4.1 Domain studies with incomplete frame arising due to qualitative change of sampling units in finite population:

In many situations incompleteness in the frame may also arise due to qualitative change of population units during passage of time. Thus the frame prepared at some prior time may become imperfect because units of the frame may undergo some qualitative change during passage of time. It is possible to discover qualitative change in the sampling units only at the time of enumeration. Therefore frame for which it was originally designed is at variance with the available frame at the time of survey. This discrepancy between sampled population and target population arising due to qualitative change of sampling units at the time of survey will certainly contribute to error and will not provide unbiased estimates for target population parameter.

The units of sampled population which have undergone qualitative change (although they may be uniquely indentifiable, exhaustive and non-overlapping) are in fact non-existent in the target population. Such units may also be treated as non-existent in the target population at the time of actual survey. For example, in a frame of irrigated land designed at some previous time may be imperfect at the time of actual survey because some land listed as irrigated land may be discovered as unirrigated at the time of enumeration i.e. when investigator visits particular land after sample is selected from sampled population. The sample from old frame of irrigated land in which same land listed as irrigated have been found as unirrigated land due to qualitative change of sampling units (here irrigated land) will not be representative of target population, if our objective were to estimate population characteristics of irrigated land. Thus estimates of target population characteristics obtained from such sample will certainly be biased.

But units of sampled population which have undergone some qualitative change in the sense that they no more belong to the target population, may be treated as non-existent in the target population.
However, such units which are non-existent in the desired target population need not mean that they are missing from the frame as explained in chapter-2. The non-existent units in the target population are those units of the sampled population which are discovered at the time of enumeration that they are, because of their qualitative change, no longer defined belonging to the target population. But such units are not physically missing from the target population so far as an investigator finds their physical presence at the time of enumeration. It is at the time of enumeration that an investigator discovers that some of the physically present sampling units do not contribute to the observation and information required for target population as defined due to qualitative change of such units. Observations obtained from such qualitatively changed sampling units due to dynamic nature of units will not correspond to the characteristics of the desired target population i.e. domain of study, hence such units may be treated as non-existent in the target population.

For example in a survey to determine some characteristics in a particular income group as listed in old frame at some previous time, the some of the individuals are discovered to be not belonging to the particular income group because of the change of the income of some of the individuals during passage of time i.e. at the time of actual survey. However the individuals are still present in the frame so far as their physical presence is concerned and hence they are not missing from the frame. But the change of income of individuals have rendered qualitative change in sampling units.

Therefore such units which are, although non-missing but are non-existent in the target population. A non-existent unit in the target population is not necessarily missing unit from target population. But a missing unit from target population is necessarily non-existent in the target population. In a study of average income of agricultural labourers, some agricultural laboures might be discovered to be belonging to non-agricultural labourers group due to qualitative
change of their occupation at the time of actual survey, but they are
still physically present and not missing from the frame but are
non-existent in the target population.

The case of imperfection in the frame can be elucidated by
following figures.

```
x x x x x x x
x x x x x x x
x x x x x x x
x x x x x x x

old frame at the time of
preparation of list of
units.
```

```
x (x) x x x x (x)
x x (x) x (x) x x
x (x) x x (x) x x
x (x) x x (x) x (x)

Sampled population at the
time of actual survey.
```

```
x x x x x
x x x x x
x x x x x
x x x x x

Target population
```

X: Sampling units belonging to sampled and target population both and
which have not gone qualitative change at the time of actual survey.
(X): Units which have gone qualitative change at the time of actual survey due to their dynamic nature and hence non-existent in the target population.

The units discovered as undergone some qualitative change which are treated as non-existent in the target population will not contribute to the desired target population characteristics and parameter the unbiased estimates in such cases can be derived as explained in chapter II.

4.2 Method of Estimation and Sampling procedure

Consider a finite sampled population of size N as listed in the frame for selecting the sample. However due to dynamic nature of some of the units, some units have have gone some qualitative change in the sense that some of the units of the original frame are not treated as existing in the target population. Let $N_2$ units of the original frame have gone qualitative change so that they are treated as non-existent for target population and $N_1$ units of the original frame actually exist in the target population so that $N = N_1 + N_2$. Therefore new up-to-date frame consists of $N_1$ units at the time of actual survey.

One procedure to select the sample is to select random sample from N units keeping old numbering as such. A random number is selected from 1 to N. The unit corresponding to this number is selected provided it is not out of $N_2$ units which have become non-existent for target population. In case, a non-existent unit is chosen, draw is rejected and procedure is repeated. It may be seen that this procedure gives equal probability of selection to the $N-N_2$ units of the target population.

But this procedure assumes that information is available about units undergoing qualitative change from the old frame at the time of sample selection. However most often we can not know about the dynamic units which have undergone qualitative change unless the actual
enumeration starts and it is only when enumerator visits a particular unit that he finds that unit have gone under qualitative change and hence be treated as not existing in the target population.

Therefore in this case also an alternative method of selecting the sample and estimation procedure is proposed.

Select a sample of size n with S.R.S.W.O.R. from old frame of n units from sampled population. As only \( N_1 \) units from the old frame are unchanged units belonging to the target population and \( N_2 \) units are changed units qualitatively in the sampled population, therefore sample selected will also contain some units which have changed their quality and hence donot exist in the target population after each unit is visited by enumerator in the sample of size \( N_1 \).

In the sample of size n, let \( n_1 \) units belong to the target population and \( n_2 \) are non-existent in the target population. However \( n_2 \) units are not missing units. Therefore observations can be obtained for \( n_2 \) units but such observations do not correspond to the characters of target population (i.e. domain of study). Hence observations for such \( n_2 \) units, though available, will not contribute for parameters of target population. Only observations of \( n_1 \) units correspond to the observations for target population.

With notations having the same meaning as given in chapter II.

\( P(n_1) \) follows a Hypergeometric distribution. Therefore an unbiased estimator for target population total is given by

\[
\hat{Y}_{N'} = \left( \frac{n_1}{n} \right) \hat{Y}_{n_1}
\]

and

\[
V(\hat{Y}_{N'}) = \frac{N(N-n)}{n} \left\{ p \hat{S}_1^2 + p(1-p) \hat{\gamma}_1^2 \right\}
\]

Estimate of variance is given by

63
\[ \hat{V}(\hat{Y}_n) = \frac{N-n}{n} \left\{ \left( \sum p_n (n-1+p_n) + n_1 (1-p_n) \right) s_1^2 \right. \\
+ \left. N n_1 p_n (1-p_n) \right\} n_1 \]

Similarly if \( N \) is very large then \( P(n_1) \) will follow binomial distribution and in such care also \( \hat{Y}_n = \left( \frac{n}{N} \right) \hat{y}_n \) and

\[ V(\hat{Y}_n) = \frac{N^2}{n} \left\{ (N-n+1)p_n-1 \right\} + \frac{N^2}{n} \left( 1-p_n \right) \hat{y}_n \]

is given by

\[ \hat{V}(\hat{Y}_n) = \left[ \frac{N}{n} \left\{ (N-n+1)p_n-1 \right\} + \frac{N(1-p_n)(N-p_n-n_1)}{(n-1)n_1} \right] s_1^2 \]

\[ + \frac{N^2 p_n (1-p_n)}{(n-1)n_1} \hat{y}_n \]

as derived in chapter 3.

