CHAPTER - III
SELECTION OF A UNIQUE ESTIMATOR FOR IKEDA-SEN'S SAMPLING PROCEDURE

3.0. Summary.

In this chapter the three unbiased estimators belonging to the three classes of linear estimators formulated by Horvitz and Thompson (1952) for Ikeda-Sen's sampling procedure are presented. Using the Bayesian estimation theory an attempt is made to select one estimator with the help of the criterion of necessary best estimator suggested by Prabhu-Ajgaonkar (1965, 1969). The optimum sampling procedures corresponding to three optimum estimators one in each class under super population concept are determined, which gives rise to three sampling strategies. It is noted that there does not exist a necessary best sampling strategy under the super-population concept for Ikeda-Sen's sampling procedure.

3.1. Introduction.

The technique of sampling with varying probabilities was brought to the light by Hansen and Hurvitz in 1943, by demonstrating the superiority of this technique over that with equal probability. However, Hansen and Hurwitz confined to
selection of only one unit and indicated the method of determining the probabilities of selection which minimized the variance of the sample estimate a fixed cost. It has been pointed out by Midzuno (1950) and Yamamoto (1955) that this paper of Hansen and Hurwitz is a mile stone in the theory of sampling techniques. Horwitz and Thompson 1952 developed the requisite theory when elements are drawn without-replacement using arbitrary probabilities of selection at each draw. An important feature of their contribution lies in the fact that they recognised different classes of linear estimators.

In this Chapter the three unbiased estimators belonging to the three classes of linear estimators formulated by Horvitz and Thompson (1952) for Ikeda-Sen's sampling procedure are presented. Using the Bayesian estimation theory an attempt is made to select one estimator with the help of the criterion of the necessary best estimator suggested by Prabhu-Ajgaonkar (1965). A well known method of utilizing related information, that of incorporating it into the sampling procedure as a basis for determining the probabilities with which various population elements enter into the sample, is the technique of sampling with varying probabilities. Accordingly the optimum sampling procedures corresponding to the three
optimum estimators one in each class under super-population concept are determined, which gives rise to three sampling strategies. It is noted that there does not exist a necessary best sampling strategy under the super population concept for Ikeda-Sen's sampling procedure.

3.2 Preliminaries.

Let us consider the finite population \( U \) of \( N \) identifiable units \( U = (U_1, \ldots, U_i, \ldots, U_N) \), and \( Y \) and \( X \) be two real valued characteristics taking values \( Y_i \) and \( X_i \) on the units \( U_i \) of \( U \) where \( Y \) is the study variable and \( X \) is the ancillary information available, closely related to \( Y \). The object is to estimate the population total

\[
Y = \sum_{i=1}^{N} Y_i \quad (3.2.1)
\]

A finite sequence of units from \( U \) is called a sample and is denoted by

\[
s = (U_1, U_2, \ldots, U_n)
\]

Let

\[
S = \{ s \}
\]

denotes the totality of all samples from \( U \).
A collection $S$ of samples $s$ from $\mathbb{U}$ with a probability measure $P$ defined on it such that corresponding to every $s \in S$ there is a probability $p(s)$ attached to it such that $p(s) \geq 0$ and $\sum_{s \in S} p(s) = 1$, is called a sampling design and is denoted by $D(S,p)$ or simply by $p$.

The conventional problem in survey sampling is to estimate the population total $Y$ by observing the values $Y_i$ for which if $s$, where $s$ is the sample drawn according to a design $p$.

An estimator $t$ is a real valued function defined on $S$ which depends on $\mathbb{U}$ only through those $Y_i$ for which if $s$.

A design $p$ together with an estimator $t$ of population total $Y$ is called a sampling strategy $H(t,p)$.

A sampling strategy $H(t,p)$ is said to be unbiased for the population total $Y$ if $t$ is $p$-unbiased for $Y$. The expectation, variance or mean square error of a strategy are defined as the expectation, variance or mean square error of the corresponding estimator.

We present below the classes of linear estimators formulated by Horvitz and Thompson (1952).

i) The estimators of $T_1$-class are defined symbolically as

$$T_1 = \sum_{i=1}^{n} a_i Y_i$$

(3.2.2)
where \( a_i, (i=1,2,\ldots,n) \) is a constant used as a weight associated with the element that turns-up at the \( i \)th draw; and \( y_i, (i=1,2,\ldots,n) \) is the value of the element for the characteristic at the \( i \)th draw.

ii) \( T_2 \)-class of estimators are represented symbolically as

\[ T_2 = \sum_i \beta_i y_i \quad (3.2.3) \]

where \( \beta_i, (i=1,2,\ldots,N) \) is the constant to be used as a weight for the \( i \)th population element whenever it is selected in the sample.

iii) \( T_3 \)-class of estimators are given by

\[ T_3 = v_{s_n} \left( \sum_{i=1}^{n} y_i \right) s_n \quad (3.2.4) \]

where \( v_{s_n} \) is the constant to be used as a weight whenever the \( s_n \)th sample is drawn.

Ikeda, a student of Midzuno suggested a very simple practical method of implementing the generalized probability function due to Midzuno (1950). Interestingly, Sen (1952) devised independently this sampling procedure in connection with an estimator belonging to the \( T_3 \)-class. The Ikeda-Sen's sampling procedure is defined below:
On the first draw a population element is selected with unequal probabilities and on the second and subsequent draws, elements are selected with equal probabilities without replacement.

A random sample of size \( n \) is drawn from a finite population of size \( N \), in accordance with the Ikeda-Sen's sampling procedure, utilizing the ancillary information \( (X_1, \ldots , X_N) \), where \( X_i \) represent a measure of \( X \)-characteristic, being ancillary information for the \( i \)th population element, \( U_i \).

Let

\[
p_i = \frac{X_i}{X}, i=1,2,\ldots ,N \tag{3.2.5}
\]

where

\[
X = \sum_{i=1}^{N} X_i \tag{3.2.6}
\]

denote the probability set of the Ikeda-Sen's sampling procedure.

A best estimator in Horvitz-Thompson's \( T_1 \)-class of linear estimators, when Ikeda-Sen's sampling procedure is adopted, is derived by Prabhu-Ajgaonkar (1967),

\[
t_1 = y_1 + \frac{N-1}{n-1} \frac{n}{\sum_{i=2}^{N} y_i} \tag{3.2.7}
\]
where \( y_i \) \((i=1,2,\ldots,n)\) denotes the outcome at the \(i\)th draw.

It is noted that

\[
\text{Var}(t_1) = \frac{N}{n-1} \sum_{i=1}^{N} \frac{(N-1) - (N-n)p_i}{2}
\]

\[
+ \frac{N}{n-1} \sum_{i \neq j} (n-2)(N-1) \frac{(N-n)(p_i + p_j)}{(n-1)(N-2)} \frac{Y_i Y_j - \bar{Y}^2}{\pi_i \pi_j}
\]

\[\text{(3.2.8)}\]

Horvitz and Thompson (1952) noted that there is only one unbiased estimator for the \(T_2\)-class, and is given by

\[
t_2 = \frac{n}{\sum_{i} \frac{y_i}{\pi_i}}
\]

\[\text{(3.2.9)}\]

where \(\pi_i\) \((i=1,2,\ldots,N)\) represents the probability of including the \(i\)th population element into the sample. For Ikeda-Sen's sampling procedure

\[
\pi_i = \frac{N-n}{N-1} p_i + \frac{n-1}{N-1}
\]

\[\text{(3.2.10)}\]

for \(i=1,2,\ldots,N\).

