A STUDY ON THE NON-RESPONSE IN TWO CHARACTERS IN A TWO STAGE DESIGN

By

R.P.S. MALIK

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

INSTITUTE OF AGRICULTURAL RESEARCH STATISTICS
(Indian Council of Agricultural Research)

NEW DELHI

1973

ACKNOWLEDGEMENTS

It is a matter of great pleasure and privilege to record my sincerest sense of gratitude and indebtedness to Shri M. Rajagopalan, Statistician-cum-Associate Professor, Institute of Agricultural Research Statistics (I.C.A.R.), New Delhi, for suggesting the problem tackled in this dissertation and his immense help, keen interest, constant encouragement, guidance and supervision throughout the course of preparation of this dissertation.

I am very much thankful to Dr. M.N. Das, Director, Institute of Agricultural Research Statistics for providing research facilities to carry out the work.

The work was supported by a Senior Research Fellowship of the Indian Council of Agricultural Research. The help is gratefully acknowledged.

I.A.R.S. (I.C.A.R.)
NEW DELHI - 12.

(R.P.S. MALIK)

CONTENTS

CONTENTS

INTRODUCTION AND REVIEW OF LITERATURE

In sample survey practice, non-response from some of the sampling units is of common occurence. This problem of 'non-response' or 'incomplete samples' is more common in Mailing Enquiries. If the non response is considerable the estimates obtained on the basis of the realised sample may not represent the characteristics of the entire population.

Conceptually, the population may be treated as divided into two strata, one representing the units which readily respond and the other not responding. Let N_1 and N_2 be the population number of units in the two strata, and, $\bar{Y}_1$ and $\bar{Y}_2$ be population means respectively. Let a sample of n units be drawn at random from N (= $N_1 + N_2$) and if n_1 units respond, then the sample mean based on n_1 units provides an estimate of $\bar{Y}_1$ and the resulting bias in estimating $\bar{Y}$ on the basis of the incomplete sample is

$$\bar{Y}_1 - \bar{Y} = \frac{N_2}{N} (\bar{Y}_1 - \bar{Y}_2)$$

Thus if $\bar{Y}_1$ and $\bar{Y}_2$ considerably differ, the bias involved will be substantial.

The only way to minimise or eliminate bias is to have 'call backs' on the non-responding units once or more than once, if need arises.

The alternate method is to devise a suitable plan of enquiry to reduce the non-responding group to the barest minimum. Several methods have been suggested to tackle the problem of non response. Some of them are discussed in the following paragraphs.

1. <u>Hansen and Hurwitz Technique</u> (1946)

This method is particularly suggested for Mailing Enquiries. This could be used with advantage in Personal Enquiry surveys also. The method consists in restoring to 'call-back' on the sub-sample of units selected from the non-responding stratum and building a suitable estimate based on this sub-sample and the sample realised at the first instance.

Suppose the population, consisting of N units, can be divided into two classes - the response class and the non-response class, with sizes N_1 and N_2. If n_1 units in the sample respond and n_2 donot, with $n_1 + n_2 = n$, then we may regard n_1 as a random sample from the response class and n_2 as a random sample from the non-response class. Out of n_2 a simple random sample of n_2' is drawn such that $n_2 = n_2' . f$ where f is the sub-sampling fraction. With this, an unbiased estimate of population mean is given by

$$\bar{y}_w = 1/n \, (n_1 \bar{y}_{n_1} + n_2 \bar{y}'_{n_2})$$

with variance $V(\bar{y}_w) = \left(\frac{1}{n} - \frac{1}{N} \right) S^2 + \frac{f-1}{n} \cdot \frac{N_2}{N} S_2^2$

The efficiancy of the method of call-backs should necessarily be considered in relation to the cost involved. The cost of the survey will be

$$C = C_0 n + C_1 n_1 + C_2 n_2'$$

where C_o represents the cost of including a sample unit in the initial sample, C_1 the cost of getting, editing and processing information per unit in the response class and C_2, the cost of interviewing and processing information in the non-response class.

The solution is then completed by determining the optimum values of n, f for a fixed variance.

2. Clausen and Ford Method (1947)

This method was suggested by Clausen and Ford (1947) for dealing with the problem of bias due to non-response in mail surveys. The main attention has been given to two aspects of the problem:

i) Maximising response by every possible means in order to cut down the size of the non-respondent group whose characteristics and attitudes are known.

ii) Making allowances or corrections for any bias that may exist in the incomplete returns.

The study was carried out by surveying segaments of the Veteran population.

In mail follow-ups of Veterans who had not responded to the initial questionnaire, it was noticed that personalised salutation and true signature did not lead to significant increases over non personalised forms in rate of response, but the use of special delivery letters markedly increased the returns. Further, it was observed that a multiphasic survey, covering several potentially interesting topics, yielded higher rates of response than a single subject survey of the same population, and also greatly lessened an interest bias in response.

3. Politz and Simmons Technique (1949)

The technique developed by Hansen and Hurwitz is primarily applicable to mail surveys. However, in most instances, particularly in the developing countries, mail surveys are not practicable. The skills and abilities of experienced enumerators are needed to obtain the information sought in the surveys.

A common difficulty with interview survey lies in the fact that only a relatively modest portion of the first calls is successful. The interviewer is therefore obliged to make several calls in order to convass the complete sample, and the more the number of calls made, the greater will be the cost of the survey. If the data obtained after the first call is used, the estimate may be biased. Politz and Simmons technique (1949) is an attempt to avoid both the need for successive calls and the danger of biases due to incomplete samples. The essence of this technique is as follows:

Interviewing of persons selected in the sample are made at randomly selected points of time. Each person is visited only once. Those who are contacted are asked to give information on the number of times in the preceeding five days they were at home at the time of visit. Questionnaires for persons interviewed are then classified into six groups according to the number of times these persons were at home. Denote by (j - 1) the number of times the i-th person in the sample was at home prior to the interview. Clearly j is a random variable taking values from 1 to 6, and $j/6$ estimates the proportion of the time the i-th person is at home during the interval of interest. Totals for the characteristics under the study are then prepared for each group and the estimate of, say, the population total is obtained by weighting each group total by the corresponding value of $6/j$ viz. by the inverse of the proportion of time the people in the corresponding group were at home. A higher weight is thus given to persons who are less frequently at home.

If we are interested in estimating the population total for a character x then

$$X' = 6 \sum_{i=1}^{n} x_i / j$$

where x_i denotes the value of the characteristic for the i-th person in the population and $\sum^{n}$ denotes the summation over the n persons interviewed.

If p_i = Probability of i-th person in the sample to be at home when the interviewer calls and $q_i = 1 - p_i$, the probability that i-th person is at home zero times during the 5 day period is = q_i^5. Similarly the probability that i-th person is at home one time during the 5 day period is = $\binom{5}{1} q_i^4 p_i$ and so on.

Also let p_{ij} = Probability of the i-th person to belong to the j-th group of population (j = 1, 2, . . . , 6)

$$\frac{6!}{j!\,(6-j)!} \, p_i^j \, q_i^{6-j}$$

then $\sum_{j=1}^{6} p_{ij} = 1 - q_i^6$

With this $E X' = \sum_{i=1}^{N_t} x_i (1 - q_i^6)$

where N_t is the number of individuals covered in the survey as respondents.

$$E X' = \sum_{i=1}^{N_t} x_i - \sum_{i=1}^{N_t} x_i q_i^6$$

The first term represents the total for persons who can be found at home at least once in the specified 6 days period. The second term (which gives bias) is not accounted for in the procedure of estimation. Thus the estimate is unbiased if the survey is arranged in such a way so that the group of those who are never at home is reduced to negligible amount i.e. q_i should be very small.

4. Birnabaum and Sirken Method (1950)

Birnabaum and Sirken (1950) presented another technique for the treatment of errors introduced into sampling surveys due to the non-availability of respondents. This procedure fixes a number of "call-backs" k, in advance for the enumeration of the second sample. All units are called back, upto a maximum number of k calls, till they respond. Here, no sub-sampling of the non-respondents is considered The assumption made about the population is also different. It is assumed that all units in the population, will have the same probability, of being available at exactly the j-th attempt, where $1 \leq j \leq k$.

The expected cost and variance of the sample survey are expressed as functions of sample size and of the number of call backs made on the non-availables. A method has then been presented which optimises precision for a given cost by playing sampling error against the bias resulting from non-availables.

5. Deming's Technique (1953)

Deming postulates a probability mechanism for the simultaneous production of the bias of non-response and for the variance of response. The non response arises from a graded series of classes of the members of the universe to be sampled. The classes range from an impregnable core of no possible response, on upto a class of complete response. Non response arises from two sources, not at home and refusals. Refusals are of two kinds, permanent and temporary. The variation in the amount of time spent at home, and the variation in the firmness of the temporary refusal, produce the graded series of classes. The bias of non response arises from the variation of any characterstic from one member to another within a single class, and from the random variation in the number of responses therefrom.

An increase in the size of the initial sample or a more efficient method of selection will decrease the variance of response, but will have no effect on the bias of non response. Successive recalls, on the other hand, decrease the bias of response, and are more effective than an increase in the size of the sample or a more efficient method of selection in decreasing the root mean square error which arises from both non response and from the variation of response.

The results show that without recalls, it is hazardous to put any confidence in the result, no matter how big the sample, even when the variation in the measured characterstic is only two-fold from the class of lowest response to the class of highest response.

With the levels of response assumed here (taken from average urban experience), and with an estimate formed by summing up the initial call and the recalls, the first two recalls effect together about a 50 per cent reduction in the initial bias of non response. Further recalls continue to be productive. In fact with this method of estimation, each recall added to a sampling plan, even to six recalls, actually increases the amount of information obtained for each dollar expanded on interviewing.

For any proposed survey, calculations based on rough advance estimates of the constants that appear in the formulae (given in the paper) will predict to a useful degree of approximation the biases and the variances to be expected from various types of plans. Figures on costs will then point out which plan is most economical, of these that are possible, for the attainment of a prescribed accuracy.

The proposed mechanism provides a theory of bias to supplement the theory of sampling. It indicates the possibility of new and more efficient methods of estimation than the simple combination of the initial attempt and the recalls, as it will provide a rational basis for extracting more information from the recalls. It also points out, for any particular method of estimation, what empirical information will be helpful in the planning of the efficient allocation of effort amongst the initial sample and the recalls.

