V.8.
V84 V95

ON PROCEDURES OF ROTATION SAMPLING

I. C. SETHI

Dissertation submitted in 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.)

DE-04

INSTITUTE OF AGRICULTURAL RESEARCH STATISTICS
NEW DELHI

1970

ACKNOWLEDGEMENTS

I am highly grateful to Dr. D. Singh, Director, Institute of Agricultural Research Statistics, New Delhi, for his valuable guidance, constant encouragement and immense help in the preparation of this thesis.

I also express my gratitude to Dr. M.N. Das, Senior Professor of Statistics and Dr. G.R. Seth, presently on an FAO assignment in Lebanon, for providing adequate facilities to carry out the work.

My thanks are also due to I.C.A.R. for providing the financial assistance in the form of a fellowship.

Sethi

(I.C. SETHI)

CONTENTS

CHAPTER I

INTRODUCTION

There are many studies, notably in agricultural, sociological and economic research which are concerned with estimating characteristics of a population on 'repeated occasions' in order to measure their trends over time. For example, in poultry surveys, we may be interested not only in studying egg production and estimation of poultry birds but also we may like to know proportion of farms adopting new feed practices and new poultry breeds developed by various scientific organisations. Further we may also be interested in estimating the trend of poultry farms viz., whether they are switching over to recent poultry breeds and feed practices or they are adopting same set of old breeds and feed practices.

In this type of surveys we have to study populations (finite or infinite) on more than one occasion ye for a set of correlated characters. If the survey is properly planned and sampling done in a suitable way, to exploit for estimation purposes the inherent correlation between different characters on different occasions, we may lessen the cost, time, and labour to a considerable extent.

Further in this type of surveys the experimenter is interested in estimating the population characters, estimating the change in values of population characters and estimating the average value of the population characters on all occasions over a given period of time. Under these circumstances experimenter has no other alternative except sampling over successive occasions. Once the decision to study the population over successive occasions is made, several alternatives in designing the sampling plan would arise, viz.,

(i) Choosing a new sample on each occasion,

(ii) retaining the same sample on all occasions, and

(iii) replacing a part of the sample on each occasion.

The relative advantages of the various types of selection procedures will obviously depend on the extent of variability among the units and the variability of changes in these units as well as on the relative importance of the population mean and the changes in these means. In large scale surveys , we are compelled to use multistage sampling designs. The reasons for use of multistage sampling designs are well-known. In the present investigation, we will consider only two stage sampling designs. The results derived can be extended to any number of stages . In a two stage design the various alternatives for drawing sample are of the following forms:

(a) Retaining all the primary stage sampling units (psu's) from occasion to occasion but selecting each time a fresh sample of second stage units (ssu's) from the selected psu's .

(b) Retaining a fraction 'p' of the psu's along with the ssu's in those psu's from occasion to occasion and selecting a fraction of 'q' of psu's afresh such that $p + q = 1$.

(c) Retaining all the psu's from the preceding occasion but keeping only a fraction 'r' of the ssu's and selecting afresh a fraction 's' of ssu's in each psu's such that $r + s = 1$.

(d) Retaining a fraction 'p' of the psu's and from each such psu retain only a fraction 'r' of the ssu's and selecting afresh a fraction 's' of the ssu's such that $p + q = 1$, $r + s = 1$.

(e) Choosing a new sample at each occasion

(f) Retaining same sample from occasion to occasion

The choice of adopting any one of the above procedures is limited partially by the available resources (funds etc.) and partially by the re-

-quirements of experiment. Suppose one is free to alter or retain the composition of the sample and the total sample size is same on all occasions, than according to Cochran (1963) if one intends to maximise the precision, the statements to be made about the replacement policy are:

(I) For estimating the average over all occasions, it is best to draw a new sample on each occasion.

(II) For estimating change, it is best to retain the same sample throughout all occasions, and

(III) For current estimates equal precision is obtained by keeping the sample or by changing it on every occasion.

Replacement of a part of the sample on each occasion may be better than these alternatives.

According to Yates (1960) there are two points which must be borne in mind in connection with sampling on successive occasions. Firstly, repeated survey of the same units may be inexpedient since resistance to the provision of the necessary information may be engendered and secondly repeated survey may result in modification of these units relative to the rest of population. Further some units may tend to be less co-operative if they are visited consistently. These errors are reduced if we rotate the sample.

Rotation of units have got following advantages over non-rotating samples :

1. Rotation spreads the burden of reporting among more respondents,

2. Rotation permits the use of data from past samples to improve current estimate. This can be easily done by means of the composite estimation procedures.

3. Rotation may afford an unbiased solution to the problem of large observations which occur in the sample.

The first, advantage, that of spreading the burden of reporting in a sample survey among more respondents, may be very important from the standpoint of maintaining the rate of response. In multistage surveys rotation may be of following forms:

(α) Rotation of sample over all stages,

(β) Rotation of sample over some of the stages including last stage,

(γ) Rotation of sample over last stage only.

Since response depends only on ultimate units which come into picture only at last stage, as a result of which, rotation at last stage is essential. All alternative forms discussed above will result in solving our problems. Obviously out of these forms we will prefer case (γ), since it involves less complications which result, when we rotate the sample.

Hence taking these points into consideration, here is an attempt to estimate a set of characters using two stage sampling design involving rotation on one or both stages. Further optimum sample size for a given cost function is also worked out. An attempt is also made to estimate value of non-responding units using knowledge of previous occasions.

CHAPTER II

REVIEW OF LITERATURE

It was Jesson (1942) who first studied the theory of sampling on successive occasions with partial replacement of units on each occasion. His study was limited to only two occasions. He obtained two independent estimates of the mean on the second occasion, one being the sample mean based on new units only and the other a regression estimate based on the units common to both occasions. These two estimates were weighted with inverse of their variances to get an estimate of the mean on the second occasion with minimum variance. In addition an over all sample mean was also obtained on the first occasion.

Yates (1949) gave a simplified method for estimating the values of the mean on successive occasions by treating each occasion separately. He considered two cases, (i) when the sample on the second occasion was a subsample of the original sample and (ii) when the sample retained was supplemented with a fresh sample on the second occasion. Yates extended Jessen's results for the study of one character on two occasions to h occasions under restrictive conditions of a constant sample size and a fixed replacement fraction at each occasion. He assumed the variability on different occasions and the correlation 'ρ' between consecutive occasions as constant. Assuming further, the correlation between i th and j th occasion to be $\rho^{|i-j|}$, he obtained the relation,

$$\bar{Y}_h = (1 - \phi_h) \left[\bar{Y}'_h + \rho (\bar{Y}_{h-1} - \bar{Y}'_{h-1}) \right] + \phi_h \bar{Y}''_h \qquad \ldots (2.1)$$

where

$\bar{Y}_h$ = precise estimate obtained for h th occasion, utilising all information upto and including the h th occasion,

$\bar{Y}_{h-1}$ = similar estimate for the previous i.e. (h-1) th occasion.

$\bar{Y}'_h$ = mean of units common to h th and (h-1)th occasion.

$\bar{Y}'_{h-1}$ = mean of units common to (h-1) th and its previous occasion.

$\bar{Y}''_h$ = mean of units taken afresh in h th occasion

ϕ_h depends on correlation (ρ), the fraction replaced 'q' on each occasion and the number of occasions 'h'. As 'h' increases ϕ_h rapidly tends to a limiting value which depends on 'ρ' and 'q'. He also established the recurrence relationship between ϕ_h and ϕ_{h-1} as

$$(\phi_h / \phi_{h-1}) = (q/p)(1-\rho^2) + \rho^2 \phi_{h-1} \qquad \ldots. \quad (2.2)$$

where $p + q = 1$.

Patterson's (1950) approach to the problem of sampling was different. He obtained the estimate as a linear function of a set of variates and developed a set of conditions for his estimate to be most efficient. Using these conditions he obtained an efficient estimate of the population mean on the h th occasion which is the same as (2.1) worked out by Yates. The recurrence relationship between ϕ_h and ϕ_{h-1} as obtained by him was

$$1 - \phi_h = \frac{p}{1 - (q-p)\rho^2 - p\rho^2(1-\phi_{h-1})} \qquad \ldots \quad (2.3)$$

which was the same as that established by Yates (vide 2.2). Patterson further put the recurrence relationship in another form as

$$(1-\phi_h)(1-\phi_{h-1}) - (\alpha+\beta)(1-\phi_h) + \alpha\beta = 0 \qquad \ldots (2.4)$$

where α and β are the roots of the quadratic equation obtained by putting $\phi_h = \phi_{h-1} = \phi$. He proved that with increasing h, $1-\phi_h$ tends numerically to the smallest root of the quadratic

$p\phi\rho^2 + \phi(1-\rho^2) - q(1-\rho^2) = 0$ and thus obtained the limiting value of ϕ as

$$\phi = \frac{-(1-\rho^2) + \sqrt{(1-\rho^2)\left[1-\rho^2(1-4pq)\right]}}{2p\rho^2} \qquad \ldots. (2.5)$$

where $p + q = 1$.

Patterson also gave an efficient estimate of the difference between the mean on th h th occasion and that on the (h-1) th occasion. The case where the sample size varies from occasion was also considered by him.

Tikkiwal (1953) studied the problem following a more general approach. He considered the correlations between units drawn on successive occasions to vary assuming that correlations follow a product model. According to him $\rho_{ij} = \prod_{t=1}^{j-1} \rho_{t,\ t+1}$ where $1 \leq i \leq j \leq h$, and ρ_{ij} is the correlation between the same set of units on i th and j th occasion. When correlations between consecutive occasions were assumed to be equal on all occasions, he proved that with limiting ϕ, the replacement fraction to be effected on different occasions tends to half from above.

Eckler (1955) developed a method of rotation sampling to obtain a minimum variance estimate of the population value (mean and total) by suitably constructing a linear function of sampling values at different times.

Hansen, M.H. and others (1955) developed another rotation plan and estimate making use of, what is referred as a composite estimate.

Their estimate for each item was a composite estimate or weighted average of rotated units and retained units. Further they used ratio estimate considering previous occasions as ancillary variates. Their results were not confined to uni-stage sampling only, but they considered two stage sampling design also.

Kathuria (1959) Studied the problem of successive sampling for two stage sampling design. Further he estimated optimum sample size for a given cost function.

Woodruff (1963) developed two procedures to increase the reliability of estimates from rotating sample. One was the composite estimation procedure which in effect weights results, from past and present rotating panels to produce an efficient estimate for the current period. The other procedure was the large observation procedure, in which unusually large observations from two or more panels are enumerated cureently with reduced weights . The latter procedure is particularly effective in reducing occasion to occasion change.

Fellgi(1963) developed the estimate for sampling over successive occasions with varying probabilities without replacement. He developed these estimates for rotating as well as non-rotating sampling designs.

Rao, and Graham (1964) developed a general rotation pattern for estimation of current level and change in level between consecutive occasions. For estimation purpose they used a composite estimate of the form

$$\bar{x}'_{o} = Q\,(\bar{x}'_{-1} + \bar{x}_{o,-1} - \bar{x}_{-1,o}) + (1-Q)\bar{x}_{o} \qquad \ldots . (2.6)$$

where Q is a constant weight factor with $0 \leq Q \leq 1$, $\bar{x}_0$ is the estimator based on the entire sample for the current occasion, O; $\bar{x}_{0,-1}$ is the estimator for the current occasion but based on the sample segment common, to both occasion O and -1 (previous occasion), $\bar{x}_{-1,0}$ is the estimator for previous occasion, -1, but based on the sample segment common to the occasion D and -1, $\bar{x}'_{-1}$ is the composite estimator for the previous occasion -1. Their composite estimator of the change, $\bar{X}_0 - \bar{X}_{-1}$ is

$$d'_0 = \bar{x}'_0 - \bar{x}'_{-1} \qquad \ldots. \quad (2.7)$$

which can also be put as

$$d'_0 = Q(\bar{x}_{0,-1} - \bar{x}_{-1,0}) + (1-Q)(\bar{x}_0 - \bar{x}'_{-1}) \qquad \ldots \quad (2.8)$$

Singh, D. (1968) studied the replacement policy for primary stage units from first occasion to second occasion while the same second stage units were observed for the selected first stage units and he presented the optimum design for two and three occasions. He further observed that for estimating the mean on third occasion it would be preferable to repeat the same sample fraction from one occasion to the next, while for estimating the mean over all occasions the sample fraction repeated on the second occasion should not be repeated on the third occasion but in its place a sub-sample of the sample selected on the second occasion should be retained.

Singh, D. and Kathuria (1969) studied the problem of successive sampling with partial replacement of units in a multistage design. They obtained estimates of the population mean (i) on the second occasion and

(ii) on the h th occasion, under the following two systems of replacement:

(a) partially replacing psu's and keeping ssu's fixed.

(b) keeping psu's fixed and partially replacing ssu's.

In generalising the results for h occasions, the pattern of variability between psu's and ssu's was assumed to be constant on all occasions.

Abraham, et al (1969) made a study similar to that made by Singh, D. and Kathuria (1969) for multistage design with a different replacement plan. Their replacement plan was :

Retain a fraction p of the psu's and from each such psu retain only a fraction r of the ssu's and select a fraction s of the ssu's afresh where p+q = 1 and r +s = 1.

Their estimate was limited to only two occasions.

T. Singh, D. Shivtar (1970) also made a study similar to that of Singh, D. and Kathuria (1969) but in generalising the results for h occasions, the patterns of variability between psu's and ssu's was not assumed to be constant over all occasions.

In the present investigation and attempt has been made to obtain m-inimum variance linear unbiased estimates for a set of correlated characters for (i) the population mean on the most recent occasion, (ii) the change in the population mean from one occasion to another, and (iii) an overall estimate of the population mean over all occasions. The sampling design considered is a two stage sampling design.

Two cases of rotation were considered, namely

1. Rotation on both stages ,

2. Rotation on second stage only.

Further in case 2, two different replacement policies were considered.

(a) Replacing a part q of the psu's and keeping a part p of the total psu's from occasion to occasion, where $p + q = 1$.

(b) Keeping a part p of the psu's same throughout all occasions but selecting entirely different part, q of psu's from one occasion to another occasion.

The rotation pattern considered is general rotation pattern.

Further optimum sample size for a given cost function is also worked out. An attempt has also been made to estimate the value of non-responding unit using knowledge of previous occasions.

CHAPTER III

SAMPLING PLAN I :- TWO STAGE ROTATION

In the present chapter an attempt is made to extend Rao and Graham's general rotation pattern to multistage design, namely, two stage sampling design with rotation on both stages.

1. Rotation Pattern:- Assume that the actual units in the population remain unchanged in time, and N and n the number of psu's in population and in the sample respectively (which are the same for all occasions). Let us (for simplicity) assume that in each psu we have L ssu's and l ssu's are selected on all occasions. Also, let N and n be multiples of n_2 ($n_2 \geqslant 1$) and L and l be multiples of l_2 ($l_2 \geqslant 1$). Where n_2 and l_2 ard units to be rotated on first and second stage respectively. Then the rotation pattern is as follows.

A group of n_2 psu's stays in the sample for r_1 occasions ($n = n_2 r_1$) leaves the sample for m_1 occasions and so on. Further in a retained psu a group of l_2 ssu's stays for r_2 occasions ($l = l_2 r_2$, $r_2 \leq r_1$), leaves the sample for m_2 occasions and so on. If a psu returns to the sample after having dropped out (k-1) previous times from the sample, we say the psu is in the k th cycle. We only consider the case $m_1 \geqslant r_1$ here ; since the case $m_1 < r_1$ is more complicated, further the case $m_1 < r_1$ is not usable in large scale surveys due to limited resources. The maximum value of m_1 is $r_1(N/n -1)$ and, if m_1 is less than the maximum value, it amounts to covering only a fraction, viz., $n_2 m_1 + n$ of the N ps units in the rotating design. The same thing holds for second stage units within a ps unit.

Illustration: - Let $N = 5$, $n = r_1 = 2$, $m_1 = 3$, $L = 5$, $l = r_2 = 2$, $m_2 = 3$, also $n_2 = l_2 = 1$, then first of all in each psu we will arrange ssu with the help of random number table and relebel them as 1, 2, 3, 4, 5, etc. (this is contrary to Rao and Graham's pattern since it ensures randomisation which was not considered by them), same thing is repeated for other ss units within other psu's. Then we relebel psu's in the same fashion. Then the sample on each occasion is as given in fig. 3.1.1.

From Fig. 3.1.1, we find that after an interval of NL occasions same units will be repeated.

2. The Composite Estimators $\bar{x}_h^M$ and $\bar{d}_h^M$: - Let $x_{a,tj}$ denote the vector of u characters under study for the j th ss unit in t th ps unit on a th occasion ($a = 1, 2, \ldots, h$; $j = 1, 2, \ldots, L$, $t = 1, 2, \ldots, N$) where h denotes the last occasion at which sampling has taken place.

Also,

$$x_{a,tj} = \begin{bmatrix} x_{a,1tj} \\ x_{a,2tj} \\ \cdot \\ \cdot \\ \cdot \\ x_{a,utj} \end{bmatrix}$$

Here $x_{a,itj}$ denotes the value of i th character for j th ss unit in t th ps unit. Then the unbiased composite estimator of the population mean vector, $\bar{X}_h$, is

$$\hat{\bar{X}}_h^M = Q\,(\bar{x}_{h-1}^M + \bar{x}_{h,h-1} - \bar{x}_{h-1,h}) + (I-Q)\,\bar{x}_h \qquad \ldots. (3.2.1)$$

where

$$\bar{x}_{h,h-1} = \sum_{t=1}^{n_1} \sum_{j=1}^{l_1} x_{h,tj}/n_1 l_1, \quad \bar{x}_{h-1,h} = \sum_{t=1}^{n_1} \sum_{j=1}^{l_1} x_{h-1,tj} / n_1 l_1$$

These $\bar{x}_{h,h-1}$ and $\bar{x}_{h-1,h}$ are mean vectors at h th and h-1 th occasion, respectively, for units common to h th and h-1 th occasion.

and $\bar{x}_h = \sum_{t=1}^{n} \sum_{l=1}^{l} x_{h,\,tj} / nl$,

Further $\bar{x}^M_{h-1}$ is the composite estimator for h-1 th occasion, also Q is a constant weightage factor matrix, whose elements lie between zero and one. Here, $l_1 n_1 = (r_1 - 1)(r_2 - 1) n_2 l_2$ is the number of matched units between i th and (i+1)th occasion, (i=1, . . . , h-1). The composite estimator of the population change, from (h-1) th occasion to h th occasion, $\bar{X}_h - \bar{X}_{h-1}$, is

$$\bar{d}^M_h = \bar{x}^M_h - \bar{x}^M_{h-1} \qquad \ldots\ldots \qquad (3.2.2)$$

The composite estimator (3.2.1) is actually of the form

$$\begin{bmatrix} \bar{x}^M_{h1} \\ \bar{x}^M_{h2} \\ \cdot\cdot \\ \cdot\cdot \\ \bar{x}^M_{hu} \end{bmatrix} = \begin{bmatrix} q_{11} & q_{21} \cdots\cdots & q_{u1} \\ q_{12} & q_{22} \cdots\cdots & q_{u2} \\ \cdot\cdot & \cdot\cdot \cdots\cdots & \cdot\cdot \\ \cdot\cdot & \cdot\cdot \cdots\cdots & \cdot\cdot \\ q_{1u} & q_{2u} \cdots\cdots & q_{uu} \end{bmatrix} \begin{bmatrix} \bar{x}^M_{h1} + \bar{x}_{h,h-1\,1} - \bar{x}_{h-1,h1} \\ \bar{x}_{h2} + \bar{x}_{h,h-1\,2} - \bar{x}_{h-1,h2} \\ \cdots\cdots \quad \cdots \\ \cdots\cdots \quad \cdots \\ \bar{x}_{hu} + \bar{x}_{h,h-1u} - \bar{x}_{h-1,hu} \end{bmatrix}$$

$$+ \begin{bmatrix} 1-q_{11} & -q_{21} \cdots\cdots & -q_{u1} \\ -q_{12} & 1-q_{22} \cdots\cdots & \cdot\cdot \\ \cdot\cdot & -q_{23} \cdots\cdots & \cdot\cdot \\ \cdot\cdot & \cdots\cdots\cdots & \cdot\cdot \\ -q_{1u} & -q_{2u} \cdots\cdots & 1-q_{uu} \end{bmatrix} \begin{bmatrix} \bar{x}_{h1} \\ \bar{x}_{h2} \\ \cdot\cdot \\ \cdot\cdot \\ \bar{x}_{hu} \end{bmatrix} \qquad \ldots \quad (3.2.1a)$$

