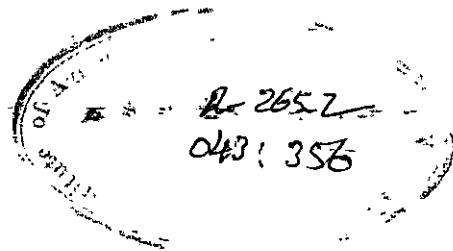


V 4  
V 95

## SOME CONTRIBUTIONS TO TWO-STAGE SAMPLING

ANAND PRAKASH



Dissertation submitted in fulfilment of the  
requirements for the award of Diploma in  
Agricultural Statistics of the Indian  
Agricultural Statistics Research  
Institute (I.C.A.R.)

DEPT OF  
9

INDIAN AGRICULTURAL STATISTICS RESEARCH INSTITUTE  
NEW DELHI  
1978

CERTIFICATE

This is to certify that the work incorporated in this thesis entitled "Some Contributions to two-stage sampling" by Nand Prakash and submitted in fulfilment of the requirements for the award of Diploma in Agricultural Statistics of the Indian Agricultural Statistics Research Institute, was done under my guidance.

Padam Singh  
(Padam Singh)  
Scientist S-3

## ACKNOWLEDGMENTS

I wish to express my deep sense of gratitude to Dr. Padea Singh, Scientist, Indian Agricultural Statistics Research Institute (ICAR) for his valuable guidance and constant encouragement during the course of investigation and writing this thesis.

My thanks are also due to Dr. Devesh Singh, Director, Indian Agricultural Statistics Research Institute (ICAR), New Delhi, for providing facilities for research work.

I am also grateful to Shri A.K. Srivastava, Scientist, Indian Agricultural Statistics Research Institute, New Delhi, for his valuable suggestions.

Anand Prakash

(Anand Prakash)  
Scientist

Indian Agricultural  
Statistics Research  
Institute, New Delhi

## CONTENTS

PAGES

### CHAPTER I

1.1	Introduction	1-3
1.2	Prerequisites	3-5
1.3	Orientation of the Problem	6-9

### CHAPTER II ALLOCATION OF SECONDARY SAMPLING UNITS

2.1	Introduction	10-11
2.2	Suggested Procedures	11-18
2.2.1	Procedure I	12-13
2.2.2	Procedure II	12-13
2.2.3	Procedure III	13-19
2.3	Relative efficiencies of suggested allocation procedures	19
2.3.1	Proportional allocation versus constant allocation	20-21
2.3.2	Optimum allocation versus proportional allocation	21-22
2.4	Numerical Illustration	22-26
2.5	Summary	26

### CHAPTER III NEW SAMPLING SCHEME

3.1	Introduction	37
3.2	The Sampling Scheme	37-39
3.3	Identification procedure	39
3.3.1	Hovisite Thompson estimator	39-41
3.3.2	Post-identification estimator	41-43
3.3.3	Estimator based on sample mean	43-45
3.4	Numerical Illustration	45
3.5	Summary	45

Summary	46-47
References	47

## CHAPTER I

### INTRODUCTION

For sound development of any industry/field, the planners should have a good background of the sector of their interest. This essentially implies the availability of the data (statistical information) about the field of interest. In dynamic population, outdated data may not play much role in planning for future development and most recent data may be needed for the purpose. Some devices have been employed to collect the required information. Of all, the method that has been widely used and found methodical and practicable is the collection of data through sample survey. It, besides, being capable of making the data available in shorter time as compared to census, result in great economy of efforts.

Through surveys, a few items (units) which may be supposed to be representative of the entirety (population) are studied and based on their values, inference about the population is drawn.

The problem of choosing a few units which could represent the population is not quite simple. The theory of probability, originated in the sixteenth century, with the study of the chance phenomena provided a basis for development of a method which could be adopted for choosing a few units to represent the population. The concept of

random numbers was off-shoot of the theory of probability. The mathematicians who had developed the theory of probability, would have never imagined at that time what could be its application and how it would help the society. In fact, the modern theory of statistics is nothing but application of the theory of probability developed by the then mathematicians in the sixteenth and seventeenth centuries. It had been demonstrated that by the application of the theory of probability it may be possible to make inference about a population through a proper sample. This approach led to the method which is now usually known as random sampling technique.

In early part of this century, Fisher and Karl Pearson gave a practical shape as to how to infer about the population from a given sample. This very concept was extended in the twenties of this century to sample a finite population with which we commonly come across in real life.

With the introduction of artificial randomisation, which implied use of devices like random number tables, the development of modern sampling techniques started with Neyman(1934) who introduced new criteria of the optimum use of resources in sampling, including the concept of optimum allocation of sampling units to different strata subject to fixed total number of sampling units. The rationale provided by Neyman through the application of probability sampling was so

convincing that the method of sampling became widely popular in a number of countries for collecting data on various aspects of economy. A distinct advantage of the method of random sampling is its ability to provide an estimate of the sampling error from the sample itself, which becomes the basis for judging the reliability of the sample value (estimator). The method, however, needed modification to fit into the actual practical conditions.

Before discussing the modifications suggested from time to time in respect of probability scheme, it would be worthwhile for clarity and completeness to discuss certain preliminaries of sampling which will be frequently used in this thesis.

## 1.2 Preliminaries

The elements of a population are those on which information is sought. The unit may be taken as well-defined and identifiable element on which observations could be made. The aggregate of the units under study is termed the population and it is said to be finite if the number of units it contains is finite.

Suppose a finite population consists of  $N$  units and let the value of a characteristic for the  $i$ -th unit be  $Y_i$ . Any function of the values of  $Y_i$ , the units of the population such as

$$Y = \sum_{i=1}^N Y_i \text{, the population total.}$$

$$\bar{Y} = \frac{1}{N} \sum_{i=1}^N Y_i , \text{ the population mean}$$

and

$$S^2 = \frac{1}{N} \sum_{i=1}^N (Y_i - \bar{Y})^2 , \text{ the population variance,}$$

are known as parameters.

A part of the population is termed as sample and the procedure of selecting the sample is known as sampling.

For sampling purposes it is essential to be able to list out all of the units in the population. Such a list is called the frame and provides the basis for the selection and identification of units in the sample.

A sample that has been selected from a population following laws of chance will be called random or probability sample. The values of the character under study for selected units are known as sample observations.

The procedure of selecting the sample unit by unit is known as sampling scheme.

All possible samples of a sampling scheme gives rise to the sample space, and all possible samples of the sample space together with their probabilities of selection is defined as the sampling design.

Let the sample observations are denoted by  $y_1, y_2, \dots, y_n$ :  $n$  being the number of units in the sample. Any function

of these values is called an estimator. The sampling design together with the estimator is termed as the sampling strategy. The different samples of a sampling scheme may give rise to different estimates with respective probabilities. Let the probability of getting the  $i$ -th sample be  $p_i$  ( $i=1, 2, \dots, N$ ) and let  $t_i$  be the estimator based on this sample. The estimator  $t_i$  is said to be an unbiased estimator of the population parameter  $\theta$  if its expected value is equal to  $\theta$  i.e.

$$E(t) = \sum_{i=1}^N t_i p_i = \theta ,$$

$N$  being total number of samples.

In case  $E(t)$  is not equal to  $\theta$ , the estimator  $t$  is said to be a biased estimator of  $\theta$  and the bias is given by  $B(t) = E(t) - \theta$ .

Since the sampling scheme gives rise to different samples, estimates will, in general, differ from sample to sample. The difference between the estimator and the parameter may be called error. A measure of the divergence of the different estimates from the true value is given by the expected value of the squared error and this is known as mean square error (MSE) of the estimator which is given by

$$MSE(t) = \sum_{i=1}^N (t_i - \theta)^2 p_i$$

It is a measure of difference of the estimator from the population parameter. If the difference is taken

From its expected value, the measure is called sampling variance and is given by

$$\text{Variance } (t) = E \left[ (t - E(t))^2 \right]$$

It can be seen that in case of  $t$  being an unbiased estimator, the MSE and sampling variance are same.

### 4.3 Orientation of the Problem

Very often, in large scale sample surveys it is extremely difficult and costly if not impossible to select directly the ultimately observable units in the sample. Such sampling in practice, essentially requires a list identifying all the elements of the population and most frequently this list does not exist or is too expensive to get. Even if such a list is available, the enumeration cost may be excessive if the sample is too widespread. Further, there are other restrictions on the sample such as the enumerators may be required to work under the close supervision of a limited number of supervisors, and as a consequence the field operations must be confined to a limited number of administrative centres. In order to meet with these restrictions and to make most effective use of available resources, various sampling techniques have been developed from time to time. Stratification of the population was the first technique suggested which consists in dividing the population before selection of the sample into a number of mutually

exclusive groups called strata. For efficient estimation of the parameters of interest, the care should be taken that there is homogeneity within each stratum. After stratification, the question is of allocation of the sample over the strata. Allocation of the sample may be done either in proportion to number of units in the stratum or proportional to  $N_1 S_1$  where  $N_1$  indicates the number of sampling units in the 1-th stratum and  $S_1$  the mean square of the values of the units in the 1-th stratum. The latter method known as optimum allocation is superior to the other method of allocation. The difficulty, however, in adopting optimum method is that the information on  $S_1$  is not available always. After allocation of units, the sampling units are independently selected from each stratum and estimators are worked out.

It is not necessary to make an actual subdivision of the population before selection of the sample. Alternatively, the units in the population may be grouped taking some nearby units and a few groups may be selected out of the groups thus formed. The groups of units formed are called clusters and the procedure of sampling in which the sampling units are clusters is known as cluster sampling. The clusters may be of equal or unequal sizes. Cluster sampling ordinarily results in greater variance than a sample of the same number of elements selected randomly.