6. Durbin's Method (1954)

In Politz and Simmons technique, it was assumed that interviewers make their call at the time selected at random within the hours of the day the survey is taking place. If the information obtained on the presence at home on the past five days is added, the result is a random sample of six points of time 24 hours apart from one another.

Durbin (1954) treated Politz and Simmons technique in a way which is free from above assumption. According to Durbin, it is sufficient to secure "that an equal number of calls is made on each of the six evenings of the week, either over the survey as a whole, or more strictly, by each investigator". With such a less restricting assumption, Durbin developed the same basic results as Politz and Simmons with some additional improvements in theory.

Durbin also compared the efficiency of Politz and Simmons technique with the procedure of recalling and found that, under some assumptions, the latter method is more efficient if the value of the characterstics on the survey programme is independent of the probability of finding the respondent at home when the call is made. If there is some correlation, the former method might be more efficient.

Some Other Developments

7. Method Suggested by El-Badry (1956)

El-Badry (1956) has extended Hansen and Hurwitz's technique based on the experience that an appreciable increase in response rates to mail questionnaires can be secured by sending repeated waves of questionnaires to the non-responding units. When the point is reached where further waves would not be effective, a sub-sample of the remaining non-responding units is selected and interviewed. Final estimates are based on the pooled results from all the attempts put together.

8. Kish and Hess Method (1959)

Kish and Hess (1959) propose that a number of addresses of non responding units from a recent and similar survey be added to the units selected for an actual survey. In the course of the field work data are collected from the two parts of the sample and added units serve as a "replacement" for the non-responding units in the actual survey. If similar surveys are repeated frequently and a record of non-response is kept this procedure might represent an economical treatement of the problem of missing data.

In another study the same authors considered the problem of non-coverage viz. failure of the enumerators to include some units in the survey. This failure could also be considered as the problem of missing data if it is known that units omitted exist. Otherwise, it makes a part of problems arising out of inaccurate frames and inaccurate listings.

9. Bedawi and Seth's Method (1961)

Bedawi and Seth (1961) extended the method suggested by Hansen and Hurwitz to few more sampling techniques so as to cope with the need in agricultural surveys. The solutions have been obtained under the assumption that all non responding units, will definitely respond at the second attempt. The different cases considered are:

a) Cluster and two stage sampling

b) Multipurpose surveys

c) Sampling with varying probabilities of slection, with replacement

For each of the above cases, a design has been developed in which the non-responding units, are sampled with a smaller sampling fraction than the responding units. Thes sub-samples are determined, by fixing in advance, the sub-sampling fractions, from each stratum. These in addition to the original sample size form the 'parameters' of the design. These parameters are selected in such a way as to provide maximum information per unit of cost.

In this dissertation, the problem of non response in multicharacter surveys in multi-stage design has been attempted.

CHAPTER - II

NON-RESPONSE IN TWO CHARACTERS IN A TWO STAGE DESIGN

Most of the work done on non-response in sample surveys, deals mainly with single stage sampling and one character is under study. Bedawi and Seth (1961) attempted the problem of non-response in two characters again in a single stage design.

The present work deals with the problem of non-response in sample surveys, where the design of the survey is one of two stage sampling and there are two characters under study.

The non-responding units, in general, can be classified into two groups:

1) Units which are temporary inaccessible, due to one reason or other, at the first attempt but would respond on contact. This group will also consist of units which would have to be persuaded to provide information by contacting them more than once.

2) This group would consist of units which would never respond even when several attempts are made to obtain the information For the purpose of this study, it is assumed that this group of units is almost non-existent. i.e. units would respond definitely at the second attempt.

In the present investigation, as mentioned above, it is assumed that units would respond definitely at the second attempt. The Hansen and Hurwitz technique is extended to the case where information is required on two characters through a two-stage sampling design.

Sampling Design and Classification of Units

Let a sample survey be conducted to estimate the mean values of two characters X and Y through a two-stage design. Let the population under study consist of N PSU's. A simple random sample of n PSU's is selected with equal probability and without replacement. In each PSU selected some SSU's may give complete information on both X and Y, some on X only, some on Y only and some may not give any information on both X and Y.

Based on the availability of information, the PSU's can conceptually be classified into 15 strata as follows:

Number	Strata	Description
1.	a	Consists of PSU's in which all SSU's give information on both X and Y.
2.	b	Consists of PSU's in which all SSU's give information on X only.
3.	c	Consists of PSU's in which all SSU's give information on Y only.
4.	d	Consists of PSU's in which all SSU's donot give any information either on X or on Y.
5.	ab	Consists of PSU's in which some of the SSU's give information on both X and Y, while the rest of SSU's give information on X only.

Number	Strata	Description
6	ac	Consists of PSU's in which some of SSU's reply for both X and Y, while the rest of SSU's give information on Y only.
7	ad	Consists of PSU's in which some of SSU's give information on both X and Y, while the rest of SSU's donot give any information.
8	bc	Consists of PSU's in which some of the SSU's give information on X only while the rest of SSU's reply for Y only.
9	bd	Consists of PSU's in which some of the SSU's give information on X only, while the rest do not given any information.
10	cd	Consists of PSU's in which some of the SSU's reply for Y only while the rest do not give any information.
11	abc	Consists of PSU's in which some of the SSU's give information on both X and Y, some give information on X only, while the rest of SSU's reply for Y only.
12	abd	Consists of PSU's in which some of the SSU's reply for both X and Y, some for X only, while the rest of SSU's do not give any information. either on X or on Y.

<u>Number</u>	<u>Strata</u>	<u>Description</u>
13	acd	Consists of PSU's in which some of the SSU's give information on both X and Y, some reply for Y only, while the rest of SSU's donot give any information, either on X or on Y.
14	bcd	Consists of PSU's in which some of the SSU's give information on X only, some on Y only, while the rest donot give any information either on X or on Y.
15	abcd	Consists of PSU's in which all the four possibilities are there, viz. some of the SSU's reply for both X and Y, some reply for X only, some reply for Y only while the rest donot given any information, either on X or on Y.

After having classified the n PSU's into 15 strata as above, let n_h denote the number of PSU's in h-th stratum ($h=1, 2, \ldots, 15$). Let M_{hi} denote the number of SSU's in the i-th PSU of h-th stratum. ($i = 1, 2, \ldots, n_h$). The second stage units are selected at random, with equal probability without replacement, from each PSU. Let m_{hi} be the sample size at the second stage drawn from M_{hi}. These m_{hi} units can again be classified into four classes according as whether the SSU gives information on both X and Y, X only, Y only, or donot give any information either on X or on Y. After the units are thus classified, let m_{hik} ($k = 1, 2, 3, 4$) denote the number of SSU's, out of m_{hi}, falling in k-th class, namely :

m_{hi1} Units respond for both X and Y

m_{hi2} Units respond for X only

m_{hi3} Units respond for Y only

m_{hi4} Units donot respond for any character

In the last three classes, second attemptw would provide information on the character which could not be recorded at the first attempt. Out of m_{hi2} units which respond for X only, draw a sub-sample of m'_{hi2} units with equal probability and without replacement for measurement of character Y. Similarly out of m_{hi3} units which reply for Y only, draw a sub sample of m'_{hi3} units with equal probability and without replacement for measurement of character X. Out of m_{hi4} units which donot respond for any character, draw a sub sample of m'_{hi4} units with equal probability and without replacement and measure X and Y on these units.

Notation :

Let N = Total number of PSU's in the population

n = Number of PSU's selected in the sample

N_h = Number of PSU's in the population falling in h-th stratum.

n_h = Number of PSU's in the sample falling in h-th stratum

(h = 1, 2, . . . , 15)

such that $\sum_{h=1}^{15} N_h = N$ and $\sum_{h=1}^{15} n_h = n$

Further let

M_{hi} = Number of SSU's in the i-th PSU

m_{hi} = Number of SSU's selected from M_{hi}

$(i = 1, 2, \ldots, n_h)$

such that $\sum_{i=1}^{N_h} M_{hi} = M_h$ = Total number of SSU's in the h-th stratum

and $\sum_{h=1}^{15} \sum_{i=1}^{N_h} M_{hi} = M_o$ = Total number of SSU's in the Population.

Also let m_{hik} denote the number of SSU's out of m_{hi}, falling in k-th class $(k = 1, 2, 3, 4)$

where

$k = 1$ denotes the class of SSU's which reply for both X and Y.

$k = 2$ denotes the class of SSU's which reply for X only.

