2\end{array}\right)$$

$$= \frac{b^2}{k^2n^2}\left(\begin{array}{l}\frac{k^2(k+1)(2k+1)}{6}\left[(n-1)^2 + 1\right] + 2(n-1)\left(\frac{k(k+1)}{2}\right)^2 \\ -\frac{2n^2k^2(k+1)^2}{4} + \frac{n^2k^2(k+1)^2}{4}\end{array}\right)$$

$$= \frac{b^2}{n^2}\left(\begin{array}{l}\frac{(k+1)(2k+1)}{6}\left[(n-1)^2 + 1\right] \\ +2(n-1)\frac{(k+1)^2}{4} - \frac{n^2(k+1)^2}{4}\end{array}\right)$$

$$= \frac{b^2}{n^2}\left(\begin{array}{l}\frac{(k+1)(2k+1)}{6}\left[(n-1)^2 + 1\right] \\ -\frac{(k+1)^2}{4}\left(n^2 - 2n + 2\right)\end{array}\right)$$

$$= \frac{b^2}{n^2}\left(\frac{(k+1)(2k+1)}{6} - \frac{(k+1)^2}{4}\right)\left((n-1)^2 + 1\right)$$

After a little algebra, the above equation can be written as

$$V(\bar{y}_{mlss}) = \left(\frac{(n-1)^2 + 1}{n^2}\right)\frac{(k-1)(k+1)}{12}b^2$$

$$V(\bar{y}_{mlss}) = \left(\frac{(n-1)^2 + 1}{n^2}\right)\frac{(k^2-1)}{12}b^2$$

Dr. J. Subramani has received the doctorate degree in Statistics from University of Madras. He is currently working as Associate Professor, Department of Statistics, Pondicherry University. He has more than 25 years of experience both in Teaching and Research. To his academic credit he has received U.S.Nair Young Statistician Award from Indian Society of Probability and Statistics and also the International Young Statistician Award from the International Statistical Institute, The Netherlands. His research interests are Estimation of Variance Components, Missing Data Analysis, Sampling Theory, Incomplete Block Designs, Control Charts and Process Capability Analysis. He has published more than 75 research papers in reputed International and National Journals. He has also participated in many conferences and workshops in India as well as abroad and presented research papers; delivered invited talks and special addresses. He has organized many International and National Conferences and also acted as an organizing committee member in many conferences. {E-mail: drjsubramani@yahoo.co.in}