Figure 3.1.1

Occasion →

psu	ssu	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
	1	x										x				x										x	x
	2	x				x										x	x										x
1	3					x	x										x				x						
	4						x				x										x	x					
	5										x	x										x				x	
	1	x										x	x										x				x
	2	x	x										x				x										x
2	3		x				x										x	x									
	4						x	x										x				x					
	5							x				x										x	x				
	1		x										x	x										x			
	2		x	x										x				x									
3	3			x				x										x	x								
	4							x	x										x				x				
	5								x				x										x	x			
	1			x										x	x										x		
	2			x	x										x				x								
4	3				x				x										x	x							
	4								x	x										x				x			
	5									x				x										x	x		
	1				x										x	x										x	
	2				x	x										x				x							
5	3					x				x										x	x						
	4									x		x									x				x		
	5											x			x										x	x	

Now (3.2.1) can also be written as

$$\bar{x}_h^M = \sum_{\alpha=1}^{h} Q^\alpha W_\alpha = \sum_{\alpha=1}^{h} \sum_{t=1}^{N} \sum_{k=1}^{L} w_{\alpha',tk}\, x_{\alpha,tk} \qquad \ldots(3.2.3)$$

$$\text{where } W_\alpha = Q(\bar{x}_{\alpha,\alpha-1} - \bar{x}_{\alpha-1,\alpha}) + (I-Q)\bar{x}_\alpha \qquad \ldots(3.2.4)$$

for $\alpha = 1,\ldots,(h-1)$, $W_h = \bar{x}_h$ and the $w_{\alpha',tk}$ are functions of Q, r_1, r_2, n_2 and l_2. From (3.2.3) and (3.2.4) it may be verified that the weights $w_{\alpha,tk}$ are as follows:

For the current occasion, $\alpha = h$

(a) $w_{h,tk} = (I-Q)/nl$ for n_2 l units (first visit of a cycle of psu's)

(b) $w_{h,tk} = (I-Q)/nl + Q/n_1l_1$ for n_1l_1 units (second to (r_2-1) th visit of ssu's within second to (r_1-1)th visit of psu's)

(c) $w_{h,tk} = (I-Q)/nl$ for n_2l_2 units (first visit of a ssu within psu's of second to (r_1-1) th visit)

(d) $w_{h,tk} = O$ for NL - nl units (not in the sample)

For occasions $\alpha = 1,\ldots,(h-1)$

(a) $w_{\alpha,tk} = Q^\alpha(I-Q)/nl - Q^\alpha/n_1l$ for n_2l units (first visit of a psu)

(b) $w_{\alpha,tk} = Q^\alpha(I-Q)/nl + Q^{\alpha+1}/n_1l_1 - Q^\alpha/n_1l_1$ for $(n_1-n_2)(l_1-l_2)$ units (second to (r_2-1) th visit of ssu within second to (r_1-1) th visit of psu)

(c) $w_{\alpha,tk} = Q^\alpha(I-Q)/nl + Q^{\alpha+1}/n_1l$ for n_2l units (r_2 th visit of ssu's within psu's of r_1 th visit)

(d) $w_{\alpha',tk} = Q^\alpha(I-Q)/nl - Q^\alpha/n_1l_1$ for $(n_1-n_2)\, l_2$ units (first visit of a ssu within second to (r_1-1) th visit of a psu)

(e) $w_{a,tk} = Q^{a}(1-Q)/\,nl - Q^{a+1}/n_1 l_1$ for $(n_1-n_2)\,l_2$ units (last visit of any ssu within second to (r_1-1) th visit of a psu)

(f) $w_{a,tk} = 0$ for $(NL-nl)$ units (not in the sample)

Variance of $\bar{x}_h^{M}$: - We have the variance of vector $\bar{x}_h^{M}$, $V(\bar{x}_h^{M})$ as

$$V(\bar{x}_h^{M}) = E(\bar{x}_h^{M}\ \bar{x}_h^{M'}) - \bar{X}_h\,\bar{X}'_h$$

$$= \sum_{a=1}^{h}\sum_{t=1}^{N}\sum_{k=1}^{L} E(w_{a,tk})\, x_{a,tk}\, x'_{a,tk}\, E(w'_{a,tk})$$

$$+ \sum_{a\neq a'=1}^{h}\sum^{h}\sum_{t=1}^{N}\sum_{k=1}^{L} E(w_{a,tk})\, x_{a,tk}\, x'_{a',tk}\, E(w'_{a',tk})$$

$$+ \sum_{a=1}^{h}\sum_{t\neq t'=1}^{N}\sum^{N}\sum_{k=1}^{L} E(w_{a,tk})\, x_{a,tk}\, x'_{a,t'k}\, E(w'_{a,t'k})$$

$$+ \sum_{a=1}^{h}\sum_{t=1}^{N}\sum_{k\neq k'=1}^{L}\sum^{L} E(w_{a,tk})\, x_{a,tk}\, x'_{a,tk'}\, E(w'_{a,tk'})$$

$$+ \sum_{a\neq a'=1}^{h}\sum^{h}\sum_{t\neq t'=1}^{N}\sum^{N}\sum_{k=1}^{L} E(w_{a,tk})\, x_{a,tk}\, x'_{a',t'k}\, E(w'_{a',t'k})$$

$$+ \sum_{a\neq a'=1}^{h}\sum^{h}\sum_{t\neq t'=1}^{N}\sum^{N}\sum_{k\neq k'=1}^{L}\sum^{L} E(w_{a,tk})\, x_{a,tk}\, x'_{a',t'k'}\, E(w'_{a',t'k'})$$

$$- \bar{x}_h^{M}\ \bar{x}_{h'}^{M}$$

Now since $\sum_{t=1}^{N}\sum_{k=1}^{L} w_{h,tk} = 1$ and $\sum_{t=1}^{N}\sum_{k=1}^{L} w_{a',tk} = 0$ $a > 0$, $a=1\ldots(h-1)$

Therefore, we have

$$V(\bar{x}_h^M) = \sum_{a=1}^{h} \sum_{t=1}^{N} \sum_{k=1}^{L} E(w_{a,tk})\, V(S_{a,tk})\, E(w'_{a,tk})$$

$$+ \sum_{a \neq a'=1}^{h} \sum^{h} \sum_{t=1}^{N} \sum_{k=1}^{L} E(w_{a,tk})\, Cov(S_{a,tk}\,;\, S_{a',tk}) E(w'_{a,tk})$$

$$+ \sum_{a=1}^{h} \sum_{t \neq t'=1}^{N} \sum^{N} \sum_{k=1}^{L} E(w_{a,tk})\, Cov(S_{a,tk}\,;\, S_{a,t'k}) E(w'_{a,t'k})$$

$$+ \sum_{a \neq a'=1}^{h} \sum^{h} \sum_{t=1}^{N} \sum_{k \neq k'=1}^{L} \sum^{L} E(w_{a,tk})\, Cov(S_{a,tk}\,,\, S_{a',tk'})\, E(w'_{a',tk'})$$

$V(S_{a,tk})$ for say $t = i$ and $k = j$ is

$$V(S_{a,ij}) = \left(\frac{1}{nm} - \frac{1}{NM}\right) \begin{bmatrix} (\bar{x}_{a,ij1} - \bar{\bar{x}}_{a,\ldots 1})(\bar{x}_{a,ij1} - \bar{\bar{x}}_{a,\ldots 1})' & (\bar{x}_{a,ij1} - \bar{\bar{x}}_{a,\ldots 1}) & (x_{a,ij2} - \bar{\bar{x}}_{a,\ldots 2})' \ldots\ldots\ldots\ldots \\ (x_{a,ij1} - \bar{x}_{a,\ldots 1}) & (x_{a,ij2} - \bar{\bar{x}}_{a,\ldots 2})'(x_{a,ij2} - \bar{\bar{x}}_{a,\ldots 2}) & (x_{a,ij2} - \bar{x}_{a,\ldots 2})' \ldots\ldots\ldots\ldots \\ \vdots & \vdots & \ldots\ldots \\ \vdots & \vdots & \ldots\ldots \\ (x_{a,iju} - \bar{x}_{a,\ldots u}) & (x_{a,ij1} - \bar{x}_{a,\ldots 1})' & \ldots\ldots\ldots\ldots\ldots\ldots \end{bmatrix}$$

where for $Cov(S_{a,tk}, S_{a',tk})$ we will replace a by a' in one of the part of elements of matrices $V(S_{a,ij})$ and for diagonal elements, we will have

$$(x_{\alpha,ij1} - \bar{x}_{\alpha,..1})(x_{\alpha',ij1} - \bar{x}_{\alpha',..1})$$

In case of Cov ($S_{\alpha,tk}$, $S_{\alpha,t'k}$) these differences will be summed over common units , viz., between α th and (α+1) th occasion the common units are $n_1 l_1$ and between α th and (α+s) th occasion the common units will be $(n - sn_2)(1 - sl_2)$ etc.

In case of say $V(S_{\alpha}, t_k)$ we have the usual S^2_{α} divided into two parts viz., first n_2 l units and rest.

<u>A Simplified Approach :-</u> Since we have,

$$\bar{x}_h^M = Q\,\bar{x}^M_{h-1} + Q(\bar{x}_{h,h-1} - \bar{x}_{h-1,h}) + (I-Q)\,\bar{x}_h$$

sothat we have

$$V(\bar{x}_h^M) = Q\,V(\bar{x}^M_{h-1})\,Q' + Q\,V(\bar{x}_{h,h-1} - \bar{x}_{h-1,h})\,Q'$$

$$+ (I-Q)\;V(\bar{x}_h)(I-Q)' + 2Q\,\mathrm{Cov}(\bar{x}^M_{h-1} ; X\,\bar{x}_{h,h-1})\,Q'$$

$$+ 2Q\,\mathrm{Cov}(\bar{x}^M_{h-1}; \bar{x}_{h-1,h})\,Q' + 2Q\,\mathrm{Cov.}(\bar{x}^M_{h-1}; \bar{x}_h)(I-Q)'$$

$$+ 2Q\,\mathrm{Cov}(\bar{x}_{h,h-1}; \bar{x}_h)(I-Q)' + 2Q\,\mathrm{Cov}(\bar{x}_{h-1,}\bar{x}_h)(I-Q)' \quad \ldots(3.2.5)$$

Now , before h th occasion we will be already having $V(\bar{x}^M_{h-1})$ and we can very easily find remaining variance, covariances matrices.

<u>Note:-</u> For say h th occasion we want to find variance, covariance matrix of estimate of $\bar{X}_h$. We will put it as composed of two S^2's, one S^2_1 will be for deviation of first n_2l units from same overall mean , second S^2_2 will be sum of squares of deviation for remaining n_1l units from same overall mean.

$$S^2_{1h,n_2} = \left(\frac{1}{n1} - \frac{1}{NL}\right)\left(\frac{I-Q}{n_1}\right)\begin{bmatrix} \sum_{t=1}^{n_2}\sum_{k=1}^{1}(x_{h,tk1}-\bar{x}_{h..1})(x_{h,tk1}-\bar{x}_{h..1})' & \sum_{t=1}^{n_2}\sum_{k=1}^{1}(x_{h,tk_1}-\bar{x}_{h,..1})(x_{h,tk_2}-\bar{x}_{h,..2})' & \cdots \\ \sum_{t=1}^{n_2}\sum_{k=1}^{1}(x_{h,tk_1}-\bar{x}_{h..1})(x_{h,tk_2}-\bar{x}_{h..,2})' & \sum_{t=1}^{n_2}\sum_{k=1}^{1}(x_{h,tk_2}-\bar{x}_{h,,,2})(x_{h,tk2}-\bar{x}_{h,..2})' & \cdots \\ \vdots & \vdots & \vdots \\ \sum_{t=1}^{n_2}\sum_{k=1}^{1}(x_{h,tk_1}-\bar{x}_{h..1})(x_{h,tku}-\bar{x}'_{h..u})' & \sum_{t=1}^{n_2}\sum_{k=1}^{1}(x_{h,tk_2}-\bar{x}_{h,..2})(x_{h,tku}-\bar{x}_{h,..u})' & \cdots \end{bmatrix} x\,(I-Q)'/n_1 \qquad \ldots.\ (3.2.6)$$

3. __Estimate of change:__ - We have from (3.2.2) estimate of change as

$$\bar{d}^M_h = \bar{x}^M_h - \bar{x}^M_{h-1}$$

but

$$\bar{x}^M_h = Q\,(\bar{x}^M_{h-1} + \bar{x}_{h,h-1} - \bar{x}_{h-1,h}) + (I-Q)\,\bar{x}_h$$

putting which we have $\bar{d}^M_h$ as

$$\bar{d}^M_h = Q\,(\bar{x}_{h,h-1} - \bar{x}_{h-1,h}) + (I-Q)\,(\bar{x}_h - \bar{x}^M_{h-1}) \qquad \ldots\ (3.3.1)$$

variance of $\bar{d}^M_h$ can be put very easily in form of type (3.2.5)

4. Modification in h th occasion estimate

h th occasion estimate can be further modified on (h+1) th occasion, using knowledge of (h+1) th occasion. In this case we can put our composite estimator as

$$\bar{x}_h^{M_o} = Q\,(\bar{x}^M_{h+1} + \bar{x}_{h,h+1} - \bar{x}_{h+1;h}) + (I-Q)\,\bar{x}^M_h \qquad \ldots\ (3.4.1)$$

5. Optimum Q

This can be very easily get from equation

$$\frac{\partial\,(V(\bar{x}_h^M))}{\partial Q} = 0$$

6 . Univariate Study:

We can shift our results of above multivariate study to univariate case, in which all vectors and matrices will shift to a scalar. Matrix Q will change to one constant weightage factor and I will be obviously one.

CHAPTER IV

SAMPLING PLAN II

One Stage Rotation With Partial Retainment of psu's From One Occasion to Another

Usually when we are having a multistage sampling design, rotation on all stages will result in cumbersome estimates. Further, it will result in increased cost for identification of units and may involve more cost for drawing a sample. As discussed in Chapter I, we do not want ultimate stage units to appear indefinitely, since, if we continue to visit same unit, it tends to be less cooperative after two or three visits. As a result of this, non-response increases. So to compromise both, we draw a sample with usual successive sampling procedure on preceding stages but by general rotation pattern at the last stage. This will hold for all occasions.

In this chapter we will consider a sampling plan for two stage sampling design involving rotation on second stage but involving partial retainment of ps units.

1. The Sampling Plan: - For simplicity, let us consider a population consisting of N psu's and in each psu we have same number L of ssu's. Suppose on first occasion n psu's are selected by simple random sampling without replacement and in each psu l ssu's are drawn by general rotation pattern. On second occasion, let us retain np of these nps units (where $0 < p < 1$) but for retained psu's we retain l_1 ssu's and rotate l_2 ssu's by new ssu's using general rotation pattern such that $l/l_2 = r$. This procedure of selecting samples can be modified from one occasion to the next. This will be clear from following illustration.

Illustration

For example, let us consider a population with N=8, and L=6. Let us draw a sample of size nl where n = 4 and l = 3. For this case we assume p = 0.5, l_1=2 and l_2=1, so that l_1+l_2=l=3. For sample over second occasion we retain np psu's (at random) from n psu's of preceding occasion and select n q psu's at random from remaining (N-n) psu's. SSU's in retained psu's are rotated by general rotation pattern. The samples drawn will be of form shown in figure 4.1.1.

Figure 4.1.1.

1st Occasion		IInd Occasion		IIIrd Occasion		IVth Occasion	
PSU	SSU	PSU	SSU	PSU	SSU	PSU	SSU
1	123	1	234	2	234	2	345
2	123	3	234	4	234	4	345
3	123	5	123	5	234	7	345
4	123	6	123	6	234	8	345

The retained psu's can be best illustrated by Figure 4.1.2

Figure 4.1.2

PSU	I Occasion	II Occasion	III Occasion	IV Occasion
1	x	x		
2	x		x	x
3	x	x		
4	x		x	x
5		x	x	
6		x	x	
7				x
8				x

2. Sampling over h occasions for u characters

An unbiased estimate of the population mean vector for the h th occasion, utilising all the information collected from first to h th occasion, including that on the h th occasion can be written as

$$\bar{x}^{(h)}_{..} = I\,\bar{x}^{(h)}_{1'..} + \sum_{i=1}^{h} \left[D_i\,\bar{x}^{(i)}_{2..} + E_i\,\bar{x}^{(i)}_{1..} - (D_i+E_i)\,\bar{x}^{(i)}_{1'..} \right] \quad \ldots(4.2.1)$$

where

$\bar{x}^{(h)}_{..} = (\bar{x}^{(h)}_{..1}\ \bar{x}^{(h)}_{..2}\ \ldots\ldots\ \bar{x}^{(h)}_{..u})'$ = modified mean vector on h th occasion using information upto, and including h th occasion.

here $\bar{x}^{(h)}_{..j}$ = modified mean per ssu for j th character using information upto, and including h th occasion.

$\bar{x}^{(i)}_{1'..} = (\bar{x}^{(i)}_{1'..1}\ \bar{x}^{(i)}_{1'..2})\ \ldots\ldots\ \bar{x}^{(i)}_{1'..u})'$ = mean vector per ssu on i th occasion for the pnl_2 units which are having psu's common to (i+1) th occasion but are different themselves (i = 1, ... , h)

here $\bar{x}^{(i)}_{1'..j}$ = mean per ssu on i th occasion for j th character based on pnl_2 units which are having same psu's on (i+1) th occasion but are different themselves (i = 1, . . . , h, j = 1, . . . , u)

$\bar{x}^{(i)}_{1..} = (\bar{x}^{(i)}_{1..1}\ \bar{x}^{(i)}_{1..2}\ \ldots\ldots\ \bar{x}^{(i)}_{1...u})'$ = mean vector per ssu on i th occasion for the pnl_1 units which are common to (i+1) th occasion (i = 1, ... , h).

here $\bar{x}^{(i)}_{1..j}$ = mean per ssu on i th occasion for the pnl_1 units which are common to (i+1) th occasion (i = 1, . . . , h, j = 1, ... , u).

$\bar{x}^{(i)}_{2..} = (\bar{x}^{(i)}_{2..1}\ \bar{x}^{(i)}_{2..2}\ \ldots\ldots\ \bar{x}^{(i)}_{2...u})'$ = mean vector per ssu on i th occasion for the qnl units which are not at all common to the (i+1)th occasion (i = 1, . . . , h)

$\bar{x}^{(i)}_{2..j}$ = mean per ssu on i th occasion for j th character for the qnl units which are not at all common to (i+1) th occasion ($i=1,\dots h$, $j = 1,\dots,u$)

Further,

$$D_i = \begin{bmatrix} d_{11i} & d_{12i} & \cdots & d_{1ui} \\ d_{21i} & d_{22i} & \cdots & d_{2ui} \\ \vdots & \vdots & & \vdots \\ d_{u1i} & d_{u2i} & \cdots & d_{uui} \end{bmatrix} \text{ is a } u \times u \text{ constant}$$

weightage matrix such that $0 \leq d_{kti} \leq 1$ ($k = 1, \dots, u$, $t = 1, \dots, u$, $i = 1, \dots, h$)

$$E_i = \begin{bmatrix} e_{11i} & e_{12i} & \cdots & e_{1ui} \\ e_{21i} & e_{22i} & \cdots & e_{2ui} \\ \vdots & \vdots & & \vdots \\ e_{u1i} & e_{u2i} & \cdots & e_{uui} \end{bmatrix} \text{ is a } u \times u$$

constant weightage matrix such that $0 \leq e_{kti} \leq 1$ ($k = 1, \dots, u$, $t = 1, \dots, u$, $i = 1, \dots, h$).

