Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF AGRICULTURAL STATISTICS 63(3) 2009 277-282

Further Results on Diagonal Systematic Sampling Scheme for Finite Populations

J. Subramani*
S. Veerasamy Chettiar College of Engineering and Technology, Tamilnadu

(Received: January 2009, Revised: September 2009, Accepted: October 2009)

SUMMARY

A generalization of diagonal systematic sampling scheme for estimation of a finite population mean is introduced. The assumption of $n \leq k$ is relaxed here and hence the proposed method is applicable for all values of n provided $N = kn$. The relative performance of proposed diagonal systematic sample mean along with those of the simple random and systematic sample means is assessed for a natural population.

Key words : Diagonal systematic sampling, Systematic sampling, Natural population, Linear trend, Trend free sampling.

1. INTRODUCTION

Consider a finite population $U = \{U_1, U_2, ..., U_N\}$ of N distinct and identifiable units. Let Y be the study variable and take a real 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 $\bar{Y} = \sum_{i=1}^{N} Y_i$ on the basis of a random sample selected from the population U. Any ordered sequence $S = \{u_1, u_2,..., u_n\} = \{U_{i_1}, U_{i_2},..., U_{i_n}\}$, $1 \leq i_l \leq N$, and $1 \leq l \leq n$ is called a random sample of size n.

In the past, several sampling schemes were suggested for selecting a random sample of size n from a finite population of size N. If there is a linear trend among the population units then the systematic sampling is recommended for selecting a sample of size n, which gives the best estimator compared to simple random sampling. Recently Subramani (2000) has introduced the diagonal systematic sampling with fixed sample size, which uses the knowledge of the labels of the population units to provide an unbiased estimator of the population mean. The explicit expressions for the diagonal systematic sample mean and its variance are obtained for certain hypothetical populations. Further, the relative performance of diagonal systematic sampling; simple random sampling and systematic sampling schemes are assessed for certain natural populations. As a result, it has been shown that the diagonal systematic sampling performs better than the simple random sampling and the linear systematic sampling for estimating the finite population mean in the presence of linear trend. For more details the readers are referred to Subramani (2000) and the references cited there in.

It is to be noted here that for the case of simple random sampling there is no restriction either on the sample size n or on the population size N. Similarly in the case of linear systematic sampling scheme it is required to have $N = kn$ and there is no restriction on the sample size n, whether $n \leq k$ or $n \geq k$. However, the drawback in diagonal systematic sampling of Subramani (2000) is that it is applicable only when the sample size is $n \leq k$, where k is the number of distinct samples and $N = kn$. It has motivated the present study and consequently the generalization of diagonal

Corresponding author : J. Subramani
E-mail address : drjsubramani@yahoo.co.in

systematic sampling scheme is introduced. The proposed sampling scheme requires only $N = kn$ as in the case of linear systematic sampling, which leads to a real comparison between the diagonal systematic sampling and linear systematic sampling schemes. The explicit expressions for the generalized diagonal systematic sample mean and its variance are presented for the hypothetical population with a perfect linear trend among the population values. The relative performance of the proposed diagonal systematic sample mean along with those of the simple random mean and systematic sample mean is assessed for a natural population.

2. GENERALIZED DIAGONAL SYSTEMATIC SAMPLING SCHEME

For the sake of simplicity and for the benefit of the readers, the steps involved in selecting a diagonal systematic sample of size n from a population of size $N = kn$ are reproduced here. Let $N = kn$ where $n \le k$, be the population size. The population units $U_1, U_2, \ldots, U_N$ are arranged in a $n \times k$ matrix **M** (say) and the j-th row of **M** is denoted by R_j, $j = 1, 2, \ldots n$. The elements of R_j are $\{U_{(j-1)k+i}, i = 1, 2, \ldots k\}$. The diagonal systematic sampling scheme consists of drawing n units from the matrix **M** systematically such that the selected n units are the diagonal elements or broken diagonal elements of the matrix **M**. Hence the selected units are from different rows and from different columns.

