V.8 184 V45

# CONTRIBUTION TO SYSTEMATIC SAMPLING

By

PADAM SINGH



CC

DISSERTATION SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF DIPLOMA IN AGRICULTURAL AND ANIMAL HUSBANDRY STATISTICS OF THE INSTITUTE OF AGRICULTURAL RESEARCH STATISTICS (I.C.A.R.) NEW DELHI

1969

## ACKNOWLEDGEMENTS

I have great pleasure in expressing my deepest sense of gratitude to Dr. D. Singh, Deputy Statistical Adviser, Institute of Agricultural Research Statistics (I.C.A.R.), New Delhi for his valuable guidance, constant encouragement and constructive criticism during the course of investigation and writing this thesis.

I am also thankful to Shri C. Ashok, Research Scholar at I.A.R.S. for his help and valuable discussions throughout the course of investigation.

My thanks are due to Dr. G.R. Seth, Director, Institute of Agricultural Research Statistics (I.C.A.R.) for providing necessary facilities in the Institute and to the Indian Council of Agricultural Research for providing financial assistance in the form of fellowship.

Padam Singh

( PADAM SINGH )

# CONTENTS

# CHAPTER I

## INTRODUCTION

Systematic sampling has got a nice feature that the whole sample is determined by the first unit selected. We know that the systematic sampling is operationally more convenient than simple random sampling and which at the same time ensures for each unit equal probability of inclusion in the sample. The technique of systematic sampling consists in selecting every k th unit starting with the unit corresponding to a number r chosen at random from 1 to resiprocal of sampling fraction. The random number r chosen from 1 to k is known as the random start and the constant k is termed, the sampling interval. A sample selected by this procedure is known as a systematic sample with a random start.

General considerations indicate that systematic sampling may except in certain special cases, give more accurate results than will be obtained from the same number of randomly located sampling points. This has been demonstrated by Madow and Madow (1944). Cochran (1946) has shown that for certain types of autocorrelated populations systematic sampling is even more efficient than stratified random sampling with one unit per stratum. Quenouille (1949) and Das (1950) following Cochran's general line of approach have independently extended Cochran's results to the two dimensional case. Yates (1948) has also drawn the conclusion that over a wide range of conditions systematic sampling of sequences and functions is more accurate than stratified random sampling. Systematic sampling has an intuitive appeal through spreading the sample evenly over the population. There are obvious advantages, for example, it is easier to take samples at fixed time intervals in sampling of the quality control type, economic times series or meteorological data, or at fixed space interval when sampling an agricultural crop or a forest area.

There are however, certain objections to systematic sampling. No valid estimate of the sampling error can be obtained from the samples themselves and only the approximate estimates of sampling error can be obtained. In some special cases of circular systematic sampling we can estimate the sampling error unbiasedly, e.g. for n = N+1 /2 . There are other disadvantages such as much less accurate results will be obtained if there are any periodicities in the population and the sampling interval is an integral multiple of the period. Also presence of trend in the material will substantially reduce the accuracy of the estimate based on systematic sampling. This does not effect much the accuracy of stratified and random sampling. End corrections proposed by Yates (1948) remove the effect of trend in the population. But this necessitates a loss of information. Centrally located systematic sampling also removes the effect of trend. But both the centrally located and sample with end corrections are biased in the sense that the mean of all the sample means does not be exactly equal to the population mean. Recently Singh, Jindal and Garg (1968) proposed modified systematic sampling which removes the effect of trend.

In the present investigation an alternative systematic sampling scheme is suggested which provides the estimate of the sampling error and at the same time ensures equal probability to all units for inclusion in the sample. A comparison is made among the usual systematic sampling, simple random sampling and modified systematic sampling. Modified systematic sampling is also extended to multi-stage designs. The use of information about the auxiliary variates is made for both the cases when information about the means for auxiliary variates is known and when it is estimated by a large sample. At the end an example is also taken to illustrate the efficiency of modified systematic sampling over simple random sampling and systematic sampling.

## CHAPTER II

## MODIFIED SYSTEMATIC SAMPLING

2. Suppose a sample of size $n$ is to be drawn from a finite population consisting of $N$ units. Defining two parameters $I$ and $J$ such that

$$(N - n)/I = J \quad \ldots \quad (2.1)$$

$$n = I + J \quad \ldots \quad (2.2)$$

select a random number from 1 to N and then select ( I + 1 ) consecutive units and then go on selecting one unit each time giving an interval ( I + 1 ). Let $r$ be the random number selected then the sample drawn is $y_r, y_{r+1}, y_{r+2}, \ldots, y_{r+I}, y_{r+(I+1)+I}, y_{r+I+2(I+1)}, \ldots, y_{r+I+(J-1)(I+1)}$. This method of selection of sample we will term as modified systematic sampling. This scheme is an alternative systematic sampling scheme in the sense that first unit selected determines the whole sample.

The relations (2.1) and (2.2) reduces to

$N = n + I(n - I)$, which imposes a restriction over N. A sample of size $n$ can be sampled only from a population having $n \geq 2(\sqrt{N+1} - 1)$

**Theorem**

Under the modified systematic sampling scheme the probabilities of inclusion of all pairs of units in the sample are given by

$\pi_{ii} = n/N$ for all $i$

$\pi_{ij} = I/N$ if $(i - j) = 1$ or $N-1$ ~~under mod N~~

$\pi_{ij} = (I-1)/N$ if $(i-j) = 2$ or $N-2$ "

.

.

$\pi_{ij} = 1/N$ if $(i-j) = I$ or $N-I$ "

$\pi_{ij}$ = J/N if ( i-j) = ( I + 1) or N-(I+1)

$\pi_{ij}$ = J-1 / N if ( i-j) = 2( I+1) or N-2(I+1)

.
.
.

$\pi_{ij}$ = 3/N if ( i-j) = ( J-2) ( I+1) or N-(J-2)(I+1)

$\pi_{ij}$ = 2/N otherwise

Proof:

Given a number u, $1 < u \leq n$, i th unit can occur as the u th unit in the sample in only once out of N possible samples, and since random number r is selected randomly from among 1 to N units, it is obvious that i th unit comes as the first unit, second unit, ..... n th unit each only once in one sample. Therefore out of N possible samples i th unit can occur in n samples showing $\pi_{ii} = \pi_i$ the inclusion probability of i th population unit is n/N for all i = 1,2,...,N.

Let $j = i + p$, $p \leq I$, given i th unit included in the sample j th unit can come as the i + p, i+p+1, .....I th unit corresponding to 1,2,....., (I-p+1) th positions of i th unit, showing $\pi_{ij}$ =(1+I-p)/N for $p \leq I$. Similarly for $j = i + (N-p)$, $\pi_{ij}$ =(1+I-p)/N.

Now let $j-i = q\,(I+1)$, $q \leq J-1$. If i th unit comes as 1,2,.....,I th unit than j th unit can be I+2, I+3,......,2I+1 th unit which is not included in the sample. If i th unit comes as (I+1), 2(I+1), ......( J-q+1 )(I+1), then j th unit can come as (q+1) (I+1), (q+2)(I+2),....., [ q + ( J-q+1) ] ( I+1) i.e. in J-q+1 possible samples showing $\pi_{ij}$ = ( J-q+1) /N.

Also let $j - i = I + \sigma$ where $0 \leq I+1$, then it can be seen very easily that in this case $\pi_{ij} = 2/N$.

Hence the proof.

## 2.1. Estimate and Variance

Let $y_1, y_2, \ldots, y_n$ be the values of n units drawn by modified systematic sampling. Considering the Horvitz Thompson estimator for estimating population mean

$$\hat{\bar{Y}}_{HT} = \frac{1}{N} \sum_{i=1}^{n} y_i / \pi_i \qquad \ldots\ldots \quad 2.1.1$$

Variance of above Horvitz Thompson estimator is given by

$$V(\hat{\bar{Y}}_{HT}) = E(\bar{Y}_{HT})^2 - \bar{Y}^2$$

$$= \frac{1}{N^2} \sum_s \left( \sum_{i=1}^{n} Y_i^2 / \pi_i^2 + \sum_{i}^{n} \sum_{i \neq i'}^{n} y_i y_{i'} / \pi_i \pi'_i \right) P(s) - \bar{Y}^2$$

$$= \frac{1}{N^2} \sum_{i=1}^{N} (y_i^2 / \pi_i^2) \sum_s P(s) + \frac{1}{N^2} \sum_{i=1}^{N} \sum_{i \neq i'}^{N} (y_i y'_i / \pi_i \pi'_i) \sum_s P(s) - \bar{Y}^2$$

where $\sum_s P(s)$ is the probabulity of inclusion of $U_i, U'_i$ together in the sample and is equal to $\pi_{ii'}$. After simplication above becomes

$$V(\hat{\bar{Y}}_{HT}) = \frac{1}{N^2} \left[ \sum^{N} \frac{1 - \pi_i}{\pi_i} y_i^2 + \sum_{i=1}^{N} \sum_{i \neq i'}^{N} (\pi_{ii'} - \pi_i \pi_{i'}) \frac{y_i \cdot y_{i'}}{\pi_i \pi_{i'}} \right]$$

$$= \frac{1}{N^2} \sum_{i=1}^{N} \sum_{i'>i}^{N} (\pi_i \pi_{i'} - \pi_{ii'}) \left( \frac{y_i}{\pi_i} - \frac{y_{i'}}{\pi_{i'}} \right)^2 \qquad \ldots (2.1.2)$$

Estimates of variance (2.1.2) suggested by Yates and Grundy (1953) is given by

$$\hat{V}_{YG}(\hat{\bar{Y}}_{HT}) = \frac{1}{N^2}\sum_{i=1}^{n}\sum_{i'>i}^{n} \frac{\pi_i\pi_{i'} - \pi_{ii'}}{\pi_{ii'}} \left(\frac{y_i}{\pi_i} - \frac{y_{i'}}{\pi_{i'}}\right)^2 \quad \ldots (2.1.3)$$

which is unbiased, since

$$\frac{1}{N^2} E \sum_{i=1}^{n}\sum_{i'>i}^{n} \frac{\pi_i\pi_{i'} - \pi_{ii'}}{\pi_{ii'}} \left(\frac{y_i}{\pi_i} - \frac{y_{i'}}{\pi_{i'}}\right)^2$$

$$= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{i'>i}^{N} \left(\frac{\pi_i\pi_{i'} - \pi_{ii'}}{\pi_{ii'}}\right)\left(\frac{y_i}{\pi_i} - \frac{y_{i'}}{\pi_{i'}}\right)^2 \pi_{ii'}$$

$$= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{i'>i}^{N} (\pi_i\pi_{i'} - \pi_{ii'})\left(\frac{y_i}{\pi_i} - \frac{y_{i'}}{\pi_{i'}}\right)^2$$

(2.1.1), (2.1.2), (2.1.3) in the case of modified systematic sampling take the following form

$$\hat{\bar{Y}}_{HT} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

$$V(\hat{\bar{Y}}_{HT}) = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{i>i}^{N} \left(1 - (N^2/n^2)\,\pi_{ii'}\right)(y_i - y_{i'})^2$$

and $$\hat{V}(\hat{\bar{Y}}_{HT}) = \frac{1}{N^2}\sum_{i=1}^{n}\sum_{i'>i}^{n} \left(\frac{1}{\pi_{ii'}} - \frac{N^2}{n^2}\right)(y_i - y_{i'})^2 \quad \ldots (2.1.4)$$

Hence in modified systematic sampling we are able to estimate the sampling error unbiasedly.