In cluster sampling all the elements of the selected clusters are enumerated. However, from many considerations it may not be desirable to enumerate all the elements of the cluster. It is logical to expect that for a given number of elements, greater precision will be attained by distributing them over a large number of clusters than by taking a small number of clusters and sampling a large number of elements from each of them. The procedure of first selecting clusters and then choosing a specified number of elements from each selected cluster is known as two stages sampling or sub-sampling. The procedure has the advantage of being flexible as different sampling procedures can be employed at different stages of selection, if necessary. A great saving in operational costs, particularly if the survey covers a large area including under-developed pockets can also be effected by adopting sub-sampling. In Indian Agricultural Statistics Research Institute, efficiency of sub-sampling and cluster sampling have been extensively studied. The theory of multi-stage sampling and its application were developed by Hansen and Hurwitz (1943, 1949), Mahalanobis (1946, 1952), Sukhatme (1947, 1950, 1953), Sukhatme and Pande (1951), Sukhatme and Karain (1952), Yates (1949), Singh (1956), Durbin (1967) etc.

In two stage sampling design the primary sampling units (p.s.u's) are generally of unequal sizes. For the choice of number of secondary sampling units (s.s.u's) from different p.s.u's, the most common method is to take

them either equal or in proportion to their sizes. The first method based on equal number of s.s.u's from each p.s.u. does not take into account the size and the variability of s.s.u's in different p.s.u's and thus results in less efficient estimators. Alternatively, if s.s.u's are taken in proportion to the size, the ultimate sample size becomes a random variable which is undesirable from many considerations.

In Chapter II, the problem of allocation of s.s.u's has been attempted and a few procedures have been suggested. The relative efficiencies of various procedures have been compared with the methods in use. A new sampling scheme for two stage design along with suitable estimators has been suggested in Chapter III.

## CHAPTER II

### ALLOCATION OF SECONDARY SAMPLING UNITS

#### 2.1 Introduction

Multi-stage sampling design is widely used in most of the surveys from many considerations. In two-stage sampling design, the population under consideration consists of certain number of primary sampling units (p.s.u's) each of which in turn consists of certain number of elements called secondary sampling units (s.s.u's). In general, for the purpose of estimation of the total or mean of the characteristics under study, a sample of suitable number of p.s.u's is selected in the first instance and thereafter from out of the s.s.u's in each of the selected p.s.u's, a sample of desired number of s.s.u's is further selected. The number of s.s.u's to be selected from the selected p.s.u's are usually taken as equal or proportional to their size. The former method based on equal number of s.s.u's from each p.s.u. does not take into account the possible importance of size and variability of different p.s.u's resulting in less efficient estimators. Alternatively, if number of s.s.u's is taken proportional to size, the total sample size (total number of s.s.u's in the sample) becomes a random variable which is undesirable from many considerations. In this Chapter some new methods of allocation of the s.s.u's over the selected primary sampling units have been suggested.

the relative efficiencies of the proposed methods with the method based on equal number of s.e.u's from each selected p.s.u. have also been investigated.

## 2.2 Suggested Procedures

Let us consider a population consisting of  $N$  p.s.u's with  $M_i$  s.e.u's in the  $i$ -th p.s.u. ( $i=1, 2, \dots, N$ ). Further, let  $y_{ij}$  be the value of the  $j$ -th s.e.u of the  $i$ -th p.s.u in the population ( $j=1, 2, \dots, M_i$ ;  $i=1, 2, \dots, N$ ). We have the following notations

$$\bar{Y}_i = \frac{1}{M_i} \sum_{j=1}^{M_i} y_{ij}, \text{ Mean of } i\text{-th p.s.u}$$

$$\bar{Y} = \frac{1}{N} \sum_{i=1}^N M_i \bar{Y}_i, \text{ Mean of population}$$

$$n_0 = \sum_{i=1}^N M_i \quad \text{Total number of s.e.u's in the population.}$$

$$s_b^2 = \frac{1}{N-1} \sum_{i=1}^N \left( \frac{M_i \bar{Y}_i}{\bar{Y}} - \frac{\bar{Y}}{N} \right)^2, \text{ Variance between p.s.u's}$$

$$s_i^2 = \frac{1}{M_i-1} \sum_{j=1}^{M_i} (y_{ij} - \bar{Y}_i)^2, \text{ Variance between the s.e.u's in the } i\text{-th p.s.u.}$$

Let  $n$  be the number of p.s.u's to be drawn and  $n_0$  be the total number of s.e.u's to be drawn from the selected p.s.u's. For this, we consider the following

Sampling schemes employing different methods of allocation of s.e.u's to the selected p.s.u's.

### 2.2.1 Procedure I

The selection of the sample under procedure I consists of the following steps:

1. Select  $n$  p.s.u's from  $N$  p.s.u's by simple random sampling without replacement,
2. Allocate the sample of  $n_0$  s.e.u's to different p.s.u's selected using

$$n_1 = \frac{M_1}{\sum_{i=1}^n M_i} n_0 \rightarrow i=1, 2, \dots, n$$

Here  $\sum_{i=1}^n M_i$  is the total number of s.e.u's of the selected p.s.u's which is a random variable as it depends upon the p.s.u's selected.

3. Select  $n_1$  s.e.u's from  $i$ -th p.s.u. by simple random sampling without replacement.

The procedure I will hereafter be known as proportional allocation procedure as the number of s.e.u's drawn from the selected p.s.u's are in proportion to their sizes. For this procedure, the unbiased estimator for population total is given by

$$\hat{T}_I = \frac{n_0}{n} \sum_{i=1}^n \frac{M_i}{N} - \frac{1}{n_1} \sum_{j=1}^{n_1} Y_{ij} \quad \text{total}$$

$$\text{where } \bar{N} = \frac{1}{n} \sum_{k=1}^n N_k$$

the expected value of the above estimator is clearly  $\bar{N}$ , the population total.

The variance of the estimator can be obtained as follows:

$$V(\hat{X}) = E_1 V_2(\hat{X}) + V_1 E_2(\hat{X}) \quad (2.1)$$

Here  $E_2$  and  $V_2$  represent respectively the expectation and variance for given sample of  $n$  units and  $E_1$  &  $V_1$  for unconditional expectation and variances for all possible samples.

(a) First considering  $E_1 E_2(\hat{X})$

$$E_2\left(\frac{N_0}{n} \sum_{k=1}^n \frac{N_k}{N} - \frac{1}{n} \sum_{k=1}^n N_{kj}\right) = \frac{N_0}{n} \sum_{k=1}^n \frac{N_k}{N} - \bar{N}$$

$$V_1\left(\frac{N_0}{n} \sum_{k=1}^n \frac{N_k}{N} - \bar{N}\right) = V_0\left(\frac{1}{n} + \frac{1}{N}\right) S_0^2$$

(b) Further

$$E_1 V_2(\hat{X}) = E_1 V_2\left(\frac{N_0}{n} \sum_{k=1}^n \frac{N_k}{N} - \bar{N}_1\right)$$

$$= 2\left[\frac{N_0^2}{n^2} \sum_{k=1}^n \frac{N_k^2}{N^2} \left(\frac{1}{N_1} + \frac{1}{N_0} - \bar{S}_0^2\right)\right]$$

$$E_1 = \left[ \frac{N_0}{n^2} \sum_{i=1}^n \frac{M_i^2}{N^2} \left( \frac{1}{a_1} - \frac{1}{m_1} \right) s_1^2 \right]$$

$$E_1 = \left[ \frac{N_0}{n^2} \sum_{i=1}^n \frac{M_i^2}{N^2} \left( \frac{N_1}{\sum_{j=1}^n M_j} - \frac{1}{a_1} \right) s_1^2 \right]$$

$$E_1 = \left[ \frac{N_0}{n^2} \sum_{i=1}^n \frac{M_i}{N^2} \left( \left( \sum_{i=1}^n \frac{M_i}{m_0} - 1 \right) s_1^2 \right) \right]$$

$$E_1 = \left[ \frac{N_0}{n^2} \sum_{i=1}^n \frac{M_i}{N^2} s_0 \left( \sum_{i=1}^n (M_1 - m_0) s_1^2 \right) \right]$$

$$E_1 = \left[ \frac{N_0}{n^2 s_0 N^2} \left[ \sum_{i=1}^n M_1 s_1^2 \sum_{j=1}^n M_1 - m_0 \sum_{i=1}^n M_1 s_1^2 \right] \right]$$

$$E_1 = \left[ \frac{N_0}{n^2 s_0 N^2} \left[ \sum_{i=1}^n M_1^2 s_1^2 + \sum_{j=1}^n M_1 s_1^2 M_j - m_0 \sum_{i=1}^n M_1 s_1^2 \right] \right]$$

$$\frac{N_0}{n^2 s_0 n^2} \left[ \frac{n}{N} \sum_{i=1}^n M_1^2 s_1^2 + \frac{n(n-1)}{N(N-1)} \sum_{i=1}^n M_1 s_1^2 M_j - m_0 \frac{n}{N} \sum_{i=1}^n M_1 s_1^2 \right]$$

$$\frac{N_0}{n^2 s_0 n^2} \left[ \frac{n}{N} \sum_{i=1}^n M_1^2 s_1^2 + \frac{n(n-1)}{N(N-1)} \left[ \sum_{i=1}^n M_1 s_1^2 \sum_{j=1}^n M_j - \sum_{i=1}^n M_1 s_1^2 \right] \right]$$

$$- m_0 \frac{n}{N} \sum_{i=1}^n M_1 s_1^2$$

$$= \frac{N_0}{n^2 s_0 n^2} \left[ \frac{n}{N} \sum_{i=1}^n M_1^2 s_1^2 + \frac{n(n-1)}{N(N-1)} \sum_{i=1}^n M_1^2 s_1^2 + \frac{n(n-1)}{N(N-1)} \sum_{i=1}^n M_1 s_1^2 \sum_{j=1}^n M_j \right]$$

$$+ \frac{m_0 n}{N} \sum_{i=1}^n M_1 s_1^2$$

$$= \frac{N_0^2}{N^2 n^2} \left[ \frac{n(\bar{x}_n, \bar{u})}{n_0(N-1)} \sum_{i=1}^N \bar{x}_i^2 \bar{u}_i^2 + \frac{n}{N} \sum_{i=1}^N \bar{x}_i \bar{u}_i^2 \left\{ \frac{n-1}{n_0(N-1)} \sum_{j=1}^N \bar{u}_j = 1 \right\} \right]$$

$$+ \frac{N_0^2}{N^2 n^2} \left[ \frac{n(\bar{x}_n, \bar{u})}{n_0(N-1)} \sum_{i=1}^N \bar{x}_i^2 \bar{u}_i^2 + \frac{n}{N} \sum_{i=1}^N \bar{x}_i \bar{u}_i^2 \left( \frac{n-1}{n_0(N-1)} \sum_{j=1}^N \bar{u}_j = 1 \right) \right]$$