These D_i's and E_i's are so chosen that $V(\bar{x}^{(h)}_{..})$ is minimised. We can put (4.2.1) in another form as

$$\bar{x}^{(h)}_{..} = \sum_{i=1}^{h-1} \left[D_i \bar{x}^{(i)}_{2..} + E_i \bar{x}^{(i)}_{1'..} - (D_i + E_i) \bar{x}^{(i)}_{1'..} \right] + D_h \bar{x}^{(h)}_{2..}$$

$$+ E_h \bar{x}^{(h)}_{1..} + (I - D_h - E_h) \bar{x}^{(h)}_{1..} \qquad \dots (4.2.2)$$

where I is u x u identity matrix, the variance of (4.2.2) can be put as

$$V(\bar{x}_{..}^{(h)}) = D_h \left[\left(\frac{1}{qn} - \frac{1}{N} \right) S_1^{2(h)} + \left(\frac{1}{l} - \frac{1}{L} \right) \frac{S_2^{2(h)}}{nq} \right] D'_h$$

$$+ E_h \left[\left(\frac{1}{np} - \frac{1}{N} \right) S_1^{1(h)} + \left(\frac{1}{l_1} - \frac{1}{L} \right) \frac{S_2^{2(h)}}{np} \right] E'_h$$

$$+ (I - D_h - E_h) \left[\left(\frac{1}{np} - \frac{1}{N} \right) S_1^{1'(h)} + \right.$$

$$\left. \left(\frac{1}{l_2} - \frac{1}{L} \right) \frac{S_2^{1'(h)}}{np} \right] (I - D_h - E_h)' + 2(I - D_h - E_h) \left[\left(\frac{1}{np} - \frac{1}{N} \right) \rho_1^{(h', h-1)} \right.$$

$$\left. S_1^{(h)} S_1^{(h-1)} \right] E'_{h-1} + \sum_{i=1}^{h-1} \left[D_i \left[\left(\frac{1}{qn} - \frac{1}{N} \right) S_1^{2(i)} + \left(\frac{1}{l} - \frac{1}{L} \right) \frac{S_2^{2(i)}}{nq} \right] D'_i \right.$$

$$+ E_i \left[\left(\frac{1}{pn} - \frac{1}{N} \right) S_1^{1(i)} + \left(\frac{1}{l_1} - \frac{1}{L} \right) \frac{S_2^{1(i)}}{np} \right] E'_i + (D_i + E_i)$$

$$\left[\left(\frac{1}{np} - \frac{1}{N} \right) S_1^{1'(i)} + \left(\frac{1}{l_2} - \frac{1}{L} \right) \frac{S_2^{1'(i)}}{np} \right] (D_i + E_i)' +$$

$$2E_i \left[\left(\frac{1}{np} - \frac{1}{N} \right) \rho_1^{(i, i+1)} S_1^{(i)} S_1^{(i+1)} + \left(\frac{1}{l_1} - \frac{1}{L} \right) \right.$$

$$\left. \rho_2^{(i, i+1)} \frac{S_2^{(i)} S_2^{(i+1)}}{np} \right] E'_{(i+1)} - 2(D_i + E_i)$$

$$\left. \left[\left(\frac{1}{np} - \frac{1}{N} \right) \rho_1^{(i', i+1)} S_1^{(i)} S_1^{(i+1)} \right] E'_{i+1} \right] \qquad \ldots\ldots (2\ 4.2.3)$$

where $S_1^{2(t)} = \sum_{i=1}^{N} \frac{(\bar{x}_{i.}^{(t)} - \bar{x}_{..}^{(t)})(\bar{x}_{i.}^{(t)} - \bar{x}_{..}^{(t)})'}{N-1}$ = true mean sum of squares and mean sum of products matrix between psu mean vector on t th occasion ($t = 1, \ldots\ldots, h$)

$$S_2^{2(t)} = \sum_{i=1}^{N} \sum_{j=1}^{L} \frac{(\bar{x}_{ij}^{(t)} - \bar{x}_{i.}^{(t)})(\bar{x}_{ij}^{(t)} - \bar{x}_{i.}^{(t)})'}{N(L-1)}$$ = true mean sum of squares and means sum of products matrix between ssu mean vectors on t th occasion ($t = 1, \ldots, h$)

$$\rho_1^{(t,t')} S_1^{(t)} S_1^{(t')} = \sum_{i=1}^{N} \frac{(\bar{x}_{i.}^{(t)} - \bar{x}_{..}^{(t)})(\bar{x}_{i.}^{(t')} - \bar{x}_{..}^{(t')})'}{N-1}$$ = true covariance matrix between psu mean vectors on t th and t'th occasion ($t \neq t' = 1, \ldots h$)

$$\rho_2^{(t,t')} S_2^{(t)} S_2^{(t')} = \sum_{i=1}^{N} \sum_{j=1}^{L} \frac{(\bar{x}_{ij}^{(t)} - \bar{x}_{i.}^{(t)})(\bar{x}_{ij}^{(t')} - \bar{x}_{i.}^{(t')})'}{N(L-1)}$$ = true covariance matrix between ssu mean vectors on t th and t' th occasion ($t \neq t' = 1, \ldots, h$)

In most of the cases encountered in practice , populations are sufficiently large, in case population is sufficiently large the terms of order $1/N$ and $1/L$ can be ignored and variance of estimate $\bar{x}_{..}^{(h)}$ i.e. (4.2.3) can be put in the form

$$V(\bar{x}_{..}^{(h)}) = \left[\frac{S_1^{1'(h)}}{np} + \frac{S_2^{1'(h)}}{np\, l_2} \right] - 2(D_h + E_h) \left[\frac{S_1^{1'(h)}}{np} + \frac{S_2^{1'(h)}}{np\, l_2} \right]$$
$$+ 2E'_{h-1} \rho_1^{(h,h-1)} \frac{S_1^{(h)} S_1^{(h-1)}}{np} + \sum_{i=1}^{h} \left[D_i \left[\frac{S_1^{2(i)}}{nq} + \frac{S_2^{2(i)}}{nql} \right] D_i' \right.$$
$$+ E_i \left[\frac{S_1^{1(i)}}{np} + \frac{S_2^{1(i)}}{np\, l_1} \right] E'_i + (D_i + E_i) \left[\frac{S_1^{1'(i)}}{np} + \frac{S_2^{1'(i)}}{np\, l_2} \right]$$
$$(D_i + E_i)' + 2E_i \left[\rho_1^{(i,i+1)} \frac{S_1^{(i)} S_1^{(i+1)}}{np} + \rho_2^{(i,i+1)} \frac{S_2^{(i)} S_2^{(i+1)}}{np\, l_1} \right] E'_{(i+1)}$$
$$\left. + 2(D_i + E_i) \rho_1^{(i',i+1)} \frac{S_1^{(i')} S_2^{(i+1)}}{np} E'_{(i+1)} \right] \quad \ldots. \quad (4.2.4)$$

Now, if for simplicity we put

$$a^{2(t)} = S_1^{2(t)} + S_2^{2(t)} / 1$$

$$a^{1(t)} = S_1^{1(t)} + S_2^{1(t)} / 1_1$$

$$a^{1'(t)} = S_1^{1'(t)} + S_2^{1'(t)} / 1_2$$

$\beta_{t,t_1} = \rho_1^{(t,t_1)} S_1^{(t)} S_1^{(t_1)} + \rho_2^{(t,t_1)} S_2^{(t)} S_2^{(t_1)} / 1_1$

and $\beta_{t,t'_2} = \rho_1^{(t,t'_2)} S_1^{(t)} S_1^{(t'_2)}$

where $t \neq t_1 \neq t'_2 = 1,\ldots,h$

Putting these in (4.2.4) we have

$$V(\bar{x}_{..}^{(h)}) = \frac{a^{1(h)}}{np} - 2(D_h + E_h)\frac{a^{1'(h)}}{np} + 2E'_{h-1}\frac{\beta_{h,h-1}}{np}$$

$$+ \sum_{i=1}^{h} \left[D_i \frac{a^{2(i)}}{nq} D'_i + E_i \frac{a^{1(i)}}{np} E'_i + (D_i + E_i)\frac{a^{1'(i)}}{np} \right.$$

$$\left. (D_i + E_i)' + 2E_i \frac{\beta_{i,i+1}}{np} E'_{i+1} + 2(D_i + E_i)\frac{\beta_{i',(i+1)}}{np} E'_{(i+1)} \right] \quad \ldots (4.2.5)$$

We can put (4.2.5) as

$$V(\bar{x}_{..}^{(h)}) = \frac{\Delta_1}{np} + \frac{\Delta_2}{nq} \quad \ldots.(4.2.6)$$

where

$$\Delta_1 = a^{1(h)} - 2(D_h + E_h)\, a^{1'(h)} + 2E'_{h-1}\beta_{h,h-1'}$$

$$+ \sum_{i=1}^{h} \left[E_i\, a^{1(i)} E'_i + (D_i + E_i)\, a^{1'(i)} (D_i + E_i)' \right.$$

$$\left. + 2E_i \beta_{i,i+1} E'_{i+1} - 2(D_i + E_i)\beta'_{i,i+1} E'_{i+1} \right]$$

and

$$\Delta_2 = \sum_{i=1}^{h} D_i \; a^{2(i)} \; D'_i$$

or (4.2.6) can be put as

$$V(\bar{x}_{..}^{(h)}) = \Delta^*_1/p + \Delta^*_2/q \qquad \ldots . \; (4.2.6a)$$

where

$$\Delta^*_1 = \Delta_1/n \qquad \text{and} \qquad \Delta^*_2 = \Delta_2/n$$

Now we have to use that value of p for which $V(\bar{x}_{..}^{(h)})$ is minimum , since $V(\bar{x}_{..}^{(h)})$ is a u x u variance, covariance matrix we cannot find optimum p from equations $\partial V(\bar{x}_{..}^{(h)}) / \partial p = 0$. Further there is no sense in minimising covariances, hence we should either minimise generalised variance or sum of variances in matrix $V(\bar{x}_{..}^{(h)})$ i.e. trace $V(\bar{x}_{..}^{(h)})$. Since generalised variance needs evaluation of determinent of matrix $V(\bar{x}_{..}^{(h)})$, hence for simplicity we will minimise trace $V(\bar{x}_{..}^{(h)})$. We have

$$\text{tr. } V(\bar{x}_{..}^{(h)}) = \text{tr } \Delta^*_1/p + \text{tr } \Delta^*_2/q \qquad \ldots \; (4.2.7)$$

since q = 1-p, hence we put q = 1-p and equate $\partial \text{tr } V(\bar{x}_{..}^{(h)}) / \partial p = 0$, equating $\partial \text{tr } V(\bar{x}_{..}^{(h)}) / \partial p = 0$ we have

$$p = \sqrt{\text{tr. } \Delta^*_1} \Big/ \left(\sqrt{\text{tr } \Delta^*_1} + \sqrt{\text{tr} \Delta^*_2}\right)$$

Putting this value of p as p_o in (4.2.5), we have,

$$V(\bar{x}^{(h)}_{..}) = \frac{a^{1(h)}}{np_o} - (2D_h + E_h) \frac{a^{1'(h)}}{np_o} + \sum_{i=1}^{h} \left(D_i \frac{a^{2(i)}}{nq_o} D_i' \right.$$

$$+ E_i \frac{a^{1(i)}}{np_o} E'_i + (D_i + E_i) \frac{a^{1'(i)}}{np_o} (D_i+E_i)' + 2E_i \frac{\beta_{i,i+1}}{np_o} E'_{i+1}$$

$$- 2 (D_i+E_i) \frac{\beta_{i',i+1}}{np_o} E'_{i+1}) \qquad \ldots\ldots(4.2.5a)$$

where $q_o = 1 - p_o$

Value of constant weightage factor matrices D_i and E_i ($i = 1,\ldots,h$) for which $V(\bar{x}_{..}^{(h)})$ is minimum can be obtained from a set of linear equations given by

$$\partial V(\bar{x}_{..}^{(h)}) / \partial D_i = O \quad i = 1,\ldots,h$$

$$\partial V(\bar{x}_{..}^{(h)}) / \partial E_i = O \quad i = 1,\ldots,h$$

we have this set of linear equations as,

$$(p_o/q_o \; a^{2(i)} + a^{1'(i)}) D'_i + a^{1'(i)} E'_i - \beta_{i',i+1} E'_{i+1} = O \qquad i = 1,\ldots,h-1$$

$$(p_o/q_o \, a^{2(h)} + a^{1'(h)}) D'_h + a^{1'(h)} E'_h = a^{1'(h)}$$

$$a^{1'(i)} D'_i + (a^{1(i)} + a^{1'(i)} - \beta_{i',i+1}) E'_i - \beta_{i',i+1} D'_i - \beta_{i,i+1} E'_{i+1} = O \qquad i = 1,\ldots,h-1$$

$$a^{1'(h)} D'_h + (a^{1(h)} + a^{1'(h)}) E'_h = a^{1'(h)}$$

This set of linear equations can be put in the form

$$TM = Z \qquad \ldots \; (4.2.8)$$

Here T is 2h x 2h matrix, M and Z are 2h x h matrices, where

$M = (D_1 D_2 \ldots\ldots D_h \; E_1 E_2 \ldots\ldots E_h)'$ $\quad Z = (O\, O, \ldots\ldots, O\; 2a^{1'(h)}\; O\, O \ldots a^{1'(h)})'$

and

$$
T = \left[\begin{array}{cccccccccc}
p_e/q_e\, a^{2(1)} + a^{1'(1)} & 0 & 0 & \cdots & 0 & a^{1'(1)} & -\beta_{1'2} & 0 & \cdots & 0 \\
0 & p_e/q_e\, a^{2(2)} + a^{1'(2)} & 0 & \cdots & 0 & 0 & a^{1'(2)} & -\beta_{2'3} & 0 \cdots & 0 \\
\cdot\cdot & \cdot\cdot & \cdot\cdot & \cdot\cdot & \cdots & \cdot\cdot & \cdots & \cdot\cdot & \cdots & \cdot\cdot \\
0 \cdots & \cdots & \cdots & \cdot\cdot & \cdots\, p_e/q_e\, a^{2(h)} + a^{1'(h)} & 0 & \cdot\cdot & & \cdots & a^{1'(h)} \\
a^{1'(1)} & 0 & \cdots & \cdot\cdot & \cdots & a^{1(1)} + a^{1'(1)} - \beta_{1'2} & -\beta_{12} & 0 & \cdots & 0 \\
-\beta_{1'2} & a^{1'(2)} & 0 & \cdot\cdot & \cdots\, 0 & 0 & a^{1(2)} + a^{1'(2)} - \beta_{2'3} - \beta_{23} & & \cdot\cdot & 0 \\
0 & -\beta_{2'3} & \cdot\cdot & \cdot\cdot & \cdots & & \cdots & \cdots & & \cdot\cdot \\
0 & 0 & & & & & & & & \\
0 & 0 & \cdot\cdot & \cdot\cdot & \cdots\; a^{1'(h)} & 0 & \cdots & \cdots & & a^{1(h)} + a^{1'(h)}
\end{array}\right]
$$

From (4.2.8) we can get unique solution of M as $M = T^{-1} Z$ if T is non-singular.

If matrix T is positive definite then **we can find its inverse by the following method.**

Let $T = K(I - T^*)$

where K is some constant, such that any element of T^* lies between zero and one. Hence transform matrix T into a matrix of form $K(I - T^*)$ where $T^* = I - T/K$. Now $T^{-1} = [K(I-T^*)]^{-1} =$

$\frac{1}{k}[I + T^* + T^{*2} + \ldots\ldots]$ so that we have approximate value of T^{-1} as,

$$T^{-1} = (I + T^*)/K \qquad \ldots. (4.2.9)$$

Since $T^* = I - T/K$ hence we have $T^{-1} = [2I - T/K]/K$

Putting this value of T^{-1} in (4.2.8) matrix M can be easily evaluated.

If T is not positive definite but is non-singular then also matrix T of any order can be easily inverted on any digital electronic computer and minimum weightage factor matrices can be easily evaluated. Using these minimum weightage factor matrices we can get minimum linear unbiased estimate of $\bar{X}_{..}^{(h)}$.

3. Estimate of Change :- We can estimate change in population from (h-1) th occasion to h th occasion as

$$\bar{d}^{(h)} = \bar{x}_{..}^{(h)} - \bar{x}_{..}^{(h-1)} \qquad \ldots. (4.3.1)$$

where $\bar{d}^{(h)}$ = estimate of change, and V

$$V(\bar{d}^{(h)}) = V(\bar{x}_{..}^{(h)}) + V(\bar{x}_{..}^{(h-1)}) - 2\,Cov(\bar{x}_{..}^{(h)}, \bar{x}_{..}^{(h-1)}), \qquad \ldots.. (4.3.2)$$

Hence we can estimate change from (h-1) th occasion to h th occasion from (4.3.1) and variance, covariance matrix of this change can be found by using expression (4.3.2). In general we can put estimate of change in population from j th to i th occasion ($i > j$) as

$$\bar{d}^{(i-j)} = \bar{x}_{..}^{(i)} - \bar{x}_{..}^{(j)} \qquad \ldots \quad (4.3.3)$$

4. <u>Overall estimate of mean vector:</u> In some cases we are required to find over all estimate over a given interval i.e. say overall estimate of for k occasion. Then this overall estimate can be put as

$$\bar{x}_{..}^{(h)} = \sum_{t=j}^{h} w_t \bar{x}^{(t)} , \quad t = j, (j+1) \ldots\ldots (j+k-1) , h$$

where $w_j + w_{j+1} \ldots\ldots w_{j+k-1} + w_h = 1$

Here w_t's ($t = j, \ldots, h$) are some suitable weightage matrix depending upon the relative importance of the occasions or they may be so chosen that variance is minimised.

5. <u>Modified Estimate of previous occasion</u>

Previous occasion estimate can be modified on following occasions, by using the knowledge of following occasion, say h th occasions estimate can be put as

$$\bar{x}^{M(h)} = A^* \bar{x}^{(h)} \ldots + (I - A^*) \bar{x}_{..}^{(h+1)} \qquad \ldots . (4.5.1)$$

where A^* is a constant weightage factor matrix, $\bar{x}_{..}^{M(h)}$ is modified mean vector of h th occasion estimate utilising the information of (h+1) th occasion also.

$\bar{x}_{..}^{(h)}$ and $\bar{x}_{..}^{(h+1)}$ are same as for estimate 4.2.1.

6. <u>Particular cases</u>

(i) <u>Two occasions</u> (h = 2). Estimate and variance can be very easily obtained from (4.2.1) and (4.2.3) by putting h = 2.

We will have estimate as

$$\bar{x}^{(2)}_{..} = D_1 \bar{x}^{(1)}_{2..} + E_1 \bar{x}^{(1)}_{1..} - (D_1 + E_1) \bar{x}^{(1)}_{1'..} + (D_2 \bar{x}^{(2)}_{2..} + E_2 \bar{x}^{(2)}_{1..} + (I - D_2 - E_2) \bar{x}^{(1)}_{1'} \quad \ldots\ldots (4.6.1)$$

Same way variance, covariance matrix can also be put. Matrix T will be 4u x 4 u matrix of the form of matrix for h occasions.

(ii) Univariate case

(a) h occasions: In this case all vectors and matrices will reduce to scalar quantities and estimate (4.2.1) etc. can be simplified. Simplification is quite easy, hence need not be put in thesis.