## 2.2. Modified Systematic Sampling For Multistage Design

Multistage design has been found to be very useful in practice and this procedure is being currently used in a number of surveys. We know that multi-stage sampling may be the only feasible procedure in a number of practical situations, where a satisfactory sampling frame of ultimate observational units is not readily available and the cost of obtaining such a frame is prohibitive or where the cost of locating and physically identifying the usu's is considerable. Modified systematic sampling can also be used in multistage designs. The procedure consists in drawing the sample by using the technique of modified systematic sampling in each stage. Here we will consider only two stage design and the extension to multistage design is simple and straightforward. Let

N be the number of psu's in the population

M be the number of ssu's in each psu of population

n be the number of psu's to be selected by modified systematic sampling.

m be the number of ssu's to be selected from each psu by modified systematic sampling.

Defining $n = I + J$, $N = I + J + IJ$

$$m = U + V, \quad M = U + V + UV \quad .$$

then the modified systematic two stage sampling procedure consists in selecting n units out of N by modified systematic sampling defined earlier and then from each psu selecting m ssu's out of M by modified systematic sampling.

Let $y_{ij}$ be the value of j th ssu of i th psu. Also $\pi_{ii'}$ be the inclusion probability of i and i' th psu's in the sample and $\lambda_{jj'}$ the inclusion probability of j and j' th ssu's in the sample. (Note: $\lambda_{(i)jj'} = \lambda_{jj'}$ ).

**Estimate and variance**

Estimate of population mean for two stage design is given

by $$\hat{\bar{Y}} = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{m}\sum_{j=1}^{m} y_{ij} \qquad \ldots (2.2.1)$$

$$V(\hat{\bar{Y}}) = E_1 V_2(\hat{\bar{Y}}) + V_1 E_2(\hat{\bar{Y}}) \qquad \ldots (2.2.2)$$

Now

$$V_2(\hat{\bar{Y}}) = \frac{1}{n^2}\sum_{i=1}^{n} V(\bar{y}_i \mid i)$$

From (2.1.2) putting

$$V(\bar{y}_i \mid i) = \sum_{j=1}^{M}\sum_{j'>j}^{M} \frac{1}{M^2}\left(1 - \frac{M^2}{m^2}\lambda_{jj'}\right)(y_{ij} - y_{ij'})^2$$

So $$E_1 V_2(\hat{\bar{Y}}) = \frac{1}{nNM^2}\sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{j'>j}^{M}\left(1 - \frac{M^2}{m^2}\lambda_{jj'}\right)(y_{ij} - y_{ij'})^2$$

$$\ldots .(2.2.3)$$

and

$$E_2(\hat{\bar{Y}}) = \sum_{i=1}^{n}\bar{Y}_i / n$$

So

$$V_1 E_2(\hat{\bar{Y}}) = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{i'>i}^{N}\left(1 - \frac{N^2}{n^2}\pi_{ii'}\right)(\bar{Y}_i - \bar{Y}_{i'})^2$$

$$\ldots .(2.2.4)$$

Hence

$$V(\hat{\bar{Y}}) = \frac{1}{nNM^2}\sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{j'>j}^{M}\left(1 - \frac{M^2}{m^2}\lambda_{jj'}\right)(y_{ij} - y_{ij'})^2$$

$$+ \frac{1}{N^2}\sum_{i=1}^{N}\sum_{i'>i}^{N}\left(1 - \frac{N^2}{n^2}\pi_{ii'}\right)(\bar{Y}_i - \bar{Y}_{i'})^2$$

The values of $\pi_{ii'}$, $\lambda_{jj'}$ can be obtained from the Theorem discussed earlier.

**Estimate of Variance**

Considering

$$\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{j'>j}^{m}\left(\frac{1}{\lambda_{jj'}}-\frac{M^2}{m^2}\right)(y_{ij}-y_{ij'})^2$$

$$E_2\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{j'>j}^{m}\left(\frac{1}{\lambda_{jj'}}-\frac{M^2}{m^2}\right)(h\,y_{ij}-y_{ij'})^2$$

$$=\sum_{i=1}^{n}\sum_{j=1}^{M}\sum_{j'>j}^{M}\left(1-\frac{M^2}{m^2}\lambda_{jj'}\right)(y_{ij}-y_{ij'})^2$$

So

$$E\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{j'>j}^{m}\left(\frac{1}{\lambda_{jj'}}-\frac{M^2}{m^2}\right)(y_{ij}-y_{ij'})^2$$

$$=\frac{n}{N}\sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{j'>j}^{M}\left(1-\frac{M^2}{m^2}\lambda_{jj'}\right)(y_{ij}-y_{ij'})^2$$

Hence

$$\text{Est. }\frac{1}{nNM^2}\sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{j'\,j}^{M}\left(1-\frac{M^2}{m^2}\lambda_{jj'}\right)(y_{ij}-y_{ij'})^2$$

$$=\frac{1}{n^2M^2}\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{j'>j}^{m}\left(\frac{1}{\lambda_{jj'}}-\frac{M^2}{m^2}\right)(y_{ij}-y_{ij'})^2 \qquad \ldots\ldots (2.2.6)$$

Considering

$$(\bar{Y}_i-\bar{Y}_{i'})^2=\bar{Y}_i^2+\bar{Y}_{i'}^2-2\bar{Y}_i\,\bar{Y}_{i'}$$

So $\text{Est. }(\bar{Y}_i-\bar{Y}_{i'})^2=\text{Est. }\bar{Y}_i^2+\text{Est. }\bar{Y}_{i'}^2-2\,\text{Est. }\bar{Y}_i\,\bar{Y}_{i'}$

$$\text{Est.}(\bar{Y}_i^2)=\hat{\bar{Y}}_i^2-\hat{V}(\hat{\bar{Y}}_i)$$

$$\text{Est.}\ (\overline{Y}_{i'}^{2}) = \hat{\overline{Y}}_{i'}^{2} - \hat{V}(\hat{\overline{Y}}_{i'})$$

$$\text{Est.}\ (\overline{Y}_i \times \overline{Y}_{i'}) = \hat{\overline{Y}}_i \ \hat{\overline{Y}}_{i'}$$

Hence

$$\text{Est.}\ (\overline{Y}_i - \overline{Y}_{i'})^2 = \hat{\overline{Y}}_i^{2} + \hat{\overline{Y}}_{i'}^{2} - 2\hat{\overline{Y}}_i \hat{\overline{Y}}_{i'} - \hat{V}(\hat{\overline{Y}}_i) - \hat{V}(\hat{\overline{Y}}_{i'})$$

$$= (\hat{\overline{Y}}_i - \hat{\overline{Y}}_{i'})^2 - \hat{V}(\hat{\overline{Y}}_i) - \hat{V}(\hat{\overline{Y}}_{i'})$$

$$= A_{ii'} \ (\text{say})$$

Now consider

$$E \sum_{i=1}^{n} \sum_{i'>i}^{n} \left( \frac{1}{\pi_{ii'}} - \frac{N^2}{n^2} \right) A_{ii'}$$

$$= E_1 \sum_{i=1}^{n} \sum_{i'>i}^{n} \left( \frac{1}{\pi_{ii'}} - \frac{N^2}{n^2} \right) (\overline{Y}_i - \overline{Y}_{i'})^2$$

$$= \sum_{i=1}^{N} \sum_{i'>i}^{N} \left( 1 - \frac{N^2}{n^2} \pi_{ii'} \right) (\overline{Y}_i - \overline{Y}_{i'})^2$$

Hence

$$\text{Est.}\ \frac{1}{N^2} \sum_{i=1}^{N} \sum_{i'>i}^{N} \left( 1 - \frac{N^2}{n^2} \pi_{ii'} \right) (\overline{Y}_i - \overline{Y}_{i'})^2$$

$$= \frac{1}{N^2} \sum_{i=1}^{n} \sum_{i'>i}^{n} \left( \frac{1}{\pi_{ii'}} - \frac{N^2}{n^2} \right) A_{ii'} \qquad \ldots\ldots (2.2.7)$$

Therefore the estimate of variance (2.2.5) is given by

$$\text{Est.}\ V(\hat{\overline{Y}}) = \frac{1}{N^2} \sum_{i=1}^{n} \sum_{i'>i}^{n} \left( \frac{1}{\pi_{ii'}} - \frac{N^2}{n^2} \right) A_{ii'}$$

$$+ \frac{1}{n^2 M^2} \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{j'>j}^{m} \left( \frac{1}{\lambda_{jj'}} - \frac{M^2}{m^2} \right) (y_{ij} - y_{ij'})^2$$

$$\ldots\ldots (2.2.8)$$

# CHAPTER III

## USE OF AUXILIARY VARIATES IN MODIFIED SYSTEMATIC SAMPLING

3. Unknown mean $\bar{Y}$ of a population can be estimated more efficiently by utilising an auxiliary variate X which is highly correlated with character under study and whose mean $\bar{X}$ is known. Utilising the information on auxiliary variate X we consider the following estimators.

3(a). Ratio estimate

Ratio estimate in case of modified systematic sampling is defined as

$$\hat{\bar{Y}}_R = \frac{\bar{y}_{MSy}}{\bar{x}_{MSy}} \times \bar{X} \qquad \ldots\ldots \quad (3.a.1)$$

Assuming $\bar{y}_{MSy} = \bar{Y} + \epsilon_O$

$$\bar{x}_{MSy} = \bar{X} + \epsilon \qquad \text{with } E(\epsilon_O) = E(\epsilon) = O,$$

where $\bar{y}_{MSy}$, $\bar{x}_{MSy}$ are the means of $y_1, y_2, \ldots, y_n$ and $x_1, x_2, \ldots, x_n$ drawn by modified systematic sampling. Variance of (3.a.1) is given by

$$V(\hat{\bar{Y}}_R) = V(\bar{y}_{MSy}) + R^2 V(\bar{x}_{MSy}) - 2R\, Cov(\bar{x}_{MSy}, \bar{y}_{MSy}) \qquad \ldots. \quad (3.a.2)$$

where $R = \bar{Y}/\bar{X}$

$$Cov(\bar{x}_{Msy}, \bar{y}_{Msy}) = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{i' \ne i}^{N} (1 - \pi_{ii'} N^2/n^2)(y_i - y_{i'}).(x_i - x_{i'})$$

and

$$V(\bar{x}_{MSy}) = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{i'>i}^{N} (1 - \pi_{ii'} N^2/n^2)(x_i - x_{i'})$$

and the estimate of (3.a.2) is given by

$$\hat{V}(\hat{\bar{Y}}_R) = \frac{1}{N^2}\sum_{i=1}^{n}\sum_{i'>i}^{n}\left(\frac{1}{\pi_{ii'}} - \frac{N^2}{n^2}\right)(y_i - y_{i'})^2$$

$$+ \frac{\hat{R}^2}{N^2}\sum_{i=1}^{n}\sum_{i'>i}^{n}\left(\frac{1}{\pi_{ii'}} - \frac{N^2}{n^2}\right)(x_i - x_{i'})^2$$

$$- \frac{2\hat{R}}{N^2}\sum_{i=1}^{n}\sum_{i'>i}^{n}\left(\frac{1}{\pi_{ii'}} - \frac{N^2}{n^2}\right)(y_i - y_{i'})(x_i - x_{i'})$$

where $\hat{R} = \bar{y}_{MSy} / \bar{x}_{MSy}$

Multivariate ratio estimate

When the information on p auxiliary variates is available we can use multivariate ratio estimate in the case of modified systematic sampling. Let y be the character under study and $X_1, X_2, \ldots, X_p$ be the p auxiliary variates. Defining

$$\hat{\bar{Y}}_{jR} = \frac{\bar{y}_{MSy}}{\bar{x}_{jMSy}} \cdot \bar{X}_j$$

where $\bar{x}_{jMSy}$ be the sample mean for j th auxiliary variate.

