The Philippine Statistician
1990, Vol. 39, Nos. 1-2 & 3-4
pages 82-96

MODIFICATIONS IN SYSTEMATIC SAMPLING METHODS IN THE PRESENCE OF A TREND

by

M.C. Agrawal*
U.P. Statistical Center
Diliman, Quezon City
and

Nirmal Jain
Hindu College,
University of Delhi, Delhi

SUMMARY

Four systematic sampling methods - Yates' end correction, centred systematic sampling, modified systematic sampling and balanced systematic sampling - which eliminate the effect of a linear trend in the population have suitably been modified to remove the effect of a quadratic trend in the population. It has been shown that the centred systematic sampling provides the most efficient estimator in the sense that it maintains its superiority vis-a-vis the other competing estimators even if a random component is superimposed on the quadratic trend model. Further, we work out some appropriate corrections for centred systematic sampling (for k even), modified systematic sampling (for n odd) and balanced systematic sampling (for n odd) so as to make them applicable for eliminating the effect of a linear trend. These methods of systematic sampling are then compared under a suitable linear trend model superimposed with a random component.

1. INTRODUCTION

Consider a finite population of size N, the units of which are identified by the labels 1,2,...,N and ordered in increasing size of the label. It is assumed that the population size N is expressible as a product of the sample size n and some integer k, i.e. N = nk. The usual systematic design classifies N units of the population into $k \geq 2$ classes $S_1, S_2 \ldots, S_k$ where S_i is then randomly selected.

* On leave from the Department of Statistics, University of Delhi, Delhi (India).

The population mean $\bar{Y}_N = N^{-1} \sum_{i=1}^{N} y_i$ is estimated by $\bar{y}_{sy}$ if S_i is selected, where $\bar{y}_{sy}$ is the mean of y values for the n units in S_i. To improve the performance of systematic sampling in the presence of a linear trend, several methods have been proposed. We consider the following four methods:

(1) End correction method has been suggested by Yates (1948). This consists of giving the weight 1/n to all the sample values except the first and the last which receive weights

$$\frac{1}{n} + \frac{(2i-k-1)}{2(n-1)k} \text{ and } \frac{1}{n} - \frac{(2i-k-1)}{2(n-1)k},$$

respectively, if S_i is selected. The sample mean $\bar{y}_{sy}$ of S_i with these end corrections provides the exact population mean.

(2) Centred systematic sampling was proposed by Madow (1953). This method is to select the class $S_{\frac{(k+1)}{2}}$ if k is odd and to randomly select one of the two classes $S_{\frac{k}{2}}$ and $S_{\frac{(k+2)}{2}}$, the probability of selection for each being 1/2, if k is even. The estimator of $\bar{Y}_N$ is equal to $\bar{y}_{\frac{(k+1)}{2}}$ for k odd, and $\bar{y}_{k/2}$ or $\bar{y}_{\frac{(k+2)}{2}}$ for k even.

(3) Modified systematic sampling method of Singh et.al (1968) chooses pairs of units equidistant from the ends of the population. If it is the random start from 1 to k, then the units selected in ith cluster S_i' will be $\{i+(j-1)k,\ nk-(j-1)k-i+1\}$, $i + \frac{(n-1)}{2}k$, $j = 1,2,\ldots,(n-1)$ for n odd. The sample mean is exactly the population mean for n even.

(4) Balanced systematic sampling is due to Sethi (1965). When n is even the population is divided into $\frac{n}{2}$ groups of size 2k each, choosing two units equidistant from the end of each group. If i the random start between 1 and k then the units selected into the ith cluster S_i''

will be $\{i+2(j-1)k,\ 2jk-i+1\},\ j = 1,2,\ldots,\frac{n}{2}$ and when n is odd, the units selected in S_i" will be $\{i+2(j-1)k,\ 2jk-i+1),\ j = 1,2,\ldots,\frac{n}{2}$ and when n is odd, the units selected in S_i" will be $\{i+2(j-1)k,\ 2jk-i+1),\ i+(n-1)k,\ j=1,2,\ldots,\frac{n-1}{2}\}$. The sample mean is exactly the population mean for n even.