It is noted that

\[
\text{Var}(t_2) = \frac{N}{\sum_{i} \frac{y_i}{\pi_i}} + \frac{N}{\sum_{i \neq j} \frac{\pi_i \pi_j}{\pi_i \pi_j}} \frac{Y_i Y_j - \bar{Y}^2}{\pi_i \pi_j}
\]

\[\text{(3.2.11)}\]
where $\pi_{ij}$ (for $i \neq j = 1, 2, \ldots, N$) represents the probability of including the $i$th and $j$th population elements into the sample. For Ikeda-Sen's sampling procedure

$$\pi_{ij} = \frac{n-1}{N-1} \left[ \frac{N-n}{N-2} (p_i + p_j) + \frac{n-2}{N-2} \right], \; i \neq j = 1, 2, \ldots, N \quad (3.2.12)$$

While presenting outline of the theory of sampling system Midzuno (1950) discovered estimator

$$t_3 = \frac{\Sigma y_i}{\Sigma p_i} \quad (3.2.13)$$

where $\Sigma y_i$ is the sample total.

Further he showed that

$$\text{Var}(t_3) = \sum_{i=1}^{N} \frac{y_i^2}{\Sigma p_i} - \frac{1}{(N-1) \Sigma p_i} + \sum_{i \neq j=1}^{N} \frac{y_i y_j}{\Sigma p_i} \sum_{i \neq j=1}^{N} \frac{1}{\Sigma p_i} - \frac{y^2}{(n-1) \Sigma p_i}$$

... \quad (3.2.14)

where $\Sigma$ denotes summation over those samples which include $s \supset i$ the $i$th population element. Similarly $\Sigma$ stands for the $s \supset i, j$ and $j$th summation over those samples which contain the $i$th population elements.
The selection of a unique estimator for a class is based on the following definition of necessary best estimator of order $r$.

**Definition 3.2.1.** Let $\mathcal{C}$ be the class of linear unbiased estimators, the variance of which can be expressed as

$$\text{Var}(\mathcal{C}) = \sum_{i=1}^{N} A_{ii} Y_i^2 + \sum_{i \neq j=1}^{N} A_{ij} Y_i Y_j$$

(3.2.15)

where the quantities $A_{ij}$'s involve the known functions of probabilities and sample and population sizes.

Further let $t'$ be unbiased estimator belonging to the class $\mathcal{C}$ and that the variance of $t'$ be given by

$$\text{Var}(t') = \sum_{i=1}^{N} B_{ii} Y_i^2 + \sum_{i \neq j=1}^{N} B_{ij} Y_i Y_j$$

(3.2.16)

where $B_{ij}$'s as before, involve the known functions of probabilities and sample and population sizes.

Consider the quantity,

$$Q = \text{Var}(\mathcal{C}) - \text{Var}(t')$$

$$= \sum_{i=1}^{N} (A_{ii} - B_{ii}) Y_i^2 + \sum_{i \neq j=1}^{N} (A_{ij} - B_{ij}) Y_i Y_j$$

(3.2.17)
The estimator $t'$ is necessary best estimator of order $r$ for the class $\mathcal{G}$, if all the leading principal minors of $Q$ up to the order $r$, are positive.

3.3. Bayes Estimation.

It was shown by Cochran (1942) that when auxiliary information on a characteristic closely related to the characteristic under consideration is available, we can utilize that information by considering a stochastic model, which is termed as the "super-population" the idea being taken from the Bayesian inference.

Cochran developed the theory in the following manner:

Whenever auxiliary information on a characteristic $X$ which takes values $X_i$ on the unit $U_i$, $i=1,2,\ldots,N$, is available closely related to the characteristic $Y$ under study, taking values $Y_i$ on $U_i$, $i=1,2,\ldots,N$, it is possible to use this information to set up a criterion of optimality.

Thus according to this super population concept, $Y = (Y_1,\ldots,Y_N)$ is assumed to be a realization of a $N$-length random vector with distribution $\delta$ depending on $X = (X_1,\ldots,X_N)$ and some unknown parameters. The role of prior distribution $\delta$ is solely to choose between different estimators and has nothing to do with final inference about $Y$ which will exclusively depend on the observed sample $s$ and variate values $Y_i$, ifs. We can talk of expectations, variances and covariances taken over $\delta$. We minimize the expected variance over $\delta$, namely
\[ \int V(H) d\delta \]  

(3.3.1)

for \( H \).

We have \( \Delta \) class of prior distributions \( \delta \) for which

\[
\begin{align*}
\mathbb{E}_{\delta_1}(Y_i | X_i) &= aX_i \\
\text{Var}_{\delta_1}(Y_i | X_i) &= 6 \ 2 \\
\text{Cov}_{\delta_1}(Y_i, Y_j | X_i, X_j) &= 0
\end{align*}
\]

(3.3.2)

where \( \mathbb{E}_{\delta_1}, \text{Var}_{\delta_1} \) and \( \text{Cov}_{\delta_1} \) denotes the conditional expectation, variance and covariance given \( X_i \)'s respectively.

Under the above super-population model the \( \text{Var}(t_i) \) can be written as

\[
\mathbb{E}_{\delta_1}[\text{Var}(t_i)] = (6 + a \sum_{i=1}^{N} \frac{(N-1) - (N-n)p_i}{n-1} - 1)X_i^2
\]

\[
+ a \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{(n-2)(N-1)+(N-n)(p_i+p_j)}{(n-1)(N-2)} - 1)X_iX_j
\]

(3.3.3)
Similarly $\ Var(t_2) \$ can be written as

$$
\mathcal{E}_{\delta_1}[\ Var(t_2) ] = (6 + a \ ) \sum_{i=1}^{N} \left( \frac{1}{\pi_i} \right) x_i^2 + a \sum_{i \neq j=1}^{N} \frac{\pi_i \pi_j}{(n-1)} x_i x_j \n$$

(3.3.4)

Also $\ Var(t_3) \$ can be denoted as

$$
\mathcal{E}_{\delta_1}[\ Var(t_3) ] = (6 + a \ ) \sum_{i=1}^{N} \left( \frac{1}{\pi_i} \right) x_i^2 \left( \sum_{s \neq i}^{N-1} \frac{1}{\sum_{j=1}^{n-1} \pi_j} - 1 \right)
+ a \sum_{i \neq j=1}^{N} x_i x_j \left( \sum_{s \neq i,j}^{N-1} \frac{1}{(n-1) \pi_j} - 1 \right) \n$$

(3.3.5)

3.4. Comparison of Estimators:

From equations (3.3.3) and (3.3.4) we have

$$
\mathcal{E}_{\delta_1}[\ Var(t_1) ] - \mathcal{E}_{\delta_1}[\ Var(t_2) ] = (6 + a \ ) \sum_{i=1}^{N} \left( \sum_{i \neq j=1}^{D_{ij}} x_i x_j + a \sum_{i \neq j=1}^{N} x_i x_j \right)
$$

(3.4.1)

where

$$
D_{ii} = \frac{(N-1) - (N-n)p_i}{n-1} - \frac{1}{\pi_i}
$$
and
\[ D_{ij} = \frac{(n-2)(N-1) + (N-n)(p_i + p_j)}{(n-1)(N-2)} - \frac{\pi_{ij}}{\pi_i \pi_j}. \]

The above equation (3.4.1) is a quadratic form in \( X_i \)'s. It is positive whatever may be \( X_i \)'s, if the quadratic form is positive definite.