Also

$$\hat{\bar{Y}}_{jR} = \frac{\bar{Y} + \epsilon_0}{\bar{X}_j + \epsilon_j} \cdot \bar{X}_j$$

where $\bar{x}_{jMSy} = \bar{X}_j + \epsilon_j$

$$\text{So } \hat{\bar{Y}}_{jR} = \frac{\bar{Y}}{\bar{X}_j}(1 + \epsilon_0/\bar{Y})(1 - \epsilon_j/\bar{X}_j)\bar{X}_j$$

$$= \bar{Y}(1 + \epsilon_0/\bar{Y} - \epsilon_j/\bar{X}_j)$$

and hence

$$\text{Cov}\ (\hat{\bar{Y}}_{jR}, \hat{\bar{Y}}_{j'R}) = \bar{Y}^2 E\ (\epsilon_O/\bar{Y} - \epsilon_j/\bar{X}_j)(\epsilon_O/\bar{Y} - \epsilon_{j'}/\bar{X}_{j'})$$

$$= \bar{Y}^2 \left[ E\ \frac{\epsilon_O^2}{\bar{Y}^2} - \frac{E\ \epsilon_O \epsilon_j}{\bar{Y}\ \bar{X}_j} - \frac{E\ \epsilon_O \epsilon_{j'}}{\bar{Y}\ \bar{X}_{j'}} + E\ \frac{\epsilon_j \epsilon_{j'}}{\bar{X}_j \bar{X}_{j'}} \right]$$

$$= \bar{Y}^2 \left[ \frac{V(\bar{y}_{MSy})}{\bar{Y}^2} - \frac{\text{Cov}\ (\bar{y}_{MSy}, \bar{x}_{jMSy})}{\bar{Y}\ \bar{X}_j} - \frac{\text{Cov}\ (\bar{y}_{MSy}, \bar{x}_{j'MSy})}{\bar{Y}\ \bar{X}_{j'}} + \frac{\text{Cov}(\bar{x}_{jMSy}, \bar{x}_{j'MSy})}{\bar{X}_j\ \bar{X}_{j'}} \right]$$

Now denoting

$$V\ (\hat{\bar{Y}}_{jR}) = V_{jj}$$

and

$$\text{Cov}\ (\hat{\bar{Y}}_{jR}, \hat{\bar{Y}}_{j'R}) = V_{jj'},$$

We define the multivariate ratio estimate as

$$\hat{\bar{Y}}_{MR} = \sum_{j=1}^{p} w_j\ \bar{Y}_{jR}$$

where

$$\sum_{j=1}^{p} w_j = 1$$

The weights $w_j$'s are chosen in such a way that the variance of $\hat{\bar{Y}}_{MR}$ is minimum. Following the algebra of Olkin, the weights $w_j$'s are given by

$w_j = R_j / D$, where $R_j$ is the sum of all the elements in the j th row of the matrix $V^{-1}$ knowing

$$V = \begin{pmatrix} V_{11} & V_{12} & \cdots\cdots & V_{1p} \\ V_{21} & V_{22} & \cdots\cdots & V_{2p} \\ \vdots & & & \\ V_{p1} & V_{p2} & \cdots\cdots & V_{pp} \end{pmatrix}$$

and D is the sum of all the elements in the matrix $V^{-1}$, $V^{-1}$ being the inverse of V. The minimum variance of $\hat{Y}_{MR}$ is equal to 1/D.

3(b). Regression Estimate

Regression estimate in the case of modified systematic sampling is given by

$$\hat{Y}_{1R} = \bar{y}_{MSy} + \frac{Cov(\bar{x}_{MSy}, \bar{y}_{MSy})}{V(\bar{x}_{MSy})} \left[ \bar{X} - \bar{x}_{MSy} \right] \qquad \ldots\ldots (3.b.1)$$

and the variance of above estimate is given by

$$V(\hat{Y}_{1R}) = V(\bar{y}_{MSy}) \left[ 1 - \frac{Cov^2(\bar{x}_{MSy}, \bar{y}_{MSy})}{V(\bar{x}_{MSy}) V(\bar{y}_{MSy})} \right] \qquad \ldots (3.b.2)$$

The estimate of (3.b.2) is given by

$$\hat{V}(\hat{Y}_{1R}) = \hat{V}(\bar{y}_{MSy}) \left[ 1 - \frac{\widehat{Cov}^2(\bar{x}_{MSy}, \bar{y}_{MSy})}{\hat{V}(\bar{x}_{MSy}) \hat{V}(\bar{y}_{MSy})} \right]$$

when the knowldge of $Cov(\bar{x}_{MSy}, \bar{y}_{MSy})$ and $V(\bar{x}_{MSy})$ is lacking then regression estimator changes to

$$\hat{\bar{Y}}_{1R} = \bar{y}_{MSy} + \frac{\hat{Cov}(\bar{x}_{MSy}, \bar{y}_{MSy})}{\hat{V}(\bar{x}_{MSy})} [\bar{X} - \bar{x}_{MSy}]$$

Assuming the sampling error in estimating $\frac{Cov(\bar{x}_{MSy}, \bar{y}_{MSy})}{V(\bar{x}_{MSy})}$ to be negligible the variance and the estimate of variance remains same as that of (3.b.1.).

## Multivariate Regression Estimate

Let Y be the dependent variate and $X_1, X_2, \ldots, X_p$ be p auxiliary variates. Then the multivariate regression estimate with pre-assigned constant regression coef_ficients, of population mean $\bar{Y}$ be defined as

$$\hat{\bar{Y}}_{M1R} = \bar{y}_{MSy} + \sum_{k=1}^{p} \beta_{0k} (\bar{X}_k - \bar{x}_{kMSy})$$

$$= \bar{y}_{MSy} + \beta (\bar{X} - \bar{x})$$

where

$\beta = (\beta_{01}, \beta_{02}, \ldots \beta_{0p})'$ is the column vector of p constants.

$\bar{X} = (\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_p)'$ is column vector of p population means.

$\bar{x} = (\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_p)'$ is column vector of p sample means based on modified systematic sampling.

$\hat{\bar{Y}}_{M1R}$ is an unbiased estimate of $\bar{Y}$ and its variance is given by

$$V(\hat{\bar{Y}}_{MIR}) = E(\hat{\bar{Y}}_{MIR} - \bar{Y})(\hat{\bar{Y}}_{MIR} - \bar{Y})'$$

$$= \left[ V(\bar{y}_{MSy}) - 2\beta' \Sigma_{21} + \beta' \Sigma_{22} \beta \right]$$

where $\Sigma_{22}$ is the variance covariance matrix of regressors means $\bar{x}_{1MSy}, \bar{x}_{2MSy}, \ldots, \bar{x}_{pMSy}$ in population.

$\Sigma_{21}$ is covariance vector between regressand mean and regressors mean in population.

It then follows that the best value $\beta$ of $\beta$ which minimises $V(\hat{\bar{Y}}_{MIR})$ is given by

$$\beta = \Sigma_{22}^{-1} \Sigma_{21}$$

and the minimum variance by

$$V(\hat{\bar{Y}}_{MIR}) = \left[ V(\bar{y}_{MSy}) - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \right]$$

$$= V(\bar{y}_{MSy}) \left[ 1 - R'^2 \right]$$

where $$R'^2 = \frac{\Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}}{V(\bar{y}_{MSy})}$$

When the $\beta$'s are estimated from the sample, assuming sampling error in estimating $\beta$'s to be negligible above expressions remain same approximately.

3(c) So far we have assumed that the population mean $\bar{X}$ of auxiliary variate is known. However in most of the situations this is not the case always. In case knowledge of $\bar{X}$ is lacking we use theory of double sampling. The technique of double sampling consists in selecting a large sample of sized $n'$ by modified systematic sampling and observing only X and then taking a subsample of size n from n' by modified systematic sampling and observing Y. Denoting $\bar{y}'_{MSy}$, $\bar{x}'_{MSy}$ the sample means based on large sample and their variances respectively; & $\bar{x}'_{MSy}$, $\bar{x}_{MSy}$

$$V(\bar{x}'_{MSy}) = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{i'>i}^{N} \left(1 - \frac{N^2}{n'^2} \pi'_{ii'}\right)(x_i - x_{i'})^2$$

and $$V(\bar{x}_{MSy}) = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{i'>i}^{N} \left(1 - \frac{N^2}{n^2} \pi_{ii'}\right)(x_i - x_{i'})^2$$

where $\pi'_{ii'}$ be the inclusion probability of $U_i$ and $U_{i'}$ for the first sample n'

$\pi_{ii'}$ the inclusion probability of $U_i$ and $U_{i'}$ from the sample of size n.

(i) In case of double sampling ratio estimate is defined as

$$\hat{\bar{Y}}_{DR} = \frac{\bar{y}_{MSy}}{\bar{x}_{MSy}} \cdot \bar{x}'_{MSy}$$

and

$$V(\hat{\bar{Y}}_{DR}) = V(\bar{y}_{MSy}) + R^2 \left[ V(\bar{x}_{MSy}) - V(\bar{x}'_{MSy}) \right]$$
$$-2R \left[ Cov(\bar{x}_{MSy}, \bar{y}_{MSy}) - Cov(\bar{x}'_{MSy}, \bar{y}'_{MSy}) \right]$$

and

$$\hat{V}(\hat{\bar{Y}}_{DR}) = \hat{V}(\bar{y}_{MSy}) + \hat{R}^2 \left[ \hat{V}(\bar{x}_{MSy}) - \hat{V}(\bar{x}'_{MSy}) \right]$$
$$-2\hat{R} \left[ \hat{Cov}(\bar{x}_{MSy}, \bar{y}_{MSy}) - \hat{Cov}(\bar{x}'_{MSy}, \bar{y}'_{MSy}) \right]$$

where

$$\hat{V}(\bar{y}_{MSy}) = \frac{1}{N^2} \sum_{i=1}^{n} \sum_{i'>i}^{n} \left( \frac{1}{\pi_{ii'}} - \frac{N^2}{n^2} \right)(y_i - y_{i'})^2$$

$$\hat{V}(\bar{y}'_{MSy}) = \frac{1}{N^2} \sum_{i=1}^{n} \sum_{i' \neq i}^{n} \left( \frac{1}{\pi'_{ii'}} - \frac{N^2}{n'^2} \frac{\pi'_{ii'}}{\pi'_{ii'}} \right) (y_i - y_{i'})^2$$

and $\hat{V}(\bar{x}_{MSy})$ and $\hat{Cov}(\bar{x}_{MSy}, \bar{y}_{MSy})$ etc. defined similarly.