Hence

$$\begin{aligned} V(\bar{x}_n) &= N_0^2 \left( \frac{1}{n^2} \frac{1}{N^2} \right) \frac{2}{n_0} + \frac{N_0^2}{N^2 n^2} \left[ \frac{n(\bar{x}_n, \bar{u})}{n_0(N-1)} \sum_{i=1}^N \bar{x}_i^2 \bar{u}_i^2 + \right. \\ &\quad \left. \sum_{i=1}^N \bar{x}_i \bar{u}_i^2 \left( \frac{n-1}{n_0(N-1)} \sum_{j=1}^N \bar{u}_j = 1 \right) \right] \end{aligned}$$

..... (2.2)

### 2.2.2 Procedure II

The selection of the sample under procedure II is as follows:

- (1) Select  $n$  p.s.u's from  $N_p$ s.p.u's by simple random sampling without replacement.
- (2) Determine the number of s.s.u's to be selected from  $i$ th selected p.s.u, using

$$m_i = m_0 \frac{\bar{x}_i \bar{u}_i}{\frac{n}{\sum_{i=1}^N \bar{x}_i \bar{u}_i}}$$

- (3) Select  $m_i$  s.s.u's from  $N_i$  s.s.u's of the  $i$ th p.s.u, by simple random sampling.

This procedure will hereafter be called as optimum

+13

allocation procedure.

In the usual two-stage sampling design, the variance of the estimate is composed of two components: the first one which depends upon the variance between the p.s.u's remains unaffected by the choice of the number of sub-sampling units while the second component based on  $n_1$ 's and the variance between s.e.u's within p.s.u's can be minimized by determining  $n_1$ 's taking into account the variability of the second stage units in the selected p.s.u's. We have minimized the conditional variance given below of the two stage sampling estimate via:

$$V_3 = \frac{n^2}{n^2} \sum_{i=1}^n \frac{s_i^2}{N_i^2} \left( \frac{1}{n_1} + \frac{1}{n_2} \right) s_i^2 \quad (2.3)$$

with the constraint  $\sum_{i=1}^n n_i = n_0$  obtaining

$$n_1 = n_0 \cdot \frac{\frac{N_1 S_1}{N_2 S_2}}{\sum_{i=1}^n \frac{N_i S_i}{N_2 S_2}} \quad (2.4)$$

which have been used at step XII.

Substituting the value of  $n_1$  from (2.4) in (2.3) we get:-

$$= E_2 \left[ \frac{n^2}{n^2} \sum_{i=1}^n \frac{s_i^2}{N_i^2} \left( \frac{\sum_{i=1}^n N_i S_i}{n_0 N_2 S_2} - \frac{1}{n_2} \right) s_i^2 \right]$$

$$= \frac{1}{n^2} \left[ \frac{\sigma_e^2 / n^2}{\sum_{i=1}^n u_i^2} \cdot \frac{n^2 - (n-1)^2}{n^2} \cdot \frac{\sum_{i=1}^n u_i q_i}{\sum_{i=1}^n u_i^2} \right] \frac{\sigma_e^2}{n}$$

$$= \frac{1}{n^2} \left[ \frac{\sigma_e^2 / n^2}{\sum_{i=1}^n u_i^2} \cdot \frac{n^2 - (n-1)^2}{n^2} \cdot \frac{u_1 u_2 u_3 u_4}{u_1^2 u_2^2} \right] =$$

$$\frac{\sigma_e^2 / n^2}{u_1^2 u_2^2} \cdot \frac{u_1 u_2 u_3 u_4}{u_1^2 u_2^2}$$

Now taking the expectation we obtain:

$$= (\sigma_e^2 / n^2) \sum_{i=1}^n u_i^2 + (\sigma_e^2 / n^2) \frac{(n-1)/n(n-1)}{\sum_{i=1}^n u_i^2}$$

$$\geq \sum_{i \neq j} u_i u_j u_i u_j - (\sigma_e^2 / n^2) \sum_{i=1}^n u_i^2$$

$$= (\sigma_e^2 / n^2) \sum_{i=1}^n u_i^2 u_j^2 + \frac{\sigma_e^2 (n-1)}{n^2 u_0^2 n(n-1)} \left[ \left( \sum_{i=1}^n u_i u_j \right)^2 \right]$$

$$= \left[ \sum_{i=1}^n u_i^2 u_j^2 \right] - (\sigma_e^2 / n^2) \sum_{i=1}^n u_i^2$$

We know  $\bar{u} = u_0/n$  or  $\bar{u} = u_0/\sqrt{n}$

$$= (1/n^2) \sum_{i=1}^n u_i^2 u_j^2 + \frac{n(n-1)}{u_0^2 n(n-1)} \left[ \left( \sum_{i=1}^n u_i u_j \right)^2 - \sum_{i=1}^n u_i^2 u_j^2 \right]$$

$$+ \frac{n}{n} \sum_{i=1}^n u_i^2 u_j^2$$

$$\begin{aligned} &= \frac{n}{n_0 n} \sum_{i=1}^n x_i^2 s_i^2 \left[ 1 + \frac{n-1}{n-1} \right] + \frac{n(n-1)}{n_0 n(n-1)} \left( \sum_{i=1}^n x_i s_i \right)^2 \\ &\quad - \frac{s^2}{n} \sum_{i=1}^n x_i s_i^2 \end{aligned}$$

$$\begin{aligned} &= \frac{n(n-1)}{n_0 n(n-1)} \sum_{i=1}^n x_i^2 s_i^2 + \frac{n(n-1)}{n_0 n(n-1)} \left( \sum_{i=1}^n x_i s_i \right)^2 - \frac{s^2}{n} \sum_{i=1}^n x_i s_i^2 \end{aligned}$$

Hence, the variance of the estimate under this procedure is as under:

$$\begin{aligned} V(\hat{x}_{III}) &= n_0 \left( \frac{1}{n} - \frac{1}{n} \right) s_0^2 + \frac{n}{n_0 n(n-1)} \left[ (n-1) \sum_{i=1}^n x_i^2 s_i^2 \right. \\ &\quad \left. + (n-1) \left( \sum_{i=1}^n x_i s_i \right)^2 \right] \\ &\quad - \frac{s^2}{n} \sum_{i=1}^n x_i s_i^2 \end{aligned}$$

... (2.5)

### 2.2.3 Procedure III

Under this method of allocation all the  $s_i, n_i$ 's to be selected from each selected p.a.u. are taken equal i.e.,  $n_i = n_0/n$ .

The variance of the estimator under the procedure III reduces to

$$V(\hat{Y}_{III}) = M_0^2 \left( \frac{1}{n} - \frac{1}{N} \right) s_b^2 + \frac{N}{M_0} \sum_{i=1}^N M_i^2 s_i^2 - \frac{N}{n} \sum_{i=1}^N M_i s_i^2$$

..(2.6)

This procedure hereafter will be called as constant allocation procedure.

It can be seen easily that the unbiased variance estimator of the population total under different allocation procedures with corresponding values of  $m_i$ 's is given by

$$\hat{V}(\hat{Y}) = M_0^2 \left[ \left( \frac{1}{n} - \frac{1}{N} \right) s_b^2 + \frac{N}{m} \left[ \sum_{i=1}^n \left( \frac{1}{m_i} - \frac{1}{M_i} \right) s_i^2 \right] \right]$$

where

$$s_b^2 = \frac{1}{n-1} \sum_{i=1}^n \left( \frac{M_i}{N} \bar{Y}_{i+} - \bar{Y}_{..} \right)^2$$

$$s_i^2 = \frac{1}{m_i-1} \sum_{j=1}^{m_i} (Y_{ij} - \bar{Y}_{i+})^2$$

### 2.3 Relative Efficiencies of the suggested allocation procedures

The merit of sampling procedure is judged from the factors such as simplicity, operational convenience and efficiency of the estimators. The proposed procedures are simple and convenient to use in practice. In this section we study the relative efficiency of proportional allocation, constant allocation and optimum allocation procedures.