But objective of the present chapter is the domain study with imperfect frame arising due to qualitative change of sampling units in finite population.

4.3 Incomplete frame arising due to dynamic nature of sampling units and its use for domain studies

Introduction

In many situations we not only need estimates of population parameter but also want separate estimates for different segment of the population. Also quite often we may be interested to study both the quantitative as well as qualitative characteristics of the population. For example, in a population survey, it may be required to estimate
both the proportions of families in different income group and also the average income in each group. Similarly in an agricultural survey, we may be interested to find the proportion of area under various crops during a reason as well as total production.

Some times we may be interested in only a particular class of the population e.g. the average income of land less labourers or average yield of only irrigated area or number of persons employed in a particular sector. In such situations it is desirable to have the frame for particular class and use the same for selecting the sample. But it is well known that there is frequent change in qualitative characteristics of units i.e. units may change from one class to other rapidly. Also there may be changes in due to coming up of some new units or dying out of some old units within each class. For example unirrigated area may become irrigated due to new canals or tubewells. Similarly land less labourers may shift from one place to other. Persons using chemical fertilizer may stop it after some time etc.

When the frame is available for the entire population but not for different groups seperately, the selection and estimation procedure as given in chapter II may easily be adopted. But in such situations when domain of interest may be only one class, the selection of sample from the entire population may be very expensive and undesirable. In such situations, it may be appropriate to have a frame of disjoint classes when ever possible.

Further any frame constructed for different classes may very soon become out-dated as units may change from one class to the other more rapidly so that some of the units listed in one class may belong to the other and vice-versa. In this chapter we consider the use of such a frame of disjoint classes to estimate the proportion of units belonging to one or more classes and also to estimate the population mean of the character under study in different classes. Again considering a suitable cost function, the optimum size of the samples to be selected and also optimum variance for a fixed cost is to be determined.
4.4 Method of Estimation

Consider a finite population of size $N$ and let $N_1$ units of the population belong to the first class and $N-N_1=N_2$ belong to other class or remaining class. Assume that seprate frame of $N_1$ and $N_2$ are available. However, due to dynamic nature of the units there is qualitative change of units from one class to other such that the target population consiste of $N'_1$ units belonging to class I and $N'_2$ units belonging to the other. Further it is assumed that $N_1 + N_2 = N'_1 + N'_2$, i.e. total number of units in the population have not changed. Thus the frame of $N_1$ units may contain some units which do not actually belong to class I and also some of the units belonging to class I might not be listed in the frame of that class.

Let $N_{11}$ denote the number of units of the target population which actually belong to the class I out of $N_1$ units belonging to the class I in the old frame, and let $N_{12}$ be the number of units which have changed from class I to II so that

$$N_1 = N_{11} + N_{12}$$

we may also write

$$N'_1 = N_{11} + N_{12} : \text{ Units actually belonging to class I and}$$

$$N'_2 = N_{22} + N_{12} : \text{ Units actually belonging to class II}$$

Further we define

$$p_{11} = \frac{N_{11}}{N_1} : \text{ The proportion of units of target population belonging to class I out of } N_1$$

units belonging to class I in the old frame.

Similarly we may have

$$p_{22} = \frac{N_{22}}{N_2}, p_{12} = \frac{N_{12}}{N_1}, p_{21} = \frac{N_{21}}{N_2}$$
Since we have only $N_1$ and $N_2$ units available from the old frame, we have to estimate

$N_1'$ i.e. $N_{11}$ and $N_{21}$.

Further let $y$ denote the characteristics under study.

We define

$$\bar{y}_{11} = \frac{1}{N_{11}} \sum_{i=1}^{N_{11}} y_i, \quad \text{Population mean of } y \text{ for } N_{11} \text{ units in class I.}$$

$$\bar{y}_{21} = \frac{1}{N_{21}} \sum_{i=1}^{N_{21}} y_i, \quad \text{Population mean of } y \text{ for } N_{21} \text{ units which have changed from class II to class I.}$$

We may write

$$\bar{y}_{N_1'} = N_{11} \bar{y}_{11} + N_{21} \bar{y}_{21} \quad (4.4.1)$$

4.5 Sampling Procedure

Now select a sample of size $n_1$ from the available frame of $N_1$ by S.R.S.W.O.R. But as some units from the population changed from class I to class II, therefore some of the units selected in the sample may not actually belong to this class. Let $N_{11}$ denote the number of units which belong to class I in target population and $N_{12}$ are those units in the sample which were once in class I but now are in class II.

Now it is well known that unbiased estimator of $N_{11}$ and $N_{12}$ are given by
\[ \hat{N}_{11} = \frac{n_{11}}{n_1} N_1 \quad [n_{11} + n_{12} = n_1] \]

\[ \hat{N}_{12} = \frac{n_{12}}{n_1} N_1 \] (4.5.1)

In order to estimate \( N_{21} \) we have to select a sample from \( N_2 \). Let a sample of size \( N_2 \) be selected with S.R.S. without replacement from class II i.e. from \( N_2 \) units. Let \( N_{21} \) denote the number of units which changed from class II to I i.e. are found belonging to class I and \( N_{22} \) are units still actually in class II so that \( n_2 = N_{21} + N_{22} \).

The unbiased estimators of \( N_{21} \) and \( N_{22} \) are given by

\[ \hat{N}_{21} = \frac{n_{21}}{n_2} N_2 \quad , \quad n_2 = n_{21} + n_{22} \]

\[ \hat{N}_{22} = \frac{n_{22}}{n_2} N_2 \]

Thus an unbiased estimator of \( N'_1 \) is easily given by

\[ \hat{N}'_1 = \hat{N}_{11} + \hat{N}_{21} \] (4.5.2)

Now an estimate of \( Y_{N'}_1 \) can be given by

\[ \hat{Y}_{N'_1} = \frac{n_{11}}{n_1} \tilde{Y}_{11} + \frac{n_{21}}{n_2} \tilde{Y}_{21} \]

\[ = \left( \frac{n_{11}}{n_1} N_1 \right) \tilde{Y}_{11} + \left( \frac{n_{21}}{n_2} N_2 \right) \tilde{Y}_{21} \] (4.5.3)

Where \( \tilde{Y}_{11} \) is an unbiased estimator of \( \bar{Y}_{11} \) given by

\[ \tilde{y}_{11} = \frac{1}{n_{11}} \sum_{i=1}^{n_{11}} Y_{1i} \quad \text{and} \quad \tilde{y}_{21} \text{ is an unbiased estimator of } \bar{Y}_{21} \text{ given by} \]
\[ \bar{y}_{21} = \frac{1}{n_{21}} \sum_{i=1}^{n_{21}} y_i. \]

Here, sampling units of size \( n_1 \) in a population are divided into two mutually exclusive classes, class I and Class II consisting of units belonging to target population and not belonging to target population respectively. Now, the probability \( P(n_{11}) \), that in a sample of \( n_1 \) selected out of \( N_1 \) by method of S.R.R.W.O.R., \( n_{11} \) will over in class I and \( N_{12} \) in class II is given by

\[
P(n_{11}) = \frac{\binom{N_1}{n_{11}} \binom{N_{12}}{n_{21}}}{\binom{N_1}{n_1}},
\]

which is a Hypergeometric distribution.