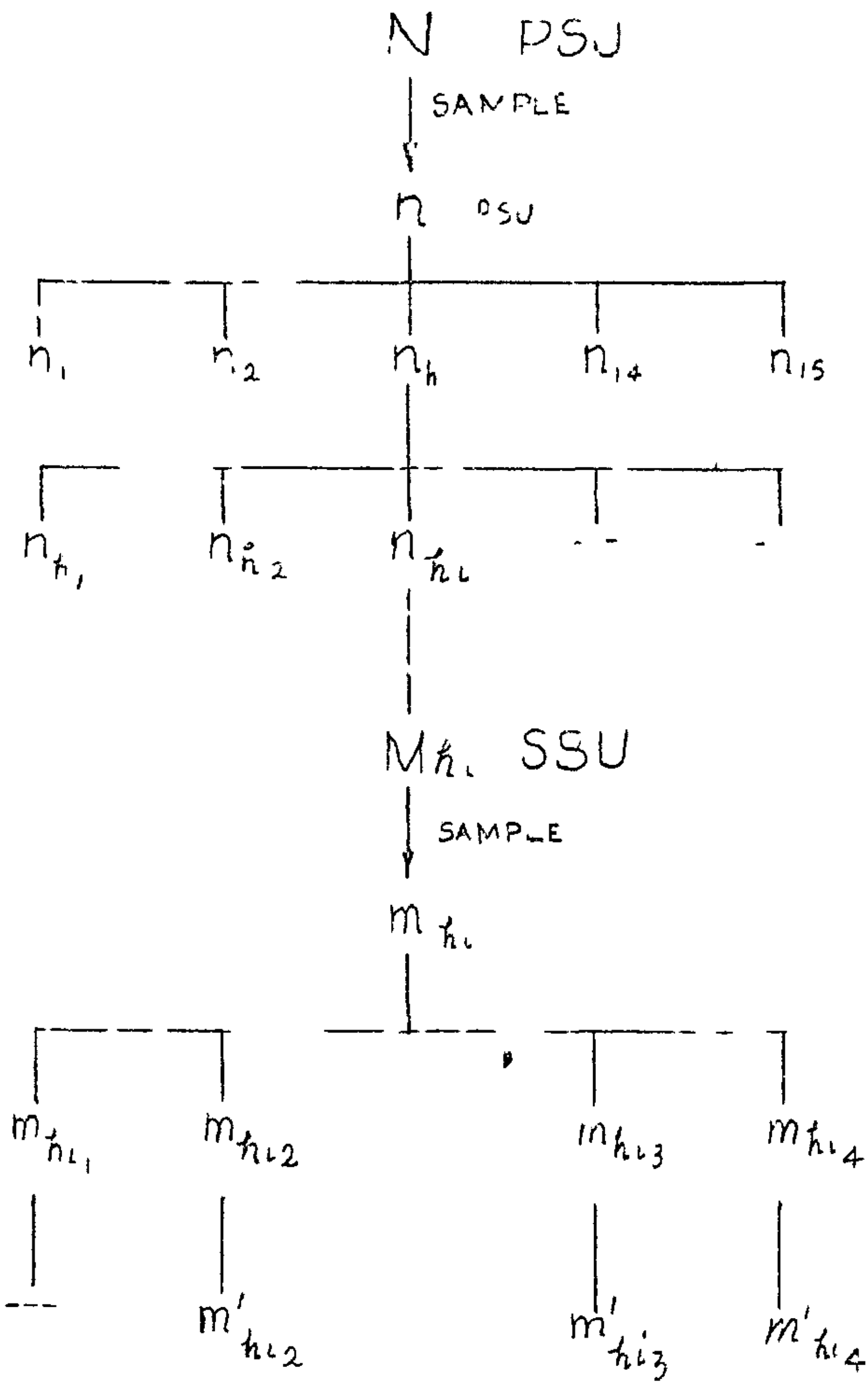
$k = 3$ denotes the class of SSU's which reply for Y only.

$k = 4$ denotes the class of SSU's which donot reply for both X and Y.

such that $\sum_{k=1}^{4} m_{hik} = m_{hi}$

Further let m'_{hik} be the sub sample to be selected from m_{hik}.

[illegible] PLAN
DIAGRAMATIC REPRESENTATION



Estimate of Population Mean

(a) Character y:

The estimate of population mean for character y can be obtained by the following steps: -

Step - I

The estimate of population mean for i-th PSU of h-th stratum is given by

$$\hat{\bar{Y}}_{hi} = \frac{m_{hi1}\,\bar{y}_{hi1} + m_{hi2}\,\bar{y}'_{hi2} + m_{hi3}\,\bar{y}_{hi3} + m_{hi4}\,\bar{y}'_{hi4}}{m_{hi}}$$

$$= \frac{1}{m_{hi}} \left[m_{hi1}\,\bar{y}_{hi1} + m_{hi2}\,\bar{y}_{hi2} + m_{hi3}\,\bar{y}_{hi3} + m_{hi4}\,\bar{y}_{hi4} - m_{hi2}\bar{y}_{hi2} - m_{hi4}\bar{y}_{hi4} + m_{hi2}\,\bar{y}'_{hi2} + m_{hi4}\,\bar{y}'_{hi4} \right]$$

$$= \frac{1}{m_{hi}} \left[\sum_{j=1}^{m_{hi1}} y_{hij} + \sum_{j=1}^{m_{hi2}} y_{hij} + \sum_{j=1}^{m_{hi3}} y_{hij} + \sum_{j=1}^{m_{hi4}} y_{hij} + m_{hi2}\,(\bar{y}'_{hi2} - \bar{y}_{hi2}) + (m_{hi4}\,(\bar{y}'_{hi4} - \bar{y}_{hi4}) \right]$$

$$= \bar{y}_{hi} + \frac{m_{hi2}}{m_{hi}}\,(\bar{y}'_{hi2} - \bar{y}_{hi2}) + \frac{m_{hi4}}{m_{hi}}\,(\bar{y}'_{hi4} - \bar{y}_{hi4}) \quad \ldots\ldots \quad (1)$$

where $\bar{y}_{hi} = \frac{1}{m_{hi}} \sum_{j=1}^{m_{hi}} y_{hij}$ and

$$m_{hi} = m_{hi1} + m_{hi2} + m_{hi3} + m_{hi4}.$$

Step - II

The estimate of population mean for SSU is thus given by

$$\hat{\bar{Y}}_{..} = \frac{N}{nM_o} \sum_{h=1}^{15} \sum_{i=1}^{n_h} M_{hi} \hat{\bar{Y}}_{hi} \qquad \ldots\ldots \qquad (2)$$

$$E(\hat{\bar{Y}}_{..}) = \frac{N}{nM_o} \sum_{h=1}^{15} E\left(\frac{n_h}{N_h} \sum_{i=1}^{N_h} M_{hi} \bar{Y}_{hi}\right)$$

$$= \frac{N}{nM_o} \sum_{h=1}^{15} \sum_{i=1}^{N_h} \frac{n}{N} M_{hi} \bar{Y}_{hi} \quad \left(\text{using the fact that } E(n_h) = \frac{N_h}{N} \cdot n\right)$$

$$= \frac{1}{M_o} \sum_{h=1}^{15} \sum_{i=1}^{N_h} M_{hi} \bar{Y}_{hi}$$

$$= \bar{Y}_{..}$$

Hence $\hat{\bar{Y}}_{..}$ is an unbiased estimate of population mean for character Y.

(b) <u>Character X</u> :

Proceeding as in the case of character Y, an unbiased estimate of population mean for character X is given by

$$\hat{\bar{X}}_{..} = \frac{N}{nM_o} \sum_{h=1}^{15} \sum_{i=1}^{N_h} M_{hi} \hat{\bar{X}}_{hi}$$

where $\hat{\bar{X}}_{hi} = \bar{x}_{hi} + \frac{m_{hi3}}{m_{hi}} (\bar{x}'_{hi3} - \bar{x}_{hi3}) + \frac{m_{hi4}}{m_{hi}} (\bar{x}'_{hi4} - \bar{x}_{hi4})$

and $\bar{x}_{hi} = \frac{1}{m_{hi}} \sum_{j=1}^{m_{hi}} x_{hij}$

<u>Variance of the Estimate</u>

(a) <u>For Character Y</u> :

We have from (2)

$$\hat{\bar{Y}}_{..} = \frac{N}{nM_o} \sum_{h=1}^{15} \sum_{i=1}^{n_h} M_{hi} \hat{\bar{Y}}_{hi} \qquad \ldots\ldots \qquad (3)$$

$$V(\hat{\bar{\bar{Y}}}.. / i) = V E(\hat{\bar{\bar{Y}}}.. / i) + E V(\hat{\bar{\bar{Y}}}.. / i) \quad ..... \quad (4)$$

We will now evaluate, separately the two terms of (4). Considering second term only, we have,

$$V(\hat{\bar{\bar{Y}}}. / i, n_h) = \frac{N^2}{n^2 M_o^2} \sum_{h=1}^{g} \sum_{i=1}^{n_h} M_{hi}^2 \, V(\hat{\bar{\bar{Y}}}_{hi}) \quad ..... \quad (4a)$$

We have from (1)

$$\hat{\bar{\bar{Y}}}_{hi} = \bar{y}_{hi} + \frac{m_{hi2}}{m_{hi}} (\bar{y}'_{hi2} - \bar{y}_{hi2}) + \frac{m_{hi4}}{m_{hi}} (\bar{y}'_{hi4} - \bar{y}_{hi4})$$

$$\therefore \; V(\hat{\bar{\bar{Y}}}_{hi}) = \left(\frac{1}{m_{hi}} - \frac{1}{M_{hi}}\right) S^2_{hiy} + \frac{m^2_{hi2}}{m^2_{hi}} \left(\frac{1}{m'_{hi2}} - \frac{1}{m_{hi2}}\right) s^2_{hi2y}$$

$$+ \frac{m^2_{hi4}}{m^2_{hi}} \left(\frac{1}{m'_{hi4}} - \frac{1}{m_{hi4}}\right) s^2_{hi4y}$$

$$= \left(\frac{1}{m_{hi}} - \frac{1}{M_{hi}}\right) S^2_{hiy} + \frac{m_{hi2}}{m^2_{hi}} (\Theta_{hi2} - 1) s^2_{hi2y} + \frac{m_{hi4}}{m^2_{hi}} (\Theta_{hi4} - 1) s^2_{hi4y}$$

where $\Theta_{hik} = \dfrac{m_{hik}}{m'_{hik}}$ $\qquad k = 1, 2, 3, 4$

Assuming $\Theta_{hik} = \Theta$, i.e., sub-sampling fraction for the second attempt is same for all classes, we get

$$V(\hat{\bar{\bar{Y}}}_{hi}) = \left(\frac{1}{m_{hi}} - \frac{1}{M_{hi}}\right) S^2_{hiy} + \frac{1}{m^2_{hi}} (\Theta - 1) \left[m_{hi2} s^2_{hi2y} + m_{hi4} s^2_{hi4y} \right]$$

$$= v_{hiy} \text{ (say)} \quad ..... \quad (5)$$

where $S^2_{hiy} = \dfrac{1}{M_{hi} - 1} \sum_{j=1}^{M_{hi}} (y_{hij} - \bar{Y}_{hi})^2$, $\bar{Y}_{hi} = \sum_{j=1}^{M_{hi}} y_{hij} / M_{hi}$

$$s^2_{hi2y} = \frac{1}{m_{hi2} - 1} \sum_{j=1}^{m_{hi2}} (y_{hij} - \bar{y}_{hi2})^2, \; \bar{y}_{hi2} = \sum_{j=1}^{m_{hi2}} y_{hij} / M_{hi2}$$

$$s^2_{hi4y} = \frac{1}{m_{hi4}-1} \sum_{j=1}^{m_{hi4}} (y_{hij} - \bar{y}_{hi4})^2, \quad \bar{y}_{hi4} = \sum_{j=1}^{m_{hi4}} y_{hij} / m_{hi4}$$