The diagonal systematic sampling scheme discussed above is applicable only when $N = kn$ and $n \le k$. If the sample size $n > k$ then the diagonal systematic sampling discussed above cannot be used to estimate the finite population mean. It is the drawback compared to linear systematic sampling, which is valid for any value of n provided the population size $N = kn$. Hence, an attempt has been made in this paper to relax the condition $n \le k$ and the resulting sampling scheme is called generalized diagonal systematic sampling. Further, explicit computable expressions for the sample mean and the variance are provided for the hypothetical population with a linear trend among the population values. Here, the sampled units need not be from different columns but are from different rows of **M**. If $n \le k$ then the generalized diagonal systematic sampling is reduced to the usual diagonal systematic sampling. Hence, we concentrate only for the case where $n > k$.

The steps involved in the generalized diagonal systematic sampling scheme for selecting a random sample of size n are given below:

Let $N = kn$ where $n > k$, be the population size. Since the sample size $n > k$, one can write the sample size $n = pk + m$, where p is any positive integer greater than one.

Step 1. Arrange the N population units $U_1, U_2, \ldots, U_N$ in a $n \times k$ matrix **M** (say).

Step 2. Select a random number r such that $1 \le r \le k$.

Step 3. Select downward the diagonal elements from r, in the direction of left to right until reaching the right most column of **M**.

Step 4. Once the right most column of **M** is reached then select the first element in the next row.

Step 5. Repeat the Steps 2 through 4 until selecting a random sample of n elements.

For example to select a generalized diagonal systematic sample of size 5 from a population of size 15 units, consider the following:

$$\mathbf{M} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \\ 10 & 11 & 12 \\ 13 & 14 & 15 \end{bmatrix}$$

As the result, (1,5,9,10,14), (2,6,7,11,15) and (3,4,8,12,13) are the generalized diagonal systematic samples of size 5 with the random starts 1, 2 and 3 respectively. It is to be noted that the first order and second order inclusion probabilities are obtained as given below:

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

and
$$\pi_{ij} = \begin{cases} \dfrac{1}{k} & \text{if } i \text{ and } j \text{ are in the same diagonal} \\ 0 & \text{otherwise} \end{cases}$$

The first order inclusion probabilities are the same for both the systematic sampling and the diagonal systematic sampling schemes but the difference is on the second order inclusion probabilities. The two units in the same column will get the same probability $\frac{1}{k}$ in the case of systematic sampling whereas the two units in the same diagonal or broken diagonal will get the same probability $\frac{1}{k}$ in the case of diagonal systematic sampling and zero for other pair of units. For more details on diagonal systematic sampling and its performance one may refer to Subramani (2000).

3. RELATIVE PERFORMANCE OF GENERALIZED DIAGONAL SYSTEMATIC SAMPLING IN THE PRESENCE OF PERFECT LINEAR TREND

It is well known that the linear systematic sampling is preferred over the simple random sampling whenever there is a linear trend among the population values. For a detailed discussion on estimation of finite population mean one may refer to Cochran (1977), Fountain and Pathak (1989) and the references cited therein. Further, Subramani (2000) has shown that the diagonal systematic sampling is more efficient than the simple random sampling and systematic sampling provided (i) $n \le k$, (ii) $N = kn$ and (iii) There exists a linear trend among the population values. In this section we have compared the relative efficiency of the generalized diagonal systematic sampling scheme with that of simple random sampling and systematic sampling for estimating the mean of finite populations with linear trend among the population values.

3.1 Population with Linear Trend

In this hypothetical population, the values of N population units are in arithmetic progression. That is,

$$Y_i = a + ib \, , \; i = 1, 2, ..., N \tag{3.1}$$

After a little algebra one may obtain the generalized diagonal systematic sample mean for the above hypothetical population with the random start r as given below:

$$\bar{y}_{gdsy} = \begin{cases} a + \frac{1}{n}\left[\frac{pk(pk^2+1)+m(m-1)(k+1)}{2} + m(pk^2+r)\right]b & \text{if } r \le k-m+1 \\ a + \frac{1}{n}\left[\frac{[pk(pk^2+1)+km(m-1) + k(k+1)+(k-m)(k-m+1)]}{2} + mpk^2 - (k-m)r\right]b & \text{if } r > k-m+1 \end{cases}$$