Similarly \( P(n_{22}) \) can be given by Hypergeometric distribution.

Therefore \( E(n_{11}) = n_1 \cdot p_{11} \) and \( E(n_{21}) = n_2 \cdot p_{21} \)

Hence expected value of \( \hat{y}_{N_1'} \) is given by

\[
E(\hat{y}_{N_1'}) = E \left\{ E \left( \frac{n_{11}}{n_1} \bar{y}_{11} \mid n_{11} \right) \right\} + E \left\{ E \left( \frac{n_{21}}{n_2} \bar{y}_{21} \mid n_{21} \right) \right\}
\]

\[
= E \left\{ \frac{N_1}{n_1} \cdot (n_{11}) \bar{y}_{11} \right\} + E \left\{ \frac{N_2}{n_2} \cdot (n_{21}) \bar{y}_{21} \right\}
\]

\[
= n_{11} \cdot p_{11} \bar{y}_{11} + n_{21} \cdot p_{21} \bar{y}_{21}
\]

Therefore \( \hat{y}_{N_1'} \) is unbiased estimator of the population total \( Y_{N_1'} \).
4.6 Variance of $\hat{Y}_{N_1}$

Now the variance of $\hat{Y}_{N_1}$ is given by

$$V(\hat{Y}_{N_1}) = V\left(\frac{n_{11}}{N_1} \bar{Y}_{11}\right) + V\left(\frac{n_{21}}{N_2} \bar{Y}_{21}\right)$$

Since $\bar{Y}_{11}$, $\bar{Y}_{21}$, $n_{11}$ and $n_2$ are independent because samples are independently selected, therefore we may write

$$V(\hat{Y}_{N_1}) = \frac{n_{11}^2}{n_1^2} V(n_{11} \bar{Y}_{11}) + \frac{n_{21}^2}{n_2^2} V(n_{21} \bar{Y}_{21})$$

(4.6.1)

Now

$$V(n_{11} \bar{Y}_{11}) = E\left\{ V\left(n_{11} \bar{Y}_{11} \mid n_{11}\right) \right\} + V\left\{ E\left(n_{11} \bar{Y}_{11} \mid n_{11}\right) \right\}$$

$$= E\left\{ \frac{n_{11}^2}{n_{11}} \left( \frac{N_{11}}{n_{11}} - \bar{Y}_{11} \right) S_{11}^2 \right\} + V\left(n_{11} \bar{Y}_{11}^2\right)$$

Where

$$S_{11}^2 = \frac{1}{N_{11} - 1} \sum_{i=1}^{N_{11}} (Y_i - \bar{Y}_{11})^2$$

Again

$$E\left\{ \frac{n_{11}^2}{n_{11}} \left( \frac{N_{11}}{n_{11}} - \bar{Y}_{11} \right) S_{11}^2 \right\} = E\left[ \frac{n_{11} N_{11}}{N_{11}} - \frac{n_{11}^2}{N_{11}} S_{11}^2 \right]$$

$$= \frac{N_1 (N_1 - n_1) (N_{11} P_{11} - 1) S_{11}^2}{n_1 (N_1 - 1)}$$

Since

$$E(n_{11}^2) = E\left\{ n_{11} (n_{11} - 1) + n_{11}\right\}$$
\[ = \sum n_{11} (n_{11} - 1)P(n_{11}) + \sum n_{11}P(n_{11}) \]

Now as we can see that

\[ E(n_{11}) = \sum n_{11}P(n_{11}) \]

\[ = \sum \frac{N_{11} p_{11} (N_{11} p_{11} - 1)}{(n_{11} - 1)! (N_{11} p_{11} - n_{11})!} \times \frac{N_{11} q_{11}!}{n_{12}! (N_{11} q_{11} - n_{12})!} \times \frac{(n_{11} - 1) (N_{11} - n_{11})!}{N_{11} (N_{11} - 1)!} \]

\[ = n_{11} p_{11} \sum \left( \begin{array}{c} N_{11} p_{11} - 1 \\ n_{11} - 1 \end{array} \right) \left( \begin{array}{c} N_{11} q_{11} - 1 \\ n_{12} - 1 \end{array} \right) \left( \begin{array}{c} N_{11} - 1 \\ n_{11} - 1 \end{array} \right) = n_{11} p_{11} \]

Again similarly

\[ E\{n_{11} (n_{11} - 1)\} = \sum n_{11} (n_{11} - 1)P(n_{11}) \]

\[ = \frac{n_{1} (n_{1} - 1) N_{11} p_{11} (N_{11} p_{11} - 1)}{N_{11} (N_{11} - 1)} \]

Therefore

\[ E\{n_{11}^2\} = \frac{n_{1} (n_{1} - 1) N_{11} p_{11} (N_{11} p_{11} - 1)}{N_{11} (N_{11} - 1)} + n_{11} p_{11} \] \tag{4.6.2}

Also we have

\[ V(n_{11}) = \bar{Y}_{11}^2 - V(n_{11}) \]

\[ = \bar{Y}_{11}^2 \{E(n_{11}^2) - \{E(n_{11})\}^2\} \]

71
\[
\hat{\bar{\gamma}}^2_{11} = \frac{N_1 (n_1 - 1)}{n_1 (N_1 - 1)} N_{p_{11}} (N_{p_{11}} - 1) + n_1 p_{11}^2 - n_1^2 p_{11}^2
\]

\[
= \bar{\gamma}^2_{11} \left( \frac{N_1}{N_1 - 1} \right) n_1 p_{11} q_{11}
\]

(4.6.3)

From (4.6.2) and (4.6.3) we obtain after simplification

\[
\frac{N_1^2}{n_1} \bar{V} \left( n_{11} \bar{Y}_{11} \right) = \frac{N_1 (N_1 - n_1)}{n_1 (N_1 - 1)} (N_{p_{11}} - 1) S_{11}^2
\]

\[
+ \frac{N_1^2 (N_1 - n_1) p_{11} (1 - p_{11}) \bar{Y}_{11}^2}{n_1 (N_1 - 1)}
\]

(4.6.4)

Similarly

\[
\frac{N_2^2}{n_2} \bar{V} \left( n_{21} \bar{Y}_{21} \right) = \frac{N_2 (N_2 - n_2)}{n_2 (N_2 - 1)} (N_{p_{21}} - 1) S_{21}^2
\]

\[
+ \frac{N_2^2 (N_2 - n_2) p_{21} (1 - p_{21}) \bar{Y}_{21}^2}{n_2 (N_2 - 1)}
\]