(b) Two occasions (h = 2): In this case also all vectors and matrices will reduce to scalars and we can put h = 2 and get estimate and its variance covariance matrix in a similar manner. We can also minimise variance in the same manner. The minimisation etc. are quite simple and hence need not be discussed. In this case T will be,

$$T = \begin{bmatrix} \frac{p_0}{q_0} a^{2(1)} + a^{1'(1)} & 0 & a^{1'(1)} & -\beta_{1'2} \\ 0 & p_0/q_0 \; a^{2(2)} + a^{1'(2)} & 0 & a^{1'(2)} \\ a^{1'(1)} & 0 & a^{1(1)} + a^{1'(1)} - \beta_{1'2} & 0 \\ -\beta_{1'2} & a^{1'(2)} & 0 & a^{1(2)} + a^{1'(2)} \end{bmatrix}$$

CHAPTER V

SAMPLING PLAN III

One Stage Rotation with a part of PSU's Retained throughout all occasions

In this chapter we will consider a sampling plan similar to that considered in Chapter IV but with the difference in retainment of psu's. In this sampling plan a part np of psu's is kept same from occasion to occasion but ... a fraction nq of psu's is selected afresh at each occasion from remaining N-np psu's. In retained psu's, ssu's are continuously rotated by general rotation pattern. This will be clear from the following illustr-ation:

Illustration: Let us assume that we have a population with N=8 as psu's and for simplicity we assume that each psu contains same number of ssu's say, L = 6 and we want to select a sample of size n = 4 and l = 3 wherein we have p = 1/2 so that np = 2 also $l_1 = 2$ and $l_2 = 1$ so that $l_1+l_2 = 3$. Then in this case the sample on all occasions may be of form shown in fig. 5.1.

Figure 5.1

Ist Occasion		IInd Occasion		IIIrd Occasion		IVth Occasion	
PSU	SSU	PSU	SSU	PSU	SSU	PSU	SSU
2	123	2	234	2	345	2	456
6	123	6	234	6	345	6	456
5	143	8	124	3	235	8	245
7	145	1	135	5	346	7	456

In this case we have 2,6 as common psu's. Position for ssu's in retained psu's will be more clear from figure 5.2

Figure 5.2

Occasion →	1	2	3	4
SSU				
1	x			
2	x	x		
3	x	x	x	
4		x	x	x
5			x	x
6				x

1. Sampling over h occasions for u characters: -

The unbiased linear estimate of mean vector of u characters on h th occasion can be written as

$$\bar{x}^{(h)}_{..} = A\ \bar{x}^{(h)}_{2..} + (I-A)\,\bar{x}^{(h)}_{2'..} \qquad \ldots\ldots (5.1.1)$$

Here A is a u x u constant weightage factor matrix such that $0 \leqslant a_{ij} \leqslant 1$ ($i = 1,\ldots,u$, $j = 1,\ldots,u$) and $\bar{x}^{(h)}_{2'..} = (\bar{x}^{(h)}_{2'..1} \ldots\ldots \bar{x}^{(h)}_{2'..u})' =$ modified mean vector at h th occasion for u characters under study for retained psu's.

where $\bar{x}^{(h)}_{2'..i}$ = modified mean at h th occasion for i th character under study for retained psu's ($i = 1,\ldots,u$) further

$$\bar{x}^{(h)}_{2'..} = Q\left(\bar{x}^{(h-1)}_{2'..} + \bar{x}^{(h,h-1)}_{1..} - \bar{x}^{(h-1,h-1)}_{1..}\right) + (I-Q)\,\bar{x}^{(h)}_{1..} \qquad \ldots (5.1.2)$$

where $\bar{x}^{(h)}_{1..}$ is estimate similar to that in Chapter IV.

Q is a constant weightage factor matrix such that $0 \leqslant q_{ij} \leqslant 1$ ($i = 1,\ldots,u$, $j = 1,\ldots,u$) and $\bar{x}^{(h-1)}_{2'..}$ is modified mean vector at (h-1)th occasion.

also $\bar{x}_{1..}^{(h,h-1)} = (\bar{x}_{1..1}^{(h,h-1)} \quad \bar{x}_{1..2}^{(h,h-1)} \quad \ldots. \bar{x}_{1..u}^{(h,h-1)})'$ = mean vector per ssu on h th occasion for np l_1 units which are common to h th occasion and (j h-1) th occasion. Here

$\bar{x}_{1..j}^{(h,h-1)}$ = mean per ssu on h th occasion for j th character for npl_1 units which are common to h th and (h-1) th occasion.

$\bar{x}_{1..}^{(h-1,h)} = (\bar{x}_{1..1}^{(h-1,h)} \quad \bar{x}_{1..2}^{(h-1,h)} \quad \ldots\ldots. \bar{x}_{1..u}^{(h-1,h)})'$ = mean vector per ssu on (h-1) th occasion for np l_1 units which are common to (h-1) th and h th occasion.

Here $\bar{x}_{1..j}^{(h-1,h)}$ = mean per ssu on (h-1) th occasion for j th character for np l_1 units which are common to h th and (h-1) th occasion.

Putting (5.1.2) in succession in (5.1.1) we will have unbiased linear estimate of u characters at h th occasion as

$$\bar{x}_{..}^{(h)} = A\,\bar{x}_{2..}^{(h)} + (I-A)\sum_{i=o}^{h-1} Q^i \left[(I-Q)\,\bar{x}_{1..}^{(h-i)} + Q\,(\bar{x}_{1..}^{(h-i,h-i-1)} - \bar{x}_{1..}^{(h-i-1,\,h-i)}) \right] + Q^{h-1}\bar{x}_{1..}^{(1)} \quad \ldots\ldots (5.1.3)$$

The variance of this , for large population at both stages (i.e. ignoring fpc at both stages) will be of the form,

$$\begin{aligned} V(\bar{x}_{..}^{(h)}) &= A\,V(\bar{x}_{2..}^{(h)})\,A' + (I-A)\left[\sum_{i=1}^{h-1} Q^i \left[(I-Q)\,V(\bar{x}_{1..}^{(h-i)})(I-Q)' \right.\right. \\ &\quad \left.\left. + Q\left[V(\bar{x}_{1..}^{(h-i,h-i-1)}) + V(\bar{x}_{1..}^{(h-i-1,h-i)}) \right] Q' \right] Q'^i \right](I-A)' \\ &\quad + (I-A)\,Q^{h-1}\,V(\bar{x}_{1..}^{(1)})\,Q'^{h-1}\,(I-A)' + 2(I-A)\left[\sum_{i=o}^{h-1} Q^i \left[(I-Q) \right.\right. \\ &\quad \mathrm{Cov}(\bar{x}_{1..}^{(h-i)}, \bar{x}_{1..}^{(1)}) + Q\,(\mathrm{Cov.}(\bar{x}_{1..}^{(h-i),\ h-i-1)}, \bar{x}_{1..}^{(1)}) \\ &\quad \left.\left. - \mathrm{Cov}(\bar{x}_{1..}^{(h-i-1),\ h-i)}, \bar{x}_{1..}^{(1)}) \right] Q'^{h} \right](I-A)' - 2\,(I-A) \\ &\quad \left[\sum_{i=o}^{h-1} Q^{i+1}\,\mathrm{Cov}(\bar{x}_{1..}^{(h-i,h-i-1)}, \bar{x}_{1..}^{(h-i-1,h-i)})\,Q'^{(i+1)} \right] \\ &\quad (I-A)' + 2(I-A)\left[\sum_{\substack{i=o \\ i\neq j}}^{h-1} \sum_{j=o}^{h-1} Q^i\,(I-Q)\left[\mathrm{Cov.}(\bar{x}_{1..}^{(h-i)}, \bar{x}_{1..}^{(h-j,h-j-1)}) \right.\right. \end{aligned}$$

$$- \mathrm{Cov}\,(\bar{x}_{1..}^{(h-i)}, \bar{x}_{1..}^{(h-j-1,h-j)}) \,\rfloor\, Q'^{(j+1)} \,\rfloor\, (I-A)' - 2(I-A)$$

$$\lfloor \sum_{\substack{i=o \\ i\neq j}}^{h-1} \sum_{j=o}^{h-1} Q^{i+1}\, \mathrm{Cov}\,(\bar{x}_{1..}^{(h-i,h-i-1)}, \bar{x}_{1..}^{(h-j-1),h-j)})\, Q'^{(j+1)} \rfloor (I-A)'$$

$$+ 2(I-A) \lfloor \sum_{\substack{i=o \\ i\neq j}}^{h-1} \sum_{j=o}^{h-1} Q^{i}(I-Q)\,\mathrm{Cov}\,(\bar{x}_{1..}^{(h-i)}, \bar{x}_{1..}^{(h-j)})\,(I-Q)'\,Q'^{i} \rfloor (I-A)' \qquad \ldots.(5.1.4)$$

In further cases also we will assume population to the large on both stages. Since $\bar{x}_{1..}^{(i)}$, $\bar{x}_{1..}^{(i,i-1)}$ are having same set of pn l_1 units common, hence we can always use the approximation,

$$\mathrm{Cov}\,(\bar{x}_{1..}^{(h-i)}, \bar{x}_{1..}^{(h-i,h-i-1)}) = V\,(\bar{x}_{1..}^{(h-i,h-i-1)})$$

$$\mathrm{Cov}\,(\bar{x}_{1..}^{(h-i-1)}, \bar{x}_{1..}^{(h-i-1),(h-i)}) = V\,(\bar{x}_{1..}^{(h-i-1,h-i)}) \text{ and}$$

$$\mathrm{Cov}\,(\bar{x}_{1..}^{(h-i)}, \bar{x}_{1..}^{(h-i-1,h-i)}) = \mathrm{Cov}\,(\bar{x}_{1..}^{(h-i-1)}, \bar{x}_{1..}^{(h-i)})$$

$$= \mathrm{Cov}\,(\bar{x}_{1..}^{(h-i-1)}, \bar{x}_{1..}^{(h-i,h-i-1)}) = \mathrm{Cov}\,(\bar{x}_{1..}^{(h-i,h-i-1)}\, \bar{x}_{1..}^{(h-i-1,h-i)})$$

Putting these in (5.1.4) we have

$$V\,(\bar{x}_{..}^{(h)}) = A\,V\,(\bar{x}_{2..}^{(h)})\,A' + (I-A)\,Q^{h-1}\,V\,(\bar{x}_{1..}^{(I)})\,Q'^{h-1}\,(I-A)'$$

$$+ (I-A) \lfloor \sum_{i=o}^{h-1} Q^{i}\,(I-Q)\,V\,(\bar{x}_{1..}^{(h-i)})\,(I-Q)'\,Q'^{i} \rfloor (I-A')$$

$$+ (I-A) \lfloor \sum_{i=o}^{h-2} Q^{i} \lfloor (2I-Q)\,V\,(\bar{x}_{1..}^{(h-i),h-i-1)})\,Q' \rfloor\, Q'^{i} \rfloor (I-A')$$

$$-(I-A) \lfloor \sum_{i=o}^{h-2} Q^{i+1}\, V\,(\bar{x}_{1..}^{(h-i-1,h-i)})\, Q'^{(i+1)} \rfloor (I-A)'$$

$$+2(I-A) \lfloor \sum_{\substack{i=o \\ i\neq j}}^{h-2} \sum_{j=o}^{h-2} \lfloor Q^{i}\,(I-Q) \lfloor \mathrm{Cov}\,(\bar{x}_{1..}^{(h-i)}, \bar{x}_{1..}^{(h-j,h-j-1)}),$$

$$- \mathrm{Cov}\,(x_{1..}^{(h-i)}, x_{1..}^{(h-j-1,h-j)}) \rfloor\, Q'^{(j+1)} \rfloor (I-A)' + 2(I-A)$$

$$\lfloor \sum_{i=o}^{h-2} Q^{i} \lfloor (I-Q)\,\mathrm{Cov}\,(\bar{x}_{1..}^{(h-i)}, \bar{x}_{1..}^{(1)}) - Q\,\mathrm{Cov}\,(\bar{x}_{1..}^{(h-i-1)}, \bar{x}_{1..}^{(1)}) \rfloor$$

$$Q'^{h} \rfloor (I-A)' + 2(I-A) \lfloor \sum_{\substack{i=o \\ i\neq j}}^{h-2} \sum_{j=o}^{h-2} Q^{i}\,(I-Q)\,\mathrm{Cov}\,(\bar{x}_{1..}^{(h-i)}, \bar{x}_{1..}^{(h-j)})(I-Q)'Q'^{i} \rfloor$$

$$(I-A)' \qquad \ldots.(5.1.5)$$

These covariances (in 5.1.5) will be zero if $|i-j|$ greater than r. Since (r+1) th occasion and first occasion are not having any common unit, further covariance between (r+m+1) th occasion and first occasion is assumed to be zero due to considerable time lag between these occasions. As in Chapter IV, here also we can put

$$V(\bar{x}_{2..}^{(h)}) = \frac{S_1^{2(h)}}{nq} + \frac{S_2^{2(h)}}{nql} \text{ and so on,}$$

further we can ~~p~~put

$$a^{2(h)} = S_1^{2(h)} + S_2^{2(h)} / l \text{ as in Chapter IV,}$$

Putting these in (5.1.5) we have

$$V(\bar{x}_{..}^{(h)}) = A \frac{a^{2(h)}}{nq} A' + (I-A) Q^{h-1} \frac{a^{1(1)}}{np} Q'^{(h-1)} (I-A)'$$

$$+ (I-A) \sum_{i=0}^{h-1} Q^i (I-Q) \frac{a^{1(h-i)}}{np} (I-Q)' Q'^i (I-A)'$$

$$- (I-A) \sum_{i=0}^{h-2} Q^{i+1} \frac{a^{1(h-i-1,h-i)}}{np} Q'^{(i+1)} (I-A)'$$

$$+ 2(I-A) \Big[\sum_{\substack{i=0 \\ i\neq j}}^{h-2} \sum_{j=0}^{h-2} Q^i (I-Q) \Big[\frac{\beta_{(h-i,h-j,h-j-1)}}{np} - \frac{\beta_{(h-i,(h-j,h-j-1))}}{np} \Big] Q'^{(j+1)} \Big) (I-A)' - (I-A)$$

$$\sum_{\substack{i=0 \\ i\neq j}}^{h-2} \sum_{j=0}^{h-2} Q^{(i+1)} \frac{\beta_{((h-i,h-i-1),(h-j,h-j-1))}}{np} Q'^{(j+1)} (I-A)' +$$

$$+ 2(I-A) \sum_{i=0}^{h-2} \Big[(I-Q) \frac{\beta_{(h-i,1)}}{np} - Q \frac{\beta_{(h-i-1,1)}}{np} \Big]$$

$$Q'^{(h-1)} (I-A)' + 2(I-A) \sum_{\substack{i=0 \\ i\neq j}}^{h-2} \sum_{j=0}^{h-2} Q^i (I-Q) \frac{\beta_{(h-i,h-j)}}{np}$$

$$(I-Q)' Q'^i (I-A)' \qquad \ldots\ldots (5.1.6)$$

We can put (5.1.6) as

$$V(\bar{x}_{..}^{(h)}) = A \frac{\Delta 3}{nq} A' + (I-A) \frac{\Delta 4}{np} (I-A)' \qquad \ldots. (5.1.7)$$

where $\Delta_3 = \sigma^{2(h)}$

and Δ_4 = sum of terms with $1/np$ as cofficients in (5.1.6)

Minimum value of matrix Q can be got from $\partial \, tr \, \Delta_4 / \partial Q = 0$.

Putting minimal value of matrix Q and putting $\Delta^*_3 = A \Delta_3 A'/n$

and

$\Delta^*_4 = (I-A) \Delta_4 (I-A)'/n$ we can put (5.1.7) as

$$V(\bar{x}_{..}^{(h)}) = \Delta^*_3/q + \Delta^*_4/p \qquad \ldots. (5.1.7a)$$

for optimum value of p we minimise trace of variance and covariance matrix ($\Delta^*_3/q + \Delta^*_4/p$) so we will get optimum p from equation

$$\partial \, tr \, (\Delta^*_3/q + \Delta^*_4/p) / \partial p = 0$$

The value will be $p_o = \sqrt{tr \, \Delta^*_4} / (\sqrt{tr \Delta^*_3} + \sqrt{tr \Delta^*_4})$

Hence we can put $p = p_o$ in (5.1.7) then we will have

$$V(\bar{x}_{..}^{(h)}) = A \frac{\Delta 3}{nqo} A' + (I-A) (\Delta_4 / npo)(I-A)' \qquad \ldots (5.1.7b)$$

where $qo = (1-p_o)$

we can find optimum weightage factor matrix A from equation

$$\partial V(\bar{x}_{..}^{(h)}) / \partial A = 0.$$

From above equation we have $A' = (p_o/q_o \Delta_3 + \Delta_4)^{-1} \Delta_4 \qquad \ldots (5.1.8)$

Hence we have

$$V(x_{..}^{(h)}) = A_o \Delta_3 A'_o / nqo + (I-A_o) (\Delta_4 / npo)(I-A_o)' \qquad \ldots (5.1.9)$$

Particular Cases

2. **Sampling over two occasions for u characters(h = 2)**

Putting h = 2 in (5.1.3) estimate of population mean vector of u characters at second occasion becomes,

$$\bar{x}_{..}^{(2)} = A\,\bar{x}_{2..}^{(2)} + (I-A)\,Q\,(\bar{x}_{1..}^{(1)} + \bar{x}_{1..}^{(21)} - \bar{x}_{1..}^{(12)}) + (I-A)(I-Q)\,\bar{x}_{1..}^{(2)} \qquad \ldots\ldots (5.2.1)$$

Putting h = 2 in (5.1.6) we get

$$\begin{aligned} V(\bar{x}_{..}^{(2)}) = {} & \frac{A\,a^{2(2)}}{nq} A' + (I-A)Q\,\frac{a^{1(1)}}{np}\,Q'(I-A)' + (I-A)(I-Q)\,\frac{a^{1(2)}}{np} \\ & (I-Q)'(I-A)' + (I-A)\,Q\,(I-Q)\,\frac{a^{1(1)}}{np}\,(I-Q)'Q'(I-A)' \\ & + (I-A)(2I-Q)\,\frac{a^{1(2,1)}}{np}\,Q'(I-A)' - (I-A)\,Q\,\frac{a^{1(1,2)}}{np}\,Q'(I-A)' \\ & + 2(I-A)(I-Q)\,[\,\beta(2,(12))/np - \beta(2,(12))/np\,]\,Q'(I-A)' \\ & -(I-A)Q\,\beta((21),(1,2))/np\,Q'(I-A)' + 2(I-A)\,[\,(I-Q)\,\frac{\beta_{(2,1)}}{np} \\ & - Q\,\frac{a^{1(1)}}{np}\,]\,Q'(I-A)' + 2(I-A)(I-Q)\,\frac{\beta_{(1,2)}}{np}\,(I-Q)'(I-A)' \end{aligned}$$

$$\ldots\ldots (5.2.2)$$

In the same way (5.1.9) will reduce to

$$V(\bar{x}_{..}^{(2)}) = A_o(\Delta_3 / nq_o)A'_o + (I-A_o)\,\frac{\Delta_4}{np_o}\,(I-A_o)' \quad \ldots\ldots (5.2.3)$$

where $\Delta_3 = a^{2(2)}$

and Δ_4 = sum of terms with $1/np$ as coefficient in (5.2.2).

Here $p_o = \sqrt{tr\,\Delta_4} \,/\, (\sqrt{tr\,\Delta_3} + \sqrt{tr\,\Delta_4})$ ..... (5.2.4)

$$A'_o = (p_o/q_o\Delta_3 + \Delta_4)^{-1}\Delta_4 \qquad \ldots\ldots (5.2.5)$$

3. Sampling over h occasions for a single character

The estimate can be modified from (5.1.3) considering all vectors and matrices as scalars.