$\therefore$ from (4 a) and (5), we get

$$V(\hat{\hat{Y}}.. / i, n_h) = \frac{N^2}{n^2 M_o^2} \sum_{h=1}^{15} \sum_{i=1}^{n_h} M^2_{hi} \, v_{hiy}$$

$$\therefore \; E \; V(\hat{\hat{Y}}../i) = \frac{N^2}{n^2 M_o^2} \sum_{h=1}^{15} \frac{n}{N} \sum_{i=1}^{N_h} M^2_{hi} \; E(v_{hiy})$$

$$= \frac{N}{n M_o^2} \sum_{h=1}^{15} \sum_{i=1}^{N_h} M^2_{hi} \; E(v_{hiy}) \quad ..... (6)$$

We will now evaluate first $E(v_{hiy})$ by using the fact that

$$E(m_{hik}) = \frac{M_{hik}}{M_{hi}} \; . \; m_{hi}$$

$$\therefore \; E(v_{hiy}) = V_{hiy} = \left(\frac{1}{m_{hi}} - \frac{1}{M_{hi}}\right) S^2_{hiy} + \frac{(\Theta - 1)}{m^2_{hi}} \left[\frac{M_{hi2}}{M_{hi}} . m_{hi} S^2_{hi2y} + \frac{M_{hi4}}{M_{hi}} m_{hi} S^2_{hi4y} \right]$$

$$= \left(\frac{1}{m_{hi}} - \frac{1}{M_{hi}}\right) S^2_{hiy} + \frac{(\Theta - 1)}{m_{hi} M_{hi}} \left[M_{hi2} S^2_{hi2y} + M_{hi4} S^2_{hi4y} \right]$$

With this, from (6) we get

$$E \; V(\hat{\hat{Y}}../i) = \frac{N}{n M_o^2} \sum_{h=1}^{15} \sum_{i=1}^{N_h} M^2_{hi} \; V_{hiy} \quad (7)$$

We now evaluate first term of (4) i.e. $V \, E(\hat{\hat{Y}}..|i)$

We have from (2)

$$\hat{\bar{Y}}_{..} = \frac{N}{n M_o} \sum_{h=1}^{15} \sum_{i=1}^{n_h} M_{hi} \hat{\bar{Y}}_{hi}$$

$$E(\hat{\bar{Y}}_{..}/i, n_h) = \frac{N}{n M_o} \sum_{h=1}^{15} \sum_{i=1}^{n_h} M_{hi} \bar{Y}_{hi}$$

$$= \frac{N}{n M_o} \sum_{h=1}^{15} n_h \cdot \frac{1}{n_h} \sum_{i=1}^{n_h} M_{hi} \bar{Y}_{hi}$$

$$\therefore V E(\hat{\bar{Y}}_{.}/i, n_h) = \frac{N^2}{n^2 M_o^2} \sum_{h=1}^{15} n_h^2 \left(\frac{1}{n_h} - \frac{1}{N_h}\right) S_{hy}^2$$

$$\text{where } S_{hy}^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} \left(M_{hi}\bar{Y}_{hi} - \frac{\sum_{i=1}^{N_h} M_{hi} \bar{Y}_{hi}}{N_h} \right)^2$$

Relaxing now the condition of fixed n_h we have

$$V E(\hat{\bar{Y}}_{..}/i) = \frac{N^2}{n^2 M_o^2} \sum_{h=1}^{15} E\left[n_h - \frac{n_h^2}{N_h} \right] S_{hy}^2 \quad \ldots\ldots \quad (8)$$

Solving separately the portion $E\left[n_h - \frac{n_h^2}{N_h} \right]$, we have

$$E\left[n_h - \frac{n_h^2}{N_h} \right] = E(n_h) - \frac{E(n_h^2)}{N_h}$$

$$= n \cdot \frac{N_h}{N} - \frac{E(n_h^2)}{N_h} \quad \ldots\ldots \quad (9)$$

Evaluating now $E(n_h^2)$:-

$$E(n_h^2) = V(n_h) + \left[E(n_h) \right]^2 \quad \ldots\ldots \quad (10)$$

$V(n_h)$ is given by

$$V(n_h) = \frac{n(N-n)}{N-1} \cdot \frac{N_h}{N} \cdot \frac{N-N_h}{N}$$

$\therefore$ from (10)

$$E(n_h^2) = \frac{n(N-n)}{N-1} \cdot \frac{N_h}{N} \cdot \frac{N-N_h}{N} + n^2 \cdot \frac{N_h^2}{N^2}$$

$$= \frac{n N_h}{N^2} \left[\frac{N-n}{N-1} (N - N_h) + n \cdot N_h \right]$$

$$= \frac{n N_h}{N^2 (N-1)} \left[(N-n)(N-N_h) + (N-1) \cdot n N_h \right]$$

$$= \frac{n \cdot N_h}{N^2 (N-1)} \quad N \left[N_h (n-1) + (N-n) \right] \quad \ldots\ldots \quad (11)$$

Assume $\frac{N}{N-1} \doteq 1$ and $\frac{N-n}{N} \doteq 1$

With these assumption (11) becomes

$$E(n_h^2) = \frac{n \cdot N_h}{N} \left[\frac{N_h}{N} (n-1) + \frac{N-n}{N} \right]$$

$$= \frac{n \cdot N_h}{N} \left[\frac{N_h}{N} (n-1) + 1 \right] \quad \ldots\ldots \quad (12)$$

∴ from (9) and (12) we get

$$E\left[n_h - \frac{n_h^2}{N_h} \right] = n \cdot \frac{N_h}{N} - \frac{n N_h}{N_h N} \left[\frac{N_h}{N} (n-1) + 1 \right]$$

$$= \frac{n}{N} N_h \left[1 - \frac{n-1}{N} \right] - \frac{n}{N} \quad \ldots\ldots (13)$$

Assuming further N to be very large, we can take

$$1 - \frac{n-1}{N} \doteq 1$$

Under this assumption (13) becomes

$$E\left[n_h - \frac{n_h^2}{N_h} \right] = \frac{n}{N} \cdot N_h - \frac{n}{N} = \frac{n}{N} (N_h - 1) \quad \ldots\ldots (14)$$

$\therefore$ from (8) and (14) we get

$$V\,E\,(\hat{\bar{\bar{Y}}}_{..}/i) = \frac{N^2}{n^2 M_o^2} \quad \frac{n}{N} \sum_{h=1}^{15} (N_h - 1)\, S_{hy}^2 \qquad (15)$$

Assuming further

$$\frac{N_h - 1}{N_h} \doteq 1$$

we get from (15)

$$V\,E\,(\hat{\bar{\bar{Y}}}_{..}/i) = \frac{N}{n M_o^2} \sum_{h=1}^{15} N_h\, S_{hy}^2 \qquad (16)$$

Therefore from (4), (7) and (16) we get

$$V\,(\hat{\bar{\bar{Y}}}_{..}) = \frac{N}{n M_o^2} \sum_{h=1}^{15} N_h\, S_{hy}^2 + \frac{N}{n M_o^2} \sum_{h=1}^{15} \sum_{i=1}^{N_h} M_{hi}^2\, V_{hiy} \qquad (17)$$

$$= \frac{N}{n M_o^2} \sum_{h=1}^{15} N_h\, S_{hy}^2 + \frac{N}{n M_o^2} \sum_{h=1}^{15} \sum_{i=1}^{N_h} M_{hi}^2 \Big[\Big(\frac{1}{m_{hi}} - \frac{1}{M_{hi}}\Big) S_{hiy}^2 +$$

$$\frac{(\Theta - 1)}{m_{hi} M_{hi}} (M_{hi2}\, S_{hi2y}^2 + M_{hi4}\, S_{hi4y}^2) \Big]$$

.... (18)

(b) <u>Character X</u> :

Proceeding similarly as in the case of character Y, we get for character X

$$V\,(\hat{\bar{\bar{X}}}_{..}) = \frac{N}{n M_o^2} \sum_{h=1}^{15} N_h\, S_{hx}^2 + \frac{N}{n M_o^2} \sum_{h=1}^{15} \sum_{i=1}^{N_h} M_{hi}^2\, V_{hix}$$

where $V_{hix} = \Big(\frac{1}{m_{hi}} - \frac{1}{M_{hi}}\Big) S_{hix}^2 + \frac{(\Theta - 1)}{m_{hi} M_{hi}} (M_{hi3}\, S_{hi3x}^2 + M_{hi4}\, S_{hi4x}^2)$

$$\therefore\ V(\hat{\bar{X}}..) = \frac{N}{n M_o^2} \sum_{h=1}^{15} N_h S_{hx}^2 + \frac{N}{n M_o^2} \sum_{h=1}^{15} \sum_{i=1}^{N_h} M_{hi}^2 \Big[\Big(\frac{1}{m_{hi}} - \frac{1}{M_{hi}}\Big) S_{hix}^2$$

$$+ \frac{(\Theta - 1)}{m_{hi} M_{hi}} \left(M_{hi3} S_{hi3x}^2 + M_{hi4} S_{hi4x}^2 \right) \Big] \quad \ldots\ldots (19)$$

Estimate of Variance :

(a) **Character** y

We have from (17)

$$V(\bar{Y}..) = \frac{N}{n M_o^2} \sum_{h=1}^{15} N_h S_{hy}^2 + \frac{N}{n M_o^2} \sum_{h=1}^{15} \sum_{i=1}^{N_h} M_{hi}^2 V_{hiy} \quad \ldots\ldots \quad (20)$$

To find estimate of variance, we obtain estimates for the two terms of (20) separately:

(i) **Estimate of** $\sum_{i=1}^{N_h} M_{hi}^2 V_{hiy}$:

Consider $E\Big[\frac{N}{n} \sum_{i=1}^{n_h} M_{hi}^2 v_{hiy} / n_h \Big] = E\Big[\frac{N}{n} \cdot \frac{n_h}{N_h} \sum_{i=1}^{N_h} M_{hi}^2 V_{hiy} \Big]$

$$= \sum_{i=1}^{N_h} M_{hi}^2 V_{hiy}$$

$$\therefore \text{ Estimate of } \sum_{i=1}^{N_h} M_{hi}^2 V_{hiy} = \frac{N}{n} \sum_{i=1}^{n_h} M_{hi}^2 v_{hiy} \quad \ldots\ldots \quad (21)$$

(ii) **Estimate of** $\sum_{h=1}^{15} N_h S_{hy}^2$

We had $S_{hy}^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} \left(M_{hi} \bar{Y}_{hi} - \frac{\sum_{i=1}^{N_h} M_{hi} \bar{Y}_{hi}}{N_h} \right)^2$

Denote $s_{hy}^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} \left(M_{hi} \hat{\bar{\bar{Y}}}_{hi} - \frac{\sum_{i=1}^{n_{hh}} M_{hi} \hat{\bar{\bar{Y}}}_{hi}}{n_h} \right)^2$

or $(n_h - 1)\, s_{hy}^2 = \sum_{i=1}^{n_h} M_{hi}^2 \hat{\bar{\bar{Y}}}_{hi}^2 - \frac{1}{n_h} \left(\sum_{i=1}^{n_h} M_{hi} \hat{\bar{\bar{Y}}}_{hi} \right)^2$

or $(n_h - 1)\, E(s_{hy}^2 / n_h) = \sum_{i=1}^{n_h} M_{hi}^2 E(\hat{\bar{\bar{Y}}}_{hi}^2) - \frac{1}{n_h} \left[E\left(\sum_{i=1}^{n_h} M_{hi} \hat{\bar{\bar{Y}}}_{hi} \right)^2 \right]$ ..... (22)

Evaluating the two terms of (22) separately ; we get

<u>Ist term</u> :

We know $V(\hat{\bar{\bar{Y}}}_{hi} / n_h) = E(\hat{\bar{\bar{Y}}}_{hi}^2 / n_h) - \left[E(\hat{\bar{\bar{Y}}}_{hi} / n_h) \right]^2$

$\therefore E(\hat{\bar{\bar{Y}}}_{hi}^2 / n_h) = V(\hat{\bar{\bar{Y}}}_{hi} / n_h) + \left[E(\hat{\bar{\bar{Y}}}_{hi} / n_h) \right]^2$

$= v_{hiy} + \bar{\bar{Y}}_{hi}^2$ ..... (23)

<u>2nd term</u> :

We know $V\left[\sum_{i=1}^{n_h} M_{hi} \hat{\bar{\bar{Y}}}_{hi} / n_h \right] = E\left(\sum_{i=1}^{n_h} M_{hi} \hat{\bar{\bar{Y}}}_{hi} / n_h \right)^2$

$- \left[E\left(\sum_{i=1}^{n_h} M_{hi} \hat{\bar{\bar{Y}}}_{hi} / n_h \right) \right]^2$

$\therefore E\left(\sum_{i=1}^{n_h} M_{hi} \hat{\bar{\bar{Y}}}_{hi} / n_h \right)^2 = V\left[\sum_{i=1}^{n_h} M_{hi} \hat{\bar{\bar{Y}}}_{hi} / n_h \right] + \left[E\left(\sum_{i=1}^{n_h} M_{hi} \hat{\bar{\bar{Y}}}_{hi} / n_h \right) \right]^2$

$= \sum_{i=1}^{n_h} M_{hi}^2 v_{hiy} + \left[\sum_{i=1}^{n_h} M_{hi} \bar{\bar{Y}}_{hi} \right]^2$ ..... (24)

$\therefore$ from (22), (23) and (24) we get

$(n_h - 1)\, E(s_{hy}^2 / n_h) = \sum_{i=1}^{n_h} M_{hi}^2 (v_{hiy} + \bar{\bar{Y}}_{hi}^2) - \frac{1}{n_h} \left[\sum_{i=1}^{n_h} M_{hi}^2 v_{hiy} + \left(\sum_{i=1}^{n_h} M_{hi} \bar{\bar{Y}}_{hi} \right)^2 \right]$

$$= \frac{n_h - 1}{n_h} \sum_{i=1}^{n_h} M_{hi}^2 v_{hiy} + \sum_{i=1}^{n_h} \left(M_{hi} \bar{Y}_{hi} - \frac{\sum_{i=1}^{n_h} M_{hi} \bar{Y}_{hi}}{n_h} \right)^2$$

$$\therefore E (s_{hy}^2 / n_h) = \frac{1}{n_h} \sum_{i=1}^{n_h} M_{hi}^2 v_{hiy} + \frac{1}{n_h - 1} \sum_{i=1}^{n_h} \left(M_{hi} \bar{Y}_{hi} - \frac{\sum_{i=1}^{n_h} M_{hi} \bar{Y}_{hi}}{n_h} \right)^2$$

$$\therefore E (s_{hy}^2) = E_{n_h} E (s_{hy}^2 / n_h) = \frac{1}{N_h} \sum_{i=1}^{N_h} M_{hi}^2 V_{hiy} + \frac{1}{N_h - 1} \sum_{i=1}^{N_h} \left(M_{hi} \bar{Y}_{hi} - \sum_{i=1}^{N_h} \frac{M_{hi} \bar{Y}_{hi}}{N_h} \right)^2$$

$$= \frac{1}{N_h} \sum_{i=1}^{N_h} M_{hi}^2 V_{hiy} + S_{hy}^2$$

$$\therefore \text{Estimate of } S_{hy}^2 = s_{hy}^2 - \text{Est.} \left[\frac{1}{N_h} \sum_{i=1}^{N_h} M_{hi}^2 V_{hiy} \right]$$

$$= s_{hy}^2 - \frac{n}{N} \frac{1}{n_h} \cdot \frac{N}{n} \sum_{i=1}^{n_h} M_{hi}^2 v_{hiy}$$

$$= s_{hy}^2 - \frac{1}{n_h} \sum_{i=1}^{n_h} M_{hi}^2 v_{hiy} \qquad \ldots\ldots \quad (25)$$

$\therefore$ from (20), (21) and (25)

$$\text{Estimate } V (\hat{\hat{Y}}..) = \frac{N}{n M_o^2} \sum_{h=1}^{15} \frac{N}{n} \sum_{i=1}^{n_h} M_{hi}^2 v_{hiy} + \frac{N}{n M_o^2} \sum_{h=1}^{15} \frac{N}{n} n_h \left(s_{hy}^2 - \frac{1}{n_h} \sum_{i=1}^{n_h} M_{hi}^2 v_{hiy} \right)$$

$$= \frac{N^2}{n^2 M_o^2} \sum_{h=1}^{15} \sum_{i=1}^{n_h} M_{hi}^2 v_{hiy} + \frac{N^2}{n^2 M_o^2} \sum_{h=1}^{15} n_h \left(s_{hy}^2 - \frac{1}{n_h} \sum_{i=1}^{n_h} M_{hi}^2 v_{hiy} \right)$$

$$= \frac{N^2}{n^2 M_o^2} \sum_{h=1}^{15} n_h \; s_{hy}^2 \qquad \ldots\ldots \quad (26)$$

CHAPTER III

THE COST FUNCTION AND OPTIMUM SAMPLE SIZES

Cost of the Survey :

Let C_o = Overhead cost

C = Cost per unit of listing and preparing a frame

C_{1x} = Cost of enumerating a unit for character X at the first attempt.

C_{1y} = Cost of enumerating a unit for character Y at the first attempt.

C_{2x} = Cost of enumerating a unit for character X at the second attempt.

C_{2y} = Cost of enumerating a unit for character Y at the second attempt.

With these notations, the total cost of the survey, C^*, can be written as:

$$C = C_o + C_N + (C_{1x} + C_{1y}) \sum_{h=1}^{15} \sum_{i=1}^{n_h} m_{hi} + C_{2y} \sum_{h=1}^{15} \sum_{i=1}^{n_h} m'_{hi2}$$

$$+ C_{2x} \sum_{h=1}^{15} \sum_{i=1}^{n_h} m'_{hi3} + (C_{2x} + C_{2y}) \sum_{h=1}^{15} \sum_{i=1}^{n_h} m'_{hi4}$$

$$= C_o + Cn + (C_{1x} + C_{1y}) \sum_{h=1}^{15} \sum_{i=1}^{n_h} m_{hi} + f\, C_{2y} \sum_{h=1}^{15} \sum_{i=1}^{n_h} m_{hi2}$$

$$+ f\, C_{2x} \sum_{h=1}^{15} \sum_{i=1}^{n_h} m_{hi3} + f\, (C_{2x} + C_{2y}) \sum_{h=1}^{15} \sum_{i=1}^{n_h} m_{hi4}$$

where $\frac{1}{\Theta} = f$

Now using the fact that

$$E(m_{hik}) = \frac{M_{hik}}{M_{hi}} m_{hi} \qquad k = 1, 2, 3, 4$$

We have

$$E(C) = C_o + Cn + (C_{1x} + C_{1y}) \sum_{h=1}^{15} \frac{n}{N} \sum_{i=1}^{N_h} m_{hi} + f\, C_{2y} \sum_{h=1}^{15} \frac{n}{N} \sum_{i=1}^{N_h} \frac{M_{hi2}}{M_{hi}} . m_{hi}$$

$$+ f\, C_{2x} \sum_{h=1}^{15} \frac{n}{N} \sum_{i=1}^{N_h} \frac{M_{hi3}}{M_{hi}} . m_{hi} + f\,(C_{2x} + C_{2y}) \sum_{h=1}^{15} \frac{n}{N} \sum_{i=1}^{N_h} \frac{M_{hi4}}{M_{hi}} . m_{hi}$$

$$= C_o + Cn + \frac{n}{N}(C_{1x} + C_{1y}) \sum_{h=1}^{15} \sum_{i=1}^{N_h} m_{hi} + f . \frac{n}{N} . \lambda \left[C_{2y} \sum_{h=1}^{15} \sum_{i=1}^{N_h} M_{hi2} \right.$$

$$\left. + C_{2x} \sum_{h=1}^{15} \sum_{i=1}^{N_h} M_{hi3} + (C_{2x} + C_{2y}) \sum_{h=1}^{15} \sum_{i=1}^{N_h} M_{hi4} \right]$$

where $\lambda = \frac{m_{hi}}{M_{hi}}$

$$= C_o + Cn + \frac{n}{N}(C_{1x} + C_{1y}) \sum_{h=1}^{15} \sum_{i=1}^{N_h} m_{hi} + f . \frac{n}{N} . \lambda \left[C_{2y} \sum_{h=1}^{15} \sum_{i=1}^{N_h} (M_{hi2} + M_{hi4}) \right.$$

$$\left. + C_{2x} \sum_{h=1}^{15} \sum_{i=1}^{N_h} (M_{hi3} + M_{hi4}) \right]$$

$$= C_o + Cn + \frac{n}{N}(C_{1x} + C_{1y}) \sum_{h=1}^{15} \sum_{i=1}^{N_h} m_{hi} + f . \frac{n}{N} . \lambda \left[C_{2y} M_y + C_{2x} M_x \right]$$

where $M_y = \sum_{h=1}^{15} \sum_{i=1}^{N_h} (M_{hi2} + M_{hi4})$ = Total number of units in the population which would not have responded for character Y.