(ii) Régréssion estimate in case of double sampling is defined as

$$\hat{\bar{Y}}_{lR} = \bar{y}_{MSy} + \frac{Cov.(\bar{x}_{MSy}, \bar{y}_{MSy})}{V(\bar{x}_{MSy})} \left[ \bar{x}'_{MSy} - \bar{x}_{MSy} \right]$$

and approximate variance of above estimate is given by

$$V(\hat{\bar{Y}}_{DlR}) = \left[ 1 - \frac{Cov^2(\bar{x}_{MSy}, \bar{y}_{MSy})}{V(\bar{x}_{MSy}) V(\bar{y}_{MSy})} \right] V(\bar{y}_{MSy}) + V(\bar{y}'_{MSy}) \frac{Cov^2(\bar{x}_{MSy}, \bar{y}_{MSy})}{V(\bar{x}_{MSy}) V(\bar{y}_{MSy})}$$

with its estimate

$$\hat{V}(\bar{Y}_{DlR}) = \hat{V}(\bar{y}_{MSy}) - \frac{\hat{Cov}^2(\bar{x}_{MSy}, \bar{y}_{MSy})}{\hat{V}(\bar{x}_{MSy}) \hat{V}(\bar{y}_{MSy})} \times \left[ \hat{V}(\bar{y}_{MSy}) - \hat{V}(\bar{y}'_{MSy}) \right]$$

## CHAPTER IV

## RELATIVE EFFICIENCY OF MODIFIED SYSTEMATIC SAMPLING

4. As general comparison of modified systematic sampling with simple random sampling and usual systematic sampling is difficult to give , we will consider the efficiency in the case of the populations , in which the variance between the elements in any group of contiguous elements increases steadily as the size of the group increases. Various mathematical models may be constructed to represent this situation. For instance, we might consider that the elements $y_i$ are drawn from different populations, the populations changing in some regular manner with i. Alternatively, the $y_i$ may be assumed to belong to the same population, but to be serially correlated. We assume further that the serial correlation between $y_i$ and $y_{i+u}$ is some quantity $\rho_u$ which depends only on u. Then if $\rho_u$ is positive and is a monotone decreasing function of u, it may be expected from intution that the variance within the group of elements $y_i, y_{i+1}, \ldots y_{i+k}$ is a monotone increasing function of k.

4.1. The elements $y_i$, $i = 1, 2, \ldots, N$, are assumed to be drawn from a super-population in which

$$E(y_i) = \mu, \quad E(y_i - \mu)^2 = \sigma^2, \quad E(y_i - \mu)(y_{i+u} - \mu) = \sigma^2 \rho_u$$

where $\rho_u > \rho_v \geq 0$ whenever $u < v$.

Now we will discuss comparisons between systematic sampling and simple random sampling made by Cochran (1944).

If $\overline{Y}$ is the mean of a finite specified population, then it can be seen that

$$N \sum_{i=1}^{N} ( y_i - \overline{Y} )^2 = \sum_{i=1}^{N} \sum_{j}^{N} ( y_i - Y_j )^2 \qquad \ldots .( 4.1.1)$$

since there are ${}^{N}C_2$ possible pairs of values ( $y_i$, $y_j$ ) this gives

$$\sum_{i=1}^{N} ( y_i - Y )^2 = (N-1/2) \; E ( y_i - y_j )^2 = (N-1/2) \; E[ (y_i - \mu) - (y_j - \mu) ]^2$$

where E is taken over all the finite population. Expanding the quadratic and averaging over all finite populations, we get

$$E \sum_{i=1}^{N} ( y_i - Y )^2 = (N-1)\sigma^2 \left[ 1 - \frac{2}{N(N-1)} \sum_{u=1}^{N-1} (N-u) \rho_u \right] \qquad \ldots (4.1.2)$$

Expected sum of siquares within a specified systematic sample can be obtained just by replacing N by n and u by ku, since there are n elements in the sample and since the correlations between successive elements are $\rho_k$, $\rho_{2k}$, ...... instead of $\rho_1, \rho_2, \ldots$ This result is the same for each of the k systematic samples. Hence

$$E \text{ ( S.S. within systematic samples) } = k (n-1) \sigma^2 \left[ 1 - \frac{2}{n(n-1)} \sum_{u=1}^{n-1} (n-u) \rho_{ku} \right]$$

## 4.2. Average variance for a random sample

For a single finite population the variance of the mean of a simple random sample is

$$\sigma_{\bar{y}}^2 = \frac{N-n}{N-1} \; \frac{1}{Nn} \sum_{i=1}^{N} ( y_i - \overline{Y} )^2 \qquad \ldots .. \quad (4.2.1)$$

From (4.1.2) we obtain

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}(1 - n/N)\left[1 - \frac{2}{N(N-1)}\sum_{u=1}^{N-1}(N-u)\rho_u\right] \qquad \ldots(4.2.2)$$

## 4.3. Average variance for the systematic sample

If $\bar{y}_{sy}$ is the mean of a typical sample, the variance for a single finite population is

$$E(\bar{y}_{sy} - \bar{Y})^2 = (1/N)\left[n\sum_{i=1}^{N}(\bar{y}_{sy} - \bar{Y})^2\right] \qquad \ldots (4.3.1)$$

where the sum is taken over the k systematic samples. Since

Total S.S. = S.S. within systematic samples + S.S. between systematic samples

$$\sigma_{sy}^2 = \frac{N-k}{N}\sigma^2\left[1 - \frac{2}{N(N-1)}\sum_{u=1}^{N-1}(N-u)\rho_u\right]$$

$$- \frac{n-1}{n}\sigma^2\left[1 - \frac{2}{n(n-1)}\sum_{u=1}^{n-1}(n-u)\rho_{ku}\right] \qquad \ldots . (4.3.2)$$

This reduces to

$$\sigma_{sy}^2 = \frac{\sigma^2}{n}(1 - n/N)\left[1 - \frac{2}{N(k-1)}\sum_{u=1}^{N-1}(N-u)\rho_u + \frac{2k}{n(k-1)}\sum_{u=1}^{n-1}(n-u)\rho_{ku}\right] \qquad \ldots (4.3.3)$$

## 4.4. Average variance of modified systematic sample

Let $y_{ri}$ denotes the value of i th unit of r th modified systematic sample then it can be easily seen that

$$\sum_{r=1}^{N}\sum_{i=1}^{n}(y_{ri} - \bar{Y})^2 = n\sum_{i=1}^{N}(y_i - \bar{Y})^2 \qquad \ldots . (4.4.1)$$

Also

$$\sum_{r=1}^{N}\sum_{i=1}^{n}(y_{ri}-\bar{Y})^2 = \sum_{r=1}^{N}\sum_{i=1}^{n}(y_{ri}-\bar{Y}_r+\bar{Y}_r-\bar{Y})^2$$

$$= \sum_{r=1}^{N}\sum_{i=1}^{n}(y_{ri}-\bar{Y}_r)^2 + n\sum_{i=1}^{N}(\bar{Y}_r-\bar{Y})^2 \qquad \ldots\ldots (4.4.2)$$

or $$n\sum_{i=1}^{N}(y_i-\bar{Y})^2 = \sum_{r=1}^{N}\sum_{i=1}^{n}(y_{ri}-\bar{Y}_r)^2 + n\sum_{r=1}^{N}(\bar{Y}_r-\bar{Y})^2$$

or $$n\sum_{r=1}^{N}(\bar{Y}_r-\bar{Y})^2 = n\sum_{i=1}^{N}(y_i-\bar{Y})^2 - \sum_{r=1}^{N}\sum_{i=1}^{n}(y_{ri}-\bar{Y}_r)^2$$

or

$$\sum_{r=1}^{N}(\bar{Y}_r-\bar{Y})^2 = \sum_{i=1}^{N}(y_i-\bar{Y})^2 - \frac{1}{n}\sum_{r=1}^{N}\sum_{i=1}^{n}(y_{ri}-\bar{Y}_r)^2$$

or $$E(\bar{Y}_r-\bar{Y})^2 = \frac{1}{N}\sum_{i=1}^{N}(y_i-\bar{Y})^2 - \frac{1}{nN}\sum_{r=1}^{N}\sum_{i=1}^{n}(y_{ri}-\bar{Y}_r)^2 \qquad \ldots\ldots (4.4.3)$$

Considering $$\sum_{i=1}^{n}(y_{ri}-\bar{Y}_r)^2 = \frac{1}{n}\sum_{i=1}^{n}\sum_{j\ i}^{n}(y_{ri}-y_{rj})^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}\sum_{j\ i}^{n}\left[(y_{ri}-\mu)-(y_{rj}-\mu)\right]^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}\sum_{j\ i}^{n}\left[(y_{ri}-\mu)^2+(y_{rj}-\mu)^2 - 2(y_{ri}-\mu)(y_{rj}-\mu)\right]$$

$$= (2\sigma^2/n)\sum_{i=1}^{n}\sum_{j\ i}^{n}\left[1-\rho_{(r_i-r_j)}\right] \qquad \ldots (4.4.4)$$

Taking the expectation over all r we get

$$E(y_{ri}-\bar{Y}_r)^2 \quad E_r(1/n)\sum_{i=1}^{n}(y_{ri}-\bar{Y}_r)^2$$

$$=(\sigma^2/n)\frac{n(n-1)}{n}\left[1-\frac{2A}{n(n-1)}\right] \quad \text{where } A = \sum_{i=1}^{n}\sum_{j>i}\rho_{(j-i)}$$

$$=\left[\sigma^2 (n-1)/n\right]\left[1-\frac{2A}{n(n-1)}\right] \qquad \ldots .(4.4.5)$$

Hence the average variance of modified systematic sampling is given by

$$\sigma^2_{MSy} = \frac{N-1}{N}\sigma^2\left[1-\frac{2}{N(N-1)}\sum_{u=1}^{N-1}(N-u)\rho_u\right]$$

$$-\frac{n-1}{n}\sigma^2\left[1-\frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{j>i}^{n}\rho_{(j-i)}\right]$$

$$=\sigma^2(1-n/N)\left[1-\frac{2}{N(K-1)}\sum_{u=1}^{N-1}(N-u)\rho_u\right.$$

$$\left.+\frac{2k}{n(k-1)}\sum_{i=1}^{n}\sum_{j>i}^{n}\rho_{(j-i)}\right] \qquad \ldots .(4.4.6)$$

Now A can be expressed as the sum of following three sums

$$A = \sum_{i=0}^{I}\sum_{i'>i}^{I}\rho_{(i'-i)} + \sum_{j=2}^{J}\sum_{j'>j}^{J}\rho_{(I+1)(j'-j)} + \sum_{i=0}^{I}\sum_{j=2}^{J}\rho_{(I+1)j-(i+1)}$$

$$\ldots .(3\ 4.4.7)$$

We will consider dach of the above sums one by one.