### 2.3.1 Proportional allocation versus constant allocation

The difference between the sampling variances of constant allocation and the proportional allocation procedures can be expressed as

$$\begin{aligned}
 V(\hat{T}_{III}) - V(\hat{T}_I) &= M_0^2 \left( \frac{1}{n} - \frac{1}{N} \right) S_b^2 + \frac{n}{n_0} \sum_{i=1}^N M_i^2 s_i^2 - \frac{N}{n} \sum_{i=1}^N M_i s_i^2 \\
 &\quad + M_0^2 \left( \frac{1}{n} - \frac{1}{N} \right) S_b^2 \frac{n}{n_0} \left[ \frac{n-n}{n_0(N-1)} \cdot \sum_{j=1}^N M_j^2 s_j^2 + \sum_{j=1}^N M_j^2 s_j^2 \right. \\
 &\quad \left. - \left( \frac{n-1}{n_0(N-1)} \sum_{j=1}^N M_j - 1 \right) \right] \\
 &= \sum_{i=1}^N M_i^2 s_i^2 \left[ \frac{\frac{2}{N(N-1)}}{nn_0(N-1)} \right] + \frac{N(N-1)}{nn_0(N-1)} \sum_{i=1}^N M_i^2 s_i^2 \sum_{j=1}^N M_j
 \end{aligned}$$

Taking  $\frac{n-1}{n_0(N-1)} \sim \frac{n}{N}$  we obtain

$$= \frac{n}{n_0} \sum_{i=1}^N M_i^2 s_i^2 - \frac{1}{n_0} \sum_{i=1}^N M_i^2 s_i^2 \sum_{j=1}^N M_j \tag{2.7}$$

From the above expression, the sufficient condition for proportional allocation procedure to be superior to constant allocation procedure is that

$$\frac{\sum_{i=1}^N M_i^2 s_i^2}{\sum_{i=1}^N M_i^2 s_i^2 \sum_{j=1}^N M_j} \geq \frac{1}{n}$$

$$\text{or } \text{Cov}(N_1, N_1 S_1^2) \geq 0 \quad . . . . . \quad (2.8)$$

Now since the relationship between  $S^2$  and  $N$  is generally of the form

$$S^2 = AN^k, \quad 0 < k < 2$$

the condition (2.8) therefore holds generally true.

### 2.3.2 Optimum allocation versus proportional allocation

It is interesting to study the comparison of these two procedures. The difference between their sampling variance may be written as

$$\begin{aligned} \hat{V}(\hat{Y}_I) - V(\hat{Y}_{II}) &= M_0^{2n} \left[ \left( \frac{1}{n} - \frac{1}{N} \right) S_D^2 \right] + \frac{N}{n} \left[ \frac{N-n}{m_0(n-1)} \sum_{j=1}^N M_j^2 S_1^2 \right. \\ &\quad \left. + \sum_{j=1}^N M_j^2 S_1^2 \left( \frac{n-1}{m_0(n-1)} \sum_{j=1}^N M_j - 1 \right) \right] \\ &= \left[ M_0^{2n} \left( \frac{1}{n} - \frac{1}{N} \right) S_D^2 \right] + \frac{N}{m_0 n(n-1)} \left[ (N-n) \sum_{j=1}^N M_j^2 S_1^2 \right. \\ &\quad \left. + (n-1) \left( \sum_{j=1}^N M_j S_1 \right)^2 \right] - \frac{N}{n} \sum_{j=1}^N M_j^2 S_1^2 \\ &= \frac{N(N-n)}{m_0 n(n-1)} \sum_{j=1}^N M_j^2 S_1^2 + \frac{N}{n} \left[ \sum_{j=1}^N M_j^2 S_1^2 \sum_{j=1}^N M_j \frac{n-1}{m_0(n-1)} \right] \\ &= \frac{N(N-n)}{m_0 n(n-1)} \sum_{j=1}^N M_j^2 S_1^2 + \frac{(n-1)N}{m_0 n(n-1)} \left( \sum_{j=1}^N M_j S_1 \right)^2 \end{aligned}$$

$$= \frac{n(n-1)}{nm_0(n-1)} \left[ \sum_{i=1}^n m_i s_i^2 \sum_{j=1}^n m_j + \frac{(n-1)n}{m_0 n(n-1)} \left( \sum_{i=1}^n m_i s_i \right)^2 \right]$$

$$= \frac{n(n-1)}{nm_0(n-1)} \left[ \sum_{i=1}^n m_i s_i^2 \sum_{j=1}^n m_j - \left( \sum_{i=1}^n m_i s_i \right)^2 \right]$$

This term being positive shows that the optimum allocation method which takes care of variability of the secondary sampling units is more efficient than the proportional allocation procedure in which the s.s.u's are apportioned over the selected p.s.u's proportionally.

## 2.4 Numerical Illustration

Data on volume of timber as obtained strip-wise by complete enumeration from ten blocks of the Blockia Mountain Experimental Forest, given as Annexure 4.2 Pages 131-132 in Murthy (1967) has been taken for illustration of efficiencies of the above allocation procedures. The variance obtained in each of the allocation procedures discussed above for  $n = 5$  and  $m_0 = 40$  together with the percentage gain in efficiency is as follows:

Table - 1

S.No.	Procedure of allocation	Variance	% gain in efficiency
1	Constant allocation	57277341.6	-
2	Proportional allocation	55774687.3	2.7 over constant allocation
3	Optimum allocation	51945204.9	2.3 over proportional allocation

The efficiency of allocation procedures has further been studied by a hypothetical population given below.

No. of p.s.u's in the population	Value of s.s.u's
1	2,4
2	3,5,6,10
3	3,5,8,9,10,13
4	1,3,5,7,8,10,15

Table - 2 gives the variance of estimate of population total of the above population for different allocation procedures together with the percentage gain in efficiency for  $n = 3$  and  $n_2 = 10$ .

Table - 2

S.No.	Procedure of allocation	Variance	% gain in efficiency
1	Constant allocation	1099.4	-
2	Proportional allocation	903.3	21.0 over constant allocation
3	Optimum allocation	874.5	3.9 over proportional allocation

The above results suggest that the optimum allocation which takes into account the distribution of s.s.u's the variability of the s.s.u's in the selected p.s.u's is superior to both

constant allocation and proportional allocation while the next best among the methods of allocation, is the proportional allocation.

It is important to comment here that for populations I and II, the contribution of second term in the variance for the estimate of the total under constant allocation is of the order 9% and 37% respectively. It may also be mentioned that the suggested allocation procedures help in reducing the second component in the variance expression which is a function of within p.s.u. variance. It is, therefore, expected that there will be relatively more gain in efficiency in the case of second population as compared to first population and the same is revealed by the results.

Hence it may be seen that the suggested allocation procedures can be used with advantage when the contribution of second component in the variance expression to the total variance is not very small. A rough idea of the magnitude of the components of variance may be had from the past surveys.

## 2.5 SUMMARY

In this chapter different allocation procedures for getting a sample of fixed number of a.s.u's have been suggested. The relative efficiencies of the suggested procedure with the constant allocation procedure have been examined. It has been observed that there is considerable gain in using the suggested allocation procedures in practice.

## CHAPTER III

### NEW SAMPLING SCHEME

#### 3.1 Introduction

In Chapter II we had discussed various procedures of allocation of the desired number of s.e.u's over the selected p.s.u's. Among the suggested procedures, the optimum allocation procedure is found to be more efficient as compared to the proportional allocation and constant allocation procedures. The optimum allocation procedure, however, poses certain difficulties such as the knowledge of the variability for all the selected p.s.u's etc., for allocation. The difficulty common with all allocation procedures is that many times specially for smaller sample sizes the allocation is never exactly achieved.

In this chapter, a new sampling scheme providing constant sample size has been suggested. Three types of estimators have also been proposed for the new sampling scheme. The efficiency of three estimators in new sampling scheme as compared to allocation procedures has been investigated empirically.

#### 3.2 The sampling scheme

The new sampling scheme for selection of  $n_0$  s.e.u's consists of the following steps:

##### Step I

Select  $n$  p.s.u's by simple random sampling without

26

replacement from  $N$  p.s.u's of the population. Let the selected p.s.u's be with serial numbers  $i_1, i_2, \dots, i_n$ .

### Step II

Pool the s.s.u's of the  $i_1, i_2, \dots, i_n$  in the p.s.u's selected at step I and get a population of  $M_{i_1} + M_{i_2} + \dots + M_{i_n}$  s.s.u's where  $M_{i_j}$  is the number of s.s.u's in the  $i_j$ -th selected p.s.u.

### Step III

Select  $n_0$  s.s.u's from the population of  $M_{i_1} + M_{i_2} + \dots + M_{i_n}$  s.s.u's of step II by simple random sampling without replacement.

### Illustration

Consider a population of 4 p.s.u's with 2, 4, 6 and 8 s.s.u's respectively. In drawing 2 p.s.u's out of 4 at step I we have the following six combinations of two p.s.u's each

$$(1,2), (1,3), (1,4), (2,3), (2,4) \text{ and } (3,4)$$

Now at step-II we pool the s.s.u's of the p.s.u's selected at Step-I in the sample. Thus we have  $(2+4=6)$ ,  $(2+6=8)$ ,  $(2+8=10)$ ,  $(4+6=10)$ ,  $(4+8=12)$ , and  $(6+8=14)$ , s.s.u's in above six possible combinations of two p.s.u's each. If 4 units are to be drawn at Step-III from the s.s.u's of the selected p.s.u's we would have  $6_{Q_6}, 8_{Q_8}, 6_{Q_{10}}, 10_{Q_{10}}, 8_{Q_{12}}$  and  $10_{Q_{14}}$  possible samples respectively corresponding to the six possible combinations of p.s.u's. Thus the

total number of samples would be

$$6_{CH} + 8_{CH} + 6_{CH} + 10_{CH} + 8_{CH} + 10_{CH}$$

### 1.3 Estimation procedure

For this sampling scheme the following estimators are considered:

- (a) Horvitz-Thompson estimator
- (b) Post-identification estimator
- (c) Estimator based on sample mean  
~~3.3.1 Horvitz-Thompson estimator~~

For using the Horvitz-Thompson estimator, the inclusion probabilities for individual as well as pair-wise units are required. Before discussing the estimation procedure we may discuss the method of calculating inclusion probabilities for individual and pair-wise units under the suggested sampling scheme.