$M_x = \sum_{h=1}^{15} \sum_{i=1}^{N_h} (M_{hi3} + M_{hi4})$ = Total number of units in the population which would not have responded for character X.

With this,

$$E(C) = C_o + Cn + \lambda . \frac{n}{N}(C_{1x}+C_{1y})\sum_{h=1}^{15}\sum_{i=1}^{N_h} M_{hi} + f . \frac{n}{N}\lambda [C_{2y} M_y + C_{2x} M_x]$$

$$= C_o + Cn + \lambda . \frac{n}{N}(C_{1x}+C_{1y})M_o + f . \frac{n}{N}\lambda [C_{2y} M_y + C_{2x} M_x] \quad \ldots (27)$$

$$= a_o + a_1 n + a_2 n \lambda + a_3 n \lambda f \qquad \ldots\ldots (28)$$

where a_o, a_1, a_2, a_3 are constants, and are given by

$$a_o = C_o \qquad a_1 = C$$

$$a_2 = (C_{1x} + C_{1y})\frac{M_o}{N}, \quad a_3 = \frac{1}{N}[C_{2y} M_y + C_{2x} M_x]$$

Variance Functions :

The variances of $\bar{Y}..$ and $\bar{X}..$ can be written in the following simplified forms. We have from (18)

$$V(\bar{Y}..) = \frac{N}{n M_o^2}\sum_{h=1}^{15}\sum_{i=1}^{N_h} M_{hi}^2 \left[\left(\frac{1}{m_{hi}} - \frac{1}{M_{hi}}\right) S_{hiy}^2 + \frac{(\Theta-1)}{m_{hi} M_{hi}}(M_{hi2} S_{hi2y}^2 + M_{hi4} S_{hi4y}^2)\right] + \frac{N}{n M_o^2}\sum_{h=1}^{15} N_h S_{hy}^2$$

$$= \left(\frac{1}{\lambda} - 1\right)\frac{N}{n M_o^2}\sum_{h=1}^{15}\sum_{i=1}^{N_h} M_{hi} S_{hiy}^2 + \frac{N}{n M_o^2}\left(\frac{1}{f} - 1\right)\frac{1}{\lambda}\sum_{h=1}^{15}\sum_{i=1}^{N_h}(M_{hi2} S_{hi2y}^2 + M_{hi4} S_{hi4y}^2) + \frac{N}{n M_o^2}\sum_{h=1}^{15} N_h S_{hy}^2$$

$$= \left(\frac{1}{\lambda} - 1\right) . \frac{1}{n} b_o + \frac{k}{n}\left(\frac{1}{f} - 1\right)\frac{1}{\lambda} . b_1 + \frac{b_2}{n}$$

$$= \frac{1}{n}(b_2 - b_o) + \frac{1}{n\lambda}(b_o - b_1) + \frac{1}{n\lambda f} b_1$$

$$= \frac{B_{oy}}{n} + \frac{B_{1y}}{n\lambda} + \frac{B_{2y}}{n\lambda f} \quad \ldots\ldots\ldots \qquad (18a)$$

where $B_{oy} = b_2 - b_o = \frac{N}{M_o^2} \left[\sum_{h=1}^{15} N_h S_{hy}^2 - \sum_{h=1}^{15} \sum_{i=1}^{N_h} M_{hi} S_{hiy}^2 \right]$

$$B_{1y} = b_o - b_1 = \frac{N}{M_o^2} \left[\sum_{h=1}^{15} \sum_{i=1}^{N_h} M_{hi} S_{hiy}^2 - \sum_{h=1}^{15} \sum_{i=1}^{N_h} (M_{hi2} S_{hi2y}^2 + M_{hi4} S_{hi4y}^2) \right]$$

$$B_{2y} = b_1 = \frac{N}{M_o^2} \left[\sum_{h=1}^{15} \sum_{i=1}^{N_h} (M_{hi2} S_{hi2y}^2 + M_{hi4} S_{hi4y}^2) \right]$$

Similarly $V(\hat{\bar{X}}..) = \frac{B_{ox}}{n} + \frac{B_{1x}}{n\lambda} + \frac{B_{2x}}{n\lambda f} \quad \ldots\ldots \qquad (19a)$

where $B_{ox} = \frac{N}{M_o^2} \left[\sum_{h=1}^{15} N_h S_{hx}^2 - \sum_{h=1}^{15} \sum_{i=1}^{N_h} M_{hi} S_{hix}^2 \right]$

$$B_{1x} = \frac{N}{M_o^2} \left[\sum_{h=1}^{15} \sum_{i=1}^{N_h} M_{hi} S_{hix}^2 - \sum_{h=1}^{15} \sum_{i=1}^{N_h} (M_{hi3} S_{hi3x}^2 + M_{hi4} S_{hi4x}^2) \right]$$

$$B_{2x} = \frac{N}{M_o^2} \left[\sum_{h=1}^{15} \sum_{i=1}^{N_h} (M_{hi3} S_{hi3x}^2 + M_{hi4} S_{hi4x}^2) \right]$$

Optimum Values :

The problem of obtaining optimum values is not so simple in this case, as in case of an ordinary two stage design. Here, we have two variance functions and one combined cost function. If we keep two of them fixed and minimise the third one, the solution obtained may not be economical. Hence, we will discard this method, and look for other methods for obtaining the optimum values. We will consider two different methods for obtaining these values. The utility of these results will depend on the various situations that may arise.

(a) Non-Linear Programming Method :

Suppose we define the optimum values of n, λ and f as those which will minimise the cost of the survey under the conditions

$$V(\hat{\bar{\bar{Y}}}..) \leq V_1 \qquad (i)$$

$$V(\hat{\bar{\bar{X}}}..) \leq V_2 \qquad (ii)$$

where V_1 and V_2 are fixed constants.

The method of non-linear programming can be applied for obtaining the solution. The first step is to obtain the values of n, λ and f which will minimise the cost of the survey under the equality of condition (i), i.e. under the condition

$$V(\hat{\bar{\bar{Y}}}..) = V_1$$

The values obtained thus, are then substituted in condition (ii) to see if it is satisfied. If this is satisfied, the above values of n, λ and f are the optimum values.

If the second condition is not satisfied, the process is repeated by considering the equality part of condition (ii). The values obtained are substituted in the first condition to see if it is satisfied.

If in both cases, the other condition is not satisfied, then there is no feasible solution.

(b) Method of Weights:

Sometimes, it is possible to give weights to the two characters according to their economic importance. In such cases, we can obtain a combined variance function by weighting the two variances with their respective weights. If W, (1 - W) are the weights given to the two characters, the combined variance is given by

$$V = W\ V(\hat{\bar{\bar{Y}}}_{..}) + (1 - W)\ V(\hat{\bar{\bar{X}}}_{..})$$

Now we minimise this variance function for a fixed cost.

Calculation of Optimum Values

(a) Non-Linear Programming Method

Applying the method of non-linear programming to our problem, we first obtain the values of n, λ, f which minimise the expected cost under the condition

$$\frac{B_{oy}}{n} + \frac{B_{1y}}{n\lambda} + \frac{B_{2y}}{n\lambda f} = V_1$$

Thus take

$$\phi = V_1 \times (C - a_o)$$

$$= \left(\frac{B_{oy}}{n} + \frac{B_{1y}}{n\lambda} + \frac{B_{2y}}{n\lambda f}\right) \quad (a_1 n + a_2 n\lambda + a_3 n\lambda f)$$

$$= (a_1\ B_{oy} + a_2 B_{1y} + a_3 B_{2y}) + \frac{a_1 B_{1y}}{\lambda} + \frac{a_1 B_{2y}}{\lambda f} + a_2 B_{oy}\lambda$$

$$+ \frac{a_2 B_{2y}}{f} + a_3 B_{oy}\lambda f + a_3 B_{1y} f$$

$$= (a_1 B_{oy} + a_2 B_{1y} + a_3 B_{2y}) + \left(\frac{a_1 B_{1y}}{\lambda} + a_2 B_{oy}\lambda\right)$$

$$+ \left(\frac{a_2 B_{2y}}{f} + a_3 B_{1y} f\right) + \left(\frac{a_1 B_{2y}}{\lambda f} + a_3 B_{oy}\lambda f\right)$$

$$= L_0 + \left(\frac{L_1}{\sqrt{\lambda}} - L_1' \sqrt{\lambda} \right)^2 + \left(\frac{L_2}{\sqrt{f}} - L_2' \sqrt{f} \right)^2$$

$$+ \left(\frac{L_3}{\sqrt{\lambda f}} - L_3' \sqrt{\lambda f} \right)^2$$

where

$$L_1^2 = a_1 B_{1y} \qquad L_2^2 = a_2 B_{2y} \qquad L_3^2 = a_1 B_{2y}$$

$$L_1'^2 = a_2 B_{ey} \qquad L_2'^2 = a_3 B_{1y} \qquad L_3'^2 = a_3 B_{ey}$$

ϕ will be minimum when each of the square term on the right hand side of above equation is minimum.

$$\therefore \quad \frac{L_1}{\sqrt{\lambda}} - L_1' \sqrt{\lambda} = 0$$

$$\lambda = L_1 / L_1' = \frac{\sqrt{a_1 B_{1y}}}{\sqrt{a_2 B_{ey}}}$$

$$f = L_2 / L_2' = \frac{\sqrt{a_2 B_{2y}}}{\sqrt{a_3 B_{1y}}}$$

$$\lambda f = L_3 / L_3' = \frac{\sqrt{a_1 B_{2y}}}{\sqrt{a_3 B_{ey}}}$$

We have $$\lambda^2 = \frac{a_1 B_{1y}}{a_2 B_{ey}}$$

Substituting the values of various constants, we have

$$\lambda^2 = \frac{CN \left[\sum_{h=1}^{15} \sum_{i=1}^{N_h} M_{hi} S'^2_{hiy} - \sum_{h=1}^{15} \sum_{i=1}^{N_h} S'^2_{hiy} (M_{hi2} + M_{hi4}) \right]}{(C_{1x} + C_{1y}) M_o \left[\sum_{h=1}^{15} N_h S^2_{hy} - \sum_{h=1}^{15} \sum_{i=1}^{N_h} M_{hi} S'^2_{hiy} \right]}$$

Let $M_{hi2} + M_{hi4} = M_{hiy}$ = Total number of SSU's in i-th PSU of h-th stratum which donot reply for Y.

