Jour. Ind. Soc. Ag. Statistics
53(2), 2000 : 187-195

Diagonal Systematic Sampling Scheme for Finite Populations

J. Subramani
Rane - TRW Steering Systems Ltd. Tamilnadu-621316
(Received : January, 2000)

SUMMARY

A modified stystematic sampling scheme, called diagonal systematic sampling, for estimation of a finite population mean is introduced. The relative performance of diagonal systematic sample mean along with those of the simple random and systematic sample means are assessed for certain labelled and natural populations. Some modifications to achieve further improvements over diagonal systematic sampling are also suggested.

Key words : Diagonal systematic sampling, Systematic sampling, Labelled populations, Linear trend, Trend free sampling.

1. Introduction

Consider a finite population $U = \{U_1, U_2, \ldots, U_N\}$ of N identifiable units. Let Y be a real variable with value Y_i, $i = 1, 2, \ldots, N$ giving a vector $Y = (Y_1, Y_2, \ldots, Y_N)$. The problem is to estimate the population mean $\overline{Y} = \frac{\Sigma Y_i}{N}$ on the basis of a sample selected from the population U. Any ordered sequence, $S = \{u_1, u_2, \ldots, u_n\} = (U_{i1}, U_{i2}, \ldots, U_{in})$, $1 \leq it \leq N$ and $1 \leq t \leq n$ is called a 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. However, if there exist a linear trend among the population units, the systematic sampling is recommended for selecting a sample of size n, which gives the best estimator compared to simple random sampling scheme. Further, the systematic sampling scheme is widely used in forest survey due to the simplicity of the scheme.

In the present study, we have introduced a modified systematic sampling scheme called diagonal systematic sampling with fixed sample size, which uses the knowledge of the labels of the population units to provide an unbiased estimate of the population mean. The proposed sampling scheme together its mean and its variance are discussed in Section 2. The explicit expressions for the variance of diagonal systematic sample means are obtained for certain labelled populations and are compared with those of simple random sampling and systematic sampling in Section 3. In Section 4, Yates type end corrections are suggested to achieve further improvements over diagonal systematic sampling. Section 5 is devoted to assess the relative performance of diagonal systematic sampling, simple random sampling and systematic sampling schemes for certain natural populations.

2. *Diagonal Systematic Sampling Scheme*

Let N (= kn, where $n \leq k$) be the population size. The population units $U_1, U_2, \ldots, U_n$ are arranged in an $n \times k$ 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 different columns. The steps involved in selecting a random sample of size n from diagonal systematic sampling scheme are as follows :

Step-1 : Arrange the N population units $U_1, U_2, \ldots, U_N$ in an $n \times k$ matrix.

Step-2 : Select a random number r such that $1 \leq r \leq k$.

Step-3 : The selected sampling units are

$$S_r = \left\{U_r, U_{(k+1)+r}, U_{2(k+1)+r}, \ldots, U_{(n-1)(k+1)+r}\right\} \text{ if } r \leq k-n+1$$

$$S_r = \left\{U_r, U_{(k+1)+r}, U_{2(k+1)+r}, \ldots, U_{t(k+1)+r\,=\,(t+1)k}, U_{(t+1)k+1}, U_{(t+2)k+2}, \ldots, U_{(n-1)k+(n-t-1)}\right\}$$

$$\text{if } r > k-n+1$$

where $0 \leq t \leq n-1$

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

$$S_i = \left\{y_{1r}, y_{2(r+1)}, \ldots, y_{n(r+n-1)}\right\} \quad i = 1, 2, \ldots, k$$

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

3. Relative Performance of Diagonal Systematic Sampling in the Presence of Linear Trend

It is well known that whenever the population consists solely of a linear trend, systematic sample mean is more efficient than simple random sample mean. For a detailed discussion on estimation of mean in the presence of linear trend, one may refer to Fountain and Pathak [2] and the references cited therein. In this section we have compared the efficiency of the diagonal systematic sample mean with the sample means obtained from simple random sampling and systematic sampling schemes for two hypothetical populations having linear trend.

3.1 Population with Linear Trend

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

$$Y_i = a + ib \qquad i = 1, 2, \ldots, N$$

For the above population with linear trend, the variances of the simple random sample mean ($\bar{y}_r$), systematic sample mean ($\bar{y}_{sy}$) and diagonal systematic sample mean ($\bar{y}_{dsy}$) are

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

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

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

The derivation of variance of diagonal systematic sample mean is given in the appendix. By comparing the above expressions, we have shown that the diagonal systematic sampling is more efficient than the simple random sampling and the systematic sampling. In fact

$$V(\bar{y}_{dsy}) \le V(\bar{y}_{sy}) \le V(\bar{y}_r)$$

Remark 3.1 : When $k = n$, $V(\bar{y}_{dsy}) = 0$. In this case, diagonal systematic sampling becomes a completely trend free sampling (See Mukerjee and Sengupta [3]).