Let  $\pi_1^t$  be the inclusion probabilities of 1-th s.e.u. from the t-th p.a.u. The inclusion probability  $\pi_1^t$  is given by

$\pi_1^t = \text{Probability of selecting } t\text{-th p.a.u} \times \text{Probability of inclusion of } 1\text{-th s.e.u. in the sample of } n_0 \text{ s.e.u's}$

$$= \frac{1}{H_{CH}} \sum_{s \ni t} \frac{n_0}{\sum_{j \neq t} H_j}$$

R-2652

where  $s$  is the sample of  $n$  p.s.u's and the summation stands for all combination of  $s$  p.s.u's containing the  $t$ -th p.s.u.

The inclusion probability of two s.s.u's from  $t$ -th p.s.u in the sample is as under:

$$\pi_{ij}^t = \frac{1}{n_{C_n}} \sum_{s \ni t} \frac{n_0(n_0-1)}{\sum_{k=1}^n n_k (\sum_{l \in s} n_l - 1)}$$

Also, the inclusion probability of two s.s.u's from different p.s.u's in the sample is as follows:

$$\pi_{ij}^{tt'} = \frac{1}{n_{C_n}} \sum_{s \ni tt'} \frac{n_0(n_0-1)}{\sum_{k=1}^n n_k (\sum_{l \in s} n_l - 1)}$$

In order to understand the calculations, let us consider the example in 3.2.

The all possible samples of 2 p.s.u's out of 6 p.s.u's are  $(1,2)$ ,  $(1,3)$ ,  $(1,4)$ ,  $(2,3)$ ,  $(2,4)$ ,  $(3,4)$  with  $\frac{1}{6}$  as the probability of selection. For calculating the inclusion probability of individual and pair-wise units from set p.s.u we consider the examples  $(1,2)$ ,  $(1,3)$  and  $(1,4)$  containing the first p.s.u with 6, 8 and 6 s.s.u's for selecting 2 s.s.u's at Step III. Thus we get

$$\pi'_{1,2} = \frac{1}{6} \left[ \frac{4}{6} + \frac{4}{6} + \frac{4}{6} \right]$$

$$\pi'_{1,3} = \frac{1}{6} \left[ \frac{4 \times 3}{6 \times 5} + \frac{4 \times 3}{6 \times 7} + \frac{4 \times 3}{6 \times 5} \right]$$

Similarly

$$\pi_{ij}^{t,2} = \frac{1}{6} \left[ \frac{4 \times 3}{6 \times 9} \right]$$

Having known the values of the units in the sample and their inclusion probabilities, the Horvitz-Thompson estimator for population total is given by

$$\hat{Y}_{H.T.} = \sum_{i=1}^n \frac{y_i}{\pi_i}$$

The variance  $\hat{V}_{H.T.}$  is given by

$$V(\hat{Y}) = \sum_{i=1}^n \sum_{j \neq i} (\pi_i \pi_j - \pi_{ij}) \left( \frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2 \quad (3.3.1.1)$$

This expression of variance in the notation of two-stage sampling reduces to:

$$\begin{aligned} & \frac{1}{2} \sum_t \left( \pi_1^t \pi_j^t - \pi_{1j}^t \right) \left( \frac{y_1^t}{\pi_1^t} - \frac{y_j^t}{\pi_j^t} \right)^2 \\ & t \neq t^* \\ & t = 1, 2, \dots, n \\ & j = 1, 2, \dots, n \\ & + \frac{1}{2} \sum_{t=1}^n \sum_{1 \neq j} \left( \pi_1^t \pi_j^t - \pi_{1j}^t \right) \times \\ & \quad \left( \frac{y_1^t}{\pi_1^t} - \frac{y_j^t}{\pi_j^t} \right)^2 \end{aligned} \quad (3.3.1.2)$$

Now  $\pi_1^t$ ,  $\pi_j^t$  and  $\pi_{1j}^t$  depend upon  $t$  only and not on  $i$  and  $j$ , we use  $\beta_t$  for  $\pi_1^t$  and  $\beta_{tt}$  for  $\pi_{1j}^t$ . As such the second term of (3.3.1.2) can be written as

$$\frac{1}{2} \sum_{t=1}^n \frac{(\beta_t \beta_t - \beta_{tt})}{\beta_t^2} \sum_{1 \neq j} (y_1^t - y_j^t)^2$$

$$= \frac{1}{2} \sum_{t=1}^N \frac{\theta_t \theta_{t+1} \rho_{tt}}{\theta_t^2} \Delta t \sum_{k=1}^{M_t} (x_k - \bar{x}_t)^2$$

$$= \sum_{t=1}^N \left(1 - \frac{\rho_{tt}}{\theta_t^2}\right) \Delta t \theta_t^2 (M_t - 1)$$

$$\text{putting } \Pi_{tj}^t = \theta_{tt} + \Pi_{tj}^t = \theta_{tt} \Pi_{tj}^{t+1} = \theta_{tt} + \Pi_{tj}^{t+1} = \theta_{tt},$$

Now considering the first term of the  $V(t)$  expression

$$\sum_{\substack{t \neq t_0 \\ t=1,2,\dots,M_t \\ j=1,2,\dots,M_j}} \left( \theta_{tt} + \theta_{tj} \right) \left( \frac{x_1^t}{\theta_{tt}} + \frac{x_j^t}{\theta_{tj}} \right)^2 = \sum_{\substack{t \neq t_0 \\ t=1,2,\dots,M_t \\ j=1,2,\dots,M_j}} \left[ \frac{\theta_{tt} (x_1^t)^2}{\theta_{tt}} \right.$$

$$+ \frac{\theta_{tj}}{\theta_{tj}} (x_j^t)^2 + 2x_1^t x_j^t - \frac{\theta_{tt}}{\theta_{tj}} (x_1^t)^2 \\ - \left. \frac{\theta_{tt}}{2} (x_j^t)^2 + \frac{2\theta_{tt}}{\theta_{tj}} x_1^t x_j^t \right]$$

$$= \sum_{t \neq t_0} \left\{ \left[ M_t \frac{\theta_{tt}}{\theta_{tt}} + M_t \frac{\theta_{tt}}{\theta_{tt}^2} \sum_{k=1}^{M_t} x_k^t \right]^2 + \left[ M_t \frac{\theta_{tt}}{\theta_{tt}} + M_t \frac{\theta_{tt}}{\theta_{tt}^2} \right] \sum_{k=1}^{M_t} x_k^t x_j^t \right. \\ \left. + 2 \sum_{k=1}^{M_t} x_k^t \sum_{j=1}^{M_t} x_j^t \left( 1 - \frac{\theta_{tt}}{\theta_{tj}} \right) \right\}$$

$$\sum_{t=1}^N \left[ \frac{N_t}{a_t} \left( \frac{x_{t1}}{a_t} - \frac{x_t}{a_t} \right) + \frac{N_t}{a_t^2} \left( \frac{x_{t1}^2}{a_t^2} - \frac{x_t^2}{a_t^2} \right) \right] \sum_{j=1}^{N_t} (x_j^t)^2$$

$$\sum_{t=1}^N \left[ \frac{N_t}{a_t} \left( \frac{x_{t1}}{a_t} - \frac{x_t}{a_t} \right) + \frac{N_t}{a_t^2} \left( \frac{x_{t1}^2}{a_t^2} - \frac{x_t^2}{a_t^2} \right) \right] \times$$

$$\sum_{j=1}^{N_t} (x_j^t)^2$$

$$-2 \sum_{t \neq t'} \sum_{j=1}^{N_t} x_{j1}^{t'} \sum_{j=1}^{N_{t'}} x_j^{t'} \left( 1 - \frac{\rho_{tt'}}{a_t a_{t'}} \right)$$

on putting  $x_t = \sum_{j=1}^{N_t} x_j^t$ , the variance expression after algebraic simplification reduces to -

$$\begin{aligned} V(\hat{x}) &= \sum_{t=1}^N \left[ \left[ \frac{N_t}{a_t} \left( \frac{x_{t1}}{a_t} - \frac{\bar{x}_t}{a_t} \right) + N_t \left( 1 - \frac{\rho_{tt}}{a_t^2} \right) \right] \sum_{j=1}^{N_t} (x_j^t)^2 \right. \\ &\quad \left. + \sum_{t=1}^N \sum_{t' \neq t} \left[ \frac{N_t}{a_t} \left( \frac{x_{t1}}{a_t} - \frac{\bar{x}_t}{a_t} \right) \right. \right. \\ &\quad \left. \left. - N_t \left( 1 - \frac{\rho_{tt'}}{a_t a_{t'}} \right) \right] \sum_{j=1}^{N_{t'}} (x_j^{t'})^2 \right. \\ &\quad \left. - 2 \sum_{t \neq t'} x_{t1} x_{t1}^{t'} \left( 1 - \frac{\rho_{tt'}}{a_t a_{t'}} \right) \right] \end{aligned}$$

$$+ 2 \sum_{t=1}^N \left(1 - \frac{\theta_{tt}}{s_t^2}\right) s_t^2 n_t (n_t - 1)$$

$$= 2 \sum_{t=1}^N n_t s_t \left[ \sum_{t=1}^N \frac{(n_t - 1)s_t^2}{s_t^2} + \sum_{t=1}^N \frac{T_t^2}{n_t s_t^2} \right]$$

$$+ 2 \sum_{t=1}^N n_t \left[ \sum_{t=1}^N \frac{(n_t - 1)}{s_t^2} s_t^2 + \sum_{t=1}^N \frac{T_t^2}{n_t s_t^2} \right]$$

$$+ 2 \sum_{t=1}^N n_t \left(1 - \frac{\theta_{tt}}{s_t^2}\right) \left[ (n_t - 1)s_t^2 + \frac{T_t^2}{n_t}\right] = 2 \sum_{t \neq t_0} T_t s_{t_0} (1 - \frac{\theta_{tt_0}}{s_{t_0}^2})$$

$$+ 2 \sum_{t=1}^N \left(1 - \frac{\theta_{tt}}{s_t^2}\right) s_t^2 n_t (n_t - 1)$$

$$v(\hat{x}_d) = 2 \sum_{t=1}^N n_t s_t \left[ \sum_{t=1}^N \frac{(n_t - 1)}{s_t^2} s_t^2 + \sum_{t=1}^N \frac{T_t^2}{n_t s_t^2} \right]$$

$$+ 2 \sum_{t=1}^N n_t \left[ \sum_{t=1}^N \theta_{tt} \left( \frac{n_t - 1}{s_t^2} \right) s_t^2 \right]$$

$$+ \sum_{t=1}^N \theta_{tt} \cdot \frac{T_t^2}{n_t s_t^2} \left[ -2 \sum_{t=1}^N T_t s_t \left(1 - \frac{\theta_{tt}}{s_t^2}\right) \right]$$

$$- 2 \sum_{t \neq t_0} T_t s_{t_0} \left(1 - \frac{\theta_{tt_0}}{s_{t_0}^2}\right)$$