Consider the first principal minor of the above quadratic form
\[ D_{11} = \frac{(N-1) - (N-n)p_i}{n-1} = \frac{1}{\pi_i} \]
\[ = \frac{(N-1) - (N-n)p_i}{n-1} = \frac{N-1}{(N-n)p_i + (n-1)} \]

\( D_{11} \) is positive if
\[ \left[ (N-1) - (N-n)p_i \right] \left[ (N-n)p_i + (n-1) \right] - (N-1)(n-1) > 0 \]
i.e.
\[ (N-n)^2 p_i (1-p_i) > 0 \]
i.e. \( p_i (1-p_i) > 0 \), which is true for \( i=1,2,\ldots,N \).

Thus we have the result that the estimator \( t_2 \) is a necessary better estimator of order one than the estimator \( t_1 \).
Considering the criterion of necessary better estimator of order two, it is noted the estimator $t_2$ is a necessary better estimator of order two than the estimator $t_1$, if the following condition is satisfied.

\[
\begin{align*}
(N-n) & \left\{ \frac{2}{p_1p_2} \left[ \frac{2}{(N-2)} \left( \frac{2}{(n-1)} - \frac{2}{(n-2)} \left( \frac{2}{N-1} \right) \right) \right] \\
+ & \frac{2}{p_1p_2} \frac{2}{N-n} \left[ \frac{2}{(N-n)} \left( \frac{2}{(n-1)} - 2 \frac{2}{(n-2)} \left( \frac{2}{N-1} \right) \right) \right] \\
+ & \frac{2}{(N-2)} \frac{2}{(N-n)} \frac{2}{p_1p_2} \left( 1 - p_1 - p_2 \right) + \frac{2}{(N-2)} \frac{2}{(N-n)} \frac{2}{p_1p_2} \\
+ & \frac{2}{p_1p_2} \left( p_1 + p_2 \right) \left( 1 - p_1 - p_2 \right) \left( n-1 \right) \left[ \frac{2}{(N-2)} \frac{2}{(N-n)} + 2 \frac{2}{(n-2)} \left( \frac{2}{N-1} \right) \right] \\
+ & \frac{2}{(n-1)} \left( 1 - p_1 - p_2 \right) \left[ \frac{2}{(N-2)} \frac{2}{p_1p_2} - \left( p_1 + p_2 \right)^2 \left( 1 - p_1 - p_2 \right) \right] \\
+ & 2 \frac{2}{p_1p_2} \frac{2}{N-n} \left( p_1 + p_2 \right)^2 \left( n-1 \right) \left( 1 - p_1 - p_2 \right) \right\} \geq 0.
\end{align*}
\]

If a pair of probabilities $p_1$ and $p_2$ satisfies the above condition, then the estimator $t_2$ is a necessary better estimator of order two than the estimator $t_1$.

For an illustration of the foregoing theory, consider the populations furnished by Yates and Grundy (1953), presented below in Table 3.4.1.
### Table 3.4.1

| Unit | \( p_i \) | Population A | | Population B | | Population C |
|------|----------|--------------| |--------------| |--------------|
|      |          | \( Y_i \) | \( X_i \) | \( Y_i \) | \( X_i \) | \( Y_i \) | \( X_i \) |
| 1    | 0.1      | 0.5          | 5    | 0.8          | 8    | 0.2          | 2    |
| 2    | 0.2      | 1.2          | 6    | 1.4          | 7    | 0.6          | 3    |
| 3    | 0.3      | 2.1          | 7    | 1.8          | 6    | 0.9          | 3    |
| 4    | 0.4      | 3.2          | 8    | 2.0          | 5    | 0.8          | 2    |
| Total| 1.0      | 26           | 26   | 10           |

It is noteworthy that Yates and Grundy (1953) have deliberately chosen these populations as being more extreme than will normally be encountered. The object is to compare the variance under super population model.

### Table 3.4.2

**Population A**

<table>
<thead>
<tr>
<th>Unit</th>
<th>( X_i )</th>
<th>( p_i )</th>
<th>( \pi_i )</th>
<th>( \frac{1}{\pi_i} )</th>
<th>( 3 - 2p_i )</th>
<th>( \frac{\sigma^2}{\left( \frac{N-1}{n-1}\right) \Sigma p_i} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>0.1923</td>
<td>0.4615</td>
<td>2.1668</td>
<td>2.6164</td>
<td>2.1765</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>0.2308</td>
<td>0.4871</td>
<td>2.0529</td>
<td>2.5384</td>
<td>2.0733</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>0.2692</td>
<td>0.5127</td>
<td>1.9504</td>
<td>2.4616</td>
<td>1.9665</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>0.3077</td>
<td>0.5384</td>
<td>1.8573</td>
<td>2.3846</td>
<td>1.8633</td>
</tr>
<tr>
<td>Total</td>
<td>26</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3.4.3
Population A

<table>
<thead>
<tr>
<th>Unit</th>
<th>$p_i^c p_j$</th>
<th>$\pi_{ij}$</th>
<th>$\pi_i \pi_j$</th>
<th>$\frac{\pi_{ij}}{\pi_i \pi_j}$</th>
<th>$\sum s_{ij}$</th>
<th>$\frac{1}{(N-1) \Sigma p_i}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$U_1 U_2$</td>
<td>0.4231</td>
<td>0.1410</td>
<td>0.2248</td>
<td>0.6272</td>
<td></td>
<td>0.7877</td>
</tr>
<tr>
<td>$U_1 U_3$</td>
<td>0.4615</td>
<td>0.1538</td>
<td>0.2366</td>
<td>0.6500</td>
<td></td>
<td>0.7221</td>
</tr>
<tr>
<td>$U_1 U_4$</td>
<td>0.5000</td>
<td>0.1666</td>
<td>0.2484</td>
<td>0.6706</td>
<td></td>
<td>0.6666</td>
</tr>
<tr>
<td>$U_2 U_3$</td>
<td>0.5000</td>
<td>0.1666</td>
<td>0.2497</td>
<td>0.6672</td>
<td></td>
<td>0.6666</td>
</tr>
<tr>
<td>$U_2 U_4$</td>
<td>0.5385</td>
<td>0.1795</td>
<td>0.2622</td>
<td>0.6845</td>
<td></td>
<td>0.6189</td>
</tr>
<tr>
<td>$U_3 U_4$</td>
<td>0.5769</td>
<td>0.2223</td>
<td>0.2761</td>
<td>0.6964</td>
<td></td>
<td>0.5777</td>
</tr>
</tbody>
</table>

Where $\pi_i$'s and $\pi_{ij}$'s are already defined.