(4.6.5)

From (4.6.3) and (4.6.4) we obtain

\[
\bar{V} \left( \bar{Y}_{N_1'} \right) = \frac{N_1 (N_1 - n_1)}{n_1 (N_1 - 1)} (N_{p_{11}} - 1) S_{11}^2 + \frac{N_1^2 (N_1 - n_1) p_{11} (1 - p_{11}) \bar{Y}_{11}^2}{n_1 (N_1 - 1)}
\]

\[
+ \frac{N_2 (N_2 - n_2)}{n_2 (N_2 - 1)} (N_{p_{21}} - 1) S_{21}^2 + \frac{N_2^2 (N_2 - n_2) p_{21} (1 - p_{21}) \bar{Y}_{21}^2}{n_2 (N_2 - 1)}
\]

(4.6.6)
Assuming \( \frac{N_1}{N_1-1} \approx 1 \) and \( \frac{N_2}{N_2-1} \approx 1 \)

We get

\[
V(\hat{y}_{N_1'}) = \frac{(N_1 - n_1)(N_1 P_{11} - 1)}{n_1} S_{11}^2 + \frac{N_1 (N_1 - n_1) P_{11} (1-P_{11}) \bar{y}_{11}^2}{n_1}
\]

\[
+ \frac{(N_2 - n_2)(N_2 P_{21} - 1)}{n_2} S_{21}^2 + \frac{N_2 (N_2 - n_2) P_{21} (1-P_{21}) \bar{y}_{21}^2}{n_2}
\]

or

\[
V(\hat{y}_{N_1'}) = \left( \frac{N_1 - n_1}{n_1} \right) \left\{ \frac{(N_1 P_{11} - 1) S_{11}^2}{n_1 (N_1 - 1)} + \frac{N_1 P_{11} (1-P_{11}) \bar{y}_{11}^2}{n_1 (N_1 - 1)} \right\}
\]

\[
+ \left( \frac{N_2 - n_2}{n_2} \right) \left\{ \frac{(N_2 P_{21} - 1) S_{21}^2}{n_2 (N_2 - 1)} + \frac{N_2 P_{21} (1-P_{21}) \bar{y}_{21}^2}{n_2 (N_2 - 1)} \right\}
\]

(4.6.7)

The notations having the same meaning as given earlier in this chapter, we can also have

\[
V(\hat{y}_{N_2'}) = \frac{N_2 (N_2 - n_2)}{n_2 (N_2 - 1)} \left( \frac{(N_2 P_{22} - 1) S_{22}^2}{n_2 (N_2 - 1)} + \frac{N_2^2 (N_2 - n_2) P_{22} (1-P_{22}) \bar{y}_{22}^2}{n_2 (N_2 - 1)} \right)
\]

\[
+ \frac{N_1 (N_1 - n_1)}{n_1 (N_1 - 1)} \left( \frac{(N_1 P_{12} - 1) S_{12}^2}{n_1 (N_1 - 1)} + \frac{N_1^2 (N_1 - n_1) P_{12} (1-P_{12}) \bar{y}_{12}^2}{n_1 (N_1 - 1)} \right)
\]

(4.6.8)

Assuming \( \frac{N_2}{N_2-1} \approx 1 \) and \( \frac{N_1}{N_1-1} \approx 1 \) we get

\[
V(\hat{y}_{N_2'}) = \frac{N_2 - n_2}{n_2} \left\{ \frac{(N_2 P_{22} - 1) S_{22}^2}{n_2 (N_2 - 1)} + \frac{N_2 P_{22} (1-P_{22}) \bar{y}_{22}^2}{n_2 (N_2 - 1)} \right\}
\]

\[
+ \frac{N_1 - n_1}{n_1} \left\{ \frac{(N_1 P_{12} - 1) S_{12}^2}{n_1 (N_1 - 1)} + \frac{N_1 P_{12} (1-P_{12}) \bar{y}_{12}^2}{n_1 (N_1 - 1)} \right\}
\]

(4.6.9)
This can be extended for other domain studies. It can be seen that variance of the estimate in case of domain studies decreases with increase of sample size \( n_1 \) and \( n_2 \).

### 4.7 Estimation of \( V(\hat{Y}_{N_1}) \)

In order to estimate the \( V(\hat{Y}_{N_1}) \), we have to determine estimate of each of the right handed side terms so that we may write

\[
V(\hat{Y}_{N_1}) = \left( \frac{N_1 - n_1}{n_1} \right) \left[ \text{Est } (N_1 P_{11} - 1) S_{11}^2 \right]
+ N_1 \text{Est } \{P_{11} (1 - P_{11}) \bar{Y}_{11}^2 \}
+ \left( \frac{N_2 - n_2}{n_2} \right) \left[ \text{Est } (N_2 P_{21} - 1) S_{21}^2 \right]
+ N_2 \text{Est } \{P_{21} (1 - P_{21}) \bar{Y}_{21}^2 \}
\]

(4.7.1)

Now \( \text{Est } (N_1 P_{11} - 1) S_{11}^2 \) = \( \text{Est } (N_1 P_{11} - 1) \text{ Est } S_{11}^2 \)

\[
= (N_1 P_{11} - 1) s_{11}^2
= N_1 \frac{n_{11}}{n_1} s_{11}^2 - s_{11}^2
\]

as \( E \left( \frac{n_{11}}{n_1} \right) = P_{11} \) and \( E \left( s_{11}^2 \right) = s_{11}^2 \)

Where

\[
s_{11}^2 = \frac{1}{n_{11} - 1} \sum_{i=1}^{n_{11}} (y_{i1} - \bar{Y}_{11})^2 \quad \text{and} \quad \hat{P}_{11} = \frac{n_{11}}{n_1}
\]
Since
\[
E \left\{ \left( \frac{N_1 n_{11}}{n} - 1 \right) s_{11}^2 \right\} = E \left[ \frac{N_1 n_{11}}{n_1} s_{11}^2 \right] - E (s_{11}^2)
\]
\[
= E \left( N_1 \ E \left( \frac{n_{11}}{n_1} s_{11}^2 \right) \bigg| n_{11} \right) - S_{11}^2
\]
\[
= N_1 E \left( \frac{n_{11}}{n_1} S_{11}^2 \right) - S_{11}^2
\]
\[
= N_1 S_{11}^2 \ E \left( \frac{n_{11}}{n_1} \right) - S_{11}^2
\]
\[
= N_1 S_{11}^2 \ P_{11} - S_{11}^2 = (N_1 P_{11} - 1) S_{11}^2
\]

Also to find \( \text{Est} \{ P_{11} (1-P_{11}) \ Y_{11}^2 \} \) we have

Now \( \text{Est} \ P_{11} (1-P_{11}) = \frac{n_1 (N_1 - 1)}{(n_1 - 1) N_1} \hat{P}_{11} (1-\hat{P}_{11}) \)

Because
\[
E \left\{ \frac{n_1}{n_1 - 1} - \hat{P}_{11} (1-\hat{P}_{11}) \right\} = \frac{N_1}{N_1 - 1} P_{11} (1-P_{11})
\]
in case of hypergeometric distribution.