Putting u = 1 the estimate of population mean at h th occasion becomes,

$$\bar{x}_{..}^{(h)} = a\,\bar{x}_{2\,..}^{(h)} + (1-a)\sum_{i=o}^{h-1} q^{i} \left[(1-q)\,\bar{x}_{1..}^{(h-i)} + q\left(\bar{x}_{1..}^{(h-i,h-i-1)} - \bar{x}_{1..}^{(h-i-1,h-i)}\right)\right] + q^{(h-1)}\,\bar{x}_{1..}^{(1)} \qquad \ldots\ldots (5.3.1)$$

and variance of this estimate is

$$V(\bar{x}_{..}^{(h)}) = a^{2}\,\frac{\alpha^{2(h)}}{nq} + (1-a)^{2}\,q^{2(h-1)}\,\frac{\alpha^{1(1)}}{np} + (1-a)^{2}\sum_{i=o}^{h-1} q^{2i}(1-q)^{2}\,\frac{\alpha^{1(h-i)}}{np} + (1-a)^{2}\sum_{i=o}^{h-2} q^{2i+1}(a-q)\,\frac{\alpha^{1(h-i,h-i-1)}}{np}$$

$$-(1-a)^{2}\sum_{i=o}^{h-2} q^{2i+1}\,\frac{\alpha^{1(h-i-1,h-i)}}{np} + 2(1-a)^{2}\sum_{\substack{i=o\\ i\neq j}}^{h-2}\sum_{j=o}^{h-2} q^{(i+j+1)}$$

$$(1-q)\left[\frac{\beta_{(h-i,(h-j,h-j-1))}}{np} - \frac{\beta_{(h-i,(h-j-1),h-j))}}{np}\right]$$

$$-(1-a)^{2}\sum_{\substack{i=o\\ i\neq j=o}}^{h-2}\sum_{j=o}^{h-2} q^{(i+j+2)}\,\frac{\beta_{((h-i,h-i-1),(h-j,h-j-1))}}{np}$$

$$+2(1-a)^{2}\,q^{h-1}\sum_{i=o}^{h-2}\left[(1-q)\,\frac{\beta_{(h-i,1)}}{np} - q\,\frac{\beta_{(h-i-1,1)}}{np}\right]$$

$$+2(1-a)^{2}\sum_{i=o}^{h-2}\sum_{j=o}^{h-2} q^{2i}(1-q)^{2}\,\frac{\beta_{(h-i,h-j)}}{np} \qquad \ldots\ldots\ldots (5.3.2)$$

In the same way (5.1.9) will reduce to

$$V(\bar{x}_{..}^{(h)}) = a_{o}^{2}\,\frac{\Delta_{3}}{nq_{o}} + (1-a_{o})^{2}\,\Delta_{4}/np_{o} \qquad \ldots(5.3.3)$$

where $\Delta_3 = \alpha^{2(h)}$

and Δ_4 = sum of terms with 1/np as coefficient in (5.3.2)

Here $p_o = \sqrt{\Delta_4^*} / (\sqrt{\Delta_3^*} + \sqrt{\Delta_4^*})$ (5.3.4)

and $q_o = 1 - p_o$

and $a_o = \Delta_4 / (p_o/q_o \, \Delta_3 + \Delta_4)$ ... (5.3.5)

4. Sampling over two occasions for a single character

Putting h = 2 in (5.3.1) estimate of population mean at second occasion becomes,

$$\bar{x}^{(2)} = a \, \bar{x}_{2..}^{(2)} + (1-a)\, q(\bar{x}_{1..}^{(1)} + \bar{x}_{1..}^{(21)} - \bar{x}_{1..}^{(12)}) + (1-a)(1-q)\bar{x}_{1..}^{(2)} \qquad \ldots (5.4.1)$$

Putting h = 2 in (5.3.2) we have

$$V(\bar{x}_{..}^{(2)}) = (1-a)^2 \frac{\alpha^{2(2)}}{nq} + (1-a)^2 q^2 \frac{\alpha^{1(1)}}{np} + (1-a)^2(1-q)^2 \frac{\alpha^{1(2)}}{np}$$

$$+(1-a)^2 q^2 (1-q)^2 \frac{\alpha^{1(1)}}{np} + (1-a)^2 q\,(2-q) \frac{\alpha^{1(2)}}{np} - (1-a)^2$$

$$q^2 \frac{\alpha^{1(1,2)}}{np} + 2(1-a)^2(1-q)q \left[\frac{\beta(2,(21))}{np} - \frac{\alpha\beta(2,(12))}{np} \right]$$

$$- (1-a)^2 q^2 \beta\,((21),(12))/np + 2(1-a)^2 q\,(1-q) \frac{\beta(21)}{np}$$

$$- q \frac{\alpha^{1(1)}}{np}) + 2(1-a)^2(1-q)^2 \beta_{12}/np \qquad \ldots (5.4.2)$$

In the same way (5.3.3) will reduce to

$$V(\bar{x}_{..}^{(2)}) = a_o^2 \, \Delta_3 / nq_o + (1-a_o)^2 \, \Delta_4 / np_o \qquad \ldots (5.4.3)$$

where $\Delta_3 = \alpha^{2(2)}$

and Δ_4 = sum of terms with 1/np as coefficient in (5.4.2).

Here

$$p_o = \sqrt{\Delta_4^*} / (\sqrt{\Delta_3^*} + \sqrt{\Delta_4^*}) \qquad \ldots (5.4.4)$$

$q_o = 1 - p_o$

and

$$a_{\hat{6}} = \Delta_4 / (p_o/q_o \, \Delta_3 + \Delta_4) \qquad \ldots (5.4.5)$$

Note: - Estimate for change and overall estimate are obvious and can be easily modified from above estimates. Hence they are not put in the thesis.

Further Considerations

6 5. Change in sample size - In some cases we may be required to increase or decrease sample size, due to high variance or due to involvment of more cost in survey than money sanctioned. In such a situation estimation procedures will become quite cumbersome for sampling plan I and sampling plan II, but for sampling plan III we can make any change in unmatched portion of sample at any occasion and increase the precision or decrease cost (as may be the demand of situation under consideration). From this we can very well conclude that sampling plan III is more flexible than sampling plan I and II.

6. Estimate of variance : In the above discussed estimates i.e. in Chapter III, IV and V estimate of variance can be easily obtained by putting the estimate for each component of variation.

Say for example, we had

$$V(\bar{x}_{1..}^{(h)}) = S_1^{1(h)} / np + S_2^{1(h)} / np\, l_1 \qquad \ldots (5.6.1)$$

New we know that $\hat{S}_1^{1(h)} = s_1^{1(h)} - s_2^{1(h)} / l_1$ and $\hat{S}_2^{1(h)} = s_2^{1(h)}$ (ignoring fpc). Putting these in (5.6.1) we have

$$V(\bar{x}_{1..}^{(h)}) = s_1^{1(h)} / np \qquad \ldots (5.6.2)$$

If we consider a finite population, than we have

$$V(\bar{x}_{1..}^{(h)}) = \left(\frac{1}{np} - \frac{1}{N}\right) S_1^{1(h)} + \left(\frac{1}{l_1} - \frac{1}{L}\right) S_2^{1(h)} \quad \ldots\ldots (5.6.3)$$

In this case we will have

$$S_2^{1(h)} = s_2^{1(h)} \quad \text{and} \quad S_1^{1(h)} = s_1^{1(h)} - \left(\frac{1}{l} - \frac{1}{L}\right) s_2^{1(h)}$$

Hence we have

$$V(\bar{x}_{2..}^{(h)}) = \left(\frac{1}{np} - \frac{1}{N}\right) s_1^{1(h)} + \left(\frac{1}{l} - \frac{1}{L}\right)\frac{s_2^{1(h)}}{N} \quad \ldots (5.6.4)$$

we can substitute these estimates of variance for different components, in variance of estimates considered, and can very easily write estimate of variance.

7. Unequal second stage units: In previous cases dealt with, we have considered equal second stage units in a given primary stage units. Usually this is not the case with populations encounte ered in practice. In a situation where number of second stage units in given psu's are different, we will draw a sample with general rotation pattern on successive occasions with np psu's retained and say in i th psu we enumerate l_i units of which l_{1i} unit are kept same as in prebious occasion but l_{2i} units are rotated in such a way that $l_i/l_{2i} = r$ for $i = 1, \ldots, N$. In this case all componenets of variation contributing to the total variance will be modified. The modification can be illustrated by taking one of the components of variance, by say $V(\bar{x}_{1..}^{(h)})$.

Now we had in (5.6.3), $V(x_{1..}^{(h)}) = \left(\frac{1}{np} - \frac{1}{N}\right) S_1^{1(h)} + \left(\frac{1}{l_1} - \frac{1}{L}\right) \frac{S_2^{1(h)}}{np}$

In this case we will have

$$V(\bar{x}_{1..}^{(h)}) = \left(\frac{1}{np} - \frac{1}{N}\right) S_1^{1(h)} + \frac{1}{nNp} \sum_{i=1}^{N} \left(\frac{1}{l_i} - \frac{1}{L_i}\right) S_{2i..}^{1(h)} \quad \ldots . (5.7.1)$$

where $S_{2i}^{1(h)} = \Sigma\,(x_{ij} - x_{1.})^2 / L_i = 1,$ here L_i is total number of ssu's in i th psu.

The estimate of variance in this case, can also be got in the manner discussed in 6.

8.. Three Stage Sampling

Upto now we have considered only two stage sampling design but if we are having, say, three stage sampling design the above scheme can be still used after slight modification. In this case we shall keep or replace same first and second stage units but at third stage we will select units by general rotation pattern. The estimate will modify as follows.

For simplicity we consider three stage sampling design with equal number of second stage units in first stage units and also we assume that we have same number of third stage units in all second stage units.

Let us assume that population consists of N psu's and in each psu there are M ssu's and in each ssu we have L third stage units. Let us assume that we have drawn a sample consisting of n psu's and in each psu we enumerate m ssu's and in eahh ssu we enumerate ltsu's . Using same notation as in previous cases we can modify all components of variance. In case of variance of $\bar{x}_{2..}^{(h)}$ modification will be

$$V(\bar{x}_{2..}^{(h)}) = \left(\frac{1}{nq} - \frac{1}{N}\right) S_1^{2(h)} + \left(\frac{1}{m} - \frac{1}{M}\right) \frac{S_2^{2(h)}}{nq} + \left(\frac{1}{l} - \frac{1}{L}\right) \frac{S_3^{2(h)}}{nmq} \quad \ldots . (5.8.1)$$

where $S_1^{2(h)} = \frac{1}{N-1} \sum_{i=1}^{N} \left(x_{i..}^{(h)} - x_{...}^{(h)} \right)^2$

$$S_2^{2(h)} = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{M} \frac{\left(\bar{x}_{ij}^{(h)} - \bar{x}_{i..}^{(h)} \right)^2}{M-1}$$

and $$S_3^{2(h)} = \frac{1}{NM} \sum_{i}^{N} \sum_{j}^{M} \sum_{k}^{L} \left(\frac{\left(x_{ijk}^{(h)} - x_{ij..}^{(h)} \right)^2}{L-1} \right.$$

and estimate of variance of (5.8.1) is

$$V(\bar{x}_{2..}^{(h)}) = \left(\frac{1}{nq} - \frac{1}{N}\right) s_1^{2(h)} + \left(\frac{1}{m} - \frac{1}{M}\right) \frac{s_2^{2(h)}}{Nq} + \left(\frac{1}{l} - \frac{1}{L}\right) \frac{s_3^{2(h)}}{NMq} \quad \ldots . (5.8.3)$$

In the same way other components of variance could be modified. This could be very easily generalised to t stage sampling design.

CHAPTER VI

A COST FUNCTION FOR TWO STAGE SAMPLING DESIGN

Cost and other resources available for collection of data in the sample survey, largely influence the sampling design, and determination of sample size andits distribution.

In large scale surveys, we cannot find exact value for different coefficients of cost function, but we can give some estimates for costs', which will be encountered for different operations. These estimates may not be satisfactory at micro-level, but they should be useful at macro-level. The sampler should either find accurate macro-level estimates of different coefficients or he should use some approximate value for different coefficients to avoid confusion during the survey. This will be very clear from following consideration.

For example, we take the case of travelling cost in a survey, conducted by some organisation. Now, in this case enumerators may use different means to travel upto the experimental units and back, as a result of this, the costs encountered will be different from enumerator to enumerator. But this type of difference can be manipulated by fixing some upper limit, for travelling a given distance.

These days, use of electronic computer for analysis of large scale survey data has become very common. Every estimation procedure takes different amount of time for programming and calculations, based on type and number of calculations to be performed on computer. Say for example, we take the case of calculations to be done in simple random sampling, ratio estimate and pps sampling. We can very easily say that, simple random sampling will take least time for programming and for data analysis on computer, ratio estimate will be more time consuming than simple random sampling and pps will be most time consuming than the other estimates discussed.

Further, during a survey, we may be even find some change in travelling cost or in any other cost, which we might not have considered during planning. These are some very difficult problems , which are not usually taken into account due to practical difficulties.

In the present investigation, we will study aspect of minimising variance for a given cost (i.e. fund sanctioned) , based on some appropriate cost function. There is one serious draw back in any cost study related to " successive sampling", i.e. time and interval is very large , due to which, coefficients for different operations involved in the survey tend to change , as a result of this, we may be even complelled to change values of different coefficients from one occasion to another. Practically it is very difficult to make a cost study taking these things into consideration , though attempts could be made.

In the present investigation, a cost study in 'successive sampling' using two stage sampling design, for the three sampling plans discussed in Chapter III, IV and V has been made . The coefficients for different costs involved in the survey have been assumed to satisfy following assumptions.

1. Travelling cost is proportional to distance travelled by enumerator, and it is same for all enumerators (even if at macro-level , it is proportional to distance travelled by enumerators, than also our problem is not affected, as total travelling cost will not be affected).

2. Computational cost of the survey comes under fixed cost, and has been assumed to be independent of estimation procedure.

3. In the time, during which the surveys is conducted , costs involved for different operations of the survey do not change. In addition , these

costs remain same from one occasion to another.

4. Costs involved for different operations in a survey, are independent of quality of work done by enumerators.

These four assumptions come under only one assumption i.e. during the survey (whether over one or h occasions) average cost (or total cost or cost at macro-level) for different operations' do not change This may not be a realistic assumption in successive sampling, but we usually use it for practical convenience. In a two stage sampling design, we encountered followings costs:

1. Total cost and fixed costs.
2. Costs that vary in proportion to the number of primary units in the sample.
3. Costs that vary in proportion to the number of listing units in the sample.

Hence, taking these things into consideration, an appropriate cost function for the three sampling designs discussed in Chapters III, IV and V for the first occasion, can be put as

$$C^{(1)} = C' + C_o\sqrt{n} + C_1 n + C_2 n\sqrt{l_1} + C_2 n\sqrt{l_2} + C_3 n l_1 + C_3 n l_2 \qquad ..(6.1.1)$$

where

$C^{(1)}$ = Total overhead cost of the survey (inclusive of fixed, overhead, expenditure at headquarters, and on the maintenance of computer, etc.,)

C' = Total fixed, overhead and expenditure at headquarters and as such is independent of sample size and sampling design.

C_o = will depend on the distance to be travelled between first stage units. If n first stage units are selected, total distance to be travelled in one round will be approximately equal to $(A n)^{1/2}$, where A is the area to be covered. If there are R rounds then

$$C_o n^{1/2} = R \ (At)^{1/2}$$

C_1 is the average cost per fsu, included in the sample and includes the cost of selection of first stage unit, locating it, listing etc.

C_2 will depend on the distance to be travelled between second stage units within first stage units. If l_1 second stage units are selected within n first stage units, total distance to be travelled in one round will be approximately equal to $n(\bar{A}\, l_1)^{1/2}$, where $\bar{A}$ is the averagearea to be covered in one psu. If there are R rounds then

$$C_o\, n\, l_1^{1/2} = R\, n\, (\bar{A}\, l_1)^{1/2}$$

C_3 is the average cost per second stage unit included in the sample and includes the cost of selection of second stage unit, locating it, listing etc.

From above cost function we find that on second and following occasions, we will not spent money on selection, identification and listing of retained psu's, and retained ssu's within retained psu's as a result of this, cost function on second occasionwill be as given below

$$C^{(2)} = C' + C_o\sqrt{n} + C_1\, nq + C_2\, n\sqrt{l_1} + C_2\, n\sqrt{l_2} + C_3\, nq\, l_1 + C_3 nq\, l_2 + C_3 npl_2 \qquad \ldots\ldots (6.1.2)$$

This is due to the fact that according to our sampling plans, we are selecting nq new psu's, l_2 new ssu's in np retained psu's. It is obvious that in only new psu's new ssu's will be selected.

Hence, total cost of survey for first two occasions, is got by adding (6.1.1) and (6.1.2), so as a result of this, we have total cost involved in the survey for first two occasions as

$$C^{(1+2)} = 2C' + 2\, C_o\sqrt{n} + C_1 1\, (n+nq) + 2C_2\, n\sqrt{l_1} + 2C_2\, n\sqrt{l_2} + C_3\,(n+nq)\, l_1 + C_3\,(n+nq)\, l_2 + C_3 npl_2 \qquad \ldots\ldots (6.1.3)$$

for sampling plan I $p = n_1 / n$,

also we had $p + q = 1$ (for all plans)

$$\therefore \; np + nq = n$$

as a result of this, we have (6.1.3) as

$$C^{(1+2)} = 2C' + 2C_o\sqrt{n} + 2C_1 n + 2C_2 n\sqrt{l_1} + 2C_2 n\sqrt{l_2} + 2C_3 nl_1 + 2C_3 nl_2 - np(C_1 + C_3 l_1) \quad \ldots\ldots \; (6.1.4)$$

In the same way, we will have total cost for h occasions (by adding costs from first to h th occasion) as

$$C^{(1+2+\ldots\ldots+h)} = hC' + hC_o\sqrt{n} + hC_1 n + hC_2 n\sqrt{l_1} + hC_2 n\sqrt{l_2} + hC_3 n l_1 + hC_3 n l_2 - (h-1)\, np\, (C_1 + C_3 l_s) \quad \ldots\ldots \; (6.1.5)$$

since C' is not depending on sampling design, we can take it on left hand side, and will use following cost function,

$$C = hC_o / n + hC_1 n + hC_2 n / l_1 + hC_2 n / l_2 + hC_3 n_1 + hC_3 n l_2 - (h-1)\, np\, (C_1 + C_3 l_1) \ldots \quad \ldots \; (6.1.6)$$

where $C = C^{(1+2++\ldots\ldots+h)} - hC'$

we can put any variance, covariance matrix of $\bar{x}_{..}^{(h)}$ (for different sampling plans discussed) as

$$V(x_{..}^{(h)}) = \frac{\Delta_3}{np} + \frac{\Delta_4}{npl_1} + \frac{\Delta_5}{npl_2} + \frac{\Delta_6}{nq} + \frac{\Delta_7}{nq_1} \quad \ldots \; (6.1.7)$$

where $\Delta_3, \Delta_4, \Delta_5, \Delta_6, \Delta_7$ are sum of variance, covariance matrices with coefficients $1/np$, $1/npl_1$, $1/npl_2$, $1/nq$, $1/nql$ respectively.

Now, we are required to find optimum value only for four scalars, n, p, l_1, l_2, which, we cannot find uniquely by using different equations of variance, covariance matrices. Hence we either minimise trace of (6.1.7) or generalised variance of (6.1.7) for a given cost. In the present case we will consider minimization of generalised variance only, the same will be applicable, when we use trace instead of generalised variance.

We have generalised variance of (6.1.7) as

$$\left| V(\bar{x}_{..}^{(h)}) \right| = \frac{|\Delta_3|}{np} + \frac{|\Delta_4|}{npl_1} + \frac{|\Delta_5|}{npl_2} + \frac{|\Delta_6|}{nq} + \frac{|\Delta_7|}{nql} \qquad \ldots\ldots (6.1.7a)$$

Here $l = l_1 + l_2$

Hence, we can put $\frac{|\Delta_7|}{nql}$ as $\frac{|\Delta_7|}{nql_1}\left(1 - \frac{l_2}{l_1} + \frac{l_2^2}{l_1^2} - \frac{l_2^3}{l_1^3} + \ldots\ldots\right)$, since in our rotation pattern we have l_2/l_1 as less than or equal to one, hence we can approximate $|\Delta_7| / nql$ by the expression

$\frac{|\Delta_7|}{nql_1} - \frac{|\Delta_7| l_2}{nql_1^2} \quad \frac{|\Delta_7| l_2^2}{nql_1^3}$ (ignoring l_2^3/l_1^3 and higher terms).

Putting above expression in (6.1.7a), we have

$$\left| V(\bar{x}_{..}^{(h)}) \right| = \frac{|\Delta_3|}{np} + \frac{|\Delta_4|}{npl_1} + \frac{|\Delta_5|}{npl_2} + \frac{|\Delta_6|}{nq} + \frac{|\Delta_7|}{nql_1} - \frac{|\Delta_7| l_2}{nql_1^2} + \frac{|\Delta_7| l_2^2}{nql_1^3} \qquad \ldots\ldots (6.1.7b)$$

In further discussion in this chapter, we will not put || bracket but it will be understood that bracket | | is there, and also it will be understood that we are dealing with generalised variance (unless otherwise stated).