where $n = pk + m$ and $p \ge 0$

For the above population with a linear trend, the variances of the simple random sample mean $V(\bar{y}_r)$ systematic sample mean $V(\bar{y}_{sy})$, diagonal systematic sample mean $V(\bar{y}_{dsy})$ and generalized diagonal systematic sample mean $V(\bar{y}_{gdsy})$ are obtained as given below:

$$V(\bar{y}_r) = \frac{(k-1)(N+1)b^2}{12} \tag{3.2}$$

$$V(\bar{y}_{sy}) = \frac{(k-1)(k+1)b^2}{12} \tag{3.3}$$

$$V(\bar{y}_{dsy}) = \frac{(k-n)[n(k-n)+2]b^2}{12n} \tag{3.4}$$

$$V(\bar{y}_{gdsy}) = \frac{m(k-m)[m(k-m)+2]b^2}{12n^2} \tag{3.5}$$

where $n = pk + m$ and $p \ge 0$

The derivation of the variance of generalized diagonal systematic sample mean is given in the Appendix. Further, one can easily show that diagonal systematic sampling is a particular case of the generalized diagonal systematic sampling. When $m = n$ in the variance expression given in (3.5), it reduces to the variance expression given in (3.4). Hence, we can hereafter consider only the generalized diagonal systematic sampling for assessing the relative performance with that of simple random sampling and systematic sampling schemes. By comparing the various variance expressions given above, one can easily show that generalized diagonal systematic sampling is more efficient than the other sampling schemes. In fact

$$V(\bar{y}_{gdsy}) \le V(\bar{y}_{sy}) \le V(\bar{y}_r) \tag{3.6}$$

and the equality sign occurs only when $n = 1$, which is the trivial case where $k = N$.

Remark 3.1. If $n = k$ or multiples of k then $V(\bar{y}_{gdsy})$ = 0. In this case the generalized diagonal systematic sampling becomes a completely trend free sampling (See Mukerjee and Sengupta 1990).

4. RELATIVE PERFORMANCE OF GENERALIZED DIAGONAL SYSTEMATIC SAMPLING FOR A NATURAL POPULATION

It has been shown in Section 3 that diagonal systematic sampling performs well, compared to simple random and systematic sampling schemes whenever there exists a perfect linear trend among the population units. However, this is an unrealistic assumption in real life situations. Consequently an attempt has been made to study the efficiency of generalized diagonal systematic sampling for a natural population considered by Subramani (2004). The data were collected for assessing the process capability of a manufacturing process from an auto ancillary manufacturing unit located in Tamil Nadu. The data were pertaining to the measurements taken continuously during turning operation performed on the component, namely Torsion bar in Frontier CNC Lathe Machine. The data were collected for estimating the mean value of the outer diameter of the Torsion bar, one of the key components in integrated power steering system. The measurements were taken continuously for the first 50 components produced in a shift. The 50 measurements based on the order of the production are given below. However, we have taken only the first 48 measurements for assessing the relative performance of the various sampling schemes, which give 8 combinations for (k, n) whereas if we take 50 measurements which lead to 4 combinations only.

Table 4.1. Data of outer Diameter of Torsion Bar (Spec. 9065 ± 25)

9050	9052	9050	9052	9052	9056	9056	9054	9056	9058
9054	9054	9060	9058	9060	9058	9056	9058	9058	9060
9062	9064	9062	9064	9066	9070	9068	9072	9072	9070
9072	9070	9070	9072	9074	9076	9078	9076	9076	9078
9078	9078	9082	9080	9082	9080	9082	9086	9086	9084

We obtain the variances of simple random sample mean, systematic sample mean and diagonal systematic sample mean for all the possible combinations of (k, n), that is, for (2,24), (3, 16), (4,12), (6,8), (8,6), (12,4), (16,3) and (24,2). Table 4.2 presents these variances for the natural population given above. It is seen now that the generalized diagonal systematic sample mean is the most efficient and also $V(\bar{y}_{gdsy}) \le V(\bar{y}_{sy}) \le V(\bar{y}_r)$ in each of the cases. An asterisk indicates the minimum variance in each of the cases in the Table 4.2.