Now substituting above values in the matrix form we have

$$
\begin{bmatrix}
0.4486(6^2 + a^2) & -0.2041 a^2 & -0.1885 a^2 & -0.1706 a^2 \\
-0.2041 a^2 & 0.4855(6^2 + a^2) & -0.1672 a^2 & -0.1460 a^2 \\
-0.1885 a^2 & -0.1672 a^2 & 0.5112(6^2 + a^2) & -0.1195 a^2 \\
-0.1706 a^2 & -0.1460 a^2 & -0.1195 a^2 & 0.5273(6^2 + a^2)
\end{bmatrix}
$$

It is noted that the first, second and third order principal minors are greater than zero, since 6 and a are positive. The fourth order principal minor is greater than zero if the following conditions is satisfied:
\[ 0.0586 \cdot 6^8 + 0.2344 \cdot 6^4 + 0.24 (0.3281 \cdot 6^2 - 0.0030) + 0.1746 \cdot 6^2 + 0.0217 a^2 - 0.0030 \geq 0. \]

Now consider the Population B.

**Table 3.4.4**

**Population B**

<table>
<thead>
<tr>
<th>Unit</th>
<th>( x_i )</th>
<th>( p_i )</th>
<th>( \pi_i )</th>
<th>( 1/\pi_i )</th>
<th>( 3-2p_i )</th>
<th>( \sum_{s \geq i,j} \frac{1}{(n-1) \Sigma \pi_i} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8</td>
<td>0.3077</td>
<td>0.5384</td>
<td>1.8573</td>
<td>2.3846</td>
<td>1.8633</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>0.2692</td>
<td>0.5127</td>
<td>1.9504</td>
<td>2.4616</td>
<td>1.9665</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>0.2308</td>
<td>0.4871</td>
<td>2.0529</td>
<td>2.5384</td>
<td>2.0733</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>0.1923</td>
<td>0.4615</td>
<td>2.1668</td>
<td>2.6164</td>
<td>2.1765</td>
</tr>
</tbody>
</table>

**Table 3.4.5**

**Population B**

<table>
<thead>
<tr>
<th>Unit</th>
<th>( p_i + p_j )</th>
<th>( \pi_{ij} )</th>
<th>( \pi_i \pi_j )</th>
<th>( \frac{\pi_{ij}}{\pi_i \pi_j} )</th>
<th>( \sum_{s \geq i,j} \frac{1}{(n-1) \Sigma \pi_i} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( u_1 u_2 )</td>
<td>0.5769</td>
<td>0.6223</td>
<td>0.2761</td>
<td>0.6964</td>
<td>0.5777</td>
</tr>
<tr>
<td>( u_1 u_3 )</td>
<td>0.5385</td>
<td>0.1795</td>
<td>0.2622</td>
<td>0.6845</td>
<td>0.6189</td>
</tr>
<tr>
<td>( u_1 u_4 )</td>
<td>0.5000</td>
<td>0.1666</td>
<td>0.2497</td>
<td>0.6672</td>
<td>0.6666</td>
</tr>
<tr>
<td>( u_2 u_3 )</td>
<td>0.5000</td>
<td>0.1666</td>
<td>0.2484</td>
<td>0.6706</td>
<td>0.6666</td>
</tr>
<tr>
<td>( u_2 u_4 )</td>
<td>0.4615</td>
<td>0.1538</td>
<td>0.2366</td>
<td>0.6500</td>
<td>0.7221</td>
</tr>
<tr>
<td>( u_3 u_4 )</td>
<td>0.4231</td>
<td>0.1410</td>
<td>0.2248</td>
<td>0.6272</td>
<td>0.7877</td>
</tr>
</tbody>
</table>

where \( \pi_i \)'s and \( \pi_{ij} \)'s are already defined.
Now substituting above values in matrix form we have

\[
\begin{bmatrix}
0.5273 \ (6^2 + a^2) & -0.1195 \ a^2 & -0.1460 \ a^2 & -0.1672 \ a^2 \\
-0.1195 \ a^2 & 0.5112 \ (6^2 + a^2) & -0.1706 \ a^2 & -0.1885 \ a^2 \\
-0.1460 \ a^2 & -0.1706 \ a^2 & 0.4855 \ (6^2 + a^2) & -0.2041 \ a^2 \\
-0.1672 \ a^2 & -0.1885 \ a^2 & -0.2041 \ a^2 & 0.4486 \ (6^2 + a^2)
\end{bmatrix}
\]

It is noted that the first three principal minors of this matrix are positive since values of \(a\) and \(b\) are positive. The principal minor of order four is positive if the following condition is satisfied.

\[
0.0598 \ b^8 + 0.0148 \ a^8 + 0.2374 \ a^6 \ b^2 + 0.3249 \ a^6 \ b^4 \\
+ a^6 \ [0.1563 \ b^2 - 0.0068] \geq 0.
\]

Now consider the Population \(C\)

<table>
<thead>
<tr>
<th>Unit</th>
<th>(X_i)</th>
<th>(p_i)</th>
<th>(\pi_i)</th>
<th>(1/\pi_i)</th>
<th>(3 \times 2p_i)</th>
<th>(E_{s+c}(N-1/S))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>0.2</td>
<td>0.4666</td>
<td>2.1432</td>
<td>2.6000</td>
<td>2.1666</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>0.3</td>
<td>0.5333</td>
<td>1.8751</td>
<td>2.4000</td>
<td>1.8888</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>0.3</td>
<td>0.5333</td>
<td>1.8751</td>
<td>2.4000</td>
<td>1.8888</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>0.2</td>
<td>0.4666</td>
<td>2.1432</td>
<td>2.6000</td>
<td>2.1666</td>
</tr>
</tbody>
</table>
Table 3.4.7

Population C

<table>
<thead>
<tr>
<th>Unit</th>
<th>$p_i + p_j$</th>
<th>$\pi_{ij}$</th>
<th>$\pi_i \pi_j$</th>
<th>$\frac{\pi_{ij}}{\pi_i \pi_j}$</th>
<th>$\Sigma_{s &gt; i, j} \frac{1}{(N-1)\Sigma_{p_i}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$U_1U_2$</td>
<td>.5000</td>
<td>.1666</td>
<td>.2488</td>
<td>6696</td>
<td>.6666</td>
</tr>
<tr>
<td>$U_1U_3$</td>
<td>.5000</td>
<td>.1666</td>
<td>.2488</td>
<td>6696</td>
<td>.6666</td>
</tr>
<tr>
<td>$U_1U_4$</td>
<td>.4000</td>
<td>.1333</td>
<td>.2177</td>
<td>.6123</td>
<td>.8333</td>
</tr>
<tr>
<td>$U_2U_3$</td>
<td>.6000</td>
<td>.2000</td>
<td>.2844</td>
<td>.7032</td>
<td>.5555</td>
</tr>
<tr>
<td>$U_2U_4$</td>
<td>.5000</td>
<td>.1666</td>
<td>.2488</td>
<td>6696</td>
<td>.6666</td>
</tr>
<tr>
<td>$U_3U_4$</td>
<td>.5000</td>
<td>.1666</td>
<td>.2488</td>
<td>6696</td>
<td>.6666</td>
</tr>
</tbody>
</table>

where $\pi_i$'s, $\pi_{ij}$'s are already defined.