From (6.1.7) and (6.1.6) we find that we have to minimise the function, (for h th occasion)

$$F = \frac{\Delta_3}{np} + \frac{\Delta_4}{npl_1} + \frac{\Delta_5}{npl_2} + \frac{\Delta_6}{nq} + \frac{\Delta_7}{nql_1} - \frac{\Delta_7 l_2}{nql_1^2}$$

$$+ \frac{\Delta_7\, l_2^2}{nq\, l_1^3} + \lambda \,(hC_o\sqrt{n} + hC_1 n + hC_2\, n\sqrt{l_1} + hC_2\, n\sqrt{l_2} + hC_3\, nl_1 + hC_3\, nl_2 - (h-1)\, np\, (C_1 + C_3\, l_1) \qquad \ldots\ldots (6.1.8)$$

In multistage sampling design using trial and error methods, we usually get fairly broad optimum sample size, at different stages. Though it serves our purpose but for accurate work we should find optimum sample size, after minimising the function F.

In practice even in second case, optimum sample sizes, at different stages is in fractions, as a result of this, we have to approximate number of sampling units to be enumerated at each stage. This reduces its utility to nearly of same order, as that of trial and error method, but optimum sample size at different stages, is less broad.

Differentiating F partially with respect to n and putting $\partial F/\partial n = 0$ we get

$$\Delta_3/np + \Delta_4/npl_1 + \Delta_5/npl_2 + \Delta_6/nql_1 - \Delta_7 l_2/nql_1^2 + \Delta_7 l_2^2/nql_1^3$$

$$= \lambda \,(\frac{hC_o}{2}\sqrt{n} + hC_1 n + hC_2\, n\sqrt{l_1} + hC_2\, n\sqrt{l_2} + hC_3\, nl_1 + hC_3 nl_2 - (h-1)\, np\, (C_1 + C_3 l_1) \qquad \ldots\ldots (6.1.9)$$

Putting (6.1.9) in (6.1.8) and then differentiating (6.1.8) partially with respect to p, l_1, l_2 (after putting $q = 1-p$) and solving these equ equations (for a given value of h, only) we get optimum values of l_1, l_2, p, n as

$$l_1^{(o)} = [\Delta_4^2/\Delta_7 + 4\Delta_5 - \Delta_4) + G(4\Delta_4 - \Delta_4^2/\Delta_7 - 10\Delta_5)$$

$$+ G^2(\Delta_4^2/\Delta_7 + 6\Delta_5 - 10\Delta_4) + G^3(6\Delta_4 + 6\Delta_3 - 9\Delta_5)$$

$$- 9G^4\Delta_4] / (9\Delta_3 G^4 + 10\Delta_3 G^2 - 4\Delta_3 G + \Delta_3 - \frac{\Delta_4^2\Delta_6}{\Delta_7}) \quad \ldots\ldots (6.2.0)$$

where $G = (\Delta_5 - \Delta_4/3) + (19\Delta_5/8\Delta_4 - \frac{2\Delta_5^2}{27\Delta_4^2} - 1)$ ... (6.2.1)

$$l_2^{(o)} = G\, l_1^{(o)} \quad \ldots (6.2.2)$$

$$p^{(o)} = \Delta_5 - l_1^{(o)3} / (\Delta_5 l_1^{(o)3} + \Delta_7 (2l_2^{(o)3} - l_1^{(o)} l_2^{(o)2})$$

$$\ldots\ldots\ldots (6.2.3)$$

and $n^{(o)1/2} = \dfrac{(h^2 C_o^2 + 4(A)^{1/2} - hC_o}{2A}$ ..... (6.2.4)

where

$$A = hC_1 + hC_2/l_1^{(o)} + hC_2/l_2^{(o)} + hC_3 l_1^{(o)} + hC_3 l_2^{(o)}$$

$$-(h-1)\,p^{(o)}\,(C_1 + C_3 l_1^{(o)}) \quad \ldots\ldots (6.2.5)$$

Using estimated value of $|\Delta_3|$, $|\Delta_4|$, $|\Delta_5|$, $|\Delta_6|$ and Δ_7 we can find optimum sample size.

Note:- When we are using trace $V(\bar{x}_{..}^{(h)})$ $/V(\bar{x}_{..}^{(h)})/$ will be replaced by V trace $V(\bar{x}_{..}^{(h)})$ [since trace and determinant are having same addition property]

CHAPTER VII

ESTIMATION OF VALUE OF NON-RESPONDING UNITS

Sampling and non-sampling errors are responsible for lessening the reliability of results derived from any sample survey. In sampling over successive occasions, we encounter non-response (a part of the non-sampling errors) more often than what we encounter , when a survey is conducted only once. The reasons for this are well known. In case of surveys involving multi-character estimation in successive sampling, non-response by respondent's are of following forms:

1. Non response by respondents over all occasions for all characters;
2. Non response by respondents over all occasions for some of the characters,
3. Non response by respondents after k th occasion (say) for all characters,
4. Non-response by respondents after k th occasion(say) for some of the characters.

Out of the these four cases, we can very well infer that for case 1 it is very difficult to estimate or say anything about non-responding units, except in very restrictive models as developed in Bayesian's decision theory. In other cases, here is an attempt to give an estimate of non-responding units. This type of estimation will hold true only under the following assumptions:

(i) Characters under study are highly correlated with each other,

(ii) Listing units under investigation continue to possess at least same set of characters from occasion to occasion.

(iii) Advancement of technology etc. has got macro-level effect over the non-responding units. By this we mean that, if on an average (say, for example) consumption of chemical fertilizers has gone up then consumption of fertilizers for non-responding units will also increase. This type of assumptions may not appeal when glancing the non-responding units at micro-level but at macro-level this type of assumptions may be quite satisfactory due to aggregation of errors of opposite sign.

Method of Estimation : - Let there be a set of highly correlated characters $x_1, x_2, \ldots, x_u$ under study for which estimate at k th occasion (utilising the information upto and including k th occasion) be $x_1^{(k)}, x_2^{(k)}, \ldots, x_u^{(k)}$. Let these estimates be based on a sample of size n. Let for (k+1) th occasion the corresponding estimates (utilising the information upto and including (k+1) th occasion) be $x_1^{(k+1)}$, $x_2^{(k+1)}, \ldots, x_u^{(k+1)}$ and the estimates of corresponding change (from k th to (k+1) th occasion) be $d_1^{(k+1)}, \ldots, d_u^{(k+1)}$. Let these estimates of (k+1) th occasion be based on a sample of size n_1 for characters with partial response and for other characters sample size is n (for simplicity). Here n_1 is less than n. As a result of this we are required to estimate the value of $n-n_1$ units for which there is non-response (for some or all characters).

1. Model (a) : Let the vector at the k th occasion (for characters under study) be $x^{(k)}$ and vector at (k+1) th occasion be $y^{(k+1)}$ (for characters to be estimated, and estimate of difference vector at (k+1) th occasion be $d^{(k+1)}$, then

$$y^{(k+1)} = x^{(k)} + d^{(k+1)} + \epsilon^{(k+1)} \qquad \ldots .(7.1.1)$$

where $E(\epsilon^{(k+1)}) = \vec{0}$ and $E(e^{(k+1)} \epsilon^{(k+1)'}) = \text{diag.}(\sigma_{11}, \sigma_{22}, \ldots, \sigma_{uu})$

This model cannot be utilised for estimation of value of non-responding units belonging to category 2 since in this actual value of character at the k th occasion should be known.

Case 3 . - We can estimate $(n-n_1)$ units from

$$\hat{y}^{(k+1)}_{n_1+i} = x^{(k)}_{(n_{1+i})} + d^{(k+1)} \quad i = 1,2,\ldots,n-n_1 \quad \ldots (7.1.2)$$

Here $x^{(k)}_{n_{1+i}}$ is the actual value of vector at k th occasion for n_1+i th unit and $d^{(k+1)}$ is the vector of change based on n_1 units, i.e.

$$x^{(k)}_{n_{1+i}} = (x^{(k)}_{1\,(n_{1+i})} , x^{(k)}_{2\,(n_1+i)} , \ldots\ldots x^{(k)}_{u(n_1+i)})'$$

$$d^{(k+1)} = (d_1^{(k+1)} , d_2^{(k+1)} , \ldots\ldots , d_u^{(k+1)})'$$

and
$$\hat{y}^{(k+1)}_{n_1+i} = (x^{(k+1)}_{1(n_1+i)} , x^{(k+1)}_{2(n_1+i)}) , \ldots\ldots , x^{(k+1)}_{u(n_1+i)})'$$

$i = 1, n, \ldots\ldots, n-n_1$

$y^{(k+1)}_{n_1+i}$ is the vector of estimate of u characters for n_1+i th non-responding unit.

Now mean of these non- responding units

$$\hat{y}^{(k+1)} = \sum_{i=1}^{n-n_1} y^{(k+1)}_{n_1+i} \Big/ n - n_1$$

or

$$\hat{y}^{(k+1)} = \sum_{i=1}^{n-n_1} x^{(k)} / n-n_1 + d^{(k+1)}$$

$$\therefore \quad V(\hat{y}^{(k+1)}) = \left(\frac{1}{(n-n_1)} - \frac{1}{N} \right) S^{2(k)} + V(d^{(k+1)})$$

Since in all cases we are dealing with populations which are large in size,

hence
$$V (y^{(k+1)}) = \frac{S^{2(k)}}{n-n_1} + V (x^{(k+1)} - x^{(k)})$$

$$= \frac{S^{2(k)}}{n-n_1} + \frac{S^{2(k+1)}}{n_1} + \frac{S^{2'(k)}}{n_1} - \frac{2\rho S^{(k+1)} S'^{(k)}}{n_1}$$

The only difference between $S^{2(k)}$ and $S^{2(k)}$ is that they refer to sum of square matrices for unit with different means. As usual $S^{2(k)}$ etc. are u x u matrices.

Hence

$$\hat{V}(y^{(k+1)}) = \frac{s^{2(k)}}{n - n_1} + \frac{s^{2(k+1)}}{n_1} + \frac{s^{2'(k)}}{n_1} - \frac{2\rho s^{(k+1)} s^{(k)}}{n_1} \qquad \ldots (7.1.4)$$

<u>Case 4:</u> - In this case let us assume (*for simplicity*) that first v characters are recorded but there is non-response for last u-v characters by $n-n_1$ respondents, then also we can estimate these u-v characters by the formula,

$$\hat{y}^{(k+1)}_{n_1+i} = x^{(k)}_{n_1+i} + d^{(k+1)} \qquad i = 1, 2, \ldots, n-n_1 \qquad \ldots (7.1.5)$$

where

$$x^{(k)}_{n_1+i} = (x^{(k)}_{v+1\,(n_1+i)} \quad \ldots\ldots\ldots \quad x^{(k)}_{u\,(n_1+i)})'$$

$$\hat{y}^{(k+1)}_{(n_1+i)} = (x^{(k+1)}_{v+1\,(n_1+i)} \quad , \ldots\ldots\ldots\ldots \quad x^{(k+1)}_{u(n_1+i)})'$$

and $d^{(k+1)} = (d^{(k+1)}_{v+1} \quad , \ldots\ldots , \quad d^{(k+1)}_{u})'$

variance covariance matrix of (7.1.5) can be estimated by the method similar to that discussed for case 3. It will be of form similar to that of (7.1.4) , with the only difference that it will be u-v x u - v variance covariance matrix.

2. Model (b): - Let $\overset{*}{x}^{(k+1)}$ be the mean vector based on n_1 units at (k+1) th occasion, then we have

$$\hat{y}^{(k+1)} = x^{*(k+1)} \left(y^{*(k)} \; x^{(k)} \right) + \epsilon \qquad \ldots\ldots (7.2.1)$$

where $\underset{n_1 \to \infty}{p \lim} E(\epsilon) \longrightarrow \vec{o}$ and $E(\epsilon\epsilon') = \text{diag}(\sigma_{11} \sigma_{22}, \ldots. \sigma_{uu})$

Here

$$y^{*(k)} = \left(w_1 / x_1^{*(k)}, w_2/x^*{}_2^{(k)}, \ldots\ldots, w_u/x_u{}^{*(k)} \right)'$$

$$x^{*(k+1)} = \left(x^*{}_1^{(k+1)}, x_2{}^{*(k+1)}, \ldots., x^*{}_u^{(k+1)} \right)'$$

$$x^{(k)} = \left(x_1^{(k)}, x_2^{(k)}, \ldots\ldots., x_u^{(k)} \right)'$$

where $w_1 + w_2 + \ldots\ldots + w_u = 1$

Case 3 In this case also we assume that we have to find an estimate for last u-v characters for $(n-n_1)$ units, then we estimate them by the formula

$$\hat{y}^{(k+1)}_{(n_1+i)} = x^{*(k+1)} \left(y^{*(k)} \; x^{(k)}_{(n_1+i)} \right) \; i = 1, \ldots., (n-n_1) \qquad \ldots (7.2.2)$$

where

$$\hat{y}^{(k+1)}_{(n_1+i} = \left(\hat{x}^{(k+1)}_{v+1(n_1+i)}, \hat{x}^{(k+1)}_{v+2(n_1+i)}, \ldots., \hat{x}^{(k+1)}_{u(n_1+i)} \right)'$$

$$x^{*(k+1)} = \left(x^{*(k+1)}_{(v+1)}, x^{*(k+1)}_{v+2}, \ldots\ldots, x^{*(k+1)}_u \right)'$$

$$x^{(k)}_{(n_1+1)} = \left(x^{(k)}_{1(n_1+i}, x^{(k)}_{2(n_1+i)}, \ldots\ldots\ldots, x^{(k)}_{v(n_1+i)} \right)'$$

and $y^* = \left(w_1 / x_1{}^{*(k)}, w_2/x^*{}_2^{(k)}, \ldots\ldots, w_v/x_v{}^{*(k)} \right)$ where $w_1, w_2, \ldots, w_v$ are constant weightage factors and are so chosen that variance is minimised, further $\sum_{i=1}^{v} w_i = 1$ and $0 < w_i < 1$, $i = 1, \ldots\ldots, v$.

Now

$$\hat{y}^{(k+1)} = \sum_{i=1}^{n-n_1} y^{(k+1)}_{(n_1+i)} / n_1 - n_1 = \sum_{i=1}^{n-n_1} x^{*(k+1)} (y^{*(k)} x^{(k)}_{(n_1+i)}) / n - n_1$$

or

$$y^{(k+1)} = x^{*(k+1)} (y^{*(k)} \; x^{(k)}) \qquad \ldots . (7.2.3)$$

but we have

$$x^{*(k+1)} (y^{*(k)} x^{(k)}) = \begin{bmatrix} w_1 x_1^{(k)} / x_1^{*(k)} \; x^{*(k+1)}_{v+1} + w_2 \, x_2^{(k)} \; x^{*(k+1)}_{v+1} / x_2^{*(k)} + \ldots \ldots \\ \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \ldots \\ \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \ldots \\ w_1 x_1^{(k)} \, x_u^{*(k+1)} / x_1^{*(k)} + w_2 x_2^{(k)} \, x_u^{*(k+1)} / x_2^{*(k)} + \ldots \ldots \end{bmatrix} \qquad \ldots .. (7.2.4)$$

Now we have

$$V(w_1 \, x^{*}{}_1^{(k)} \; x^{*(k+1)}_{v+1} / x_1^{*(k)} + w_2 x_2^{(k)} \; x^{*(k+1)}_{v+2} / x_2^{*(k)} + \ldots \ldots)$$

$$= \sum_{i=1}^{v} w_i^2 \left[\text{Cov.} (x_i^{(k)}, x^{*(k+1)}_{v+i}) + \text{Cov.} (x_i^{(k)}, x_i^{*(k)}) - \right.$$

$$- \text{Cov} (x_i^{(k)}, x_i^{*(k)}) + V (x^{*(k+1)}_{q+i}) + V (x^{*}{}_i^{(k)}) + V (x_i^{(k)}) \Big]$$

$$+ \sum_{\substack{i=1 \\ i \neq j}}^{v} \sum_{j=1}^{v} w_i w_j \Big[\text{Cov} . (x_i^{(k)}, x^{*}{}_j^{(k)}) + \text{Cov}(x^{*}{}_i^{(k)}, x_j^{(k)})$$

$$+ 2 \, \text{Cov.} (x_i^{(k)} , x^{*(k+1)}_{v+j})$$

$$- 2\text{Cov} (x^{*}{}_i^{(k)} , x^{*(k)}_{v+j}) + 2\text{Cov.} (x_j^{(k)}, x^{*(k+1)}_{v+i})$$

$$-2\text{Cov.} (x^{*}{}_j^{(k)}, x^{*(k+1)}_{v+i} \Big] \qquad \ldots . (7.2.5)$$

In the same way we can find other elements of variance, covariance matrix of $V (x^{*(k+1)} (y^{*(k)} \; x^{(k)}))$ and thus we will be able to find variance, covariance matrix for non-responded characters under study.

Case 3: - In this case also we estimate value of non-responding units as

$$\hat{y}^{(k+1)}_{(n_1+i)} = x^{*(k+1)} \left(y^{*(k)} \; x^{(k)}_{(n_1+i)} \right) \quad i = 1, 2, \ldots, n-n_1 \quad \ldots (7.2.6)$$

where

$$\hat{y}^{(k+1)}_{(n_1+i)} = \left(\hat{x}^{(k+1)}_{1(n_1+i)}, \hat{x}^{(k+1)}_{2(n_1+i)}, \ldots\ldots\ldots, \hat{x}^{(k+1)}_{u(n_1+i)} \right)'$$

$$x^{*(k+1)} = \left(x^{*(k+1)}_{1}, x^{*(k+1)}_{2}, \ldots\ldots, x^{*(k+1)}_{u} \right)'$$

$$x^{(k)}_{(n_1+i)} = \left(x^{k}_{1(n_1+i)}, x^{(k)}_{2(n_1+i)} \ldots\ldots\ldots, x^{(k)}_{u(n_1+i)} \right)'$$

and

$$y^{*(k)} = \left(w_1/x^{*(k)}_1, w_2/x^{*(k)}_2, \ldots\ldots, w_u/x^{*(k)}_u \right)$$

also $\sum_{i=1}^{u} w_i = 1$

variance, covariance matrix of vector $x^{*(k+1)} \left(y^{*(k)} \; x^{(k)} \right)$ can be found by procedure similar to that used for (7.2.4)

Case 4: - Here also we will deal with a situation similar to model (a) case 4. We estimate value of u-v characters for $n-n_1$ units from formula

$$\hat{y}^{(k+1)}_{(n_1+i)} = x^{*(k+1)}, \left(y^{*(k)} \; x^{(k)}_{(n_1+i)} \right), \quad i = 1, \ldots, n-n_1 \quad \ldots (7.2.7)$$

where

$$\hat{y}^{(k+1)}_{(n_1+i)} = \left(\hat{x}^{(k+1)}_{v+1(n_1+i)}, \hat{x}^{(k+1)}_{v+2(n_1+i)} \ldots\ldots, \hat{x}^{(k+1)}_{u(n_1+i)} \right)'$$

$$x^{*(k+1)} = \left(x^{*(k+1)}_{v+1} \; x^{*(k+1)}_{v+2}, \ldots\ldots, x^{*(k+1)}_{u} \right)'$$

$$x^{(k)}_{(n_1+i)} = \left(x^{(k)}_{1(n_1+i)} \ldots\ldots, x^{(k)}_{u(n_1+i)} \right)' \text{ and}$$

$$y^{*(k)} = \partial\left(w_1/x^{*(k)}_1 \quad w_2/x^{*(k)}_2, \ldots\ldots, w_u/x^{*(k)}_u \right)$$

also $\sum_{i=1}^{u} w_i = 1$

variance covariance matrix of vector $x^{*(k+1)}$ ($y^{*(k)}$ $x^{(k)}$) can be found by the procedure similar to that used for (7.2.4)

<u>Optimum Constant Weightage factors :</u> - They can be very easily estimated from equations $\partial V\left(x^{*(k+1)}\left(y^{*(k)} x^{(k)} \right)\right) / \partial w_i = 0,\ i = 1, \ldots, u(\text{or} v)$ and they can be substituted in vector $y^{*(k)}$ to get an estimate of value of non-responding units.