2. IMPROVED METHODS OF ELIMINATING LINEAR TREND

Neither the method of balanced systematic sampling suggested by Sethi (1965) nor the method of modified systematic sampling due to Singh et.al (1968) eliminates the effect of a linear trend when n is odd. The centred systematic sampling also fails to eliminate the linear trend when k is even. We suggest suitable corrections to remove these restrictive features in respect of modified, balanced and centred sytematic sampling in the sense that they eliminate the effect of the linear trend completely.

In balanced systematic sampling, if i is the random start then, for eliminating the linear trend for n odd, we give the weight 1 to all the sample observations (before taking the average) except the first observation i and the last observation $i+(n-1)k$ to which we assign the weights $(1+\eta)$ and $(1-\eta)$, respectively, where

$$\eta = \frac{(2i-k-1)}{2(n-1)k}$$

For eliminating a linear trend, modified systematic sampling for n odd is improved in the sense that all the sample values have weight unity (before division by n) except the random start i and observation $i + \frac{n-1}{2}k$ which receive the weights $(1+\eta')$ and $(1-\eta')$, respectively, where

$$\eta' = \frac{(2i-k-1)}{(n-1)k}.$$

Centred systematic sampling for k even can eliminate the linear trend completely if we assign the weights 1 to all sample observation (before division by n) except the first observation identified by label $\frac{k}{2}$ or $\frac{k+2}{2}$ and the last observation identified by label $\frac{(2n-1)k}{2}$ or $\frac{2(n-1)k+2}{2}$ to which the end corrections suggested by Yates (1948) for the first and the last observations are applied using $i = \frac{k}{2}$ or $\frac{k+2}{2}$ according as $\bar{y}_{\frac{k}{2}}$ or $\bar{y}_{\frac{k+2}{2}}$ is selected as an etimator of $\bar{Y}_N$.

For the cases considered above, the resulting sample means, say $\bar{y}^{(1)}$, $\bar{y}^{(2)}$, $\bar{y}^{(3)}$ and $\bar{y}^{(4)}$ under balanced, modified and centred sampling with $\frac{k}{2}$ and $\frac{k+2}{2}$, respectively, coincide with the population mean $\bar{Y}_N$.

Now we shall compare these modified sampling schemes for eliminating the linear trend under the superpopulation linear trend model

$$y_h = \mu + h\,\theta_1 + e_h \quad (h=1,2,\ldots,N) \qquad (2.1)$$

where μ and θ_1 are constants and the random component e_h has the properties

$$E(e_h) = 0,\ E(e_h\, e_k) = 0 \quad (h \neq k) \text{ and } E(e_h^2) = \sigma_e^2.$$

the symbol E denotes the expectation with respect to the superpopulation.

Let V(.) denote the variance of $\bar{y}^{(i)}$ (i=1,2,3,4). Now the average variances of $\bar{y}^{(i)}$ (i=1,2,3,4) under the model (2.1) with $E(e_h)^2 =$ are

$$EV(\bar{y}^{(1)}) = \frac{(k-1)}{nk}\lambda + \frac{(k^2-1)}{6n^2(n-1)^2k^2}\lambda \qquad (2.2)$$

$$EV(\bar{y}^{(2)}) = \frac{(k-1)}{nk}\lambda + \frac{2(k^2-1)}{3n^2(n-1)^2k^2}\lambda \qquad (2.3)$$

and

$$EV(\bar{y}^{(3)}) = EV(\bar{y}^{(4)}) = \frac{(k-1)}{nk}\lambda + \frac{\lambda}{2n^2(n-1)^2k^2} \qquad (2.4)$$

Bellhouse and Rao (1975) have compared the performance of modified systematic sampling (for n even), balanced systematic sampling (for n even), centred systematic sampling (for k odd) and Yates' end correction method which eliminate the linear trend in the population. In the light of the above discussion we can augment the conclusions arrived at by them and say the four systematic sampling methods (whatever be n or k) perform almost similarly under the model (2.1) with $E(e_h^2) =$ as (2.2), (2.3) & (2.4) could be approximated to $\frac{(k-1)}{nk}$ which is also the average variance derived by Bellhouse and Rao (1975).