Table 4.2. Comparison of simple random, systematic and generalized diagonal systematic sample means for the population considered by Subramani (2004)

Popln. Number	k	n	$V(\bar{y}_r)$	$V(\bar{y}_{sy})$	$V(\bar{y}_{gdsy})$
1	2	24	2.03531	0.17361	0.00000
2	3	16	4.07063	0.17014	0.00347
3	4	12	6.10594	0.42361	0.20139
4	6	8	10.17657	0.90972	0.61806
5	8	6	14.24719	3.71528	0.43750
6	12	4	22.38844	4.11806	3.07639
7	16	3	30.52970	15.21528	11.10417
8	24	2	46.81221	19.07639	15.82639

ACKNOWLEDGEMENTS

The author wishes to thank the referee and the editor for their useful comments which have helped to improve the presentation of this paper.

REFERENCES

Cochran, W.G. (1977). *Sampling Techniques.* 3rd Edition, John Wiley and Sons, New York.

Fountain, R.L. and Pathak, P.L. (1989). Systematic and non-random sampling in the presence of linear trends. *Comm. Statist. – Theory Methods*, **18**, 2511-2526.

Mukerjee, R. and Sengupta, S. (1990). Optimal estimation of a finite population means in the presence of linear trend. *Biometrika*, **77**, 625-630.

Subramani, J. (2000). Diagonal systematic sampling scheme for finite populations. *J. Ind. Soc. Agril. Statist.,* **53(2)**, 187-195.

Subramani, J. (2004). Diagonal systematic sampling for process control in the presence of linear trend. Presented at the International Conference on "Future of Statistical Theory, Practice and Education" held in Indian School of Business, Hyderabad, during 29 December 2004 to 1 January 2005.

APPENDIX

Derivation of Variances of Generalized Diagonal Systematic Sample Means in the Presence of Linear Trend

Let $N = kn$, where $n \geq k$, be the population size. The population units $U_1, U_2, ..., U_N$ are arranged in a $n \times k$ matrix **M** (say) and the j-th row of **M** is denoted by R_j, $j = 1, 2, ..., n$. The elements of R_j are $\{U_{(j-1)k+i}, i = 1, 2, ..., k\}$. The diagonal systematic sampling scheme consists of drawing n units from the matrix **M** systematically such that the selected n units are the diagonal elements or broken diagonal elements of the matrix **M**. Hence the selected units are from different rows and need not be from different columns.

Let Y_{ij} be the observation corresponding to the unit in i-th row and j-th, column, that is corresponding to the unit, $U_{(i-1)k+j}$ then the sample observations are denoted by

$$Sr = \{Y_{1r}, Y_{2(r+1)}, Y_{3(r+2)}, ..., Y_{n(r+n-1)}\}, r = 1, 2, ..., k$$

If $r + n - 1 > k$ then $r + n - 1$ has to be reduced to mod k.

A.1 Population with Linear Trend

In this hypothetical population, the values of N population units are in arithmetic progression. That is,

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

For the above population with a linear trend, the diagonal systematic sample mean with the random start r and the population mean are obtained as given below:

$$\bar{y}_{gdsy} = \begin{cases} a + \dfrac{1}{n}\left[\dfrac{pk(pk^2+1) + m(m-1)(k+1)}{2} + m(pk^2 + r)\right] b & \text{if } r \leq k - m + 1 \\ a + \dfrac{1}{n}\left[\dfrac{[pk(pk^2+1) + km(m-1) + k(k+1) + (k-m)(k-m+1)]}{2} + mpk^2 - (k-m)r\right] b & \text{if } r > k - m + 1 \end{cases}$$

$$\bar{Y} = \frac{pk^2 + mk + 1}{2}$$

where $n = pk + m$ and $p \geq 0$.

The variance of generalized diagonal systematic sample mean $\bar{y}_{gdsy}$ is obtained as

$$V(\bar{y}_{gdsy}) = E(\bar{y}_{gdsy} - \bar{Y})^2$$

$$= \frac{1}{k}\sum_{i=1}^{k}(\bar{y}_{gdsy} - \bar{Y})^2$$

$$= \frac{1}{k}\left[\sum_{i=1}^{k-m+1}(\bar{y}_{gdsy} - \bar{Y})^2 + \sum_{i=k-m+2}^{k}(\bar{y}_{gdsy} - \bar{Y})^2\right]$$

$$= \frac{1}{k}[SS1 + SS2] \quad \text{(say)}$$

To compute the $V(\bar{y}_{gdsy})$ given above consider the two parts separately. Further the constant b occurs both in sample mean and the population mean as a multiplier and it will be considered at the end of this derivation and is not included in each of the steps involved in the derivation of variance.