Here \( \frac{n_1}{n_1 - 1} \hat{P}_{11} (1-\hat{P}_{11}) \) is the variance of the sampling proportion

in case of hypergeometric distribution. As it is clear that \( E(s^2) = S^2 \), therefore we obtain that an unbiased estimate of the product
\[
P_{11} (1-P_{11}) = \frac{n_1 (N_1 - 1)}{(n_1 - 1) N_1} \hat{P}_{11} (1-\hat{P}_{11}).
\]

Again \( \text{Est} \ Y_{11}^2 = \cdot Y_{11}^2 - \text{Est} V(\bar{Y}_{11}) \)

75
Because \( V(\hat{y}_{11}) = E(\hat{y}_{11}^2) - \overline{y}_{11}^2 \)

Thus Est \( \overline{y}_{11}^2 = \overline{y}_{11}^2 - \frac{N_{11} - n_{11}}{N_{11} \cdot n_{11}} s_{11}^2 \)

Similarly

\[
\text{Est } \left( N_2 p_{21} - 1 \right) s_{21}^2 = \left( N_2 \hat{p}_{21} - 1 \right) s_{21}^2,
\]

\[
\text{Est } p_{21} \left( 1 - p_{21} \right) = \frac{n_2 (N_2 - 1)}{(n_2 - 1) N_2 \cdot p_{21} \cdot (1 - p_{21})} \text{ and}
\]

\[
\text{Est } \overline{y}_{21}^2 = \overline{y}_{21}^2 - \frac{N_{21} - n_{21}}{N_{21} \cdot n_{21}} s_{21}^2
\]

Thus estimate of \( V(\hat{Y}_{N_1}^{'}) \) can be written as

\[
\hat{V}(\hat{Y}_{N_1}^{'}) = \frac{N_1 - n_1}{n_1} \left[ \left( N_1 \hat{p}_{11} - 1 \right) s_{11}^2 + \frac{N_1 n_1 (N_1 - 1)}{(n_1 - 1) N_1 \cdot p_{11} \cdot (1 - p_{11})} \left\{ \overline{y}_{11}^2 - \frac{N_{11} - n_{11}}{N_{11} \cdot n_{11}} s_{11}^2 \right\} \right]
\]

\[
+ \frac{N_2 - n_2}{n_2} \left( N_2 \hat{p}_{21} - 1 \right) s_{21}^2 + \frac{n_2 (N_2 - 1)}{(n_2 - 1) N_2 \cdot p_{21} \cdot (1 - p_{21})} \left\{ \overline{y}_{21}^2 - \frac{N_{21} - n_{21}}{N_{21} \cdot n_{21}} s_{21}^2 \right\}
\]

\[
(4.7.2)
\]

or

\[
\hat{V}(\hat{Y}_{N_1}^{'}) = \frac{(N_1 - n_1)(N_1 \hat{p}_{11} - 1)s_{11}^2}{n_1}
\]

\[
+ \frac{(N_1 - n_1)(N_1 - 1)}{(n_1 - 1) \hat{p}_{11} \cdot (1 - \hat{p}_{11})} \left\{ \overline{y}_{11}^2 - \left( \frac{N_{11} - n_{11}}{N_{11} \cdot n_{11}} \right)^2 s_{11}^2 \right\}
\]

\[
76
\]
\[
\frac{(N_2-n_2)(N_2-1)}{n_2} s^2_1 + \frac{(N_2-n_2)(N_2-1)}{n_2} p_{21} (1-p_{21}) \left( \frac{\hat{Y}^2_{21} - N_{21} \bar{Y}^2_{21}}{N_{21} n_{21}^2} \right) \frac{s^2_{21}}{s^2_{21}}
\] (4.7.3)

Thus estimate of variance is function of \(n_1, n_2, N_1, N_2, \hat{p}_{11}, \hat{p}_{21}, s_{11}, s_{21}, \bar{Y}_{11}, \bar{Y}_{21}, N_{11}, N_{21}\). Estimate of variance decreases as \(n_1\) and \(n_2\) of imperfect frame increases.

It is interesting to note that if \(n_1 = n_{11}\) and \(n_2 = n_{21}\) i.e. when there is no imperfection of frame in class I and frame of class II totally converges to class I then \(\hat{p}_{11} = 1\) and \(\hat{p}_{21} = 1\).

Here Again we see that \(s^2_{11} = s^2_1\) and \(s^2_{21} = s^2_2\).

Therefore
\[
\hat{V}(\hat{Y}_{N_1}) = \frac{(N_1-n_1)(N_1-1)}{n_1} s^2_1 + \frac{(N_2-n_2)(N_2-1)}{n_2} s^2_2
\]

Which is case of S.R.S.W.O.R. with no imperfection in frame I but frame II completely belonging to class I due to its dynamic nature. However if no unit of frame of class II is dynamic in nature so that \(n_{21} = 0\) and \(s_{21} = 0\) and \(n_1 = n_{11}\) i.e. no imperfection in frame of class I. Then we have
\[
\hat{V}(\hat{Y}_{N_1}) = \frac{(N_1-n_1)(N_1-1)}{n_1} s^2_1. \text{ [If there is no imperfection in the frame of class I].}
\]

Similarly if \(n_{11} = 0\) i.e. \(\hat{p}_{11} = 0\) and \(s^2_1 = 0\) then also we can have a situation when \(n_2 = n_{21}\) so that \(\hat{p}_{21} = 1\) so that
\[
\hat{V}(\hat{Y}_{N_1}) = \frac{(N_2-n_2)(N_2-1)}{n_2} s^2_2 \text{ if } n_2 \text{ units of frame are found to be belonging to class I in the sample of } n_2. \text{ Similarly } \hat{V}(\hat{Y}_{N_1}) \text{ can be}
\]
found for \( n_{11} = 0 \) or \( n_{21} = 0 \) from (4.7.3).

4.8 estimation of \( \frac{1}{N_{11}} \) and \( \frac{1}{N_{22}} \):

In the case of imperfect frame arising due dynamic nature of sampling units, the size of \( N_{11} \) and \( N_{21} \) are not known due to lack of information prior to sampling regarding target population. We have prior knowledge of \( N_1 \) and \( N_2 \) and in the sample, we have information of \( n_1, n_2, \bar{y}_{11}, s_{11}, \bar{y}_{21}, n_{11} \) and \( n_{21} \) after each unit in the sample of size \( n_1 \) and \( n_2 \) are enumerated from two classes I and II respectively for desired domain studies. It is possible to discover at the enumeration stage and found that \( n_{11} \) units belong to class and \( n_{21} \) units from class II transform it self to class I due to its dynamic nature in quality. In the formula for estimate of variance of target population as given in (4.7.3) and as the sizes of \( N_{11} \) and \( N_{21} \) are not known, most often, it has to be substituted with its estimate from sample of size \( n_1 \) and \( n_2 \).