3.2 Two Dimensional Population with Linear Trend

A two dimensional hypothetical population with linear trend may be represented by the relation (Cochran [1])

$$Y_{ij} = i + j \qquad i = 1, 2, \ldots, k \text{ and } j = 1, 2, \ldots, n$$

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

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

$$V(\bar{y}_{sy}) = \frac{(k-1)(k+1)}{12}$$

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

The derivation of variance of diagonal systematic sample mean is given in the appendix. By comparing the above expressions, we have shown that the diagonal systematic sampling is more efficient than the simple random sampling and the systematic sampling. In fact

$$V(\bar{y}_{dsy}) \leq V(\bar{y}_{sy}) \text{ for all values of k and n}$$

$$V(\bar{y}_{dsy}) \leq V(\bar{y}_r) \text{ iff } 2(N-1) \geq n(k-n)^2$$

Remark 3.2 : When $k = n$, $V(y_{dsy}) = 0$. In this case, diagonal systematic sampling becomes a completely trend free sampling (See Mukerjee and Sengupta [3]).

4. Some Modifications on Diagonal Systematic Sampling

It has been shown in Remark 3.1 and Remark 3.2 that the diagonal systematic sampling is a completely trend free sampling whenever there exists a perfect linear trend in the population and the population size $N = n^2$ where n is the sample size. When $N = kn$ and $k \neq n$, and there exists a perfect linear trend in the population, diagonal systematic sampling is not a completely trend free sampling. However in this case, improvement may be achieved by modifying the usual diagonal systematic sampling scheme, as in the case of systematic sampling.

4.1 Yates Type End Corrections

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

$$\bar{y}^{*}_{dsy} = \bar{y}_{dsy} + a(y_1 - y_n)$$

that is, the units selected first and last are given the weights $n^{-1} + a$ and $n^{-1} - a$ respectively, whereas the remaining units get the weight n^{-1}.

For a population with a linear trend defined in Section 3.1

$$a = \begin{cases} \dfrac{(2r - k + n - 2)}{N - k + n - 1} & \text{if } r \le k - n + 1 \\[2ex] \dfrac{(k - n)(2k - 2r + 2 - n)}{2n(N - 2k + n - 1)} & \text{if } r > k - n + 1 \end{cases}$$

where r is the random start.

In the presence of a perfect linear trend, the modified estimator $\bar{y}^{*}_{dsy}$ becomes the population mean $\bar{Y}$ and hence the variance $V(\bar{y}^{*}_{dsy}) = 0$.

For a two dimensional population with a linear trend defined in Section 3.2, the modified estimator is defined by

$$\bar{y}^{**}_{dsy} = \bar{y}_{dsy} + b(y_1 - y_n)$$

where

$$b = \begin{cases} \dfrac{(2r - k + n - 2)}{4(n - 1)} & \text{if } r \le k - n + 1 \\[2ex] \dfrac{(k - n)(2r - 2k - 2 + n)}{2n(k - 2n + 2)} & \text{if } r > k - n + 1 \end{cases}$$

where r is the random start.

In the presence of a perfect linear trend, the modified estimator $\bar{y}^{**}_{dsy}$ becomes the population mean $\bar{Y}$ and hence the variance $V(\bar{y}^{**}_{dsy}) = 0$.

5. Relative Performance of Diagonal Systematic Sampling for Certain Natural Populations

It has been shown in Section 3 that diagonal systematic sampling performs well, compared to simple random and systematic sampling schemes whenever

Table 5.1. Comparison of simple random, systematic and diagonal systematic sample means for 7 populations considered by Sukhatme *et al.* [4]

Population no.	k	n	$V(\bar{y}_r)$	$V(\bar{y}_{sy})$	$V(\bar{y}_{dsy})$
1	6	2	178.75*	263.45	290.62
	4	3	107.25	65.37	58.02*
2	6	2	107.10	79.14	56.89*
	4	3	64.26	21.43*	36.36
3	6	2	39.96	19.61	14.37*
	4	3	23.98	13.21*	32.08
4	6	2	71.62	56.16*	56.39
	4	3	42.97	21.80*	34.94
5	6	2	117.81	196.15	84.73*
	4	3	70.69	3.63*	18.74
6	6	2	72.65	16.67	8.92*
	4	3	43.59	24.05*	62.39
7	6	2	9.35	4.28*	11.04
	4	3	5.62	3.74	3.24*

Table 5.2. Comparison of simple random, systematic and diagonal systematic sample means for 7 populations considered by Sukhatme *et al.* [4] arranged in ascending order