$$\sum_{i=0}^{I}\sum_{i'>i}^{I}\rho_{(i'-i)} = \rho_1+\rho_2+\ldots\ldots+\rho_I$$

$$+\rho_1+\rho_2+\ldots\ldots+\rho_{I-1}$$

$$+\rho_1+\rho_2+\ldots\ldots+\rho_{I-2}$$

$$\vdots$$

$$+\rho_1+\rho_2$$

$$+\rho_1$$

$$= I\rho_1 + (I-1)\rho_2 + (I-2)\rho_3 + \ldots + 2\rho_{I-1} + \rho_I \qquad \ldots (4.4.8)$$

Similarly

$$\sum_{j=2}^{J} \sum_{j'>j}^{J} \rho_{(I+1)(J j'-j)} = (J-2)\rho_{I+1} + (J-3)\rho_{2(I+1)} + \ldots + \rho_{(J-2)(I+1)} \qquad \ldots (4.4.9)$$

and

$$\begin{aligned} \sum_{i=0}^{I} \sum_{j=2}^{J} \rho_{(I+1)j-(i+1)} = \; & \rho_{2(I+1)-1} + \rho_{3(I+1)-1} + \ldots + \rho_{J(I+1)-1} \\ & + \rho_{2(I+1)-2} + \rho_{3(I+1)-2} + \ldots + \rho_{J(I+1)-2} \\ & \quad \vdots \\ & + \rho_{2(I+1)-(I+1)} + \rho_{3(I+1)-(I+1)} + \ldots + \rho_{J(I+1)-(I+1)} \end{aligned}$$

$$\ldots\ldots (4.4.10)$$

Hence A can be put as

$$\sum_{u=0}^{I-1} (I-u)\,\rho_{u+1} + \sum_{u=1}^{J-2} (J-2-u)\,\rho_{u(I+1)} + \sum_{u=0}^{(J-2)(I+1)} \rho_{(I+1)+u}$$

$$= a_1 + a_2 + a_3$$

where $a_1 = \sum_{u=0}^{I-1} (I-u)\rho_{u+1}$, $a_2 = \sum_{u=1}^{J-2} (J-u-1)\rho_{u(I+1)}$

$$a_3 = \sum_{u=0}^{(J-2)(I+1)} \rho_{(I+1)+u} \qquad \ldots\ldots (4.4.111)$$

4.5. From above results it is evident that no general conclusion can be established about the relative efficiency of modified systematic sampling unless further assumptions are made on the form of the population. We shall consider the efficiency for the two types of populations where

(a) $\rho_u = 1 - u/L$ , where L is some positive constant

(b) $\rho_u = \rho^u$, where $\rho$ is less than 1

Case (a) <u>Linear correlogram</u>

Here $\rho_{u_I} = 1 - u/L_I$ ... (4.5.1)

Now, $a_1 = \sum_{i=1} \sum_{i'>i} \rho_{(i'-i)}$

$$= I(I+1)/2 - \frac{I(I+1)(I+2)}{6L} \quad \text{..... (4.5.2)}$$

$$a_2 = \sum_{j=2}^{J} \sum_{j'>j}^{J} \rho_{(I+1)(j'-j)}$$

for $J$ even $= (J-2)\rho_{I+1} + (J-3)\rho_{2(I+1)} \ldots\ldots + (J/2-1)\rho_{(J/2)(I+1)}$

$$+ (J/2-2)\rho_{N-(J/2+1)(I+1)} + \ldots \rho_{N-(J-2)(I+1)} \quad \text{..... (4.5.3)}$$

because the distance can not exceed N/2 in the circular type of systematic sampling. Putting $J = 2m$, above sum becomes

$$(2m-2)\rho_{(I+1)} + (2m-3)\rho_{2(I+1)} + \ldots\ldots + (m-1)\rho_{m(I+1)}$$

$$+ (m-2)\rho_{N-(m+1)(I+1)} + (m-3)\rho_{N-(m+2)(I+1)} + \ldots + \rho_{N-(2m-2)(I+1)}$$

which under model $\rho_u = 1 - u/L$ takes the form

$$(2m-2)\left[1 - \frac{I+1}{L}\right] + (2m-3)\left[1 - \frac{2(I+1)}{L}\right] + \ldots\ldots$$

$$+ (m-1)\left[1 - \frac{m(I+1)}{L}\right] + (m-2)\left[1 - \frac{N+I}{2L} + (I+1)/L\right]$$

$$+ (m-3)\left[1 - \frac{N+I}{2L} + \frac{2(I+1)}{L}\right] + \ldots\ldots\ldots$$

$$+ \left[1 - \frac{N+I}{2L} + \frac{(m-2)(I+1)}{L}\right] \quad \text{.... (4.5.4)}$$

as $N - (m+1)(I+1) = N - mI - m - I - 1$

$$= N - \frac{N-I}{2} - (I+1) = (N+I)/2 - (I+1)$$

Considering the sum

$$(2m-2)\left[1-\frac{(I+1)}{L}\right]+(2m-3)\left[1-\frac{2(I+1)}{L}\right]+\ldots\ldots$$

$$+(m-1)\left[1-\frac{m(I+1)}{L}\right]$$

$$=(2m-2)+(2m-3)+\ldots\ldots+(m-1)-\frac{(I+1)}{L}\left[(2m-2)+2(2m-3)+\right.$$

$$\left.3(2m-4)+\ldots\ldots\ldots+m(m-1)\right]$$

$$=(m/2)\left[m+2m-2\right]-\frac{(I+1)}{L}\left[\sum_{t=1}^{m}t(2m-t-1)\right]$$

$$=(m/2)(3m-3)-\frac{(I+1)}{L}\left[\frac{(2m-1)m(m+1)}{2}-\frac{m(m+1)(2m+1)}{6}\right]$$

$$=3m(m-1)/2-\frac{(I+1)}{L}\left[\frac{m(m+1)}{6}(6m-3-2m-1)\right]$$

$$=3m(m-1)/2-\frac{(I+1)}{L}\,\frac{2}{3}\,m(m-1)(m+1)\qquad\ldots\ldots\qquad(4.5.5)$$

Considering the sum

$$(m-2)\left[1-\frac{(N+1)/2-(I+1)}{L}\right]+(m-3)\left[1-\frac{(N+1)/2-2(I+1)}{L}\right]$$

$$+\ldots\ldots\ldots+\left[1-\frac{(N+1)/2-(m-2)(I+1)}{L}\right]$$

$$=\frac{(m-2)(m-1)}{2}\left[1-\frac{(N+1)}{2L}\right]+\frac{(I+1)}{L}\left[(m-2)+2(m-3)+\ldots(m-2).1\right]$$

$$=\frac{(m-1)(m-2)}{2}\left[1-\frac{(N+1)}{2L}\right]+\frac{(I+1)}{L}\left[\sum_{t=1}^{m-2}t(m-t-1)\right]$$

$$=\frac{(m-1)(m-2)}{2}\left[1-\frac{N+1}{2L}\right]$$

$$+(I+1)\left[\frac{(m-1)(m-2)(m-1)}{2}-\frac{(m-2)(m-1)(2m-3)}{6}\right]/L$$

$$= \frac{(m-1)(m-2)}{2} \left[ 1 - \frac{(N+I)}{2L} \right] + \frac{(I+1)}{L} m(m-1)(m-2)/6 \quad \ldots (4.5.6)$$

Adding (4.5.5) and (4.5.6) we get

$$a_2 = 3m(m-1)/2 - 2(I+1)\, m(m-1)(m+1) / 3L$$

$$+ \frac{(m-1)(m-2)}{2} \left[ 1 - \frac{(N+I)}{2L} \right] + (I+1)\, m(m-1)(m-2)/6L \quad \ldots (4.5.7)$$

Considering the sum

$$a_3 = \sum_{i=o}^{I} \sum_{j=2}^{J} \rho_{(I+1)j-(i+1)}$$

$$= \rho_{(I+1)} + \rho_{(I+2)} + \ldots\ldots\ldots + \rho_{J(I+1)-1} \quad \ldots (4.5.8)$$

Total number of terms in above sum = N-2I-1. Here we will consider both the cases when N is odd and N is even.

Case (i) N is even, then N -2I - 1 is odd.

Let N - 2I-1 = 2u + 1 , or u = (N-2I-2) /2 or u+1 = (N-2I) /2

As in the sampling scheme distance never exceeds N/2, above sum becomes

$$2\left[ \left(1 - \frac{I+1}{L}\right) + \left(1 - \frac{I+2}{L}\right) + \ldots\ldots + \left(1 - \frac{N-2I-2+2I}{2L}\right)\right]$$

$$+ 1 - \frac{(N-2I+2I)}{2L}$$

$$= 2\left[ \frac{N-2I-2}{2} - \frac{1}{L} \left[ (I+1)+(I+2)+\ldots\ldots+\left(I + \frac{N-2I-2}{2}\right)\right]\right]$$

$$+ (1 - N/2L)$$

$$= ( N-2I-2) + ( 1 - N/2L) -(2/L)\left[ \frac{2\ N-2I-2}{2 \times 2} \right]\left[ I+1 +(N-2I-2)/2 + 1\right]$$

$$= (N-2I-2) + ( 1 - N/2L) - (2/L)\ (N-2I-2)(I+1+ \frac{N-2}{2}\ )\ /\ 4$$

$$= (N-2I-1) - \frac{2}{L}\ \frac{(N-2I-2)}{4}\ \frac{(N+2I)}{2} - \frac{N}{2L}$$

$$= (N-2I-1) - (1/4L) \left[ (N-2I-2)(N+2I) + 2N \right]$$

$$= (N-2I-1) - \left[ N^2-2IN + 2IN - 4I^2 - 2N - 4I + 2N \right] / 4L$$

$$= (N-2I-1) - ( N^2 - 4I^2 - 4I ) /4L \qquad \ldots\ldots (4.5.9)$$

Case (ii) N is odd

i.e., N-2I-1 is even

let $N - 2I - 1 = 2v$ or $v = (N-2I-1)/2$

Now $e_3$ becomes

$$2\left[ (1 - \frac{I+1}{L}) + (1 - \frac{I+2}{L}) + \ldots\ldots\ldots + 1- ( \frac{N-2I-1}{2} +I )/L\right]$$

$$= 2 \left[ \frac{(N-2I-1)}{2} - \left[ (I+1)+(I+2)+\ldots\ldots + I + \frac{(N-2I-1)}{2} \right]/L\right]$$

$$=(N-2I-1) - \frac{2}{L}\left[ ( I+1) + I + \frac{(N-2I-1)}{2} \right] ( \frac{N-2I - 1}{2\times 2} )$$

$$= ( N-2I-1) - \frac{2}{4L} \left[ 2I + 1 + \frac{N - 2I-1}{2} \right] ( N -2I +-1 )$$

$$= ( N- 2I-1) - ( 4I + 2 + N -2I -1) ( N-2I - 1) / 4L$$

$$= (N-2I-1) - (N-2I-1) (N+2I + 1 ) / \rho\ 4L$$

$$= (N-2I -1) - (N^2-4I^2 -4I - 1 ) / 4L \qquad \ldots\ldots (4.5.10)$$

So 4.5.10 $\sim$ 4.5.9 is 1/4L.

Hence $e_3 = ( N-2I-1) - (N^2-4I^2 - 4I) / 4L$ approximately.