Thus 
\[
\frac{N_{11} - n_{11}}{N_{11}, n_{11}} = \left( \frac{1}{n_{11}} - \frac{1}{N_{11}} \right) \text{ and its estimate is given by}
\]

\[
\left\{ \frac{1}{n_{11}} - \text{Est} \left( \frac{1}{N_{11}} \right) \right\}
\]

Now assume that
\[
\hat{N}_{11} = N_{11} + \varepsilon \quad \text{Where } E(\varepsilon) = 0
\]

Hence
\[
E \left( \frac{1}{\hat{N}_{11}} \right)^{-1} = E \left[ N_{11}^{-1} \left( 1 + \frac{\varepsilon}{N_{11}} \right)^{-1} \right]
\]

or
\[
E \left( \frac{1}{\hat{N}_{11}} \right)^{-1} = E \left[ N_{11}^{-1} \left( 1 - \frac{\varepsilon}{N_{11}} \right) \right] \text{ neglecting the power of order higher than one.}
\]

\[
E \left[ \frac{1}{N_{11}} - \frac{\varepsilon}{N_{11}^2} \right]
\]
\[ \approx E \left( \frac{1}{N_{11}} \right) = 0 \]

\[ \approx \frac{1}{N_{11}} \]

Therefore \( \text{Est} \left( \frac{1}{N_{11}} \right) = \frac{1}{\hat{N}_{11}} = \frac{1}{N_{11} \hat{p}_{11}} \) as \( \hat{N}_{1} = N_{1} \hat{p}_{11} \)

Hence \( \text{Est} \left( \frac{1}{n_{11}} - \frac{1}{N_{11}} \right) = \frac{N \hat{p}_{11} - n_{1}}{N_{1} \hat{p}_{11} n_{11}} \)

Similarly \( \text{Est} \left( \frac{1}{n_{21}} - \frac{1}{N_{21}} \right) = \frac{N \hat{p}_{21} - n_{21}}{N_{2} \hat{p}_{21} n_{21}} \)

Putting these values in (4.7.3) we obtain

\[ \hat{V}(Y'_{N'}) = \frac{(N_{1} - n_{1})(N_{11} \hat{p}_{11} - 1)}{n_{11}^2} + \frac{(N_{2} - n_{2})(N_{21} \hat{p}_{21} - 1)(1 - \hat{p}_{21})}{n_{21}^2} \frac{\hat{y}_{21}^2}{\hat{p}_{21}} \]

\[ - \frac{(N_{1} - n_{1})(N_{11} \hat{p}_{11} - n_{11})}{(n_{11} - 1) N_{1} \hat{p}_{11} n_{11}} \frac{(1 - \hat{p}_{11}) (N_{11} \hat{p}_{11} - n_{11})}{\hat{p}_{11}^2} \frac{s_{11}^2}{n_{11}} \]

\[ - \frac{(N_{2} - n_{2})(N_{21} \hat{p}_{21} - 1)(1 - \hat{p}_{21}) (N_{21} \hat{p}_{21} - n_{21})}{(n_{21} - 1) N_{2} \hat{p}_{21} n_{21}} \frac{s_{21}^2}{n_{21}} \]

\[ + \frac{(N_{2} - n_{2})(N_{21} \hat{p}_{21} - 1)}{n_{21}^2} + \frac{(N_{2} - n_{2})(N_{21} - 1) \hat{p}_{21}^2 (1 - \hat{p}_{21}) \hat{y}_{21}^2}{n_{21}^2} \]

(4.8.1)
Assuming $\frac{1}{n_1} \equiv \frac{1}{n_2 - 1}$, and $\frac{1}{n_2} \equiv \frac{1}{n_2 - 1}$ and again Assuming

$\frac{N_1 - 1}{N_1} \equiv 1$ and $\frac{N_2 - 1}{N_2} \equiv 1$ we have

$\hat{V}(\hat{Y}_{11}') = \frac{(N_1 - n_1)s_{11}^2}{n_1 n_{11}} \left\{ n_{11} (N_1 \hat{p}_{11} - 1) - (1 - \hat{p}_{11}) (N_1 \hat{p}_{11} - n_{11}) \right\} + \frac{(N_2 - n_2)s_{21}^2}{n_2 n_{21}} \left\{ n_{21} (N_2 \hat{p}_{21} - 1) - (1 - \hat{p}_{21}) (N_2 \hat{p}_{21} - n_{21}) \right\}$

$\left( \frac{N_1 - n_1}{n_1} \right) \left( \frac{N_2 - n_2}{n_2} \right) \left( \frac{N_1 - 1}{N_1} \right) \left( \frac{N_2 - 1}{N_2} \right) \frac{\hat{p}_{11}}{\hat{p}_{11}} \left( 1 - \hat{p}_{11} \right) \frac{\hat{p}_{21}}{\hat{p}_{21}} \left( 1 - \hat{p}_{21} \right) \frac{\hat{Y}_{11}^2}{\hat{Y}_{21}^2}$

(4.8.2)

or

$\hat{V}(\hat{Y}_{11}') = \frac{N_1 - n_1}{n_1 n_{11}} s_{11}^2 \hat{p}_{11} \left\{ n_{11} (N_1 - 1) - N_1 (1 - \hat{p}_{11}) \right\} + \frac{(N_1 - n_1)(N_1 - 1)\hat{p}_{11}}{n_1} \left( 1 - \hat{p}_{11} \right) \frac{\hat{Y}_{11}^2}{\hat{Y}_{11}^2}$

$+ \frac{N_2 - n_2}{n_2 n_{21}} s_{21}^2 \hat{p}_{21} \left\{ n_{21} (N_2 - 1) - N_2 (1 - \hat{p}_{21}) \right\} + \frac{(N_2 - n_2)(N_2 - 1)\hat{p}_{21}}{n_2} \left( 1 - \hat{p}_{21} \right) \frac{\hat{Y}_{21}^2}{\hat{Y}_{21}^2}$

(4.8.3)

It can be seen that estimate of variance of $\hat{Y}_{11}'$ depends on $N_1, N_2, n_1, n_2, s_{11}, s_{21}, n_{11}, n_{21}, \hat{p}_{11}, \hat{p}_{21}$ which can be determined...
on the basis of sample selected from imperfect frame arising due to
dynamic nature of population units. Estimate of variance increases as
\( n_{11} \) and \( n_{21} \) increases.