Now substituting above values in matrix, we have,

\[
\begin{pmatrix}
0.4568 (6^2 + a^2) & -0.1696 a^2 & -0.1696 a^2 & -0.2123 a^2 \\
-0.1696 a^2 & 0.5249 (6^2 + a^2) & -0.1032 a^2 & -0.1696 a^2 \\
-0.1696 a^2 & -0.1032 a^2 & 0.5249 (6^2 + a^2) & -0.1696 a^2 \\
-0.2123 a^2 & -0.1696 a^2 & -0.1696 a^2 & 0.4568 (6^2 + a^2)
\end{pmatrix}
\]

Here it is noted that all the principal minors upto order four are positive, since $a$ and $6$ are positive.

From equation (3.3.4) and (3.3.5) we have

\[
\mathcal{E}_{\delta_1} \left[ \text{Var}(t_3) \right] - \mathcal{E}_{\delta_1} \left[ \text{Var}(t_2) \right] = (6^2 + a^2) \sum_{i=1}^{N} B_{ii} x_i^2 + a^2 \sum_{i+j=1}^{N} \Sigma B_{ij} x_i x_j
\]

(3.4.2)
where

\[ B_{ii} = \sum_{s \neq i} \frac{1}{(n-1) \pi_i \Sigma p_s} - \frac{1}{\pi_i} \]

and

\[ B_{ij} = \sum_{s \neq i, j} \frac{1}{(n-1) \Sigma p_s} - \frac{\pi_{ij}}{n_i n_j} \]

The equation (3.4.2) is positive if the value of the determinant of the following matrix is positive

\[
\begin{bmatrix}
B_{11} & B_{12} & B_{13} & B_{14} & \cdots & \cdots & B_{1N} \\
B_{21} & B_{22} & B_{23} & B_{24} & \cdots & \cdots & B_{2N} \\
B_{31} & B_{32} & B_{33} & B_{34} & \cdots & \cdots & B_{3N} \\
\vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\
B_{N1} & B_{N2} & B_{N3} & B_{N4} & \cdots & \cdots & B_{NN}
\end{bmatrix}
\]

Now first order principal minor of the above matrix is

\[ B_{11} = \sum_{s \neq i} \frac{1}{(n-1) \Sigma p_s} - \frac{1}{\pi_i} \]

from the cauchy-inequality we have

\[
\left[ \sum_{s \neq i} \frac{1}{(n-1) \Sigma p_s} \right] \left[ \sum_{s \neq i} \frac{\Sigma p_s}{(n-1)} \right] > 1.
\]
where
\[
\sum_{s=1}^{n} \frac{1}{(N-1)} \left( \frac{N-1}{n-1} \right) p_i + \left( \frac{N-2}{n-2} \right) (1-p_i)
\]
\[
\sum_{s=1}^{n} \frac{1}{(N-1)} (N-1) = p_i + \frac{n-1}{N-1} (1-p_i)
\]
\[
\frac{(N-n)p_i + (n-1)}{N-1} = \pi_i
\]
Thus we have proved that
\[
\sum_{s=1}^{n} \frac{1}{(N-1)} \Sigma p_i - \frac{1}{\pi_i} > 0.
\]
Hence we have the result that the estimator \( t_2 \) is a necessary better estimator of order one than the estimator \( t_3 \).

Consider the populations defined in Table 3.4.1. Using the values in Table No. 3.4.2 and Table No. 3.4.3 the expression (3.4.2) can be written as
\[
\begin{bmatrix}
.0097 (6^2 + a^2) & .1605 a^2 & .0721 a^2 & -.0040 a^2 \\
.1605 a^2 & .0204 (6^2 + a^2) & -.0006 a^2 & -.0656 a^2 \\
.0721 a^2 & -.0006 a^2 & .0161 (6^2 + a^2) & -.1187 a^2 \\
-.0040 a^2 & -.0656 a^2 & -.1187 a^2 & .0060 (6^2 + a^2)
\end{bmatrix}
\]
1) The value of the principal minor of order one is
\[
.0097 (6^2 + a^2)
\]
which is positive.
ii) The value of the principal minor of order two is

\[ .000197 \, b^4 + .000394 \, a^2 \, b^2 - .2556237 \, a^4 \]

The value of this equation will be positive if

\[ \frac{b^2}{a^2} > 9.1598 \]

iii) The value of principal minor of order three is

\[ .00000318 \, b^4 + .00000317 \, a^4 + .00000635 \, a^2 \, b^2 \]

\[ +.00052072 \, a^2 \, b^4 - .00052769 \, a^6 \]

This equation will be positive if the value of \[ \frac{b^2}{a^2} > .013804 \]

iv) The value of principal minor of order four is

\[ .00000000189 \, b^8 + .0000000076 \, a^2 \, b^6 - .0000018302 \, a^4 \, b^2 \]

\[ -.0000050378 \, a^6 - .0001610839 \, a^8 \]

Now considering the Population B, we have the matrix of the quadratic form in equation (3.4.2) as follows:

\[
\begin{bmatrix}
.0060 (b^2 + a^2) & -.1187 a^2 & -.6560 a^2 & -.0040 a^2 \\
-.1187 a^2 & .0161 (b^2 + a^2) & -.0006 a^2 & .0721 a^2 \\
-.6560 a^2 & -.0006 a^2 & .0204 (b^2 + a^2) & .1605 a^2 \\
-.0040 a^2 & .0721 a^2 & .1605 a^2 & .0097 (b^2 + a^2)
\end{bmatrix}
\]

i) The value of principal minor of order one is

\[ .0060 \, (b^2 + a^2) \]

which is positive.
ii) The value of principal minor of order two is
\[ 0.00009666 \, 6^2 + 0.0019320 \, a^2 \, 6^2 - 0.1399309 \, a^4, \]
and it will be positive if \( \frac{6^2}{a^2} \geq 18.5295. \)

iii) The value of principal minor of order three is
\[ 0.00000197 \, 6^6 - 0.0035080 \, a^4 \, 6^2 + 0.00000591 \, 6^4 \, a^2 \\
-0.0035474 \, a^6. \]

iv) The value of principal minor of order four is
\[ 0.0000000196 \, 6^8 + 0.0000000760 \, a^2 \, 6^6 + 0.0000000052 \, a^4 \, 6^2 \\
+ 0.000000052 \, a^6 - 0.00064496 \, a^4 \, 6^4 - 0.0000222111 \, a^2 \, 6^2 \\
-0.0005282941 \, a^8. \]