Substituting back $2m = J$ in 4,5.7 we get

$$a_2 = (J-1)(J-2)/2 - (J-2)(NJ-4I)/8L$$

Hence

$$a_1 + a_2 + a_3 = I(I+1)/2 - I(I+1)(I+2)/6L$$
$$+ (J-1)(J-2)/2 - (J-2)(NJ-4I)/8L$$
$$+ (N-2I-1) - (N^2-4I^2-4I)/4L$$
$$= (I^2+I+J^2-3J+2+2N-4I-2)/2 - I(I+1)(I+2)/6L$$
$$- (J-2)(NJ-4I)/8L - (N^2-4I^2-4I)/4L$$
$$= (n^2-2N+2N-n)/2 - I(I+1)(I+2)/6L - (J-2)(NJ-4I)/8L$$
$$- (N^2-4I^2-4I)/4L$$
$$= n(n-1)/2 - [3Nn^2 + I(4I^2 - 3NI+6N-12n+8)] / 24L \quad \ldots(3.5.11)$$

4.6. Considering $\sum_{u=1}^{N-1} (N-u)\rho_u$

There are (N-1) terms in above sum. When N is odd above sum is

$$= (N-1)\rho_1 + (N-2)\rho_2 \ldots\ldots\ldots + \frac{N+1}{2} \rho_{\frac{N-1}{2}}$$
$$+ \rho_1 + 2\rho_2 + \ldots\ldots\ldots + \frac{N-1}{2} \rho_{\frac{N-1}{2}} \quad \ldots (4.6.1)$$
$$= N[\rho_1 + \rho_2 + \ldots\ldots\ldots + \rho_{(N-1)/2}]$$

which under the model $\rho_u = 1 - u/L$ takes the form

$$N[1 - \frac{1}{L} + 1 - \frac{2}{L} + \ldots\ldots 1 - \frac{N-1}{2L}]$$
$$= N[\frac{N-1}{2} - \frac{1}{L}(1+2+\ldots\ldots(N-1)/2)]$$

$$= N\left[\frac{N-1}{2} - \frac{1}{2L}\left(\frac{N-1}{2}\right)\left(\frac{N-1}{2}+1\right)\right]$$

$$= N(N-1)/2 - N(N-1)(N+1) / 8L \;\; = N(N-1)/2 - N(N^2-1)/8L \quad \ldots\ldots (4.6.2)$$

Similarly when N is even $\Sigma\ (N-u)\,\rho_u$ is

$$(N-1)\rho_1 + (N-2)\rho_2 + \ldots\ldots + (N/2+1)\,\rho_{\frac{N}{2}-1} + (N/2)\,\rho_{\frac{N}{2}}$$

$$+ \rho_1 + 2\rho_2 + \ldots\ldots + (N/2-1)\,\rho_{\frac{N}{2}-1} \qquad \ldots\ldots (4.6.3)$$

$$= N\left[\rho_1 + \rho_2 + \ldots\ldots + (\rho_{\frac{N}{2}-1})\right] + (N/2)\,\rho_{\frac{N}{2}}$$

$$= N\left[1 - 1/L + 1-2/L + \ldots\ldots + 1-(N-2)/2L)\right] + (N/2)\left[1 - N/2L\right]$$

$$= N\left[\frac{N}{2} - 1 - \frac{1}{L}\left[1 + 2 + \ldots\ldots + \frac{N-2}{2}\right]\right] + (N/2)\,(1-N/2L)$$

$$= N(N/2-1) - N(N-2)(N)/8L + (N/2)(1 - N/2L)$$

$$= N(N-1)/2 - N(N^2-2N+2N)/8L$$

$$= N(N-1)/2 - N^3/8L \qquad \ldots\ldots (4.6.4)$$

So (4.6.2) and (4.6.4) are approximately same and we will take any one out of those two.

4.7. Considering

$$\sum_{u=1}^{n-1} (n-u)\,\rho_{ku}$$

$$= (n-1)\rho_k + (n-2)\rho_{2k} + \ldots\ldots + \rho_{(n-1)k} \qquad \ldots\ldots (4.7.1)$$

when n is odd above sum takes the form

$$(n-1)\rho_k + (n-2)\rho_{2k} + \ldots\ldots + [(n+1)/2]\,\rho_{\frac{n-1}{2}k}$$

$$+ \rho_k + 2\rho_{2k} + \ldots\ldots\ldots + [(n-1)/2]\,\rho_{\frac{n-1}{2}k}$$

$$= n\,(\rho_k + \rho_{2k} + \ldots\ldots + \rho_{\frac{n-1}{2}k}) \qquad \ldots (4.7.2)$$

which under the model $\rho_u = 1 - u/L$, takes the form

$$n\,[\,1 - \frac{k}{L} + 1 - \frac{1-2k}{L} + \ldots\ldots + 1 - \frac{(n-1)k/2}{L}\,]$$

$$= \frac{n(n-1)}{2} - \frac{nk}{L}\,[\,1 + 2 + \ldots\ldots + (n-1)/2\,]$$

$$= n(n-1)/2 - k(n-1)(n+1)\,n/8L \quad = n(n-1)/2 - k(n^2-1)n/8L$$

$$= n(n-1)/2 - k(n-1)(n+1)n\,/\,8L \qquad \ldots (4.7.3)$$

Similarly for n even above sum is

$$n(\rho_k + \rho_{2k} + \ldots\ldots + \rho_{(n/2-1)k}) + (n/2)\rho_{nk/2} \qquad \ldots (4.6.4)$$

$$= n\,[\,1 - \frac{k}{L} + 1 - \frac{2k}{L} + \ldots\ldots 1 - \frac{(n-2)k/2L}{L}\,] + (n/2)(1 - \frac{nk}{2L})$$

$$= n(n-1)/2 - nk\,(n-2)(n/2x2x2\,L) - n^2k\,/4L$$

$$= n(n-1)/n - n^2k/(n-2)/8L - n^2k\,/4L$$

$$= n(n-1)/2 - n^3k/8L \qquad \ldots (4.7.5)$$

So (4.7.3) and (4.7.5) are approximately same, and will take any one out yof these two.

Slu Substituting the values of $\sum_{u=1}^{n-1}(n-u)\rho_{ku}$, $\sum_{u=1}^{N-1}(N-u)\rho_u$ and $a_1 + a_2 + a_3$ we get

$$\sigma_{\bar{y}}^2 = \sigma^2(1-n/N)\frac{1}{n}\left[1-\frac{2}{N(N-1)}\left[\frac{N(N-1)}{2}-\frac{N(N-1)(N+1)}{8L}\right]\right]$$

$$= \sigma^2/(1-n/N)\frac{(N+1)}{4Ln}$$

$$\sigma_{sy}^2 = (\sigma^2/n)(1-n/N)\left[1-\frac{2}{N(k-1)}\left[\frac{N(N-1)}{2}-\frac{N(N-1)(N+1)}{8L}\right]\right.$$

$$\left.+\frac{2k}{n(k-1)}\left[\frac{n(n-1)}{2}-\frac{(n^2-1)nk}{8L}\right]\right]$$

$$= \frac{\sigma^2}{n}(1-n/N)\left[1-\frac{2}{Nn(k-1)}\left[\frac{nN(N-1)}{2}-\frac{N(n-1)nk}{2}\right.\right.$$

$$\left.\left.-\frac{nN(N^2-1)}{8L}+\frac{Nnk^2(n^2-1)}{8L}\right]\right]$$

$$= \frac{\sigma^2}{n}(1-n/N)\left[1-\frac{2}{Nn(k-1)}\left[\frac{N^2n-Nn-N^2n+N^2}{2}\right.\right.$$

$$\left.\left.-\frac{nN^3+N^2k-nN-N^3n}{8L}\right]\right]$$

$$= \frac{\sigma^2}{n}(1-n/N)\left[1-\frac{2}{Nn(k-1)}\left[\frac{N(N-n)}{2}-\frac{nN(1-k^2)}{8L}\right]\right]$$

$$= \frac{\sigma^2}{n}(1-n/N)\left[1-1+\frac{2nN(1+k)}{Nn(8L)}\right]$$

$$= \frac{\sigma^2}{n}(1-n/N)\left[\frac{1+k}{4L}\right] = \frac{\sigma^2}{n}(1-n/N)\left(\frac{N+n}{4nL}\right)$$

$$\sigma_{MSy}^2 = \frac{\sigma^2}{n}(1-n/N)\left[1-\frac{2k}{N(k-1)}\left[\frac{N(N-1)}{2}-\frac{N^3}{8L}\right]\right.$$

$$\left.+\frac{2N}{n(N-n)}\left[\frac{n(n-1)}{2}-\frac{1}{24L}\left[3Nn^2+I(4I^2-3NI+6N-12n+8)\right]\right]\right]$$

$$= \frac{\sigma^2}{n}(1-n/N)\left[1-\frac{2n}{N-n}\left[\frac{N-1}{2}-\frac{N^2}{8L}\right]\right.$$

$$\left.+\frac{2N}{n(N-n)}\left[\frac{n(n-1)}{2}-\frac{1}{24L}\left[3N^2n^2+I(4I^2-3NI+6N-12n+8)\right]\right]\right]$$

$$= \frac{\sigma^2}{n}(1-n/N)\Big[1 - \frac{2}{n(N-n)}\Big[\frac{-nN(n-1)}{2} + \frac{n^2(N-1)}{2}$$

$$+ \frac{n^2N^2}{8L} - \frac{N^2n^2}{8L} + \frac{N}{24L}I(4I^2-3NI+6N-12n+8)\Big]\Big]$$

$$= \frac{\sigma^2}{n}(1-n/N)\Big[1 - \frac{1}{n(N-n)}\Big[nN-n^2 + \frac{2IN}{24L}(4I^2-3NI+6N-12n+8)\Big]\Big]$$

$$= \frac{\sigma^2}{n}(1-n/N)\Big[\frac{-IN}{12\,nIJL}(4I^2-3NI+6N-12n+8)\Big]$$

$$= \frac{\sigma^2}{n}(1-n/N)\frac{-(4I^2+6N-3NI-12n+8)N}{12n(n-I)L}$$

$$= \frac{\sigma^2}{n}(1-n/N)\Big[\frac{N+1}{4L} - \frac{1}{4L} + \frac{\{I^2+3I^2(J+1)+6n-8-3n^2\}}{12n(n-I)L}\Big]$$

So it can be seen that the modified systematic sampling is almost always more efficient than simple random sampling for the population with linear correlogram $\rho_u = 1-u/L$ and the relative gain in efficiency is

$$\Big[-\frac{1}{4L} + \frac{N\{3I^2(J+1)+6n-8+I^2-3n^2\}}{12\,n\,J\,L}\Big]\cdot\frac{4L}{N+1}$$

$$= \frac{1}{N+1} + \frac{\{3I^2(J+1)+6n-8+I^2-3n^2\}N}{3nJ(N+1)}$$

When n increases I and J also increases with the result the relative gain in efficiency also increases. However for $I=J=n/2$ relative gain in efficiency is

$$\frac{1}{N+1} - \frac{\frac{3n^2}{4}(n+2)/2+6n-8+\frac{n^2}{4}-3n^2}{\frac{3n^2}{2}\frac{(N+1)}{N}}$$

which is a function increasing with n. (up to n=4)

*Comparing the efficiency of modified systematic sampling with* usual systematic sampling, the relative gain in efficiency is

$$+ \frac{2N}{24nJL} \frac{(4I^2 - 3NI + 6N - 12n + 8)}{(N+n)} \cdot 4nL$$

$$= + \frac{N}{3J(N+n)} (4I^2 - 3NI + 6N - 12n + 8)$$

From above nothing can be concluded about the relative efficiency of modified systematic sampling over systematic sampling. However for $I = J = n/2$ we have relative gain in efficiency

$$= + \frac{N}{3n(N+n)/2} \left[ 4n^2 - 6n + 8 - 3n^3/8 \right]$$

which is positive for $n \leq 12$ and negative for $n > 2.10$.