(3.3.1.3)

The use of Yates and Grundy form of the variance estimator with known  $\pi_{1j}$  and  $\pi_{2j}$  presents no problem

**Particular Case** Let  $M_1 = M_2 = M_3 = \dots = M_N = M$ . Then  $a_t = c_t = C_1$

Substituting  $C_1$ ,  $C_2$  and  $C_2'$  for  $a_t$ ,  $b_{tt}$ , and  $b_{tt}'$  respectively in the above expression we get:-

$$\begin{aligned}
 V(\hat{x}) &= 2 \sum_{t=1}^N M C_1 \left[ \sum_{k=1}^N \frac{(k-1) S_k^2}{C_1} + \frac{\sum_{t=1}^N T_t^2}{M C_1} \right] - 2 \sum_{t=1}^N M \left[ \sum_{k=1}^N \frac{C_2(k-1) S_k^2}{C_1^2} \right] \\
 &\quad + C_2 \left[ \sum_{t=1}^N \frac{T_t^2}{M C_1^2} \right] + 2 \sum_{t=1}^N T_t^2 \left( 1 - \frac{C_2}{C_1^2} \right) - 2 \sum_{t=1}^N T_t S_t \left( 1 - \frac{C_2}{C_1^2} \right) \\
 &= 2M(N-1)\left(1 - \frac{C_2}{C_1^2}\right) \sum_{t=1}^N S_t^2 + 2M\left(1 - \frac{C_2}{C_1^2}\right) \sum_{t=1}^N T_t^2 \\
 &\quad - 2\left(1 - \frac{C_2}{C_1^2}\right) \left( \sum_{t=1}^N S_t T_t \right)^2 + 2\left(1 - \frac{C_2}{C_1^2}\right) \sum_{t=1}^N S_t^2 \\
 &\quad + 2\left(1 - \frac{C_2'}{C_1^2}\right) \sum_{t=1}^N T_t^2 \\
 \\ 
 &= 2\left(1 - \frac{C_2}{C_1^2}\right) \left[ M(N-1) \sum_{t=1}^N S_t^2 + \frac{M(N-1)}{N^2} S_0^2 \right] \\
 &\quad + 2 \sum_{t=1}^N T_t^2 \left[ \frac{C_2}{C_1^2} - \frac{C_2'}{C_1^2} \right] \\
 \\ 
 V(\hat{x}) &\approx 2\left(1 - \frac{C_2}{C_1^2}\right) \left[ M(N-1) \sum_{t=1}^N S_t^2 + \frac{(N-1)}{N^2} S_0^2 \right] + \frac{C_2' - C_2}{C_1^2} 2 \sum_{t=1}^N T_t^2
 \end{aligned}$$

### 3.3.2 Post-stratification estimator

As is evident from the discussion in 3.3.1, the Horvitz-Thompson estimator requires the inclusion probabilities for each ultimate sampling unit in the population. The calculations of these probabilities and variance are, clearly, not quite simple. Therefore, the search for an alternative simple estimator is worth attempting. Alternatively, the selected s.e.u's may be identified to which p.s.u. they belong. Let  $n_1$  be the number of s.e.u's selected at step III in a sample of  $n_0$  from the  $N_1$  s.e.u's of the  $i$ -th p.s.u. Here  $n_1$  is a random variable. Now we consider the following estimator for population total.

$$\hat{x}_b = \frac{n_0}{n} \sum_{i=1}^n \frac{n_1 \bar{Y}_i}{N} \quad \dots \dots \dots \quad (3.3.2.1)$$

where  $\bar{Y}_i$  is the sample mean for  $i$ -th p.s.u.

It may be mentioned here that this estimator is of the form used in post-stratified sampling.

The sampling variance of this estimator is given by

$$V(\hat{x}_b) = V_1 E_2 V_3(\hat{Y}) + V_4 V_2 E_3(\hat{Y}) + V_4 E_2 E_3(\hat{Y}) \quad (3.3.2.2)$$

Let us first consider the third term of the above equality,

$$V_4 E_2 E_3 = V_4 E_2 \frac{n_0}{n} \sum_{i=1}^n \frac{n_1 \bar{Y}_i}{N}$$

-3-

$$\text{or } \sigma^2 = V_3 \left[ \frac{n_0}{n} \sum_{i=1}^n \frac{x_i \bar{x}_i}{\bar{x}} \right]$$

$$\text{or } V_3 \left[ \frac{n_0}{n} \sum_{i=1}^n \frac{x_i \bar{x}_i}{\bar{x}} \right] = n_0^2 \left( \frac{1}{n} - \frac{1}{\bar{x}} \right) s_0^2$$

$$\text{where } s_0^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i \bar{x}_i - \bar{x})^2$$

Now considering the second term of (3.3.2.2)

$$E_1 V_2 E_3 (\hat{x}_0) = E_2 V_2 \left( \frac{n_0}{n} \sum_{i=1}^n \frac{x_i \bar{x}_i}{\bar{x}} \right)$$

$V_2$  of the quantity in bracket is zero.

Hence

$$E_1 V_2 E_3 (\hat{x}) \approx 0$$

Lastly, the first term of equation (3.3.2.2) is as given below:-

$$E_1 E_2 V_3 (\hat{x}_0) = E_1 E_2 V_3 \left( \frac{n_0}{n} \sum_{i=1}^n \frac{x_i \bar{x}_i}{\bar{x}} \right) = E_1 E_2 \left[ \frac{n_0^2}{n^2 \bar{x}^2} \sum_{i=1}^n x_i^2 \left( \frac{1}{x_i} - \frac{1}{\bar{x}} \right)^2 s_1^2 \right]$$

$$\text{where } s_1^2 = \frac{1}{n-1} \sum_{j=1}^{n-1} (x_{1j} - \bar{x}_1)^2$$

$$E_1 = \frac{x_0^2}{n^2 M^2} \sum_{i=1}^n x_i^2 \left[ \frac{s_2}{s_0} \frac{1}{s_1} - \frac{1}{s_2} \right] s_2^2 \quad (a)$$

we know  $s_2 \left( \frac{1}{s_1} \right)$  is equal to

$$\frac{\frac{1}{s_1}}{\frac{\sum_{i=1}^n x_i}{\sum_{i=1}^n x_1}} = \frac{\frac{n}{\sum_{i=1}^n x_1}}{s_0^2 \left( \frac{n}{\sum_{i=1}^n x_1} \right)^2}$$

Vide page 95 of statistics  
and probability (1970).

substituting  $s_2 \left( \frac{1}{s_1} \right)$  in (a) above we get:-

$$E_1 = \frac{x_0^2}{n^2 M^2} \sum_{i=1}^n x_i^2 \left[ \frac{\sum_{i=1}^n x_i}{\frac{x_0^2}{s_0^2} s_1} + \frac{\left( \sum_{i=1}^n x_i - \bar{x}_1 \right) \sum_{i=1}^n x_i}{\frac{x_0^2}{s_0^2} M_1^2} - \frac{1}{s_2^2} \right]$$

$$E_1 = \frac{x_0^2}{n^2 M^2} \left[ \sum_{i=1}^n x_i^2 \frac{s_2^2 \left( \sum_{i=1}^n x_i \right)^2}{s_0^2} + \sum_{i=1}^n x_i s_1^2 \sum_{i=1}^n x_i \frac{(s_1 - 1)}{s_0^2} - \sum_{i=1}^n x_i s_1^2 s_2^2 \right]$$

$$E_1 = \frac{x_0^2}{n^2 M^2} \left[ \frac{1}{s_0^2} \sum_{i=1}^n x_i^2 s_1^2 + \frac{1}{s_0^2} \sum_{i=1}^n x_i^2 s_j^2 \right]$$

$$+ \left[ \sum_{i=1}^n s_i^2 \sum_{j \neq i}^n \frac{n_1 n_j}{2} + \frac{(n-1)}{s_0^2} \left( \sum_{i=1}^n s_i^2 n_1^2 + \sum_{i,j}^n n_1 n_j n_i^2 + \sum_{i=1}^n n_1 s_i^2 \right) \right]$$

$$+ \frac{s_0^2}{n^2 \bar{y}^2} \left\{ \frac{n}{n s_0^2} \sum_{i=1}^n n_1^2 s_i^2 + \frac{n(n-1)}{s_0^2 n(n-1)} \sum_{i,j}^n n_1^2 s_i^2 \right\}$$

$$+ \frac{2n(n-1)}{s_0^2 n(n-1)} \sum_{1 \neq j \neq k}^n s_i^2 n_1 n_j + \frac{n(n-1)(n-2)}{s_0^2 n(n-1)(n-2)} \sum_{1 \neq j \neq k}^n s_i^2 n_j n_k$$

$$+ \frac{n_0(n-1)}{s_0^2} \left[ \frac{n}{n} \sum_{i=1}^n s_i^2 n_1^2 + \frac{n(n-1)}{n(n-1)} \left\{ \sum_{i=1}^n n_1 s_i^2 \sum_{j=1}^n n_j \right. \right.$$

$$\left. \left. - \sum_{1 \neq i \neq k}^n n_1^2 s_i^2 \right\} \right] = \frac{n}{n} \sum_{i=1}^n n_1 s_i^2$$

$$+ \frac{s_0^2}{n^2 \bar{y}^2} \left[ \sum_{i=1}^n n_1^2 s_i^2 + \frac{n}{n s_0^2} \left\{ \frac{n_0(n-n_0-2)(n-1)}{n-1} \right\} \right]$$

$$+ \sum_{i=1}^n n_1 s_i^2 \left\{ \frac{n(n-1)(n_0+1) n_0}{n(n-1) s_0^2} + \frac{n}{n} \right\} + \frac{n(n-1)}{s_0^2 n(n-1)} \sum_{i=1}^n n_1^2 \sum_{j=1}^n s_i^2$$

$$+ \frac{n(n-1)(n-2)}{s_0^2 n(n-1)(n-2)} \sum_{1 \neq j \neq k}^n s_i^2 n_j n_k$$