3. <u>Model C :</u> With notations similar to those for model b, we have model C as

$$\hat{y}^{(k+1)} = x^{*(k+1)} + W\left(x^{(k)} - y^{*(k)} \right) + \epsilon \qquad \ldots. (7.3.1)$$

where $E(\epsilon) = \vec{O}$ and $E(\epsilon\epsilon') = \text{diag}\ (\sigma_{11}, \sigma_{22}, \ldots\ldots, \sigma_{uu})$

Here also W is a constant weightage factor matrix.

<u>Case 2:</u>- In this case also we are dealing with situation similar to that for case 2 model(b). Here we estimate $(n-n_1)$ units by the formula

$$\hat{y}^{(k+1)}_{(n_1+i)} = x^{*(k+1)} + W\left(x^{(k)}_{(n_1+i)} - y^{*(k)} \right) \quad i = 1, 2, \ldots, (n-n_1) \quad \ldots(7.3\ 2)$$

where

$$y^{(k+1)}_{(n_1+i)} = \left(x^{(k+1)}_{v+1(n_1+i)} \quad x^{(k+1)}_{v+2(n_1+i)} \ldots\ldots, x^{(k+1)}_{u(n_1+i)} \right)'$$

$$x^{*(k+1)} = \left(x^{*(k+1)}_{v+1},\ x^{*(k+1)}_{v+2}, \ldots\ldots, x^{*(k+1)}_{u} \right)'$$

$$y^{*(k)} = \left(x^{*(k)}_1,\ x^{*(k)}_2, \ldots\ldots\ldots, x^{*(k)}_v \right)'$$

$$x^{(k)}_{(n_1+i)} = (x^{(k)}_{1(n_1+i)} \quad x^{(k)}_{2(n_1+i)} \quad , \ldots\ldots , x^{(k)}_{v(n_1+i)} \quad)'$$

$$W = \lfloor\overline{\ } w_{ij} \ \underline{\ }\rfloor \qquad i = 1, \ldots u-v, \ j = 1, \ldots\ldots, v$$

$\sum_{j=1}^{v} w_{ij} = 1$, and these constant weightage factor lie between zero and one.

$$\text{Now} \quad y^{(k+1)} = \sum_{i=1}^{n-n_1} y^{(k+1)}_{(n_1+i)} \ / \ (n - n_1) = x^{*(k+1)} + W (x^{(k)} - y^{*(k)}) \qquad \ldots. (7.3.3)$$

Hence

$$V (y^{(k+1)}) = V (x^{*(k+1)}) + W (x^{(k)}) W' + W V(y^{*(k)}) W' - 2W \ \text{Cov.} (x^{(k)}, y^{*(k)'}) W' \qquad \ldots \ (7.3.4)$$

or

$$\hat{V} (y^{(k+1)}) = \hat{V} (x^{*(k+1)}) + W\hat{V}(x^{(k)}) W' + W \hat{V}(y^{*(k)}) W' - 2W \ \hat{\text{Cov}}.(x^{(k)}, y^{*(k)'}) W' \qquad \ldots.(7.3.5)$$

where

$V(y^{*(k)})$ = is v x v variance covariance matrix.
Same is true for other matrices i.e they are also VXV variance covariance matrices

<u>Case 3:</u> - In this case also we estimate non-responding units by the formula

$$\hat{y}^{(k+1)}_{(n_1+i)} = x^{*(k+1)} + W (x^{(k)}_{(n_1+i)} - y^{*(k)}) \ i = 1, 2, \ldots., n-n_1 \qquad \ldots (7.3.6)$$

where

$$y^{(k+1)}_{(n_1+i)} = (x^{(k+1)}_{1(n_1+i)} , \ x^{(k+1)}_{2(n_1+i)} \ , \ldots\ldots, x^{(k+1)}_{u(n_1+i)} \quad)'$$

$$x^{*(k+1)} = (x^{*(k+1)}_{1} , \ x^{*(k+1)}_{2} \ , \ldots\ldots, x^{*(k+1)}_{u} \ )'$$

$$* \ x^{(k)}_{(n_1+i)} = x^{(k)}_{1(n_1+i)} \ , \ x^{(k)}_{2(n_1+i)} \ , \ldots\ldots, x^{(k)}_{u(n_1+i)} \quad)'$$

$$y^{*(k)} = (x^{*}_1{}^{(k)},\ x_2^{*(k)},\ \ldots\ldots, x^{*}_u{}^{(k)})' \quad \text{and} \quad W = [w_{ij}] \quad \begin{array}{l} i = 1,\ldots,u \\ j = 1,\ldots,u \end{array}$$

where w's are constant weightage factors such that $\sum_{j=1}^{u} w_{ij} = 1, (i = 1,\ldots,u)$ $(j = 1,\ldots,u)$ and $O < w_{ij} < 1$ and are so chosen that variance covariance matrix becomes minimum.

Here also estimate of variance covafiance matrix is of the form shown in (7.3.5) but with the differente in order. The order of this variance covariance matrix being u x u.

<u>Case 4:</u> In this case also non-respondents value can be estimated from

$$\hat{y}^{(k+1)}_{(n_1+i)} = x^{*(k+1)} + W(x^{(k)}_{(n_1+i)} - y^{*(k)}) \quad i = 1,2,\ldots,n-n_1 \quad \ldots (7.3.8)$$

where

$$\hat{y}^{(k+1)}_{(n_1+i)} = (x^{(k+1)}_{v+1(n_1+i)},\ x^{(k+1)}_{v+2(n_1+i)},\ \ldots\ldots\ldots, x^{(k+1)}_{u(n_1+i)})'$$

$$x^{*(k+1)} = (x^{*(k+1)}_{v+1},\ x^{*(k+1)}_{v+2},\ \ldots\ldots, x^{*(k+1)}_{u})'$$

$$x^{(k)}_{(n_1+i)} = (x^{(k)}_{1(n_1+i)}\ \ x_{2\,(n_1+i)}^{(k)}\ \ \ldots\ldots, x_{u\ (n_1+i)}^{(k)})'$$

$$y^{*(k)} = (x^{*}_1{}^{(k)}\ \ x_2{}^{*(k)},\ldots\ldots, x_u{}^{*\,(k)})' \quad \text{and} \quad W = [w_{ij}]$$

$$i = 1,\ldots,u\ ,\ j = 1,\ldots u\text{-}v$$

where w_{ij}'s are so chosen that $V (y^{(k+1)})$ is minimum and are such that $\sum_{j=1}^{u} w_{ij} = 1$, and $O < w_{ij} < 1$ $(i=1,\ldots,n\text{-}v)$.

Here also estimate of variance covariance matrix is of the form shown in (7.3.5) but with the difference in order. The order of this variance covariance matrix being u x u.

<u>Optimum constant weightage factors:</u> - Optimum values of w_{ij}'s are obtained from a set of linear equations generated by

$$\partial \hat{V} (y^{(k+1)}) / \partial w = 0$$

NOTE: (A) Before putting this type of estimates for any non-respondents, we should enquire from enumerator, the cause of nonresponse (or enumerators should be instructed to write the cause of non-response in schedules prepared for recording data) and should also enquire that non-responding units is still in our picture or not. If the non-responding unit is not in picture we should not put such an estimate. This can be illustrated with the help of following example.

Forexample, consider a survey involving, use of different fertilizers , different technical machines and multi crop-cutting experiments in India. We should make it clear from investigators (or from schedule) that the non-responding cultivator is cultivating his land or not (since cultivator will use fertilizer etc. only when he is cultivating his land).

(B) In same survey any one or all types of non-responses can occur. In that case the value of non-responding units can be estimated by grouping them based on type of non-response they possess and then estimating them using any one or all models (as the situation may demand).

CHAPTER VIII

AN ILLUSTRATION

For illustration, data of Intensive Agricultural District Programme for District Aligarh, collected from year 1962-63 to 1964-65 was studied.

The Intensive Agricultural District Programme popularly known as the " Package Programme " was developed on the "10-Point Pilot Programme". It involves the selection of favourable areas with maximum irrigation facilities and minimum of natural hazards, providing simultaneously all the essential elements, such as full supplies, credit, etc. needed to increase agricultural production. With this view State Governments selected the following districts for the implementation of the programme:

1. Thanjavur (Madras)
2. West Godavari (Andhra Pradesh)
3. Shahabad (Bihar)
4. Raipur (Madhya Pradesh)
5. Aligarh (Uttar Pradesh)
6. Ludhiana (Punjab)
7. Pali (Rajasthan)

Of the districts selected by the first seven States for implementation of the programme, four are predominantly rice growing viz., Thanjavur (Madras), West Godavari (Andhra Pradesh), Shahabad (Bihar) and Raipur (Madhya Pradesh), two wheat growing viz., Aligarh (Uttar Pradesh) and Ludhiana (Punjab), and one district, namely, Pali(Rajasthan) has preponderence of millets and wheat.

Between Aligarh and Ludhiana districts which are predominately wheat growing areas, the former has a much larger size having 17 blocks and a gross cropped area of 5.34 lakh hectares as compared to

Ludhiana which had 9 blocks and a gross cropped area of 3.41 lakh hectares. Ludhiana, however, had a much higher proportion of assured irrigation (58 per cent) than Aligarh where irrigated area was only 43 per cent. In Aligarh, wheat whichis mostly irrigated yields on an average of 10 quintals per hectare and bajra, the most important kharif crop yields less than half of this quantity. In Ludhiana also yield of wheat was more or less the same as in Aligarh. The large variety of crops grown in these two districts make farm planning rather difficult.) To avoid this difficulty, only area under wheat, together with , rate of sowing , green manure, farm yard manure and other organic manure was studied. All these five characters have been considered as main characters under study. No study of use of chemical fertilizer could be made since very little chemical fertilizer was in use at the time of introduction of this programme. The data collected for Bench mark surveys under the I.A.D.P., for wheat only, from year 1962-63 to 1965-66 was studied. The sampling design of the survey is as follows:

Sampling Plan: - The sampling plan was suggested by I.A.R.S. and was based on stratified multistage random sampling technique . A zone consisting of 2 to 4 blocks constitutes a stratum. A village and a cultivators holding in the village were first and second stage units respectively. From the whole population about 100 villages were selected and from each selected village eight cultivators' holdings were convassed every year. Sampled cultivators in a total number of 24 villages selected on first occasion were kept fixed for convassing from year to year. In addition, a fresh sample of about 76 villages was selected each year from thosevillages which were not selected on previous occasion and from each selected village eight cultivators were canvassed in order to keep sample size same on all occasions.

Before the start of Intensive Agricultural District Programme, Aligarh was comprised of 6 talukas and 1,746 villages covering a geographical area of about 5 thousand sq. kms. and was delimited into 17 community development blocks. The population in 1961 was of the order of 17.65 lakhs which indicates a density of about 351 persons per sq. km. About 84 per cent of the population was classified as rural. The district was served by 648 kms. of main roads and 550 kms. of arterial roads. About 47 per cent of the roads were kutcha and 288 kms. of arterial roads were suitable only for fair weather transport. Two railway lines - one broad and other meter gauge of about 64 kms. - intersect the district.

I. Tables 8.1.1 to 8.1.4 contain estimates of mean vector for years 1962-63, 1963-64, 1964-65, 1965-66 respectively, along with their variance, covariance matrix without using the knowledge of any other year. The sampling plan used was Two-stage rotation pattern, after considering the 24 retained villages along with retained cultivators as population to be studied. The sample selected was of 10 villages with four cultivators selected in each village by general rotation pattern. Since the sample size (as far as Aligarh is concerned) becomes very small and it can not be a valid estimate for Aligarh. From these tables we find that area under wheat increased from year to year, which means that cultivators were interested in increasing their wheat yield and for that reason, they might have started growing wheat in more area. Rate of sowing increased from year 1962-63 to 1963-64 considerably but afterwards increase in rate of sowing was not much though area under wheat increased considerably. This may be due to the fact that Irrigation and knowledge of techniques of scientific cultivation became more popular, which might have resulted in reducing vauge ideas in the minds of farmers, that if, they sow more, they will get more yield. The use of green manure increased considerably from year 1962-63 to 1963-64 due to the fact that 1962-63 was the starting year of "package programme" and as a result of which cultivators in 1963-64 might have tried to use knowledge given to them in 1962-63. The use of green manure reduced a little bit in 1964-65 and then it increased a little bit in 1965-66. The average consumption of farm yard manure remained very low throughout all years but average consumption of other organic manures increased considerably from 1962-63 to 1963-64 due to the fact that farmers carried out instructions given to them in 1962-63 in next year i.e. 1963-64. The average use of other organic manure per cultivator decreased considerably in year 1964-65 and further reduced in year 1965-66 as farmers were getting more encouragement and intensives for using chemical fertilizers.

2. Tables. 8.2.1 to 8.2.4 contain the estimates for years 1962-63, 1963-64, 1964-65, 1965-66 respectively, (without using the knowledge of of any other year) along with their variance and covariance matrix, efficiency matrix as well as efficiency of generalised variance with respect to corresponding estimates of sampling plan I. From this also trend of different components of mean vector was found to be of the form discussed in 1, but for generalised variance and efficiency matrix (in most of the cases) it was found that sampling plan II is more efficient than sampling plan I.

3. Tables 8.3.1 to 8.3.4 contain the estimates of mean vector from year (1962-63 to 1965-66) respectively (without using the knowledge of any other year), along with their variance covariance matrix, efficiency matrix and efficiency of generalised variance with respect to corresponding estimates of sampling plan I. From these tables also, we find trend of different components of mean vector to be of the form similar to that in case 1. But for generalised variance and efficiency matrix (in most of the cases) sampling plan III is more efficient than sampling plan I. Comparing the respective efficiency matrices of different years of Tables 8.2.1 to 8.2.4 to those in Tables 8.3.1 to 8.3.4 we find that sampling plan III is more efficient than sampling plan II even. From this we conclude that for the independent estimates of mean vector evdry year, sampling plan III is most efficient among the plans discussed for the data used.

4. Tables 8.4.1 and 8.4.2 contain estimates of mean vector for year 1964-65 using the knowledge of previous year 1963-64 and previous two years 1962-63 and 1963-64 respectively, using sampling plan I for drawing sample. Looking on their efficiency matrices (with respect to independent estimates of variance, covariance matrix given in table 8.1.3) and efficiency of generalised variance in sampling design I, we find that they are considerably less efficient than the independent estimate for year 1964-65 using sampling plan I. Further estimates using knowledge of previous two occasions was less efficient than the estimate in which only one previous occasions information was utilised. Table 8.4.3 contains modified estimate for year 1964-65 utilising the knowledge of following year, 1965-66. This estimate was also (in most of the cases) less efficient than estimate dealt with, in

Table 8.1.3, but this estimate was found to be more efficient than those dealt in tables 8.4.1 and 8.4.2. Same type of trend was shown by estimates (similar to these in 8.4.1 and 8.4.2 but under sampling plan II) in tables 8.4.4 and 8.4.5. The estimates dealt in table 8.4.4. and 8.4.5 were more efficient than the estimates in table 8.4.1 and 8.4.2 respectively. Modified estimate utilising information of the following occasion is given in table 8.4.6. It was found to be more efficient than estimates given from 8.4.1 to 8.4.5 but this was also less efficient than the estimate dealt in table 8.1.3. Tables 8.4.7 and 8.4.8 dealt with estimates similar to those in table 8.4.1 and 8.4.2 respectively but for sampling plan III. These were found to be slightly more efficient than estimates dealt in table 8.4.4 and 8.4.5 but they were also less efficient than estimate dealt in table 8.1.3 . Modified estimate (under sampling plan III) utilising information of following occasion was found to be slightly more efficient than estimate dealt in table 8.4.6 but this was also less efficient than the estimate dealt in table 8.1.3. This happened due to the fact that characters under study had low correlation among themselves and from one year to another. The estimate of correlation matrix for year 1964-65 was of the order.

$$\begin{bmatrix} 1.000 & 0.152 & 0.351 & 0.002 & -0.162 \\ 0.152 & 1.000 & 0.398 & 0.006 & 0.031 \\ 0.351 & 0.398 & 1.000 & 0.016 & 0.041 \\ 0.002 & 0.006 & 0.016 & 1.000 & 0.461 \\ -0.162 & 0.031 & 0.041 & 0.461 & 1.000 \end{bmatrix}$$

5. Tables 8.5.1, 8.5.2, 8.5.3 contain the estimates of change in mean vectors from 1963-64 to 1964-65 together with their variance and covariance matrices using sampling plan I, II, III respectively. Tables 8.5.2 and 8.5.3 further contained the efficiency matrix and efficiency of generalised

variance with respect to estimate of table 8.5.1. In general , for change, sampling plan I was foudd to be more efficient for the data used.

6. The study of cost and non-response could not be made on the data used for illustration, since it was not satisfying assumptions underlying the type of data to which results of chapter VI and VII could be implemented.

NOTE: - Mean vector contain the following characters in the order, given below (in all tables) , Area under wheat (cents), Rate of sowing (kg/acre), Green manure (kg/acre), Farm Yard manure (quintals) and other organic manures (10 kg/aere).

TABLE 8.1.1

ESTIMATE OF MEAN VECTOR FOR YEAR 1962-63

86.92
35.98
62.76
.21
5.14

VARIANCE COVARIANCE MATRIX FOR MEAN VECTOR OF YEAR 1962-63

1364.78	88.41	-8.48	-3.61	-45.81
88.41	62.74	141.27	1.21	19.36
-8.48	141.27	6941.43	38.40	121.41
-3.61	1.21	3.84	1.01	2.36
-45.81	19.36	121.41	2.36	49.07

Generalised Variance = 809072.611

TABLE 8.1.1

ESTIMATE OF MEAN VECTOR FOR YEAR 1963-64

148.04
378.49
5880.41
.74
169.02

VARIANCE COVARIANCE MATRIX FOR MEAN VECTOR OF YEAR 1963-64

1441.76	451.61	-3602.71	-5.98	51.45
451.61	7142.41	1123.48	5.76	106.73
-3602.71	1123.48	21417.63	35.26	5.61
-5.98	5.76	35.26	3.21	7.78
51.45	106.73	5.61	7.78	21.41

Generalised Variance = 169854.681

TABLE 8.1.3

ESTIMATE OF MEAN VECTOR FOR YEAR 1964-65

2169.58
571.78
4557.86
.37
.66

134171.45	641.71	-23121.12	.24	-131.41
641.71	98.72	5012.34	.51	-10.41
-23121.12	5012.34	80191.76	2.21	2.23
.24	.51	2.21	2.23	2.16
-131.41	-10.41	2.23	2.16	2.38

Generalised Variance = 2143211.632

TABLE 8.1.4

ESTIMATE OF MEAN VECTOR FOR YEAR 1965-66

2326.76
511.71
4641.21
.18
.16

VARIANCE COVARIANCE MATRIX FOR MEAN VECTOR OF YEAR 1965-66

312112.31	-22481.61	642121.12	.71	.82
-22481.61	6741.71	-4282.71	1.43	.71
642121.12	-4282.71	55412.41	1.61	1.72
.71	1.43	1.61	2.31	3.28
.82	.71	1.72	3.28	5.67

Generalised Variance = 2571426.827

TABLE 8.2.1

ESTIMATE OF MEAN VECTOR FOR YEAR 1962-63

88.94
32.51
60.79
2.18
7.16

VARIANCE COVARIANCE MATRIX FOR MEAN VECTOR OF YEAR 1962-63

1272.55	87.87	-6.46	-2.29	-43.78
87.87	58.37	132.16	.03	17.26
-6.46	132.16	6752.35	2.93	118.48
-2.29	.03	2.96	.04	1.05
-43.78	17.26	118.48	1.05	46.08

EFFECINCEY MATRIX WITH RESPECT TO
VARIANCE COVARIANCE MATRIX OF TABLE 8.1.1

1.07	1.00	1.31	1.57	1.04
1.00	1.07	1.06	40.33	1.12
1.31	1.06	1.02	13.10	1.02
1.57	40.33	1.29	25.25	2.24
1.04	1.12	1.02	2.24	1.06

Efficiency of generalised variance = 1.08

TABLE 8.2.2.