Therefore, we conclude that for the populations with linear correlogram modified systematic sampling is more efficient than simple random sampling and even sometimes more efficient than the usual systematic sampling.

Case (b) : Now we will take the case of exponential correlogram where $\rho_u = \rho^u$

$$\sum_{u=1}^{N-1} (N-u)\rho_u = N(\rho_1 + \rho_2 + \ldots\ldots + \rho_{(N-1)/2})$$

$$= N\left[ (\rho + \rho^2 + \ldots\ldots\ldots + \rho^{(N-1)/2} \right]$$

$$= N \left[ \frac{\rho - \rho^{(N-1)/2}}{1 - \rho} \right]$$

$$\Sigma (n-u)\rho_{uk} = n \frac{\left[ \rho - \rho^{(n-1)k/2} \right]}{1 - \rho^{k}}$$

Also

$$a_1 = I\rho + (I-1)\rho^2 + \ldots\ldots + \rho^I$$

$$\rho a_1 = \quad I\rho^2 + \ldots\ldots\ldots + 2\rho^I + \rho^{I+1}$$

Subtracting

$$a_1 (\rho-1) = \rho^2 + \rho^3 + \ldots\ldots + \rho^I + \rho^{I+1} - I\rho$$

$$= \rho^2 \left[ 1 + \rho + \rho^2 + \ldots\ldots + \rho^{I-1} \right] - I\rho$$

$$= \rho^2 \frac{1 - \rho^{I-1}}{1 - \rho} - I\rho$$

or

$$a_1 = \frac{\rho^2}{\rho - 1} \frac{(1 - \rho^{I-1})}{1 - \rho} - \frac{I\rho}{\rho - 1}$$

$$= \frac{I\rho}{1 - \rho} - \frac{\rho^2 (1 - \rho^{I-1})}{(1 - \rho)^2}$$

$$a_2 = (2m-2)\, \rho_{(I+1)} + 2m-3)\, \rho_{2(I+1)} + \ldots\ldots\ldots + (m-1)\, \rho_{m(I+1)}$$

$$+ (m-2)\, \rho_{\frac{N+I}{2} - (I+1)} + (m-3)\, \rho_{\frac{N+I}{2} - 2(I+1)} + \ldots\ldots$$

$$= (2m-2)\, \rho^{I+1} + (2m-3)\, \rho^{2(I+1)} + \ldots\ldots + (m-1)\, \rho^{m(I+1)}$$

$$+ m-2)\, \rho^{\frac{N+I}{2} - (I+1)} + (m-3)\, \rho^{\frac{N+I}{2} - 2(I+1)} + \ldots\ldots$$

$$+ \rho^{\frac{N+I}{2} - (m-2)(I+1)}$$

Now considering the sum

$$a_{21} = (2m-2)\,\rho^{I+1} + (2m-3)\,\rho^{2(I+1)} + \ldots + (m-1)\,\rho^{m(I+1)}$$

$$a_{21}\,\rho^{I+1} = \qquad (2m-2)\,\rho^{2(I+1)} + \ldots + (m)\rho^{m(I+1)} + (m-1)\rho^{(m+1)(I+1)}$$

$$\text{or}\quad a_{21} = (\rho^{I+1}-1) = \rho^{2(I+1)} + \rho^{3(I+1)} + \ldots + \rho^{m(I+1)} + (m-1)\,\rho^{(m+1)(I+1)} - (2m-2)\,\rho^{I+1}$$

$$= \rho^{2(I+1)} \left[ 1 + \rho^{I+1} + \rho^{2(I+1)} + \ldots + \rho^{(m-2)(I+1)} \right] + (m-1)\rho^{(m+1)(I+1)} - (2m-2)\,\rho^{I+1}$$

$$= \rho^{2(I+1)} \left[ \frac{1-\rho^{(m-2)(I+1)}}{1-\rho^{I+1}} \right] + (m-1)\rho^{(m+1)(I+1)} - (2m-2)\rho^{I+1}$$

$$\text{or}\quad a_{21} = \frac{\rho^{2(I+1)}}{\rho^{I+1}-1} \left[ \frac{1-\rho^{(m-2)(I+1)}}{1-\rho^{I+1}} \right] + \frac{(m-1)\rho^{(m+1)(I+1)}}{\rho^{I+1}-1} - \frac{(2m-2)\rho^{I+1}}{\rho^{I+1}-1}$$

Similarly

$$a_{22} = (m-2)\,\rho^{\frac{N+I}{2}-(I+1)} + (m-3)\,\rho^{\frac{I+N}{2}-2(I+1)} + \ldots + \rho^{\frac{N+I}{2}-(m-2)(I+1)}$$

$$\rho^{-(I+1)} a_{22} = \qquad (m-2)\rho^{\frac{N+I}{2}-2(I+1)} + \ldots + \rho^{\frac{N+I}{2}-(m-3)(I+1)}$$

By subtracting, we get

$$a_{22}\,(\rho^{-(I+1)}-1) = \rho^{\frac{N+I}{2}-2(I+1)} + \rho^{\frac{N+I}{2}-3(I+1)} + \ldots + \rho^{\frac{N+I}{2}-(m-3)(I+1)} - (m-2)\rho^{\frac{N+I}{2}-(I+1)}$$

$$= \rho^{\frac{N+I}{2} - (I+1)} \left[ 1 + \rho^{-(I+1)} + \rho^{-(I+1)2} + \ldots\ldots + \rho^{-(m-2)(I+1)} \right]$$

$$- (m-2)\rho^{\frac{N+I}{2} - (I+1)}$$

$$a_{22} = \frac{\rho^{\frac{N+I}{2} - (I+1)}}{\rho^{-(I+1)} - 1} \left[ \frac{\rho^{-(I+1)} - \rho^{-(m-2)(I+1)}}{1 - \rho^{-(I+1)}} - (m-2) \right]$$

Hence

$$a_2 = a_{21} + a_{22} = \frac{(2m-2)\,\rho^{I+1}}{1 - \rho^{I+1}} - \frac{(m-1)\,\rho^{(m+1)(I+1)}}{1 - \rho^{I+1}}$$

$$- \rho^{2(I+1)} \left[ \frac{1 - \rho^{(m-2)(I+1)}}{1 - (1 - \rho^{I+1})^2} \right]$$

$$+ \frac{\rho^{\frac{N+I}{2} - (I+1)}}{1 - \rho^{-(I+1)}} \left[ (m-2) - \frac{\rho^{-(I+1)} - \rho^{-(m-2)(I+1)}}{1 - \rho^{-(I+1)}} \right]$$

Also

$$a_3 = \rho_{(I+1)} + \rho_{(I+2)} + \ldots\ldots + \rho_{J(I+1)-1}$$

$$= 2 \left[ \rho_{(I+1)} + \rho_{(I+2)} + \ldots\ldots + \rho_{(I+\frac{N-2I-1}{2})} \right]$$

$$= 2 \left[ \rho^{I+1} + \rho^{I+2} + \ldots\ldots \rho^{I + \frac{(N-2I-1)}{2}} \right]$$

$$= 2 \left[ 1 + \rho + \rho^2 + \ldots + \rho^{\frac{N-2I-1}{2} - 1} \right] \rho^{I+1}$$

$$= 2\rho^{I+1} \left[ \frac{1 - \rho^{\frac{N-2I-1}{2} - 1}}{1 - \rho} \right]$$

Aftering ignoring the terms of order higher than $\rho^2$ we get

$$a1 = I\rho + (I-1)\rho^2, \quad a_2 = 0, \quad a_3 = 0$$

$$\sum_{u=1} (N-u)\rho_u = N\rho(1+\rho), \quad \sum_{u=1}^{n-1} (n-u)\rho_{kn} = 0$$

Therefore under exponential model $\rho_u = \rho^u$

$$\sigma_r^2 = \frac{\sigma^2}{n}(1-n/N)\left[1-\frac{2\rho(1+\rho)}{N-1}\right]$$

$$\sigma_{sy}^2 = \frac{\sigma^2}{n}(1-n/N)\left[1-\frac{2\rho(1+\rho)}{k-1}\right]$$

$$\sigma_{MSy}^2 = \sigma^2(1-n/N)\left[1-\frac{2\rho(1+\rho)}{k-1} - \frac{2k}{n(k-1)}\left[I\rho+(I-1)\rho^2\right]/n\right]$$

$$\sigma_{MSy}^2 = (1-n/N)\frac{\sigma^2}{n}\left[1-\frac{2}{N(N-1)}N\rho(1+\rho)\right]$$

$$+\sigma^2/n)\left[2\rho(1+\rho)\left[\frac{n-1}{N-1}-\frac{I}{n}\right]+2\rho^2/n\right]$$

Comparing the efficiency of modified systematic sampling with systematic sampling for the population with exponential correlogram we conclude that modified systematic sampling will be less efficient then systematic sampling unless $\rho$ is negative and the relative loss in efficiency is

$$\frac{2k}{n(k-1)}\left[I\rho+(I-1)\rho^2\right] \times \frac{1}{1-\frac{2\rho(1+\rho)}{k-1}}$$

$$= 2\left[I\rho+(I-1)\rho^2\right]/n \qquad \text{for large } N.$$

which decreases when n increases.

Comparing the efficiency of modified systematic sampling with simple random sampling we conclude that modified systematic sampling will be more efficient than simple random sampling when $\frac{n-1}{N-1} \geq \frac{I}{n}$ ($\rho$ positive) i.e. the fraction of sample to be clustered is less than equal to the sampling fraction approximately.

# CHAPTER IV (b)

## RELATIVE EFFICIENCY OF MODIFIED SYSTEMATIC SAMPLING IN TERMS OF INTRA-CLASS CORRELATIONS

4.8 . Estimation of population total for modified systematic scheme is given by

$$\hat{Y} = \sum_{i=1}^{n} y_i / w_i$$

$$= \frac{N}{n} \sum_{i=0}^{I} y_{r+i} + \frac{N}{n} \sum_{j=2}^{J} y_{r+(I+1)j-1}$$

$$V(\hat{Y}) = \frac{N^2}{n^2} \Big[ V\Big(\sum_{i=0}^{I} y_{r+i}\Big) + V\Big(\sum_{j=2}^{J} y_{r+(I+1)j-1}\Big)$$

$$+ 2\,Cov\Big(\sum_{i=0}^{I} y_{r+i}, \sum_{j=2}^{J} y_{r+(I+1)j-1}\Big) \Big]$$

Now

$$\frac{N^2}{n^2} V\Big(\sum_{i=0}^{I} y_{r+i}\Big) = \frac{N^2}{n^2} (I+1)^2 \frac{N-1}{N} \frac{[N(I+1)-1]\, S^2}{(I+1)(N-1)(I+1)} [1+I\rho_c]$$

$$= \frac{N^2}{n^2} \frac{N(I+1)-1}{N} S^2 [1+I\rho_c]$$

where $\rho_c$ is the intraclass correlation of I consecutive units.