3. MODIFIED METHODS OF ELIMINATING QUADRATIC TREND

Let us suppose that the values of the successive units of the population increase in accordance with a quadratic model

$$y_h = \mu + h\theta_1 + h^2\theta_2 \quad (h = 1,2,\ldots,N) \qquad (3.1)$$

where μ, θ_1 and θ_2 are constants. We would modify the four systematic sampling schemes (discussed in section 1) with a view to eliminating the quadratic trend in the population.

<u>Modified Yates' end Correction Method</u>: If we apply end correction method to remove the trend effect in the case of a population with quadratic trend modelled as (3.1) then we have to modify it for two cases: (i) n odd (ii) n even. We then propose suitable estimators under end correction method (for both the cases) in which all but some of the members of the sample have weight unity (before division by n). If i is the random start between 1 and k then the specified labels of observations, whose weights are different from unity, are shown alongwith their assigned weights in the following table (taking i as the random start):

Table 3.1

Case (i)

label	i	$i+\frac{(n-1)k}{2}$	$i+(n-1)k$
Weight	$(1+Z_1+\eta_1)$	$(1-2Z_1)$	$(1+Z_1-\eta_1)$

Case (ii)

label	i	$i+\frac{(n-2)k}{2}$	$i+(n-2)k.$	$i+(n-1)k$
Weight	$(1+Z_2+\eta_1)$	$(1-2Z_2)$	$(1+Z_2)$	$(1-\eta_1)$

where $$Z_1 = \frac{n\,[6\{i^2-i(k+1)\} + (k+1)\ (2k+1)]}{3(n-1)^2k^2}$$

$$Z_2 = \frac{n\,[6\{i^2-i(k+1)\} + (k+1)(k+1)]}{3(n-2)^2k^2}$$

and

$$\eta_1 = \frac{n(2i-k-1)}{2(n-1)k}\ .$$

The weighted sample means, say $\bar{y}_e^{(1)}$ and $y_e^{(2)}$ for cases (i) and (ii coincide .with the population mean $\bar{Y}_N$.

<u>Centred Systematic Sampling</u>: If we employ centred systematic sampling in order to eliminate the effect of a quadratic trend represetned by (3.1), we distinguish six cases (i) k odd, n odd; (ii) k odd, n even; (ii) k even, $S_{\frac{k}{2}}$ selected, n odd; (iv) k even, $S_{\frac{k}{2}}$ selected, n even; (v) k even, $S_{\frac{k+2}{2}}$ selected, n odd; (vi) k even, $S_{\frac{k+2}{2}}$ selected, n even.

All but some labels selected in the sample under the centred systematic sampling are assigned the weight 1 (before division by n) for all the six cases, and such labels as have weights different from 1 have been specified along with their assigned weights in the following table:

Table 3.2

Case (i)	Label	$\frac{k+1}{2}$	$\frac{nk+1}{2}$	$\frac{(2n-1)k+1}{2}$	
	Weight	$1+\phi_1$	$1-2\phi_1$	$1+\phi_1$	
Case (ii)	label	$\frac{k+1}{2}$	$\frac{(n-1)k+1}{2}$	$\frac{(2n-3)k+1}{2}$	
	Weight	$1+\phi_2$	$1-2\phi_2$	$1+\phi_2$	
Case (iii)	label	$\frac{k}{2}$	$\frac{nk}{2}$	$\frac{(2n-1)k}{2}$	
	Weight	$1+\phi_3+\eta_2$	$1-2\phi_3$	$1+\phi_3-\eta_2$	
Case (iv)	label	$\frac{k}{2}$	$\frac{(n-1)k}{2}$	$\frac{(2n-3)k}{2}$	$\frac{(2n-1)k}{2}$
	Weight	$1+\phi_4+\eta_2$	$1-2\phi_4$	$1+\phi_4$	$1-\eta_2$
Case (v)	label	$\frac{k+2}{2}$	$\frac{nk+2}{2}$	$\frac{(2n-1)k+2}{2}$	
	Weight	$1+\phi_5+\eta_2$	$1-2\phi_5$	$1+\phi_5-\eta_2$	
Case (vi)	label	$\frac{k+2}{2}$	$\frac{(n-1)k+2}{2}$	$\frac{(2n-3)k+2}{2}$	$\frac{(2n-1)k+2}{2}$
	Weight	$1+\phi_6+\eta_2$	$1-2\phi_6$	$1+\phi_6$	$1-\eta_2$

where

$$\phi_1 = \frac{n(k^2-1)}{6k^2(n-1)^2}, \quad \phi_2 = \frac{n(k^2-1)}{6k^2(n-2)^2}, \quad \phi_3 = \frac{n(k^2+2)}{6k^2(n-1)^2} = \phi_5$$

$$\phi_4 = \frac{n(k^2+2)}{6k^2(n-2)^2}, = \phi_6, \quad \eta_2 = \frac{n}{2k(n-1)}$$