ESTIMATE OF MEAN VECTOR FOR YEAR 1963-64

150.28
411.51
5718.43
2.38
174.04

VARIANCE COVARIANCE MATRIX FOR MEAN VECTOR OF YEAR 1963-64

1322.55	335.51	-3201.58	-4.91	51.25
335.51	7044.73	1088.56	4.72	106.49
-3201.58	1088.56	19597.35	30.20	4.36
-4.91	4.72	30.20	.53	2.36
51.25	106.49	4.36	2.36	16.63

EFFECINCEY MATRIX WITH RESPECT TO
VARIANCE COVARIANCE MATRIX OF TABLE 8.1.2

1.09	1.34	1.12	1.21	1.00
1.34	1.01	1.03	1.22	1.00
1.12	1.03	1.09	1.16	1.28
1.21	1.22	1.16	6.05	3.29
1.00	1.00	1.28	3.29	1.28

Efficiency of generalised variance = 1.12

TABLE 8.2.3

ESTIMATE OF MEAN VECTOR FOR YEAR 1964-65

2178.46
568.36
4615.42
.21
2.66

VARIANCE COVARIANCE MATRIX FOR MEAN VECTOR OF YEAR 1964-65

129764.39	495.37	-22337.11	.21	-121.31
495.37	91.71	1216.82	.41	8.84
-22337.11	1216.82	61832.78	.61	.78
.21	.41	.61	.38	.38
-121.31	8.84	.78	.38	8.28

EFFECINCEY MATRIX WITH RESPECT TO
VARIANCE COVARIANCE MATRIX OF TABLE 8.1.3

1.03	1.29	1.03	1.14	1.08
1.29	1.07	4.11	1.24	-1.17
1.03	4.11	1.29	3.62	2.85
1.14	1.24	3.62	5.86	5.68
1.08	-1.17	2.85	5.68	.28

Efficiency of generalised variance = 1.06

TABLE 8. 2.4

ESTIMATE OF MEAN VECTOR FOR YEAR 1965-66

2437.78
528.62
4661.71
.14
1.66

VARIANCE COVARIANCE MATRIX FOR MEAN VECTOR OF YEAR 1965-66

296797.73	-21497.55	632439.16	.41	.48
-21497.55	5846.69	-4171.66	.43	.58
632439.16	-4171.66	28335.83	.71	.76
.41	.43	.71	1.46	1.47
.48	.58	.76	1.47	3.56

EFFECINCEY MATRIX WITH RESPECT TO
VARIANCE COVARIANCE MATRIX OF TABLE 8.1.4

1.05	1.04	1.01	1.73	1.70
1.04	1.15	1.02	3.32	1.22
1.01	1.02	1.95	2.26	2.26
1.73	3.32	2.26	1.58	2.23
1.70	1.22	2.26	2.23	1.59

Efficiency of generalised variance = 1.21

TABLE 8. 3.1

ESTIMATE OF MEAN VECTOR FOR YEAR 1962-63

91.56
36.42
63.71
.51
5.71

VARIANCE COVARIANCE MATRIX FOR MEAN VECTOR OF YEAR 1962-63

1164.41	89.96	-9.43	-4.62	-45.87
89.96	68.77	151.43	2.31	2.12
-9.43	151.43	6621.66	3.92	131.45
-4.62	2.31	3.92	4.71	2.68
-45.87	2.12	131.45	2.68	3.79

EFFECINCEY MATRIX WITH RESPECT TO
VARIANCE COVARIANCE MATRIX OF TABLE 8.1.1

1.17	.98	.89	.78	.99
.98	.91	.93	.52	9.13
.89	.93	1.04	9.79	.92
.78	.52	.97	.21	.88
.99	9.13	.92	.88	12.94

Efficiency of generalised variance = 1.02

TABLE 8.3.2

ESTIMATE OF MEAN VECTOR FOR YEAR 1963-64

153.43
421.44
6101.48
.67
171.39

VARIANCE COVARIANCE MATRIX FOR MEAN VECTOR OF YEAR 1963-64

1231.41	346.48	-3412.41	-7.46	56.24
346.48	7048.78	1091.76	4.92	111.71
-3412.41	1091.76	18741.46	34.26	4.53
-7.46	4.92	34.26	6.17	8.42
56.24	111.71	4.53	8.42	207.56

EFFECINCEY MATRIX WITH RESPECT TO
VARIANCE COVARIANCE MATRIX OF TABLE 8.1.2

1.17	1.30	1.05	.80	.91
1.30	1.01	1.02	1.17	.95
1.05	1.02	1.14	1.02	1.23
.80	1.17	1.02	.52	.92
.91	.95	1.23	.92	.10

Efficiency of generalised variance = 1.15

TABLE 8.3.3

ESTIMATE OF MEAN VECTOR FOR YEAR 1964-65

2214.73
584.61
4667.87
.21
.36

VARIANCE COVARIANCE MATRIX FOR MEAN VECTOR OF YEAR 1964-65

117123.41	483.46	-21341.13	.31	-131.41
483.46	87.46	1217.31	.49	9.16
-21341.13	1217.31	51421.70	1.38	2.76
.31	.49	1.38	3.16	2.28
-131.41	9.16	2.76	2.28	5.46

EFFECINCEY MATRIX WITH RESPECT TO
VARIANCE COVARIANCE MATRIX OF TABLE 8.1.3

1.14	1.32	1.08	.77	1.00
1.32	1.12	4.11	1.04	-1.13
1.08	4.11	1.55	1.60	.80
.77	1.04	1.60	.70	.94
1.00	-1.13	.80	.94	.43

Efficiency of generalised variance = 1.16

TABLE 8.3.4

ESTIMATE OF MEAN VECTOR FOR YEAR 1965-66

2437.76
535.48
4741.31
.19
.26

VARIANCE COVARIANCE MATRIX FOR MEAN VECTOR OF YEAR 1965-66

287426.64	-22461.61	61213.26	.48	1.21
-22461.61	5846.71	-1271.71	.45	1.61
61213.26	-1271.71	27371.67	1.42	2.76
.48	.45	1.42	1.81	1.72
1.21	1.61	2.76	1.72	1.40

EFFECINCEY MATRIX WITH RESPECT TO
VARIANCE COVARIANCE MATRIX OF TABLE 8.1.4

1.08	1.00	10.48	1.47	.67
1.00	1.15	3.36	3.17	.44
10.48	3.36	2.02	1.13	.62
1.47	3.17	1.13	1.27	1.90
,67	.44	.62	1.90	4.05

Efficiency of generalised variance = 1.23

TABLE 8.4.1

ESTIMATE OF MEAN VECTOR FOR YEAR 1964-65

2273.42
641.31
4712.43
1.41
2.58

VARIANCE COVARIANCE MATRIX FOR MEAN VECTOR OF YEAR 1964-65

185213.42	943.79	-16713.18	1.52	-88.41
943.79	1211.41	1381.49	2.46	-13.28
-16713.18	1381.49	24373.61	3.12	5.96
1.52	2.46	3.12	4.16	6.29
-88.41	-13.28	5.96	6.29	8.49

EFFECINCEY MATRIX WITH RESPECT TO
VARIANCE COVARIANCE MATRIX OF TABLE 8.1.3

.72	.67	1.38	.15	1.48
.67	.08	3.62	.20	.78
1.38	3.62	3.29	.70	.37
.15	.20	.70	.53	.34
1.48	.78	.37	.34	.28

Efficiency of generalised variance = 0.83

TABLE 8. 4.2

ESTIMATE OF MEAN VECTOR FOR YEAR 1964-65

2581.76
721.46
4269.47
2.32
1.48

VARIANCE COVARIANCE MATRIX FOR MEAN VECTOR OF YEAR 1964-65

194212.58	1031.76	-12712.36	2.21	-96.42
1031.76	1312.47	1471.64	4.27	-21.19
-12712.36	1471.64	27495.64	4.76	5.12
2.21	4.27	4.76	3.21	2.73
-96.42	-21.19	5.12	2.73	3.79

EFFECINCEY MATRIX WITH RESPECT TO
VARIANCE COVARIANCE MATRIX OF TABLE 8.1.3

.69	.62	1.81	.10	1.36
.62	.07	3.40	.11	.49
1.81	3.40	2.91	.46	.43
.10	.11	.46	.69	.79
1.36	.49	.43	.79	.62

Efficiency of generalised variance = 0. 79

TABLE 8. 4.3

ESTIMATE OF MEAN VECTOR FOR YEAR 1964-65

2584.76
702.46
4926.48
.39
6.84

VARIANCE COVARIANCE MATRIX FOR MEAN VECTOR OF YEAR 1964-65

174216.61	1132.78	-11619.41	3.46	-78.32
1132.78	1421.79	1321.23	5.26	-10.19
-11619.41	1321.23	25412.71	4.81	4.73
3.46	5.26	4.81	5.98	3.21
-78.32	-10.19	4.73	3.21	6.76

EFFECINCEY MATRIX WITH RESPECT TO
VARIANCE COVARIANCE MATRIX OF TABLE 8.1.3

.77	.56	1.98	.06	1.67
.56	.06	3.79	.09	1.02
1.98	3.79	3.15	.45	.47
.06	.09	.45	.37	.67
1.67	1.02	.47	.67	.35

Efficiency of generalised variance = 0.86

TABLE 8. 4.5

ESTIMATE OF MEAN VECTOR FOR YEAR 1964-65

2238.46
541.94
4568.92
4.36
4.83

VARIANCE COVARIANCE MATRIX FOR MEAN VECTOR OF YEAR 1964-65

191628.43	947.78	-11681.42	4.71	-98.76
947.78	1428.73	1162.43	4.92	-11.79
-11681.42	1162.43	24916.73	5.43	6.71
4.71	4.92	5.43	8.42	7.68
-98.76	-11.79	6.71	7.68	10.41

EFFECINCEY MATRIX WITH RESPECT TO
VARIANCE COVARIANCE MATRIX OF TABLE 8.1.3

.70	.67	1.97	.05	1.33
.67	.06	4.31	.10	.88
1.97	4.31	3.21	.40	.33
.05	.10	.40	.26	.28
1.33	.88	.33	.28	.22

Efficiency of generalised variance = 0.85

TABLE 8.4.6

ESTIMATE OF MEAN VECTOR FOR YEAR 1964-65

2007.86
601.34
4223.69
4.68
5.69

VARIANCE COVARIANCE MATRIX FOR MEAN VECTOR OF YEAR 1964-65

171032.16	1431.76	-12621.47	5.68	-70.32
1431.76	1164.47	1231.49	10.11	-21.43
-12621.47	1231.49	26718.71	6.41	5.92
5.68	10.11	6.41	10.19	12.78
-70.32	-21.43	5.92	12.78	19.76

EFFECINCEY MATRIX WITH RESPECT TO
VARIANCE COVARIANCE MATRIX OF TABLE 8.1.3

.78	.44	1.83	.04	1.86
.44	.08	4.07	.05	.48
1.83	4.07	3.00	.34	.37
.04	.05	.34	.21	.16
1.86	.48	.37	.16	.12

Efficiency of generalised variance = 0.89

TABLE 8.4.7

ESTIMATE OF MEAN VECTOR FOR YEAR 1964-65

2342.79
541.76
4946.79
6.41
1.43

VARIANCE COVARIANCE MATRIX FOR MEAN VECTOR OF YEAR 1964-65

164127.63	831.28	-12631.49	5.73	-24.76
831.28	1271.63	1312.67	4.71	-12.69
-12631.49	1312.67	25412.76	5.43	4.98
5.73	4.71	5.43	6.11	5.92
-24.76	-12.69	4.98	5.92	10.11

EFFECINCEY MATRIX WITH RESPECT TO
VARIANCE COVARIANCE MATRIX OF TABLE 8.1.3

.81	.77	1.83	.04	5.30
.77	.07	3.81	.10	.82
1.83	3.81	3.15	.40	.44
.04	.10	.40	.36	.36
5.30	.82	.44	.36	.23

Efficiency of generalised variance = O.[illegible]

TABLE 8.4.8

ESTIMATE OF MEAN VECTOR FOR YEAR 1964-65

2104.12
521.53
4876.31
9.42
2.56

VARIANCE COVARIANCE MATRIX FOR MEAN VECTOR OF YEAR 1964-65

192173.41	1127.68	-13421.76	6.41	-100.12
1127.68	1368.47	1211.79	2.91	-2.46
-13421.76	1211.79	26738.91	6.42	5.38
6.41	2.91	6.42	11.21	4.76
-100.12	-2.46	5.38	4.76	8.47

EFFECINCEY MATRIX WITH RESPECT TO
VARIANCE COVARIANCE MATRIX OF TABLE 8.1.3

.69	.56	1.72	.03	1.31
.56	.07	4.13	.17	4.23
1.72	4.13	2.99	.34	.41
.03	.17	.34	.19	.45
1.31	4.23	.41	.45	.28

Efficiency of generalised variance = 0.86

TABLE 8.4.9.

ESTIMATE OF MEAN VECTOR FOR YEAR 1964-65

2543.68
596.80
4919.78
8.93
2.43

VARIANCE COVARIANCE MATRIX FOR MEAN VECTOR OF YEAR 1964-65

169437.70	1011.98	-11731.42	7.69	-84.43
1011.98	1429.76	1329.76	9.42	-41.71
-11731.42	1329.76	29761.06	7.32	6.78
7.69	9.42	7.32	14.39	4.71
-84.43	-41.71	6.78	4.71	2.92

EFFECINCEY MATRIX WITH RESPECT TO
VARIANCE COVARIANCE MATRIX OF TABLE 8.1.3

.79	.63	1.97	.03	1.55
.63	.06	3.76	.05	.24
1.97	3.76	2.69	.30	.32
.03	.05	.30	.15	.45
1.55	.24	.32	.45	.81

Efficiency of generalised variance = 0.93

TABLE 8. 5.1

ESTIMATE OF CHANGE IN MEAN VECTOR FROM 1963-1964 TO 1964-65

1672.41
192.76
-1441.92
-.76
-4.29

VARIANCE COVARIANCE MATRIX FOR VECTOR OF ESTIMATE OF ABOVE CHANGE

102162.79	479.81	-11897.43	.42	-7.82
479.81	1246.31	2218.71	4.37	-3.81
-11897.43	2218.71	21786.73	1.96	4.72
.42	4.37	1.96	2.19	3.82
-7.82	-3.81	4.72	3.82	11.71

Generalised variance = 2768421.73

TABLE 8.5.2

ESTIMATE OF CHANGE IN MEAN VECTOR FROM 1963-1964 TO 1964-65

1543.75
201.81
-1251.68
-1.28
-4.31

VARIANCE COVARIANCE MATRIX FOR VECTOR OF ESTIMATE OF ABOVE CHANGE

123147.81	546.87	-13296.48	2.16	-8.46
546.87	1471.28	1104.76	3.28	-2.12
-13296.48	1104.76	22837.68	2.27	4.91
2.16	3.28	2.27	3.16	5.12
-8.46	-2.12	4.91	5.12	6.91

EFFECINCEY MATRIX WITH RESPECT TO
VARIANCE COVARIANCE MATRIX OF TABLE 8.5.1

.82	.87	.89	.19	.92
.87	.84	2.00	1.33	1.79
.89	2.00	.95	.86	.96
.19	1.33	.86	.69	.74
.92	1.79	.96	.74	1.69

Efficiency of generalised variance = 0.93

TABLE 8.5.3

ESTIMATE OF CHANGE IN MEAN VECTOR FROM 1963-1964 TO 1964-65

1743.35
185.61
-1532.19
-.92
-5.96

VARIANCE COVARIANCE MATRIX FOR VECTOR OF ESTIMATE OF ABOVE CHANGE

114312.73	425.76	-11286.71	3.27	-9.78
425.76	1298.78	2473.26	1.26	-4.39
-11286.71	2473.26	22037.67	2.77	3.68
3.27	1.26	2.77	3.09	4.26
-9.78	-4.39	3.68	4.26	5.61

EFFECINCEY MATRIX WITH RESPECT TO
VARIANCE COVARIANCE MATRIX OF TABLE 8.5.1

.89	1.12	1.05	.12	.79
1.12	.95	.89	3.46	.86
1.05	.89	.98	.70	1.28
.12	3.46	.70	.70	.89
.79	.86	1.28	.89	2.08

Efficiency of generalised variance = O. 89

SUMMARY

The reliability of any repeated occasion survey is reduced due to marked increase in non-response, which occurs due to increased burden on some of the respondents. To reduce this defect, in the present investigation a rotation scheme has been developed to select sample on successive occasions in multi-stage sampling designs. For this three different type of sampling plans have been developed to estimate one or more characters at the same time. Using these sampling designs an attempt has been made to obtain the minimum variance linear unbiased estimates of:

1. the population mean vector of characters at the most recent occasion,
2. the change in population mean vector of characters from one occasion to another,
3. an overall estimate of the population mean vector of characters overall occasions for a dynamic population, and
4. the modified population mean vector of characters of past occasions using knowledge of estimate of population mean vector of characters for following occasions, in a two stage sampling design.

Optimum sample size in different sampling plans for a given cost (after developiing the appropriate cost function) has also been worked out. Further an attempt has also been made to estimate the missing value of the characters of non-responding units, using the information on previous occasions and recent occasions, under some restrictions.

For illustration a study of I.A.D.P. data, collected for Bench mark surveys in Aligarh district from 1962-63 to 1965-66 has been made.

REFERENCES

1. Abraham, T.P., Khosla, R.K. and Kathuria, O.P. (1969). "Some investigations on the use of successive sampling in pest and disease surveys". J. of the Indian Society of Agricultural Statistics.

2. Anderson, T.W. (1958). " An Introduction to Multivariate analysis", John Wiley and Sons.

3. Cochran, William, G. (1963), "Sampling Techniques", 2nd edition, John Wiley and Sons.

4. Eckler, A.R. (1955). "Rotation sampling". Ann. Math. Stat. 26.

5. Fellgi, Ivan, P. (1963). "Sampling with varying probabilities without replacement, rotating and non-rotating samples". J. Amer. Stat. Assoc.

6. Hall, H.S., and Knight, S.R. (1960). "Higher Algebra" Macmillan and Co. Ltd.

7. Hansen, M.H., Hurwitz, W.N. and Madow, W.G. (1953). "Sample survey methods and theory," Vol. 1 and 2; John Wiley and Sons.

8. Hansen, M.H., Hurwitz, W.N., Nisselson H., and Steinberg, J. (1955). "The Redesign of the census current population survey". J. Amer. Stat. Assoc.

9. Jessen, R.J. (1942). " Statistical investigations of a sample survey for obtaining farm facts". Iowa Agri. Exp. Station, Res. Bull. 304.

10. Kathuria, O.P. (1959). " Some aspects of successive sampling in multistage designs". Unpublished thesis ssubmitted for the award of I CAR diploma.

11. Murty, M.N. (1967). "Sampling theory and methods"

12. Patterson, H.D. (1950). "Sampling on successive occasion, with partial replacement of units". J. Roy. Stat. Soc. Series B-12.

13. Rao, J.N.K. and Graham, Jack E. (1964). " Rotation Design for sampling over repeated occasions", J. Amer. Stat. Assoc.

14. Singh, D. (1968). " Estimates in successive sampling using a multistage design". J. Amer. Stat. Assoc.

15. Singh, D. and Kathuria, O.P. (1969).
"On two stage sacessive sampling". Australian J. of Statistics, Vol.II, No.2.

16. Singh, Shivtar. (1970)
"Some studies in two stage successive sampling". Unpublished thesis submitted for award of ICAR Diploma.

17. Sukhatme, P.V. (1954).
"Sampling theory of surveys with application"

18. Tikkiwal, B.D. (1951).
"Theory of successive sampling". Unpublished Diploma thesis submitted to ICAR.

19. Woodruff, Ralph. S. (1963).
"The use of rotating samples in the census Bureau's monthly surveys". J. Amer. Stat. Assoc.

20. Yates. F. (1960).
"Sampling methods for censuses and surveys". Charles Griffin and Co. London.