If \( n_1 = n_{11} \) i.e. \( p_{11} = 1 \) so that there is no imperfection in
the frame of class I, then

\[
\hat{V}(Y'_{N_1}) = \frac{(N_1-n_1)(N_1-1)}{n_1} s_{11}^2 + \frac{n_2 n_2}{n_2 n_{21}} s_{21}^2 \hat{p}_{21} \left\{(n_{21} (N_2-1)
\right. \\
\left. - N_2 (1-\hat{p}_{21}) \right\} + \frac{(N_2-n_2)(N_2-1) \hat{p}_{21} (1-\hat{p}_{21}) \hat{y}_{21}^2}{n_2-1}
\]

However, if \( n_1 = n_{11} \) and \( n_2 = n_{21} \) i.e. case of no imperfection in the
frame of classes I and total conversion of units of class II into class
I due to dynamic qualitative change of units then \( p_{11} = 1 \) and \( p_{21} = 1 \) so
that we have

\[
\hat{V}(Y'_{N_1}) = \frac{(N_1-n_1)(N_1-1)}{n_1} s_{11}^2 + \frac{(N_2-n_2)(N_2-1)}{n_2} s_{21}^2
\]

But if there is no imperfection in the frame so that \( n_{11} = n_1 \)
and \( n_{21} = 0 \) Then

\[
\hat{V}(Y'_{N_1}) = \frac{(N_1-n_1)(N_1-1)}{n_1} s_{11}^2 \text{ Where } s_{11}^2 = s_1^2
\]

But if \( n_{11} = 0 \) and \( n_{21} = n_2 \) then

\[
\hat{V}(Y'_{N_1}) = \frac{(N_2-n_2)(N_2-1)}{n_2} s_{21}^2 \text{ where } s_{21}^2 = s_2^2
\]

Thus different cases of imperfection in the frame can be
considered for domain studies.
4.9 Cost function

In order to obtain the optimum sample size \( n_1 \) and \( n_2 \) to obtain estimator for class I with maximum precision for a fixed given cost, we have to consider the optimum cost function. Th total cost of survey depends upon many factors e.g. overhead cost, travelling cost, enumeration cost etc. In this situation we consider a simple cost function given by

\[
C_0 = C_1 n_1 + C_2 n_2
\]
or

\[
E(C_0) = C_1 n_1 + C_2 n_2
\]

where \( C_1 \) denotes the cost per unit of selection, travel and enumeration for the class I. \( C_2 \) denote cost per unit of selection, travel and enumeration for the frame of class II. To determine \( n_1 \) and \( n_2 \) for a fixed cost \( C_0 \) so that variance of the estimator \( \hat{Y}_{N_1} \) is least, we consider the expression

\[
\phi = \hat{v}(\hat{Y}_{N_1}) + \mu (C_1 n_1 + C_2 n_2 - C_0)
\]

where \( \mu \) is a lagranian multiplier. On differentation \( \phi \) partially with respect to \( n_1 \) and \( n_2 \) and equating the results zero we get.

\[
n_{1 \text{opt}} = N_1 \left[ \frac{(N_1 P_{11} - 1) S^2 + N_1 P_{11} (1 - P_{11}) \bar{y}_1^2}{\mu C_1 (N_1 - 1)} \right]^{1/2}
\]  (4.9.1)

and

\[
n_{2 \text{opt}} = N_2 \left[ \frac{(N_2 P_{21} - 1) S^2 + N_2 P_{21} (1 - P_{21}) \bar{y}_1^2}{\mu C_2 (N_2 - 1)} \right]^{1/2}
\]  (4.9.2)

We can draw following conclusions:-

For fixed cost \( C = C_1 n_1 + C_2 n_2 \) we have optimum value of \( n_1 \) such that

(i) \( n_1 \propto N_1 \) i.e. if population size of \( N_1 \) is large then sample selected should also be large.

(ii) \( n_1 \propto \frac{1}{\sqrt{C_1}} \) i.e. if cost is large then \( n_1 \) should be small.
(iii) If $S_{11}^2$ and $\bar{Y}_{11}$ are large then sample selected should also be large.

Similarly we see that

(i) $n_2 \propto N_2$ i.e. if population size of $N_2$ is large sample selected should also be large.

(ii) $n_2 \propto \frac{1}{\sqrt{C_2}}$ i.e. if cost is large then $n_2$ should be large small.

(iii) If $S_{21}^2$ and $\bar{Y}_{21}$ are large then sample $n_2$ selected should also be large.

Putting the optimum values of $n_1$ and $n_2$ we can determine $C_0$ given by

$$C_0 = C_1 n_{\text{opt}}^1 + C_2 n_{\text{opt}}^2$$

or

$$C_0 = C_1 N_1 \left[ \frac{(N P_{11} - 1) S_{11}^2 + N P_{11} (1-P_{11}) \bar{Y}_{11}^2}{\mu (N^2 - 1) C_1} \right]^{1/2}$$

$$+ C_2 N_2 \left[ \frac{(N P_{21} - 1) S_{21}^2 + N P_{21} (1-P_{21}) \bar{Y}_{21}^2}{\mu (N^2 - 1) C_2} \right]^{1/2}$$

$$= \frac{\sqrt{C_1}}{\sqrt{N_1}} \left[ \frac{N_1}{\sqrt{N_1}} \left( (N P_{11} - 1) S_{11}^2 + N P_{11} (1-P_{11}) \bar{Y}_{11}^2 \right) \right]^{1/2}$$

$$+ \frac{\sqrt{C_2}}{\sqrt{N_2}} \left[ \frac{N_2}{\sqrt{N_2}} \left( (N P_{21} - 1) S_{21}^2 + N P_{21} (1-P_{21}) \bar{Y}_{21}^2 \right) \right]^{1/2}$$

83
From above we have

\[
\sqrt{\mu} = \frac{\sqrt{C_1} N_1}{\sqrt{N_1-1} C_0} \left\{ (N_1 P_{11} - 1) S_{11}^2 + N_1 P_{11} (1-P_{11}) \gamma_{11}^2 \right\}^{\frac{1}{2}} \\
+ \frac{\sqrt{C_2} N_2}{C_0 \sqrt{N_2-1}} \left\{ (N_2 P_{21} - 1) S_{21}^2 + N_2 P_{21} (1-P_{21}) \gamma_{21}^2 \right\}^{\frac{1}{2}} \\
= \frac{\sqrt{C_1}}{C_0} \frac{N_1}{\sqrt{(N_1-1)}} A + \frac{\sqrt{C_2}}{C_0} \frac{N_2}{\sqrt{(N_2-1)}} B 
\]  

(4.9.3)

where

\[ A = \left[ (N_1 P_{11} - 1) S_{11}^2 + N_1 P_{11} (1-P_{11}) \gamma_{11}^2 \right]^{\frac{1}{2}} \]

and

\[ B = \left[ (N_2 P_{21} - 1) S_{21}^2 + N_2 P_{21} (1-P_{21}) \gamma_{21}^2 \right]^{\frac{1}{2}} \]