The weighted sample means, say $\bar{y}_c^{\langle 1 \rangle}$, $\bar{y}_c^{\langle 2 \rangle}$, $\bar{y}_c^{\langle 3 \rangle}$, $\bar{y}_c^{\langle 4 \rangle}$, $\bar{y}_c^{\langle 5 \rangle}$ and $\bar{y}_c^{\langle 6 \rangle}$ under the cases (i), (ii), (iii), (iv), (v) and (vi) would yield a value equal to the population mean $\bar{Y}_N$.

<u>Balanced Systematic Sampling</u>: When the population exhibits a quadratic trend indicated in (3.1), we identify four cases:

(i) n even, $\frac{n}{2}$ even; (ii) n even, $\frac{n}{2}$ odd;

(iii) n odd, $\frac{n-1}{2}$ even; (iv) n odd, $\frac{n-1}{2}$ odd while

using balanced systematic sampling to eliminate the trend. In all the four cases (i), (ii), (iii) and (iv), we allot the weight unity (before division by n) to all but some members of the sample under balanced systematic sampling. These members of the sample having the weight other than unity are indicated along with the suggested weights in the following table (taking i the random start):

Table 3.3

Case					
Case(i)	**Label**	i	$i+\frac{(n-4)k}{2}$	$i+(n-4)k$	
	Weight	$1+\Psi_1$	$1-2\Psi_1$	$1+\Psi_1$	
Case(ii)	Label	i	$i+\frac{(n-2)k}{2}$	$i+(n-2)k$	
	Weight	$1+\Psi_2$	$1-2\Psi_2$	$1+\Psi_2$	
Case(iii)	Label	i	$1+\frac{(n-5)k}{2}$	$i+(n-5)k$	$i+(n-1)k$
	Weight	$1+\ {}_3+\eta_3'$	$1-2\Psi_3$	$1+\Psi_3$	$1-\eta_3$
Case(iv)	Label	i	$+\frac{(n-3)k}{2}$	$i+(n-3)k$	$i+(n-1)k$
	Weight	$1+\Psi_4+\eta_3'$	$1-2\Psi_4$	$1+\Psi_4$	$1-\eta_3$

where

$$\Psi_1 = \frac{2n}{3k^2(n-4)^2}[3i(2k-i+1) - (k+1)\ (2k+1)]$$

$$\Psi_2 = \frac{2n}{3k^2(n-2)^2}[3i(2k-i+1) - (k+1)\ (2k+1)]$$

$$\Psi_3 = \frac{1}{3(n-5)^2k^2}[(k+1)(3k+3-nk-2n)-6ni(n-2)(i-1-k)]$$

$$\Psi_4 = \frac{1}{3(n-3)^2k^2}[(k+1)(3k+3-nk-2n)\ -6ni(n-2)(i-1-k)]$$

and

$$\eta_3 = \frac{(2i-k-1)}{2(n-1)k} \quad \text{and} \quad \eta_3' = \frac{2i-k-1}{(n-1)k}$$

The weighted sample means, say, $\bar{y}_b^{(1)}$, $\bar{y}_b^{(2)}$, $y_b^{(3)}$ and $\bar{y}_b^{(4)}$ for the cases (i), (ii), (iii) and (iv) yield the population mean $\bar{Y}_N$.