On consolidating the results of the three terms, the variance of the table is given by

$$V(\hat{Y}) = n_0^2 \left( \frac{1}{n} - \frac{1}{N} \right) s_0^2 + \frac{n^2}{n^2 N^2} \left[ \sum_{i=1}^N x_i^2 s_i^2 \frac{n}{n s_0^2} \left[ \frac{s_0(n-n) \cdot 2(n-1)}{n-1} \right] \right]$$

$$+ \sum_{i=1}^N n_1 s_i^2 \left[ \frac{n(n-1)(n_0+1)}{n(n-1)s_0^2} n_0 = \frac{n}{N} \right] + \frac{n(n-1)}{s_0^2 n(n-1)} \sum_{i=1}^N x_i^2 \frac{n}{n s_1^2} \sum_{i=1}^N s_i^2$$

$$+ \frac{n(n-1)(n-2)}{\frac{2}{s_0^2 n(n-1)(n-2)}} \sum_{i=1}^N s_i^2 n_1 n_k$$

$$+ n_0^2 \left( \frac{1}{n} - \frac{1}{N} \right) s_0^2 + \frac{n_0^2}{n^2 N^2} \left[ \sum_{i=1}^N n_1^2 s_i^2 \frac{n}{n s_0^2} \left[ \frac{s_0(n-n) \cdot 2(n-1)}{n-1} \right] \right]$$

$$+ \sum_{i=1}^N n_1 s_i^2 \left[ \frac{n(n-1)(n_0+1)}{n(n-1)s_0^2} n_0 = \frac{n}{N} \right] + \frac{n(n-1)}{s_0^2 n(n-1)} \sum_{i=1}^N n_1^2 \sum_{j=1}^N s_j^2$$

$$+ \frac{n(n-1)(n-2)}{\frac{2}{s_0^2 n(n-1)(n-2)}} \left[ \sum_{i=1}^N s_i^2 \left( \sum_{j=1}^N x_{ij} \right)^2 - 2 \sum_{i=1}^N n_1 s_i^2 \sum_{j=1}^N n_{ij} \right]$$

$$+ \sum_{i=1}^N s_i^2 \sum_{j=1}^N n_{ij}^2 + 2 \sum_{i=1}^N s_i^2 n_1^2$$

$$\begin{aligned}
 V(\hat{x}_b) &= n_0^2 \left( \frac{1}{n} - \frac{1}{N} \right) s_b^2 + \frac{s_b^2}{n^2 N^2} \left[ \frac{n(n-a)}{n_0^2 n(n-1)(n-2)} \sum_{i=1}^n n_i^2 s_i^2 \right. \\
 &\quad \left. + \frac{n(n-1)(n-a)}{n_0^2 n(n-1)(n-2)} \sum_{i=1}^n n_i^2 \sum_{j=1}^n s_j^2 \right] \\
 &\quad + \left[ n_0 \frac{n(n-1)}{n_0^2 n(n-1)(n-2)} \left[ (n-2)(n_0+1)-2(n-2) \right] - \frac{n}{n} \right] \sum_{i=1}^n n_i^2 s_i^2 \\
 &\quad + \frac{n(n-1)(n-2)}{n_0^2 n(n-1)(n-2)} \sum_{i=1}^n s_i^2 \left( \frac{n}{\sum_{i=1}^n n_i} \right)^2
 \end{aligned}$$

The unbiased variance estimator in the present case following the theory of Post-stratified sampling takes the form

$$= n^2 \left[ \left( \frac{1}{n} - \frac{1}{N} \right) s_b^2 + \frac{1}{nN} \sum_{i=1}^n \frac{n_i^2}{N} \left( \frac{1}{n_i} - \frac{1}{N} \right) s_i^2 \right]$$

### 3.3.3 Estimator based on sample mean

We have seen that the Horvitz-Thompson estimator involves calculation of inclusion probabilities for individual and pair-wise units making the computation lengthy whereas the post-identification of the selected s.a.u's. We consider in this section another estimator which is based on the sample mean of the units selected.

The proposed estimator is

$$\hat{Y}_c = \frac{N}{n} \bar{Y}_n \bar{Y}_{n_0} \dots \dots \dots \quad (343.3.1)$$

Estimator  
requires  
the  
identifi-  
cation

where  $E_p = \sum_{i=1}^n E_i$ ,  $E_i$  being the number of s.e.u's in the  $i$ -th p.s.u.

and

$$\bar{Y}_{m_0} = \frac{1}{m_0} \sum_{i=1}^n \sum_{j=1}^{m_i} y_{ij} , \text{ the mean of } m_0 \text{ units.}$$

It can be seen easily that (3.3.3.1) is unbiased for the population total.

The sampling variance of this estimator is given by

$$V(\hat{T}_0) = V_1 V_2(\hat{T}) + V_1 E_2(\hat{T}_0) \quad (3.3.3.2)$$

First considerin: the second term of (3.3.3.2)

$$\begin{aligned} V_1 E_2(\hat{T}) &= V_1 V_2 \left( \frac{N}{n} E_n \bar{Y}_{m_0} \right) \\ &= V_1 \left( \frac{N}{n} E_n \bar{Y}_{m_0} \right) = V_1 \left( \frac{N}{n} \sum_{i=1}^n Y_{i*} \right) \end{aligned}$$

where  $Y_i = \sum_{j=1}^{m_i} y_{ij}$  ; and  $\bar{Y}_{m_0} = \frac{1}{m_0} \sum_{i=1}^n Y_i$  , i.e.

the mean of the psu's selected.

Hence

$$V_1 \left( \frac{N}{n} \sum_{i=1}^n Y_i \right) = N^2 \left( \frac{1}{n} - \frac{1}{N} \right) \cdot \frac{1}{N-1} \sum_{i=1}^n \left( Y_i - \frac{1}{N} \sum_{i=1}^N Y_i \right)^2 \quad \dots (A)$$

Now considering the first term of (3.3.3.2) we gets-

$$E_1 V_2 \left( \frac{S}{n} \bar{Y}_{M_0} \bar{Y}_{M_0} \right) = E_1 \left[ \frac{n^2}{n^2} n_n^2 \left( \frac{1}{n_0} - \frac{1}{M_0} \right) \frac{1}{M_0} \sum_{i=1}^n \sum_{j=1}^{M_1} (y_{ij} - \bar{Y}_{M_0})^2 \right]$$

$$= E_1 \left[ \frac{n^2 n_0 (M_0 - n_0)}{n^2 n_0 (M_0 - 1)} \sum_{i=1}^n \sum_{j=1}^{M_1} (y_{ij} - \bar{Y}_{M_0})^2 \right]$$

$$= E_1 \left[ \frac{n^2 n_0 (M_0 - n_0)}{n^2 n_0 (M_0 - 1)} \left[ \sum_{i=1}^n \sum_{j=1}^{M_1} (y_{ij} - \bar{Y}_{i*})^2 + \sum_{i=1}^n \sum_{j=1}^{M_1} (\bar{Y}_{i*} - \bar{Y}_{M_0})^2 \right] \right]$$

where  $\bar{Y}_{i*} = \frac{1}{M_1} \sum_{j=1}^{M_1} y_{ij}$ , the mean of  $i$ -th psu

$$= E_1 \left[ \frac{n^2}{n^2} \frac{M_0 (M_0 - n_0)}{n_0 (M_0 - 1)} \left[ \sum_{i=1}^n (M_1 - 1) s_i^2 + \sum_{i=1}^n n_i (\bar{Y}_i - \bar{Y}_{M_0})^2 \right] \right]$$

Taking  $\frac{M_0 - n_0}{M_0 - 1} \leq 1$  we get

$$= E_1 \left[ \frac{n^2}{n^2 n_0} \left[ M_0 \sum_{i=1}^n (M_1 - 1) s_i^2 + M_0 \sum_{i=1}^n \frac{s_i^2}{M_1} + \left( \sum_{i=1}^n Y_i \right)^2 \right] \right]$$