When $r \leq k - m + 1$ consider the following:

$$SS1 = \sum_{i=1}^{k-m+1}(\bar{y}_{gdsy} - \bar{Y})^2$$

$$= \sum_{i=1}^{k-m+1}\left(\frac{1}{n}\left[\frac{pk(pk^2+1) + m(m-1)(k+1)}{2} + m(pk^2 + r)\right] - \frac{pk^2 + mk + 1}{2}\right)^2$$

$$= \sum_{i=1}^{k-m+1}\left(\frac{1}{n}\left[\frac{pk(pk^2+1) + m(m-1)(k+1)}{2} + m(pk^2 + r) - \frac{n(pk^2 + mk + 1)}{2}\right]\right)^2$$

After a little algebra we have obtained the value of $SS1$ as given below:

$$SS1 = \frac{m^2}{4n^2}\sum_{i=1}^{k-m+1}\left[2r-(k-m+2)\right]^2$$

$$\Rightarrow \quad SS1 = \frac{m^2}{12n^2}(k-m+1)(k-m+2)(k-m)$$

When $r \geq k-m+2$ consider the following:

$$SS2 = \sum_{i=k-m+2}^{k}(\bar{y}_{gdsy}-\bar{Y})^2$$

$$= \sum_{i=k-m+2}^{k}\left(\frac{1}{n}\left[\frac{[pk(pk^2+1)+km(m-1)+k(k+1)+(k-m)(k-m+1)]}{2} + mpk^2-(k-m)r\right] - \frac{pk^2+mk+1}{2}\right)^2$$

$$= \sum_{i=k-m+2}^{k}\left(\frac{1}{n}\left[\frac{[pk(pk^2+1)+km(m-1)+k(k+1)+(k-m)(k-m+1)]}{2} + mpk^2-(k-m)r-\frac{n(pk^2+mk+1)}{2}\right]\right)^2$$

After a little algebra, we have obtained the value of $SS2$ as given below:

$$SS2 = \frac{(k-m)^2}{4n^2}\sum_{i=k-m+2}^{k}(2k-m+2-2r)^2$$

$$\Rightarrow \quad SS2 = \frac{(k-m)^2}{4n^2}\left[m(m-1)(m-2)\right]$$

Now by combining the expressions obtained for $SS1$ and $SS2$, we can obtain the variance of the generalized diagonal systematic sample mean as follows:

$$V(\bar{y}_{gdsy}) = E(\bar{y}_{gdsy}-\bar{Y})^2 = \frac{1}{k}\sum_{i=1}^{k}(\bar{y}_{gdsy}-\bar{Y})^2$$

$$= \frac{1}{k}\left[\sum_{i=1}^{k-m+1}(\bar{y}_{gdsy}-\bar{Y})^2 + \sum_{i=k-m+2}^{k}(\bar{y}_{gdsy}-\bar{Y})^2\right]$$

$$= \frac{1}{k}[SS1 + SS2]$$

$$= \frac{1}{k}\left[\frac{m^2}{12n^2}(k-m+1)(k-m+2)(k-m) + \frac{(k-m)^2}{12n^2}m(m-1)(m-2)\right]$$

$$= \frac{m(k-m)}{12kn^2}\,[m(k-m+1)(k-m+2) + (k-m)(m-1)(m-2)]$$

After a little algebra we have obtained the value of $V(\bar{y}_{gdsy})$ as given below:

$$\Rightarrow \quad V(\bar{y}_{gdsy}) = \frac{m(k-m)}{12n^2}\left[m(k-m)+2\right]$$

That is, the variance of the generalized diagonal systematic sample mean is obtained as given below:

$$V(\bar{y}_{gdsy}) = \frac{m(k-m)\left[m(k-m)+2\right]b^2}{12n^2}$$

where $n = pk + m$ and $p \geq 0$.