Modified Systematic Sampling: If the population is expected to have a quadratic trend represented by (3.1) and if we use modified systematic sampling to eliminate the trend, then we distinguish four cases:

(i) n even, $\frac{n}{2}$; (ii) n even, $\frac{n}{2}$ odd; (iii) n odd, $\frac{n-1}{2}$; even (iv) n odd, $\frac{n-1}{2}$ odd. We allot the weight 1 to all but the following labels in the sample, under modified systematic sampling, for all the four cases (i), (ii), (iii) and (iv) (before taking average). If i is the random start from 1 to k, then the labels having weights other than 1 are exhibited along with their attached weights in the table given below:

Table 3.4

Case (i)	Label	i	$i+\frac{(n-4)k}{4}$	$i+\frac{(n-4)k}{2}$
	Weight	$1+\xi_1$	$1-2\xi_1$	$1+\xi_1$
Case (ii)	Label	i	$i+\frac{(n-2)k}{4}$	$i+\frac{(n-2)k}{2}$
	Weight	$1+\xi_2$	$1-2\xi_2$	$1+\xi_2$

Case (iii)					
	Label	i	$i+\frac{(n-5)k}{4}$	$i+\frac{(n-5)k}{2}$	$i+\frac{(n-1)k}{2}$
	Weight	$1+\xi_3+\eta'_3$	$1-2\xi_3$	$1+\xi_3$	$1-\eta'_3$

Case (iv)					
	Label	i	$i+\frac{(n-3)k}{4}$	$i+\frac{(n-3)k}{2}$	$i+\frac{(n-1)k}{2}$
	Weight	$1+\xi_4+\eta_3$	$1-2\xi_4$	$1+\xi_4$	$1-\eta'_3$

where

$$\xi_1 = \frac{2n}{3(n-4)^2k^2}[6i(nk+2k+2-2i) - (k+1)(3nk+2k+4)]$$

$$\xi_2 = \frac{2n}{3(n-2)^2k^2}[6i(nk+2k+2-2i) - (k+1)(3nk+2k+4)]$$

$$\xi_3 = \frac{2n}{3(n-5)^2k^2}[12(n-2)(i-i^2) + 6ik(n^2+n-4)$$
$$+(k+1)(nk-4n-3n^2k+6+6k)]$$

$$\xi_4 = \frac{2n}{3(n-3)^2k^2}[12(n-2)(i-i^2)+6ik(n^2+n-4)$$
$$+(k+1)(nk-4n3n^2k+6+6k)]$$

and

$$\eta'_3 = \frac{(2i-k-1)}{(n-1)k} \quad \text{and} \quad \eta_3 = \frac{2i-k-1}{2(n-1)k}$$

The weighted sample means for the cases (i), (ii), (iii) and (iv), say, $\bar{y}_{m(1)}$, $\bar{y}_{m(2)}$ $\bar{y}_{m(3)}$ and $y_{m(4)}$ will coincide with the population mean $\bar{Y}_N$.

4. A COMPARISON OF THE FOUR MODIFIED SYSTEMATIC SAMPLING SCHEMES

In this section, we would compare the four modified methods of systematic sampling (viz., the end correction method, centred syste-matic sampling, modified systematic sampling and balanced systematic sampling) which are now capable of eliminating the quadratic trend in the population.

If we assume the finite population to be a sample from a superpopulation, the quadratic trend model (3.1) could be express as

$$y_h = \mu + h\theta_1 + h^2\theta_2 + e_h \quad (h = 1,2,\ldots,N) \qquad (4.1)$$

where μ, θ_1 and θ_2 are constants and the random error e_h has the properties as stated in (2.1).

Subject to the model (4.1), the average variances under the four sampling schemes, for n odd, are obtained as

$$E\,V(\bar{y}_c^{(1)}) = (A+6\,B^2)\,\lambda,\ k \text{ odd}$$

$$E\,V(\bar{y}_c^{(3)}) = EV(\bar{y}_c^{(5)}) = (A+\frac{1}{2k^2(n-1)^2n^2} + \frac{(k^2+2)^2}{6k^4(n-1)^4})\,\lambda,\ k \text{ even}$$

$$E\,V(\bar{y}_e^{(1)}) = (A + B + \frac{72\,B^2(2k^2-3)}{5(k^2-1)})\,\lambda,$$

$$E\,V(\bar{y}_b^{(3)}) = (A+\frac{B}{n^2} + \frac{D}{(n-5)^4})\,\lambda,\quad \frac{n-1}{2} \text{ even}$$