Population no.	k	n	$V(\bar{y}_r)$	$V(\bar{y}_{sy})$	$V(\bar{y}_{dsy})$
1	6	2	178.75	83.03	44.39*
	4	3	107.25	37.18	1.91*
2	6	2	107.10	81.55	61.14
	4	3	64.26	26.59	3.64*
3	6	2	39.96	24.36	13.62*
	4	3	23.98	6.07	0.58*
4	6	2	71.62	54.32	40.31*
	4	3	42.97	21.07	4.94*
5	6	2	117.81	62.48	38.81*
	4	3	70.69	30.17	8.19*
6	6	2	72.65	48.58	27.67*
	4	3	43.59	16.68	3.22*
7	6	2	9.35	5.78	4.45*
	4	3	5.62	2.57	0.52*

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 diagonal systematic sampling for a set of natural populations considered by Sukhatme *et al.* [4]. The data were collected for estimating the catch of marine fish in a sample of landing centres on the Malabar coasts of India. At each landing centre, the number of boats landing every hour from 6 A.M. to 6 P.M. is counted for 7 consecutive Mondays. Assuming the data collected every Monday as a population of 12 elements, we obtain the variances of simple random sample mean, systematic sample mean and diagonal systematic sample mean for the cases k = 6, n = 2 and k = 4, n = 3. Table 5.1 presents these variances for the natural populations. It is seen that, out of 14 cases considered the systematic, diagonal systematic and simple random sampling estimators performs well in 7, 6 and 1 cases respectively. The minimum variance in each case is indicated by an asterisk in the table. Next, we arrange the 12 elements of each population in an ascending order, presuming that this produces a population with linear trend. In Table 5.2, we present the variances for the various sample means for these populations. It is seen now that the diagonal systematic sample mean is the most efficient and also

$$V(\bar{y}_{dsy}) \leq V(\bar{y}_{sy}) \leq V(\bar{y}_r)$$

in each case. The minimum variance in each case in indicated by an asterisk in the table.

REFERENCES

[1] Cochran, W.G. (1977). *Sampling Techniques.* 3rd Edn, John Wiley and Sons. New York.

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

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

[4] Sukhatme, P.V., Sukhatme, B.V., Sukhatme, S. and Asok, C. (1984). *Sampling Theory of Surveys with Applications*, Iowa State University Press, Ames, Iowa.

[5] Yates, F. (1948). Systematic Sampling, *Phil. Trans. Royal Soc.,* **A 241,** 345-377.

APPENDIX

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

Let N (= kn, where $n \leq k$) be the population size. The population units $U_1, U_2 \ldots, U_N$ are arranged in an $n \times k$ 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 different columns. As discussed earlier the selected diagonal systematic sampling units are :

$$S_r = \{U_r, U_{(k+1)+r}, U_{2(k+1)+r}, \ldots, U_{(n-1)(k+1)+r}\} \text{ if } r \leq k-n+1$$

$$S_r = \{U_r, U_{(k+1)+r}, U_{2(k+1)+r}, \ldots, U_{t(k+1)+r = (t+1)k}, U_{(t+1)k+1}, U_{(t+2)k+2}, \ldots, U_{(n-1)k+(n-t-1)}\} \text{ if } r > k-n+1$$

where $0 \leq t \leq n-1$

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

$$S_i = \{y_{1r}, y_{2(r+1)}, \ldots, y_{n(r+n-1)}\} \quad i = 1, 2, \ldots, k$$

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 the N population units are in arithmetical progression. That is

$$Y_i = a + ib \qquad i = 1, 2, \ldots, N$$

For the above population with linear trend, the diagonal systematic sample mean and the population mean are obtained as

$$\overline{Y}_{dsy} = \begin{cases} a + \{r + (n-1)(k+1)/2\}\, b & \text{if } r \le k-n+1 \\ a + \dfrac{1}{n}\left\{\dfrac{[kn(n-1) + k(k+1) + (k-n)(k-n+1)]}{2} - (k-n)\, r\right\} b & \text{if } r > k-n+1 \end{cases}$$

$$\overline{Y} = a + (N+1)\frac{b}{2}$$

The variance of the diagonal systematic sample mean ($\bar{y}_{dsy}$) is obtained as

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

After a little algebra, we can obtain

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

A.2 Two Dimensional Population with Linear Trend

A two dimensional hypothetical population with linear trend may be represented by the relation (Cochran [1])

$$Y_{ij} = i + j \qquad i = 1, 2, \dots, k \text{ and } j = 1, 2, \dots, n$$

For the above population with linear trend, the diagonal systematic sample mean and the population mean are obtained as

$$\overline{Y}_{dsy} = \begin{cases} n + r & \text{if } r \le k-n+1 \\ \dfrac{\{(k-n)^2 + k(n+1) + (k-n)\, r\}}{n} & \text{if } r > k-n+1 \end{cases}$$

$$\overline{Y} = \frac{(k+n+2)}{2}$$

The variance of the diagonal systematic sample mean ($\bar{y}_{dsy}$) is obtained as

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

After a little algebra, we can obtain

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