$M_{hi3} + M_{hi4} = M_{hix}$ = Corresponding number for X

With this

$$\lambda^2 = \frac{CN \left[\sum_{h=1}^{15} \sum_{i=1}^{N_h} (M_{hi} - M_{hiy}) S'^2_{hiy} \right]}{(C_{1x} + C_{1y}) M_o \left[\sum_{h=1}^{15} N_h S^2_{hy} - \sum_{h=1}^{15} \sum_{i=1}^{N_h} S'^2_{hiy} M_{hi} \right]} \quad \ldots. (29)$$

Similarly $f^2 = \frac{a_2 \; B_{2y}}{a_3 \; B_{1y}}$

Substituting the values of various constants, we have

$$f^2 = \frac{(C_{1x}+C_{1y}) \frac{M_o}{N} \cdot \frac{N}{M_o^2} \left[\sum_{h=1}^{15} \sum_{i=1}^{N_h} (M_{hi2} S^2_{hi2y} + M_{hi4} S^2_{hi4y}) \right]}{\frac{1}{N} (C_{2y} M_y + C_{2x} M_x) \cdot \frac{N}{M_o^2} \left[\sum_{h=1}^{15} \sum_{i=1}^{N_h} M_{hi} S^2_{hiy} - \sum_{h=1}^{15} \sum_{i=1}^{N_h} (M_{hi2} S^2_{hi2y} + M_{hi4} S^2_{hi4y}) \right]}$$

$$= \frac{(C_{1x} + C_{1y}) M_o \left[\sum_{h=1}^{15} \sum_{i=1}^{N_h} M_{hiy} S'^2_{hiy} \right]}{(C_{2y} M_y + C_{2x} M_x) \left[\sum_{h=1}^{15} \sum_{i=1}^{N_h} (M_{hi} - M_{hiy}) S'^2_{hiy} \right]} \quad \ldots.. (30)$$

Thus by knowing the values of λ and f from (29) and (30), we get n from

$$n = (C - a_o)/(a_1 + a_2\lambda + a_3 \lambda f) \quad \ldots\ldots \quad (31)$$

The values obtained in (29), (30) and (31) are taken as the Optimum values if they satisfy the condition (11). If it is not satisfied then we find the optimum values of n, λ, f which minimise the expected cost under the condition

$$\frac{B_{ox}}{n} + \frac{B_{1x}}{n\lambda} + \frac{B_{2x}}{n\lambda f} = V_2$$

Taking $\phi = V_2 \times (C - a_o)$ and proceeding exactly in the same way as above, the optimum values of n, λ, f are given by

$$\lambda^2 = \frac{CN \left[\sum_{h=1}^{15} \sum_{i=1}^{N_h} (M_{hi} - M_{hix}) \, S'^2_{hix} \right]}{(C_{1x} + C_{1y}) M_o \left[\sum_{h=1}^{15} N_h S^2_{hx} - \sum_{h=1}^{15} \sum_{i=1}^{N_h} M_{hi} \, S'^2_{hix} \right]} \quad \ldots\ldots\ldots (32)$$

$$f^2 = \frac{(C_{1x} + C_{1y}) M_o \left[\sum_{h=1}^{15} \sum_{i=1}^{N_h} M_{hix} \, S^2_{hix} \right]}{(C_{2y} M_y + C_{2x} M_x) \left[\sum_{h=1}^{15} \sum_{i=1}^{N_h} (M_{hi} - M_{hix}) \, S'^2_{hix} \right]} \quad \ldots\ldots \quad (33)$$

Thus by knowing the values of λ and f, n is given by

$$n = (C - a_o) / (a_1 + a_2 \lambda + a_3 \lambda f) \quad \ldots\ldots \quad (34)$$

Then the values obtained in (32), (33) and (34) are the optimum values, if condition (1) is satisfied. Otherwise there is no feasible solution.

(b) <u>Method of Weights</u>

$$V = W(V(\bar{Y}_{..}) + (1-W)\,V(\bar{X}_{..})$$

$$= W\left[\frac{B_{oy}}{n} + \frac{B_{1y}}{n\lambda} + \frac{B_{2y}}{n\lambda f}\right] + (1-W)\left[\frac{B_{ox}}{n} + \frac{B_{1x}}{n\lambda} + \frac{B_{2x}}{n\lambda f}\right]$$

$$= \frac{1}{n}\left[W B_{oy} + (1-W) B_{ox}\right] + \frac{1}{n\lambda}\left[W B_{1y} + (1-W) B_{1x}\right] +$$

$$\frac{1}{n\lambda f}\left[W B_{2y} + (1-W) B_{2x}\right]$$

or $V = \frac{K_o}{n} + \frac{K_1}{n\lambda} + \frac{K_2}{n\lambda f}$

where $K_o = W B_{oy} + (1-W) B_{ox}$

$K_1 = W B_{1y} + (1-W) B_{1x}$

$K_2 = W B_{2y} + (1-W) B_{2x}$

Also we had $C = a_o + a_1 n + a_2 \cdot n\lambda + a_3 n\lambda f$

$\therefore C - a_o = a_1 n + a_2 n\lambda + a_3 n\lambda f$

Take $\phi = V \times (C - a_o)$

$$= \left(\frac{K_o}{n} + \frac{K_1}{n\lambda} + \frac{K_2}{n\lambda f}\right)(a_1 n + a_2 n\lambda + a_3 n\lambda f)$$

$$= (K_o a_1 + K_1 a_2 + K_2 a_3) + \left(K_o a_2 \lambda + \frac{K_1 a_1}{\lambda}\right) +$$

$$\left(K_1 a_3 f + \frac{K_2 a_2}{f}\right) + \left(K_o a_3 \lambda f + \frac{K_2 a_1}{\lambda f}\right)$$

$$= L_0 + \left(L_1\sqrt{\lambda} - \frac{L_1'}{\sqrt{\lambda}} \right)^2 + \left(L_2\sqrt{f} - \frac{L_2'}{\sqrt{f}} \right)^2$$

$$+ \left(L_3\sqrt{\lambda f} - \frac{L_3'}{\sqrt{\lambda f}} \right)^2$$

where $L_1^2 = K_o a_2$ $\quad L_2^2 = K_1 a_3$ $\quad L_3^2 = K_o a_3$

$L_1'^2 = K_1 a_1$ $\quad L_2'^2 = K_2 a_2$ $\quad L_3'^2 = K_2 a_1$

ϕ will be minimum when each of the square term on the right hand side of the above equation is minimum.

$$\therefore \quad L_1\sqrt{\lambda} - \frac{L_1'}{\sqrt{\lambda}} = 0$$

$$\lambda = L_1'/L_1 = \frac{\sqrt{K_1 a_1}}{\sqrt{K_o a_2}}$$

$$f = L_2'/L_2 = \frac{\sqrt{K_2 a_2}}{\sqrt{K_1 a_3}}$$

$$\lambda f = L_3'/L_3 = \frac{\sqrt{K_2 a_1}}{\sqrt{K_o a_3}}$$

Thus we have

$$\lambda^2 = \frac{K_1 a_1}{K_o a_2} = \frac{a_1 \left[W B_{ly} + (1-W) B_{lx} \right]}{a_2 \left[W B_{oy} + (1-W) B_{ox} \right]}$$

Substituting the values of constants, and solving numerator and denominator separately, we have

$$\text{Num} = C \left[W.\frac{N}{M_o^2}\left\{ \sum_{h=1}^{15}\sum_{i=1}^{N_h} M_{hi}\, S^2_{hiy} - \sum_{h=1}^{15}\sum_{i=1}^{N_h}\left(M_{hi2}\, S^2_{hi2y} + M_{hi4}\, S^2_{hi4y}\right)\right\}\right.$$

$$\left. + (1-W)\frac{N}{M_o^2}\left\{ \sum_{h=1}^{15}\sum_{i=1}^{N_h} M_{hi}\, S^2_{hix} - \sum_{h=1}^{15}\sum_{i=1}^{N_h}\left(M_{hi3}\, S^2_{hi3x} + M_{hi4}\, S^2_{hi4x}\right)\right\}\right]$$

Assuming now

$$S^2_{hiy} = S^2_{hi2y} = S^2_{hi4y} = S'^2_{hiy}$$

$$S^2_{hix} = S^2_{hi3x} = S^2_{hi4x} = S'^2_{hix}$$

$$\text{Num.} = C.\frac{N}{M_o^2}\left[W\left\{ \sum_{h=1}^{15}\sum_{i=1}^{N_h} M_{hi}\, S'^2_{hiy} - \sum_{h=1}^{15}\sum_{i=1}^{N_h}\left(M_{hi2}\, S'^2_{hiy} + M_{hi4}\, S'^2_{hiy}\right)\right\}\right.$$

$$\left. + (1-W)\left\{ \sum_{h=1}^{15}\sum_{i=1}^{N_h} M_{hi}\, S'^2_{hix} - \sum_{h=1}^{15}\sum_{i=1}^{N_h}\left(M_{hi3}\, S'^2_{hix} + M_{hi4}\, S'^2_{hix}\right)\right\}\right]$$

Let $M_{hi2} + M_{hi4} = M_{hiy}$ = Total number of SSU's in the i-th PSU of h-th stratum which donot respond for Y.