Considering the population C. The matrix of the quadratic form (3,4,2) is as follows:

\[
\begin{bmatrix}
0.0234 \, (6^2+a^2) & -0.0030 \, a^2 & -0.0030 \, a^2 & -0.2170 \, a^2 \\
-0.0030 \, a^2 & 0.0137 \, (6^2+a^2) & -0.1477 \, a^2 & -0.0030 \, a^2 \\
-0.0030 \, a^2 & -0.1477 \, a^2 & 0.0137 \, (6^2+a^2) & -0.0030 \, a^2 \\
-0.2170 \, a^2 & -0.0030 \, a^2 & -0.0030 \, a^2 & 0.0234 \, (6^2+a^2) \\
\end{bmatrix}
\]

i) The value of the principal minor of order one is
\[ 0.0234 \, (6^2+a^2) \]
which is positive.

ii) The value of the principal minor of order two is
\[ 0.00032 \, (6^4+2a^26^2) + 0.00031 \, a^4 \]
which is positive.
iii) The value of principal minor of order three is

\[ 0.00000439 \ 6^6 - 0.00049755 \ a^{4.4} - 0.00001318 \ a^{4.4} - 0.00051338 \ a^6. \]

iv) The value of principal minor of order four is

\[ 0.000001024 \ 6^8 + 0.000004096 \ a^{2.6} - 0.0000025073 \ a^{4.6} - 0.0000057707 \ a^{6.2} - 0.0000037546 \ a^8. \]

From equations (3.3.3) and (3.3.5) we have

\[ \varepsilon_{\delta_1} \left[ \text{Var}(t_1) \right] - \varepsilon_{\delta_1} \left[ \text{Var}(t_3) \right] = \left( 6^2 + a^2 \right)^N \sum_{i=1}^{N} D_{i1} x_i^2 + a^2 \sum_{i<j=1}^{N} D_{ij} x_i x_j \]

where

\[ D_{i1} = \frac{(N-1) - (N-n)p_i}{n-1} - \sum_{s \neq i}^{1} \frac{1}{(N-1) \Sigma p_i} \]

and

\[ D_{ij} = \frac{(N-1)(n-2) + (N-n)(p_i + p_j)}{(n-1)(N-2)} - \sum_{s \neq i, j}^{1} \frac{1}{(N-1) \Sigma p_i} \]

The matrix of the above quadratic form in equation (3.4.3) can be written as

\[
\begin{bmatrix}
0.4389 \ (6^2 + a^2) & -3.646 \ a^2 & -2.606 \ a^2 & -1.666 \ a^2 \\
-3.646 \ a^2 & 4.651 \ (6^2 + a^2) & -1.666 \ a^2 & -0.0804 \ a^2 \\
-2.606 \ a^2 & -1.666 \ a^2 & 4.951 \ (6^2 + a^2) & -0.0008 \ a^2 \\
-1.666 \ a^2 & -0.0804 \ a^2 & -0.0008 \ a^2 & 0.5213 \ (6^2 + a^2)
\end{bmatrix}
\]
Here the first two principal minors of orders respectively one and two are positive since $a$ and $b$ are positive.

i) The value of the principal minor of order three is
\[
.1010 \, b^5 - .0279 \, a^5 - .0037 \, b^2a^4 + .3030 \, b^4a^2 - .0121 \, a^6b^2 \\
-.0121 \, a^4.
\]

ii) The value of the principal minor of order four is given by
\[
.0520 \, b^8 + .0344 \, a^8 + .1562 \, b^6a^2 + .2475 \, b^4a^4 + .0572 \, a^6b^2.
\]

Now for the population $B$, the matrix of the quadratic form in equation (3.4.3) can be noted as
\[
\begin{bmatrix}
.5213 \, (b^2+a^2) & -.0008 \, a^2 & -.0804 \, a^2 & -.1666 \, a^2 \\
-.0008 \, a^2 & .4951 \, (b^2+a^2) & -.1666 \, a^2 & -.2606 \, a^2 \\
-.0804 \, a^2 & -.1666 \, a^2 & .4651 \, (b^2+a^2) & -.3646 \, a^2 \\
-.1666 \, a^2 & -.2606 \, a^2 & -.3646 \, a^2 & .4389 \, (b^2+a^2)
\end{bmatrix}
\]

Here also the first two principal minors of orders respectively one and two are positive since $a$ and $b$ are positive.

i) The value of the principal minor of order three is
\[
.1200 \, b^6 + .3600 \, a^2b^4 + a^4(.35109976 \, b^2 - .00000804) \\
+ .11108912 \, a^6.
\]

ii) The value of the principal minor of order four is
\[
.0526 \, b^8 + a^2b^4 (.2104 \, b^2 - .0164) + a^4b^2(.3176 \, b^2 - .0326) \\
+ .2455 \, a^6b^2 + .0677 \, a^8.
\]

Now for the population $C$, the matrix of the quadratic form in equation (3.4.3) can be noted as,
Here also the first two principal minors of orders respectively one and two are positive since $a$ and $b$ are positive.

i) The value of the principal minor of order three is given by

$$0.4334 (b^2 + a^2) - 0.1666 a^2 - 0.1666 a^2 - 0.4333 a^2$$

$$-0.1666 a^2 - 0.5112 (b^2 + a^2) - 0.0445 a^2 - 0.1666 a^2$$

$$-0.1666 a^2 - 0.0445 a^2 - 0.5112 (b^2 + a^2) - 0.1666 a^2$$

$$-0.4333 a^2 - 0.1666 a^2 - 0.1666 a^2 - 0.4334 (b^2 + a^2)$$

ii) The value of the principal minor of order four is given by

$$0.1132 b^6 + 0.3255 b^2 a^4 - 0.0008 a^4 - 0.0141 b^6 a^2 - 0.0141 a^8$$

$$0.0491 b^8 + 0.0257 b^2 a^6 + 0.2203 b^4 a^4 + 0.1472 b^6 a^2 + 0.355 a^8$$

3.5 Optimum Sampling Procedure.

Under a prior distribution $\delta$, we minimize the expected variance $V[H(S,p,t)]$ uniformly in $\gamma$ for variation of $H$ over the class of all equally costly strategies. Thus we minimize the expected loss.

$$\mathbb{E}_\delta V(H) = \int V(H) d\delta$$

over $\delta$

which helps us in arriving at optimal sampling strategies.

Using this concept Godambe (1955) proved that under the particular model $\Delta_1$, there exists an $H_1$ optimum sampling strategy for which
i) The inclusion probability for the $i$th unit, $π_i$, is proportional to the value $X_i$ taken by the auxiliary characteristic on that unit.

ii) Every sample has $n$ distinct units.

iii) The estimator used is the corresponding Horvitz and Thompson estimator.

Now we have to minimize $E_δ_1 [\text{Var}(t_1)]$ with respect to $p_i$'s.