Also

$$\frac{N^2}{n^2} V\Big[\sum_{j=2}^{J} y_{r+(I+1)j-1}\Big] = \frac{N^2}{n^2} (J-1)^2 \frac{(I+1)(J-1)-1}{(I+1)(J-1)} \frac{S^2}{(J-1)} [1+\rho_s(J-2)]$$

$$= \frac{N^2}{n^2} \frac{(I+1)(J-1)-1}{(I+1)} S^2 [1+(J-2)\rho_s]$$

where $\rho_s$ is the intraclass correlation of units (I+1) distance apart.

Let us find

$$Cov\Big[\sum_{i=0}^{I} y_{r+i}, \sum_{j=2}^{J} y_{r+(I+1)j-1}\Big]$$

Assuming $Cov(y_i, y_j) = \rho_{(j-i)} S^2$ we get

$$\mathrm{Cov}\left[\sum_{i=0}^{I} y_{r+i},\ \sum_{j=2}^{J} y_{r+(I+1)j-1}\right] = \sum_{j=2}^{J}\ \sum_{i=0}^{I} \mathrm{Cov}\left(y_{r+i},\ y_{r+(I+1)j-1}\right)$$

$$= \sum_{i=0}^{I}\ \sum_{j=2}^{J} \rho_{(I+1)j-i-1}$$

Hence the variance of modified systematic estimate

$$V\left(\sum_{i=1}^{n} y_i / \pi_i\right) = \frac{N}{n^2}\left[N(I+1)-1\right] S^2 \left(1 + I\rho_c\right)$$

$$+ \frac{N^2}{n^2}\ \frac{(I+1)(J-1) - 1}{(I+1)}\ S^2\left[1 + (J-2)\rho_s\right]$$

$$+ \frac{2N}{n^2}\,(N-1)S^2\left[\sum_{i=0}^{I}\ \sum_{j=2}^{J} \rho_{(I+1)j-i-1}\right]$$

$$\doteq \frac{N^2}{n^2}\,(I+1)\,S^2\left[1 + I\rho_c\right] + \frac{N^2}{n^2}\,(J-1)\,S^2\left[1+(J-2)\rho_s\right]$$

$$+ 2\,\frac{N^2}{n^2}\,S^2\left[\sum_{i=0}^{I}\ \sum_{j=2}^{J} \rho_{(I+1)j-i-1}\right]$$

Also we know

$$V\left(\hat{Y}_{srs}\right) = N^2\, V\left(\hat{\bar{Y}}_{srs}\right) = \frac{N^2(N-n)}{Nn}\, S^2 \doteq \frac{N^2}{n}\, S^2$$

So efficiency of modified systematic sampling over simple random sampling is given by

$$E = \frac{n}{n + \left[I(I+1)\rho_c + (J-1)(J-2)\rho_s + 2\sum_{i=0}^{I}\ \sum_{j=2}^{J} \rho_{(I+1)j-i-1}\right]}$$

~~Mofied~~

Mofied

Modified systematic sampling will be more efficient to simple random sampling if

$$I(I+1)\rho_c + (J-1)(J-2)\rho_s + 2 \sum_{i=0}^{I} \sum_{j=2}^{J} \rho_{(I+1)j-i-1} \quad 0$$

Also we know the variance of usual systematic sampling is

$$V(\hat{Y}_{sys}) = N^2 V(\hat{\bar{Y}}_{sys}) = N^2 \frac{N-1}{n} \frac{S^2}{N} \left[ 1 + \rho(n-1) \right]$$

$$\doteq \frac{N^2}{n} S^2 \left[ 1 + (n-1)\rho \right]$$

So the efficiency of modified systematic sampling over usual systematic sampling is

$$E' = \frac{n\left[1 + \rho(n-1)\right]}{n + I(I+1)\rho_c + (J-1)(J-2)\rho_s + 2 \sum_{i=0}^{I} \sum_{j=2}^{J} \rho_{(I+1)j-i-1}}$$

$E' > 1$

if $\quad n(n-1)\rho > I(I+1)\rho_c + (J-1)(J-2)\rho_s + 2 \sum_{i=0}^{I} \sum_{j=2}^{J} \rho_{(I+1)j-i-1}$

# CHAPTER V

## NUMERICAL ILLUSTRATIONS

With a view to study the relative efficiency of modified systematic sampling, simple random sampling and systematic sampling we may consider the following examples. In each of these examples N is 24 and for a sample of size 8 is relative efficiency of modified systematic sampling ( with I = 4, J = 4 ) systematic sampling ( with k = 3 ) and simple random sampling has been calculated.

**Example 1:**

For a group of 24 operational holdings, whose sizes are given in Kachha Bigha variances have been calculated for these types of sampling, for sample size n = 8.

| S.No. of village | Total area able land in Kacha Bigha | S.No. of village | Total area able land in Kacha Bigha |
|---|---|---|---|
| 1 | 60 | 13 | 50 |
| 2 | 50 | 14 | 10 |
| 3 | 14 | 15 | 85 |
| 4 | 10 | 16 | 30 |
| 5 | 150 | 17 | 30 |
| 6 | 150 | 18 | 70 |
| 7 | 100 | 19 | 30 |
| 8 | 22 | 20 | 35 |
| 9 | 25 | 21 | 30 |
| 10 | 190 | 22 | 10 |
| 11 | 25 | 23 | 30 |
| 12 | 13 | 24 | 70 |

Table 1

| Estimator | Variance |
|---|---|
| 1. Simple random sampling | 201.83 |
| 2. Systematic sampling | 576.36 |
| 3. Modified systematic sampling | 222.80 |

Example 2

The data given below relates to the number of fruit bearing trees of lime orchards for 24 villages of Venkatagiri taluk of Andhra Pradesh. The data were collected during the third round 1963-64 of a pilot survey conducted by I.C.A.R. in Nellore district for the study of yield of lime.

| S.No. of village | No.of fruit bearing trees | S.No.of village | No. of fruit bearing trees |
|---|---|---|---|
| 1 | 9089 | 13 | 607 |
| 2 | 4889 | 14 | 1515 |
| 3 | 1002 | 15 | 3568 |
| 4 | 419 | 16 | 3619 |
| 5 | 4341 | 17 | 269 |
| 6 | 2299 | 18 | 3619 |
| 7 | 1762 | 19 | 2480 |
| 8 | 508 | 20 | 1624 |
| 9 | 1905 | 21 | 1640 |
| 10 | 5916 | 22 | 2486 |
| 11 | 1011 | 23 | 731 |
| 12 | 3568 | 24 | 2480 |

Table 2

| Estimator | Variance |
|---|---|
| 1. Simple random sampling | 350571.51 |
| 2. Systematic sampling | 344860.00 |
| 3. Modified systematic sampling | 338611 .20 |

Example 3

The data given below relates to the area under pepper enumerated completely in each of 24 villages, selected randomly in the state of Kerala during round I of 1966-67.

| S.No. of village | Area under pepper in hectares | S.No. of village | Area under pepper in hectares |
|---|---|---|---|
| 1 | 252 | 13 | 310 |
| 2 | 42 | 14 | 118 |
| 3 | 44 | 15 | 128 |
| 4 | 27 | 16 | 334 |
| 5 | 528 | 17 | 257 |
| 6 | 1056 | 18 | 90 |
| 7 | 458 | 19 | 242 |
| 8 | 238 | 20 | 143 |
| 9 | 42 | 21 | 660 |
| 10 | 185 | 22 | 163 |
| 11 | 209 | 23 | 69 |
| 12 | 129 | 24 | 164 |

Table 3

| Estimator | Variance |
|---|---|
| 1. Simple random sampling | 4583.33 |
| 2. Systematic sampling | 1309.58 |
| 3. Modified systematic sampling | 4548.83 |

Although these are typical examples and it is difficulty to generalise on the basis of these results, it is evidently clear that the performance of the modified systematic sampling may not be worse than the usual systematic sampling. In fact in some situations it may do better than the usual systematic procedure. Since it is a combination of cluster sampling and systematic sampling its efficiency will be between the cluster sampling and usual systematic sampling. In cases where usual systematic sampling performs better than the simple random sampling, the modified systematic sampling may also lead to approximately similar results.

## SUMMARY

Systematic sampling has got a nice feature of selecting the whole sample with one random start. There are some disadvantages of systematic sampling viz., no valid estimate of sampling error can be obtained from sample itself. In the present investigation an alternative systematic sampling is suggested. The alternative method of systematic sampling is superior to the usual systematic sampling method in the sense that a it besides providing unbiased estimate of population mean, suggests procedures for calculating unbiased estimate of the sampling variance from the sample itself. The method has, however got some limitations. For a given population size there is some restriction on the sample size. If the population size be N and sample size n then $n \geq 2(\sqrt{N+1} - 1)$. The limitation will not effect very much this sampling procedure from a large population but the method may not be useful if a small fraction to be sampled from a population of small size.

Modified systematic sampling is extended to multi-stage designs. Use of auxiliary variates is also made to improve the estimate o in case of modified systematic sampling.

Relative efficiency of modified systematic sampling is discussed for the populations with linear correlogram and it is observed that modified systematic sampling is always more efficient than simple random sampling and sometimes it may prove better than usual systematic sampling. For the populations with exponential correlogram modified systematic sampling is however noted to be less efficient than the usual systematic sampling but generally more effizient than simple random sampling. In the end some numerical examples are taken to illustrate the efficiency of modified systematic sampling.

taken to illustrate the efficiency of modified systematic sampling.

## REFERENCES


1. Buckland, W.R. (1950). A review of the literature of systematic Sampling. Jour. Roy. Stat. Soc. Series A, Vol. 12, 208-215.

2. Cochran, W.G. (1946). Relative assucracy of systematic and stratified random samples for a certain class of populations. Ann. Math. Stat. Vol. 17, 164-177.

3. Cochran, W.G. (1953). Sampling techniques. John Wiley and Sons, New York.

4. Das, A.C. (1950). Two dimensional systematic sampling and the associated stratified and random sampling. Sankhya, Vol. 10, 95-108.

5. Madow, W.G. and L.H. (1944). On the theory of systematic sampling. Ann. Math. Stat. Vol. 15, 1-24.

6. Quenouille M.H. (1949). Problems in plane sampling. Ann. Math. Stat., Vol. 20, 355-375.

7. Yates, F. (1948). Systematic sampling. Phil. Trans. of Roy. Soc. Lond., Series A, Vol. 241, 345-377.

8. Singh, D., Jindal, K.K. and Garg, J.N. (1968). Modified systematic sampling. Biometrics.

9. Talukdar, A.H. (1960). Multivariate Regression Estimates For Finite Populations A.I.M.S. Vol. 20, 441-445.

10. Olkin, I. (1958) Multivariate Ratio Estimation For Finite Population Biometrika 45, 154-165.