Taking  $M_1 - 1 = M_1$

$$= E_1 \frac{n^2}{n^2 n_0} \left[ \sum_{i=1}^n M_1 \sum_{j=1}^n r_{ij} s_i^2 + \sum_{i=1}^n M_1 \sum_{j=1}^n \frac{r_{ij}^2}{M_1} - \left( \sum_{i=1}^n Y_i \right)^2 \right]$$

$$= E_1 \left[ \frac{n^2}{n^2 n_0} \left[ \sum_{i=1}^n M_1^2 s_i^2 + \sum_{i=1}^n M_1 M_j s_i^2 + \sum_{i=1}^n r_{ij}^2 + \sum_{i=1}^n \frac{r_{ij}^2}{M_1} \right] \right]$$

$$\left[ - \sum_{i=1}^n Y_i^2 - \sum_{i \neq j}^n Y_i Y_j \right]$$

$$= \frac{n^2}{n^2 m_0} \left[ n \sum_{i=1}^n n_1^2 s_1^2 + \frac{n(n-1)}{n(n-1)} \sum_{i \neq j}^n n_1 n_2 s_1^2 s_2^2 \right]$$

$$+ \frac{n(n-1)}{n(n-1)} \sum_{i \neq j}^n n_1 n_2 s_1^2 s_2^2 - \frac{n}{n} \sum_{i \neq j}^n Y_i Y_j$$

$$= \frac{n^2}{n^2 m_0} \left[ \frac{(n-n)}{(n-1)n} \sum_{i=1}^n n_1^2 s_1^2 + \frac{n(n-1)}{n(n-1)} \sum_{i=1}^n n_1 s_1^2 \sum_{i=1}^n n_1 \right]$$

$$+ \frac{n(n-1)}{n(n-1)} \sum_{i=1}^n n_1 \sum_{i=1}^n Y_i^2 = \frac{n(n-1)}{n(n-1)} \left( \sum_{i=1}^n Y_i \right)^2$$

$$= \frac{n}{n m_0 (n-1)} \left[ (n-n) \sum_{i=1}^n n_1^2 s_1^2 + (n-1) \left[ \sum_{i=1}^n n_1 s_1^2 \sum_{i=1}^n n_1 \right] \right]$$

$$+ \left[ \sum_{i=1}^n n_1 \sum_{i=1}^n Y_i^2 - \left( \sum_{i=1}^n Y_i \right)^2 \right] \quad \dots (B)$$

Combining (A) and (B) we get

$$V(Y_C) = n^2 \left( \frac{1}{n} + \frac{1}{n} \right) \frac{1}{n-1} \sum_{i=1}^n \left( Y_i - \frac{1}{n} \sum_{i=1}^n Y_i \right)^2 + \frac{n}{n m_0 (n-1)} \left[ (n-n) \sum_{i=1}^n n_1^2 s_1^2 + (n-1) \left[ \sum_{i=1}^n n_1 s_1^2 \sum_{i=1}^n n_1 + \sum_{i=1}^n n_1 \sum_{i=1}^n Y_i^2 - \left( \sum_{i=1}^n Y_i \right)^2 \right] \right]$$

The estimate of the variance is given by

$$\begin{aligned}
 &= N^2 \left[ \left( \frac{1}{n} - \frac{1}{N} \right) s_0^2 + \frac{1}{n^2} M_n^2 \left( \frac{1}{m_0} - \frac{1}{M_n} \right) \frac{1}{n-1} \sum_{i=1}^n \sum_{j=1}^{m_i} (y_{ij} - \bar{y}_{m_0})^2 \right] \\
 &= N^2 \left[ \left( \frac{1}{n} - \frac{1}{N} \right) s_0^2 + \frac{1}{n} \sum_{i=1}^n M_i^2 \left( \frac{1}{m_i} - \frac{1}{M_n} \right) s_i^2 \right. \\
 &\quad \left. + \frac{M_n^2}{n^2} \left( \frac{1}{m_0} - \frac{1}{M_n} \right) \frac{1}{m_0-1} \sum_{i=1}^n \sum_{j=1}^{m_i} (y_{ij} - \bar{y}_{m_0})^2 \right]
 \end{aligned}$$

... (3.3.3.4)

#### Numerical Illustration

It is difficult to give a general comparison of different estimators considered in this chapter. However, in order to compare the different procedures discussed in this chapter and in Chapter II, two populations have been considered which are being presented in Table 3.1. These populations are such that the variance within p.s.u.'s increases with the number of s.s.u's. Further, the variability in the values of the s.s.u's for second population is more than that of population 1.

Table 3.1  
Table showing two populations

Population	Value of the s.s.u's in each of the p.s.u's			
	1	2	3	4
1	2, 3	3, 4, 5, 6	4, 5, 6, 6, 7, 8	5, 5, 6, 8, 8, 9, 10, 12
2	2, 4	3, 5, 6, 10	3, 5, 8, 9, 10, 13	1, 3, 5, 6, 7, 8, 10, 16

The variances of the estimator of population total obtained for various estimators and also of different methods of allocation for  $n = 3$  and  $n_0 \geq 10$  are given in Table 3.2.

Table 3.2  
Variances of the estimator

Estimators	Population	
	1	2
Nerwitz-Thompson Estimators	312.16	437.64
Post-Identification Estimators	921.00	1122.68
Estimator based on sample mean	1105.08	1469.78
Usual estimator under optimum allocation	876.22	1874.50
Usual estimator under proportional allocation	903.27	908.30
Usual estimator under constant allocation	957.39	1099.40

It is evident from the above results that the Nerwitz-Thompson estimator is the best of all the estimators considered.

for these two populations and the gain in efficiency is substantial. The next best is the usual estimator under optimum allocation procedure which takes into account the variability within p.s.u's in deciding the number of s.s.u's to be drawn from different p.s.u's.

It may be mentioned that the estimator based on optimum allocation assumes the knowledge of the variability within p.s.u's whereas the use of Horvitz-Thompson estimator involves the computation of inclusion probabilities for individual as well as pairwise units. Generally, the knowledge about the variability is not readily available and as such the method of optimum allocation cannot be applied always. If the computational facilities are available, the values of inclusion probabilities can be calculated and in that case the Horvitz-Thompson estimator can be used with considerable advantage.

### 3.5 SUMMARY

In this chapter a new sampling scheme for two stage sampling has been suggested which provides always a sample of fixed size. Three different types of estimators have been proposed for the new sampling scheme. The efficiencies of these estimators among themselves and also with the allocation procedures of Chapter II have been investigated for two small populations. It has been observed that the Horvitz-Thompson estimator performs the best of all the procedures considered for two populations.

## CHAPTER

In two-stage sampling design, the primary sampling units are generally of unequal size. For the choice of number of s.s.u's to be drawn from different p.s.u's, the most common method is to take them either equal or in proportion to their sizes. The drawback in taking s.s.u's equal in number from each p.s.u. is that the size and the variability of s.s.u's in different p.s.u's which are important for efficient estimation, are ignored. The method of taking the number of s.s.u's proportionate to the total number of s.s.u's in the p.s.u's is also not desirable because it renders the ultimate sampling size a random variable which poses serious problems from field operation point of view.

In this dissertation, the problem of allocation and selection of a sample of fixed size in two stage sampling has been attempted.

In Chapter II, two procedures of allocation of s.s.u's to the selected p.s.u's have been suggested. The efficiencies of these procedures as compared to constant allocation procedures have been studied theoretically as well as empirically. It has been observed that of these, optimum allocation procedure is superior to the other two, followed by proportional allocation procedure. Further, the gain in efficiency is also substantial.

In Chapter III, a new sampling scheme of selection of a sample of fixed number of s.s.u's has been suggested. Under this sampling scheme, three estimators namely Horvitz-Thompson estimator, Post-Identification estimator and estimator based on sample mean are considered.

The relative efficiencies of these estimators together with the usual estimators under different allocation procedures have been investigated empirically on two small populations and it has been observed that the performance of the Horvitz-Thompson estimator is the best of all estimates for these two populations.

## REFERENCES

1. Cochran, G.W. (1963), "Sampling Techniques", John Wiley & Sons, Inc., New York.
2. Durbin, J. (1967), "Design of multi-stage surveys for the estimation of sampling errors", *Applied statistics*, Vol. 16, 152-164.
3. Hansen, M.H. & Hurwitz, W.H. (1943), "On the theory of sampling from finite populations" *Ann. of Math. Stat.*, Vol. 14, 333-362.
4. Hansen, M.H., and Hurwitz, W.H. (1949), "On the determination of the optimum probabilities in sampling". *Ann. of Math. Stat.*, Vol. 20, 426-432.
5. Horvitz, D.G. and Thompson, J., (1952) "Generalization of sampling without replacement from a finite universe", *Jour. of Amer. Stat. Assoc.*, Vol. 47, 663-685.
6. Mahalanobis, P.C. (1946), "Sample surveys of crop yield in India", *Sankhya*, 7, 269-280.
7. Mahalanobis, P.C. (1952), "One aspects of the design of sample surveys", *Sankhya*, 12, 1-7.
8. Murthy, M.N. (1967), "Sampling theory and methods" Statistical Publishing Society.
9. Neyman, J. (1934), "On the two different aspects of the representative method. - A method of stratified sampling and the method of purposive sampling", *Jour. of Roy. Stat. Soc.*, 97, 558-625.
10. Ghosh, M. (1956), "On efficiency of cluster sampling", *Jour. of Ind. Acad. of Agric. Stat.*, 8, 45-59.
11. Sukhatme, P.V. (1967), "The problem of plot size in large scale yield surveys", *Jour. of Amer. Stat. Assoc.*, Vol. 42, 297-310, 460.
12. Sukhatme, P.V. (1950), "Sample surveys in agriculture", Presidential Address to the Statistics Section, 37th Session, Ind. Sci. Congr., Poona.
13. Sukhatme, P.V. (1953), "Measurement of observational errors in surveys", *Rev. of the International Stat. Inst.*, 20,(2-3), 121-134.

14. Sukhatme, P.V. and Varain, P.D. (1952) "Sampling with replacement", Jour. of Ind. Soc. of Agri. Stat., 4, 42-49.
15. Sukhatme, P.V. and Panse, V.G. (1951) "Crop surveys in India - II", Jour. of Ind. Soc. of Agri. Stat., 3, 97-168.
16. Sukhatme, P.V. and Sukhatme, B.V. (1970), "Sampling theory of surveys with applications", Asia Publishing House.
17. Yates, F. (1949), "Sampling methods for censuses and surveys", Charles Griffin and Co., Ltd., London, Third Edition (1960).
18. Yates, F. and Grundy, P.M. (1953), "Selection without replacement from within strata with probability proportionate to size", Jour. of Roy. Stat. Soc., Series B, Vol. 15, 253-261.