$M_{hi3} + M_{hi4} = M_{hix}$ = Corresponding number for x.

With this, numerator becomes

$$\text{Num.} = \frac{CN}{M_o^2}\left[W\left\{ \sum_{h=1}^{15}\sum_{i=1}^{N_h} M_{hi}\, S'^2_{hiy} - \sum_{h=1}^{15}\sum_{i=1}^{N_h} M_{hiy}\, S'^2_{hiy}\right\} + (1-W)\right.$$

$$\left.\left\{ \sum_{h=1}^{15}\sum_{i=1}^{N_h} M_{hi}\, S'^2_{hix} - \sum_{h=1}^{15}\sum_{i=1}^{N_h} M_{hix}\, S'^2_{hix}\right\}\right]$$

$$= \frac{CN}{M_o^2}\left[W \sum_{h=1}^{15}\sum_{i=1}^{N_h}\left(M_{hi} - M_{hiy}\right) S'^2_{hiy} + (1-W)\sum_{h=1}^{15}\sum_{i=1}^{N_h}\left(M_{hi} - M_{hix}\right) S'^2_{hix}\right]$$

$$\text{Denom.} = (C_{1x} + C_{1y}) \frac{M_o}{N} \Big[\frac{WN}{M_o^2} \Big\{ \sum_{h=1}^{15} N_h S_{hy}^2 - \sum_{h=1}^{15} \sum_{i=1}^{N_h} M_{hi} S_{hiy}^2 \Big\}$$

$$+ (1-W) \frac{N}{M_o^2} \Big\{ \sum_{h=1}^{15} N_h S_{hx}^2 - \sum_{h=1}^{15} \sum_{i=1}^{N_h} M_{hi} S_{hix}^2 \Big\} \Big]$$

$$= \frac{(C_{1x} + C_{1y})}{M_o} \Big[W \Big\{ \sum_{h=1}^{15} N_h S_{hy}^2 - \sum_{h=1}^{15} \sum_{i=1}^{N_h} M_{hi} S'^2_{hiy} \Big\} + (1-W)$$

$$\Big\{ \sum_{h=1}^{15} N_h S_{hx}^2 - \sum_{h=1}^{15} \sum_{i=1}^{N_h} M_{hi} S'^2_{hix} \Big\} \Big]$$

$$\therefore \lambda^2 = \frac{\text{Num.}}{\text{Denom.}}$$

$$= \frac{CN \Big[W \sum_{h=1}^{15} \sum_{i=1}^{N_h} (M_{hi} - M_{hiy}) S'^2_{hiy} + (1-W) \sum_{h=1}^{15} \sum_{i=1}^{N_h} (M_{hi} - M_{hix}) S'^2_{hix} \Big]}{(C_{1x}+C_{1y}) M_o \Big[W \Big\{ \sum_{h=1}^{15} N_h S_{hy}^2 - \sum_{h=1}^{15} \sum_{i=1}^{N_h} M_{hi} S'^2_{hiy} \Big\} + (1-W) \sum_{h=1}^{15} N_h S_{hx}^2 - \sum_{h=1}^{15} \sum_{i=1}^{N_h} M_{hi} S'^2_{hix} \Big]}$$

$$\text{Similarly} \quad f^2 = \frac{a_2 K_2}{\beta a_3 K_1} = \frac{a_2 \big[W B_{2y} + (1-W) B_{2x} \big]}{a_3 \big[W B_{1y} + (1-W) B_{1x} \big]}$$

Substituting the values of various constants and solving numerator and denominator separately, we have

$$\text{Num.} = (C_{1x} + C_{1y}) \frac{M_o}{N} \Big[\frac{WN}{M_o^2} \Big\{ \sum_{h=1}^{15} \sum_{i=1}^{N_h} (M_{hi2} S_{hi2y}^2 + M_{hi4} S_{hi4y}^2) \Big\} + (1-W) \frac{N}{M_o^2}$$

$$\Big\{ \sum_{h=1}^{15} \sum_{i=1}^{N_h} (M_{hi3} S_{hi3x}^2 + M_{hix} S_{hi4x}^2) \Big\} \Big]$$

$$\frac{(C_{1x} + C_{1y})}{M_o} \Big[W \sum_{h=1}^{15} \sum_{i=1}^{N_h} M_{hiy} S'^2_{hiy} + (1-W) \sum_{h=1}^{15} \sum_{i=1}^{N_h} M_{hix} S'^2_{hix} \Big]$$

Den. $\dfrac{(C_{2y} M_y + C_{2x} M_x)}{N} \Big[W . \dfrac{N}{M_o^2} \Big\{ \sum_{h=1}^{15} \sum_{i=1}^{N_h} M_{hi} S_{hiy}^2 - \sum_{h=1}^{15} \sum_{i=1}^{N_h} (M_{hi2} S_{hi2y}^2 +$

$M_{hi4} S_{hi4y}^2) \Big\}$

$+ (1-W) \dfrac{N}{M_o^2} \Big\{ \sum_{h=1}^{15} \sum_{i=1}^{N_h} M_{hi} S_{hix}^2 - \sum_{h=1}^{15} \sum_{i=1}^{N_h} (M_{hi3} S_{hi3x}^2 + M_{hi4} S_{hi4x}^2) \Big\} \Big]$

$$= \frac{(C_{2y} M_y + C_{2x} M_x)}{M_o^2} \Big[W \Big\{ \sum_{h=1}^{15} \sum_{i=1}^{N_h} M_{hi} S_{hiy}^{'2} - \sum_{h=1}^{15} \sum_{i=1}^{N_h} S_{hiy}^{'2} (M_{hi2} + M_{hi4}) \Big\}$$

$$+ (1-W) \Big\{ \sum_{h=1}^{15} \sum_{i=1}^{N_h} M_{hi} S_{hix}^{'2} - \sum_{h=1}^{15} \sum_{i=1}^{N_h} S_{hix}^{'2} (M_{hi3} + M_{hi4}) \Big\} \Big]$$

$$= \frac{(C_{2y} M_y + C_{2x} M_x)}{M_o^2} \Big[\bar{W} \sum_{h=1}^{15} \sum_{i=1}^{N_h} (M_{hi} + M_{hiy})' S_{hiy}^2 + (1-W) \sum_{h=1}^{15} \sum_{i=1}^{N_h} (M_{hi} + M_{hix}) S_{hix}^{'2} \Big]$$

$$\therefore \; f^2 \; \frac{(C_{1x} + C_{1y}) M_o \Big[\bar{W} \sum_{h=1}^{15} \sum_{i=1}^{N_h} M_{hiy} S_{hiy}^{'2} + (1-W) \sum_{h=1}^{15} \sum_{i=1}^{N_h} M_{hix} S_{hix}^{'2} \Big]}{(C_{2y} M_y + C_{2x} M_x) \Big[\bar{W} \sum_{h=1}^{15} \sum_{i=1}^{N_h} (M_{hi} + M_{hiy}) S_{hiy}^{'2} + (1-W) \sum_{h=1}^{15} \sum_{i=1}^{N_h} (M_{hi} + M_{hix}) S_{hix}^{'2} \Big]}$$

Having known the values of λ and f, n can be found from the following equation

$$n = (C - a_o) / (a_1 + a_2 \lambda + a_3 \lambda f).$$

CHAPTER VI

SUMMARY

In sample survey practice, non-response from some of the sampling units is of common occurence. This problem of 'non-response' or 'incomplete samples' is more common in Mailing Enquires. If the non-response is considerable the estimates obtained on the basis of the realised sample may not represent the characterstics of the entire population. The only way to minimise or eliminate bias is to have 'call-backs' on then non-responding units, once or more than once, if need arises.

Various attempts have been made to develop suitable plan of enquiry, which will take both bias and the cost of the survey into consideration, whenever a problem of 'non-response' arises. Most of the work that has been done in this direction deals with single stage sampling and one character under study. The present thesis deals with the problem of non-response in two stage sampling and with two characters under study. The procedure consists in classifying the first sample of PSU's into 15 strata on the basis of information supplied by SSU's. The final sample of SSU's is again classified into four classes according to the information supplied by them.

An estimate for the population mean has been built up. The expressions for variance and estimate of variance has been derieved. A combined cost function has been developed, to determine the optimum sizes of samples drawn from the different classes.

REFERENCES

1. Hansen M.H. and Hurwitz W.N. (1946) — "The problem of non response in sample surveys". JASA, Vol. 41, p. 517-529.

2. Clausen J.A. and Ford N. (1947) — "Bias in mail questionnaires" JASA, Vol. 42, p 497-511.

3. Politz A.N. and Simmon W.R. (1949) — "An attempt to get the 'not at homes' in the sample without call-backs". JASA, Vol. 44, p 9-31.

4. Birnbaum Z.W. and Sirken M.G. (1950) — "Bias due to non-availability in Sampling Surveys". JASA, Vol. 45, p 98-111.

5. Deming W.E. (1953) — "On a probability mechanism to attain an economic balance between the resultant error of response and bias of non-response". JASA, Vol. 48, p 743-772.

6. Durbin (1954 a) — "Non response and call-backs in surveys". Bull. Int. Stat. Inst., Vol. 34, p 72-89.

7. EL - Badry M.A. (1956) — "A sampling procedure for mail questionnaire". JASA, Vol. 51, p 209-227.

8. Kish L. and Hess I. (1958) — "On non coverage of sample dwellings." JASA, Vol. 53, p 509-524.

9. Bedawi M.A. and Seth G.R. (1961) — "Non response in sample surveys". Unpublished thesis submitted to the Institute of Agricultural Research Statistics, New Delhi, for the award of Post Graduate Diploma.

10. Zarkovich S.S. — "Sampling methods and Censuses". Vol. II, Draft, FAO publication.

11. Sukhatme P.V. and Sukhatme B.V. (1970) — "Sampling theory of surveys with applications".

12. Cochran W.G. (Vol. II) — Sampling Techniques.