As noted in the previous section

\[
E_δ_1 [\text{Var}(t_1)] = (6^2 + a^2) \sum_{i=1}^{N} \frac{(N-1) - (N-n)p_i}{n-1} X_i^2 \\
+ a^2 \sum_{i \neq j}^{N} \frac{(n-2)(N-1)+(N-n)(p_i+p_j)}{(n-1)(N-2)} X_i X_j \\
- (6^2 + a^2) \sum_{i=1}^{N} X_i^2 - a^2 \sum_{i \neq j}^{N} X_i X_j 
\]

(3.5.1)

Min $E_δ_1 [\text{Var}(t_1)] = (6^2 + a^2) \frac{N-1}{n-1} \sum_{i=1}^{N} X_i^2 + a^2 \frac{(n-2)(N-1)}{(n-1)(N-2)} \sum_{i \neq j}^{N} X_i X_j \\
- (6^2 + a^2) \sum_{i=1}^{N} X_i^2 - a^2 \sum_{i \neq j}^{N} X_i X_j \\
- \text{Max}[ (6^2 + a^2) \frac{N-n}{n-1} \sum_{i=1}^{N} p_i X_i^2 - a^2 \frac{N-n}{(n-1)(N-2)} \sum_{i \neq j}^{N} (p_i+p_j) X_i X_j ]$

Now we will minimize the term

\[
(6^2 + a^2) \frac{N-n}{n-1} \sum_{i=1}^{N} p_i X_i^2 - a^2 \frac{N-n}{(n-1)(N-2)} \sum_{i \neq j}^{N} (p_i+p_j) X_i X_j 
\]

(3.5.2)

with respect to $p_i$'s.
Since
\[
\sum_{i \neq j=1}^{N,N} (p_i + p_j)x_ix_j = \sum_{i \neq j=1}^{N,N} p_ix_jx_i + \sum_{i \neq j=1}^{N,N} p_jx_ix_j
\]
\[
= \sum_{i=1}^{N} p_i x_i (x_i - x_i) + \sum_{j=1}^{N} p_j x_j (x_i - x_j)
\]
\[
= 2x \sum_{i=1}^{N} p_i x_i - 2 \sum_{i=1}^{N} p_i x_i
\]
Thus equation (3.5.2) can be written as
\[
(6^2 + a^2)^{N-n} \sum_{i=1}^{N} p_i x_i^2 - a^2 \frac{2(N-n)}{(n-1)(n-2)} \left[ 2x \sum_{i=1}^{N} p_i x_i - 2 \sum_{i=1}^{N} p_i x_i^2 \right]
\]
\[
= a^2 \sum_{i=1}^{N} p_i x_i^2 + a^2 \frac{2(N-n)}{(n-1)(n-2)} \sum_{i=1}^{N} p_i x_i
\]
\[
+ 6^2 \sum_{i=1}^{N} p_i x_i^2 - a^2 \frac{2(N-n)}{(n-1)(n-2)} x \sum_{i=1}^{N} p_i x_i
\]
\[
= a^2 \frac{N(N-n)}{(n-1)(n-2)} \sum_{i=1}^{N} p_i x_i^2 + 6^2 \sum_{i=1}^{N} p_i x_i - a^2 \sum_{i=1}^{N} p_i x_i^2 + \frac{2(N-n)}{(n-1)(n-2)} x \sum_{i=1}^{N} p_i x_i
\]
\[
= -a^2 \sum_{i=1}^{N} p_i x_i^2 \left[ 2x - N x_i - \frac{6^2}{a^2} (N-2)x_i \right]
\]
Let \(Z_1, Z_2, \ldots, Z_N\) be the ranking when the \(N\) quantities
\[
X_i \left[ 2X - N X_i - \frac{6^2}{a^2} (N-2)X_i \right]
\]
are ranked in the decreasing order of magnitude. So accordingly we get
\[
\text{Min} \sum_{i=1}^{N} \delta_1 \left[ \text{Var}(t_1) \right] = (6^2 + a^2)^{N-1} \sum_{i=1}^{N} x_i^2 + a^2 \frac{(n-2)(n-1)}{(n-1)(n-2)} \sum_{i=1}^{N} p_i x_i x_j
\]
\[- (6^2 + a^2)^{N} \sum_{i=1}^{N} x_i - a \sum_{i=1}^{N} \sum_{i \neq j=1}^{N} x_i x_j - \min \sum_{i=1}^{N} p_i Z_i \].
The minimum occurs when $p_N$ corresponding to $Z_N$ is one and accordingly the rest of $p_i$'s are zero. This is noted as follows

$$\sum_{i=1}^{N} p_i Z_i = \sum_{i=1}^{N-1} p_i Z_i + (1 - \sum_{i=1}^{N-1} p_i) Z_N$$

$$= \sum_{i=1}^{N-1} p_i (Z_i - Z_N) + Z_N.$$

Apparently the minimum occurs for the variations of $p_i$'s when $p_i = 0$, for $i=1,2,\ldots,N-1$, and subsequently $p_N = 1$. Since the determination of $p_i$'s depend on the values of $X_i$'s. Let $X_i$ be the minimum among the $N$ quantities $X_i \left[ 2X - NX_i - \frac{6}{a^2} (N-2)X_i \right]$. Accordingly the minimum value of $\mathcal{C}_{\delta_i} \left[ \text{Var}(t_i) \right]$ is noted as

$$\text{Min } \mathcal{C}_{\delta_i} \left[ \text{Var}(t_i) \right] = (6^2 + a^2) \sum_{i=1}^{N-1} \frac{X_i^2}{(n-1)(n-2)} - \frac{a^2}{(n-1)(n-2)} \sum_{i=1}^{N} X_i X_j$$

$$+ a^2 \sum_{i=1}^{N-n} X_i \left[ 2X - NX_i - \frac{6}{a^2} (N-2)X_i \right] - (6^2 + a^2) \sum_{i=1}^{N} X_i^2 a \sum_{i+j=1}^{N} X_i X_j.$$

$$= (6^2 + a^2) \sum_{i=1}^{N} \left( \frac{N-1}{n-1} - 1 \right) X_i + a^2 \sum_{i=j=1}^{N} \left( \frac{N-2}{(n-1)(n-2)} - 1 \right) X_i X_j$$

$$+ 2a^2 \sum_{i=1}^{(N-n)} X_i X - a^2 \sum_{i=1}^{N} X_i^2 \left( \frac{N(N-n)}{(n-1)(n-2)} - 6 \right) \frac{n-1}{n-1} X_i^2.$$
\[
= 6^2 \sum_{i=1}^{N} \left( \frac{N-1}{n-1} - 1 \right) X_i^2 + a^2 \sum_{i=1}^{N} \left( \frac{N-1}{n-1} - 1 \right) X_i^2 - a^2 \frac{N(N-n)}{(n-1)(n-2)} X' X^2 \\
- 6^2 \frac{N-n}{n-1} X^2 + a^2 \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{(n-2)(N-1)}{(n-1)(N-2)} X_i X_j \\
+ a^2 \frac{(N-n)}{(n-1)(N-2)} \left[ 2 \sum_{i=1}^{N} \left( \frac{X_i + X'}{n} \right) \right] X' \\
\text{Since } X = \sum_{i=1}^{N} \frac{X_i + X'}{n} \]