Putting from (4.9.1) and (4.9.3) we have

\[ \frac{N_1}{n_{\text{opt}}} = \sqrt{\frac{N_1-1}{N_1}} \sqrt{\frac{\mu}{C_1}} \]

\[
= \frac{N_1}{\sqrt{N_1-1} \sqrt{C_1}} \left\{ \frac{\sqrt{C_1} N_1}{C_0 \sqrt{(N_1-1)}} A + \frac{\sqrt{C_2} N_2}{C_0 \sqrt{(N_2-1)}} B \right\} 
\]

\[ = \frac{1}{\sqrt{N_1-1} \left\{ \frac{C_1}{C_0} \frac{N_1}{\sqrt{N_1-1}} + \frac{\sqrt{C_1 C_2}}{C_0 \sqrt{N_2-1}} \right\}} \]

(4.9.4)

\[
= \frac{N_2}{n_{\text{opt}}} = \frac{B}{\sqrt{N_2-1} \mu C_2} 
\]

(4.9.5)
\[
\begin{align*}
= \frac{N_2}{\sqrt{N_2-1}} \frac{B}{C_2} & \left\{ \frac{\sqrt{C_1}}{C_0} \frac{N_1}{\sqrt{N_2-1}} A + \frac{\sqrt{C_2}}{C_0 \sqrt{N_2-1}} B \right\} \\
= \frac{N_2}{\sqrt{N_2-1}} & \left\{ \frac{C_2 N_2}{C_0 \sqrt{N_2-1}} + \frac{\sqrt{C_1} C_2}{C_0} \frac{N_1}{\sqrt{N_1-1}} \right\} (A/B) \\
= \frac{1}{\left\{ \frac{C_2}{C_0} + \frac{\sqrt{C_1} C_2}{C_0} \frac{N_1 \sqrt{(N_1-1)}}{N_2 \sqrt{(N_1-1)}} \frac{A}{B} \right\}} \\
& \text{(4.9.6)}
\end{align*}
\]

Thus (4.9.5) and (4.9.6) give optimum values for the sample sizes \( n_1 \) and \( n_2 \). Putting these optimum values in variance term we can determin optimum value of variance of the estimator \( \hat{Y}_{n_1} \).

Therefore we have

\[
\begin{align*}
V(\hat{Y}_{n_1}) = \frac{N_1 (N-n_1)(N_1 p_{11}^1 -1) S_{11}^2}{N_1 (N_1-1)} + \frac{N_1^2 (N_1 - n_1)}{N_1 (N_1-1)} p_{11} (1-p_{11}) \frac{\bar{Y}_{11}^2}{n_1 (N_1-1)} \\
+ \frac{N_2 (N_2-n_2)}{N_2 (N_2-1)} (N_2 p_{21} -1) S_{21}^2 + \frac{N_2^2 (N_2 - n_2)}{n_2 (N_2-1)} p_{21} (1-p_{21}) \frac{\bar{Y}_{21}^2}{n_2 (N_2-1)} \\
= \frac{N_1^2}{n_1 (N_1-1)} \left[ (N_1 p_{11}^1 -1) S_{11}^2 + N_1 p_{11} (1-p_{11}) \frac{\bar{Y}_{11}^2}{n_1 (N_1-1)} \right] \\
- \left[ \frac{N_1}{(N_1-1)} \right] \left[ N_1 p_{11} S_{11}^2 + N_1 p_{11} (1-p_{11}) \frac{\bar{Y}_{11}^2}{n_1 (N_1-1)} \right]
\end{align*}
\]
\[
\frac{N_2^2}{n_2(N_2 - 1)} \left[ (N_2^2 p_{21} - 1) S_{21}^2 + N_2^2 p_{21} (1-p_{21}) \bar{Y}_{21}^2 \right]
\]

\[
- \frac{N_2}{(N_2 - 1)} \left[ (N_2^2 p_{21} - 1) S_{21}^2 + N_2^2 p_{21} (1-p_{21}) \bar{Y}_{21}^2 \right]
\]

\[
= \frac{N_1^2}{n_1(N_1 - 1)} A^2 - \frac{N_1}{N_1 - 1} A^2 + \frac{N_2^2}{n_2(N_2 - 1)} B^2 - \frac{N_2}{N_2 - 1} B^2
\]

or

\[
V(\hat{Y}_{N_1'}) = \frac{N_1^2}{n_1(N_1 - 1)} A^2 + \frac{N_2^2}{n_2(N_2 - 1)} B^2 - \left[ \frac{N_1 A^2}{N_1 - 1} + \frac{N_2 B^2}{N_2 - 1} \right]
\]

(4.9.8)

Putting the optimum values of \( n_1 \) and \( n_2 \) we obtain

\[
V_{\text{opt}}(\hat{Y}_{N_1'}) = \frac{N_1^2}{N_1 - 1} \left\{ \frac{C_1}{C_0} + \frac{\sqrt{C_{12}}}{C_0} \frac{N_2}{N_1} \left[ \sqrt{\frac{N_1 - 1}{N_2 - 1}} B \right] \right\}
\]

\[
+ \frac{N_2^2}{N_2 - 1} \left\{ \frac{C_2}{C_0} + \frac{\sqrt{C_{12}}}{C_0} \frac{N_1}{N_2} \left[ \sqrt{\frac{N_1 - 1}{N_2 - 1}} A \right] \right\}
\]

\[
- \left[ \frac{N_1}{N_1 - 1} A^2 + \frac{N_2}{N_2 - 1} B^2 \right]
\]

If \( N_1 \) and \( N_2 \) are large such that

\[
\frac{N_1}{N_1 - 1} \approx 1 \quad \text{and} \quad \frac{N_2}{N_2 - 1} \approx 1
\]
Then we have after simplification

\[
V_{\text{opt}}(\hat{Y}_{N_1}) = N_1 A^2 \left\{ \frac{C_1}{C_0} + \frac{\sqrt{C_2}}{C_0} \left[ \frac{N_2}{N_1} - \frac{B}{A} \right] \right\} \\
+ N_2 B^2 \left\{ \frac{C_2}{C_0} + \frac{\sqrt{C_1}}{C_0} \frac{N_1}{N_2} \frac{A}{B} \right\} - (A^2 + B^2)
\]

\[
= \frac{1}{C_0} \left[ A \sqrt{\frac{C_1}{N_1}} + B \sqrt{\frac{C_2}{N_2}} \right]^2 - (A^2 + B^2)
\]  \hspace{1cm} (4.9.9)

Where

\[
A = \left[ (N_{11} - 1) S_{11}^2 + N_1 p_{11}(1-p_{11}) \bar{Y}_{11}^2 \right]^{1/2}
\]

\[
B = \left[ (N_{21} p_{21} - 1) S_{21}^2 + N_2 p_{21}(1-p_{21}) \bar{Y}_{21}^2 \right]^{1/2}
\]

Thus optimum variance depends on \(C_1, C_2, C_0, \bar{Y}_{11}, \bar{Y}_{21}, S_{11}^2, S_{21}^2\) and also on \(N_1\) and \(N_2\). Given these values we can determine optimum variance for given fixed cost \(C_0\). Estimate of optimum value of variance of \(\hat{Y}_{N_1}\) can be determined by estimates of \(A\) and \(B\) which has been derived earlier in this chapter.