$$E\,V(\bar{y}_b^{(4)}) = (A+\frac{B}{n^2} + \frac{D}{(n-3)^4})\,\lambda,\quad \frac{n-1}{2} \text{ odd}$$

$$E\,V(\bar{y}_m^{(3)}) = (A+\frac{B}{n^2} + \frac{D'}{(n-5)^4})\,\lambda,\quad \frac{n-1}{2} \text{ even}$$

$$E\,V(\bar{y}_m^{(4)}) = (A+\frac{B}{n^2} + \frac{D'}{(n-3)^4})\,\lambda\quad \frac{n-1}{2}$$

where

$$A = \frac{k-1}{nk}, \quad B = \frac{k^2 - 1}{6k^2(n-1)^2}$$

$$D = \frac{2(k^2-1)\,[(n-2)^2\,(k^2-4) + 5(k^2-1)]}{15k^4n^2}$$

and

$$D = \frac{8(k^2-1)}{15k^4}\,[15n^3(n-2)k^2 + (n-2)^2(19k^2-16)$$

$$+ 60(n-2)k^2 + 20(4k^2-1)].$$

The average variances of four sampling schemes, subject to model (4.1) with $E(e_h^2) = \lambda$, for n even, are as follows:

$$E\,V(\bar{y}_c^{(2)}) = (A + 6B'^2)\,\lambda\,, \quad k \text{ odd}$$

$$E\,V\,(\bar{y}_c^{(4)}) = EV(\bar{y}_c^{(6)}) = (A + \frac{1}{2k^2(n-1)^2n^2} + \frac{(k^2+2)^2}{6k^4\,(n-2)^4}\lambda, \quad k \text{ even}$$

$$E\,V\,(\bar{y}_e^{(2)}) = (A+B+ \frac{2(2k^2-3)\,B'^2}{5(k^2-1)})\,\lambda$$

$$E\,V\,(\bar{y}_b^{(1)}) = (A+ \frac{8\,F}{(n-4)^4})\lambda, \quad \frac{n}{2} \text{ even}$$

$$E\,V\,(\bar{y}_b^{(2)}) = (A + \frac{F}{(n-2)^4})\lambda, \quad \frac{n}{2} \text{ odd}$$

$$E\,V\,(\bar{y}_m^{(1)}) = (A+ \frac{F'}{(n-4)^4})\,\lambda\,, \quad \frac{n}{2} \text{ even}$$

and

$$EV\ (\bar{y}_m^{(2)}) = \left(A + \frac{F'}{(n-2)^4}\right) \lambda\ , \quad \frac{n}{2} \quad \text{odd}$$

where

$$F = \frac{8(k^2-1)(4k^2-1)}{15k^4}$$

$$F' = \frac{8(k^2-1)}{15k^4} \ [15(n^2-4)k^2 + 16\ (4k^2 - 1)]$$

and

$$B = \frac{(k^2-1)}{6k^2(n-2)^2}$$

A and B have the same value as defined above.

A close scrutiny of the above average variances worked out for four systematic sampling methods under the model (4.1) reveals the fact that the modified centred systematic sampling dominates the other competing sampling schemes in the sense that it provides the maximum efficiency. The next best sampling scheme is Yates' end correction method followed by balanced systematic sampling and modified systematic sampling in the decreasing order of their efficiency. This finding coupled with the work of Agrawal and Jain (1988) should be taken to mean that centred systematic sampling is the best bet in the presence of linear or quadratic trend underlying the population.

REFERENCES

Agrawal, M. C. and Jain, Nirmal (1988). "Comparison of some sampling strategies in the presence of a trend", Journal of the Indian Society of Agricultural Statistics, XL, 191-201.

Bellhouse, D.R. and Rao, J.N.K. (1975). "Systematic sampling in the presence of a trend", Biometrika, 62, 694-697.

Mada, W.G. (1953). "On the Theory of Systematic Sampling, III. Annals of Mathematical Statistics, 24, 101-106.

Sethi, V.K. (1965). "On optimum pairing of units", Sankhya, ser. B, 27, 315-320.

Singh, D., Jindal, K.K. and Garg, J.N. (1968). "Modified Systematic Sampling. Biometrika, 55, 541-546.

Yates, F (1968). "Systematic sampling", Phil. Trans. Roy. Soc., London, A241, 345-377.