\[
= 6 \sum_{i=1}^{N} \left( \frac{N-1}{n-1} - 1 \right) X_i^2 + a^2 \left[ \sum_{i=1}^{N-n} X_i^2 + \frac{2(N-n)}{(n-1)(N-2)} X^2 \\
- \frac{N(N-n)}{(n-1)(N-2)} X^2 \right] - \frac{N-n}{n-1} 6^2 X^2 + a^2 \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{(n-2)(N-1)}{(n-1)(N-2)} X_i X_j \\
+ a^2 \frac{2(N-n)}{(n-1)(N-2)} X' \sum_{i=1}^{N} \frac{X_i}{n} \\
= 6^2 \left( \frac{N-1}{n-1} - 1 \right) \sum_{i=1}^{N} X_i^2 + a^2 \frac{N-n}{n-1} \left[ \sum_{i=1}^{N} X_i^2 + \frac{2X^2}{n-2} - \frac{NX'^2}{n-2} \right] \\
+ a^2 \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{(n-2)(N-1)}{(n-1)(N-2)} X_i X_j + \frac{2a^2(N-n)}{(n-1)(N-2)} X' \sum_{i=1}^{N} X_i \\
- \frac{N-n}{n-1} 6^2 X^2 \\
\text{(3.5.5.3)}
\]

Under the super-population model the \( \text{Var}(t_2) \) can be written as
\[ \mathcal{C}_{\delta_1} [\text{Var}(t_2^2)] = \sum_{i=1}^{N} \left( \frac{1}{\pi_i} - 1 \right) \mathcal{C}_{\delta_1} (Y_{i1}^2 | X_1) + \sum_{i \neq j=1}^{N} \sum_{i \neq j}^{N} \left( \frac{\pi_{ij}}{\pi_i \pi_j} - 1 \right) \mathcal{C}_{\delta_1} (Y_{ij} Y_{ij} | X_1, X_j) \]

\[ = \sum_{i=1}^{N} \left( \frac{1}{\pi_i} - 1 \right) \left( 6 \frac{2}{X_1} + a \frac{2}{X_1} \right) + \sum_{i \neq j=1}^{N} \sum_{i \neq j}^{N} \left( \frac{\pi_{ij}}{\pi_i \pi_j} - 1 \right) \frac{2}{X_1} \frac{X_j}{X_j} \]

\[ = 6 \sum_{i=1}^{N} \left( \frac{1}{\pi_i} - 1 \right) \frac{2}{X_1} + a^2 \text{Var} \left( \frac{X_1}{\text{ifs } \pi_i} \right) \]

Godambe (1955) demonstrated that the above expression is minimum when \( \pi_i \) (the probability of inclusion of the \( i \)-th population unit in the sample) is proportional to \( X_1 \). i.e.

\[ \pi_i = np_i = n \frac{X_i}{\bar{X}} \quad \text{and} \quad p_i = \frac{X_i}{\bar{X}} \quad \text{assumed less than } \frac{1}{n} \]

Now if \( \pi_i = np_i \), then in the above equation the second term

\[ a^2 \text{Var} \left( \frac{X_1}{\text{ifs } \pi_i} \right) \]

vanishes, since \( \sum_{\text{ifs } \pi_i} \frac{X_1}{\pi_i} \) is a constant for \( \pi_i \propto X_1 \).

Thus we have

\[ \text{Min } \mathcal{C}_{\delta_1} [\text{Var}(t_2^2)] = 6 \sum_{i=1}^{N} \frac{X_i}{\bar{X} np_i} \left( \frac{1}{np_i} - 1 \right) \quad (3.5.4) \]

Now under the super-population model the \( \text{Var}(t_3^2) \) can be written as
\[ \mathcal{E}_{\delta_1}[\text{Var}(t_3)] = \sum_{i=1}^{N} \mathcal{E}_{\delta_1}(Y_i|X_i) \sum_{s \neq i} \frac{1}{(N-1) \Sigma \pi_i} \\
+ \sum_{i \neq j=1}^{N} \mathcal{E}_{\delta_1}(Y_iY_j|X_i,X_j) \sum_{s \neq i,j} \frac{1}{(N-1) \Sigma \pi_i} - \sum_{i=1}^{N} \mathcal{E}_{\delta_1}(Y_i|X_i) \\
- \sum_{i \neq j=1}^{N} \mathcal{E}_{\delta_1}(Y_iY_j|X_i,X_j) \\
= 6 \sum_{i=1}^{N} \frac{X_i^2}{s_{\pi i}} \left( \frac{1}{(N-1) \Sigma \pi_i} - 1 \right) \\
+ a^2 \left[ \sum_{i=1}^{N} \frac{X_i^2}{s_{\pi i}} \left( \frac{1}{(N-1) \Sigma \pi_i} - 1 \right) + \sum_{i \neq j=1}^{N} \frac{X_iX_j}{s_{\pi i,j}} \left( \frac{1}{(N-1) \Sigma \pi_i} - 1 \right) \right] \\
(3.5.5) \\
\]

It is noted that if \( \Sigma \frac{p_i}{\pi} \propto \Sigma \frac{x_i}{\pi} \) then the second term of the above equation vanishes and we obtain the equation

\[ \mathcal{E}_{\delta_1}[\text{Var}(t_3)] = 6 \sum_{i=1}^{N} \frac{X_i^2}{s_{\pi i}} \left( \frac{1}{(N-1) \Sigma \pi_i} - 1 \right) \]

(3.5.6)

And if \( \Sigma \frac{p_i}{\pi} \propto \Sigma \frac{x_i}{\pi} \) then the first term of the above equation will be minimum. It has been noted that under certain conditions the estimator \( t_3 \) will have the less variance if \( \Sigma \frac{p_i}{\pi} \propto \Sigma \frac{x_i}{\pi} \). Thus for our comparison purpose we consider the variance given by equation (3.5.6).
3.6. Comparison of Estimators.

In this section we shall discuss the choice of a suitable estimator under super population model.

**Theorem 3.6.1.** The estimator $t_2$ is a necessary better estimator of order $r$ than the estimator $t_1$.

**Proof:**

$$
E \delta_1 \left[ \frac{\text{Var}(t_1) - \text{Var}(t_2)}{6^2} \right] = \sum_{i=1}^{a^2(N-n)} \left[ \frac{N}{\sum X_i^2} \frac{X_i^2}{N-2} - \frac{N}{X^2} \right]
$$

$$
+ \frac{a^2(N-n)}{6^2(n-1)} \left[ \sum_{i=1}^{N} \sum_{j=1}^{X_i} \frac{(n-2)(N-1)}{(n-1)(N-2)} \frac{X_iX_j}{(N-1)(N-2)} \right]
$$

$$
+ \frac{a^2}{6^2} \sum_{i=1}^{X_i} \sum_{j=1}^{X_i} \frac{(N-n)}{(n-1)(N-2)} \frac{X_iX_j}{(N-1)(N-2)} \frac{X_i^2}{X_j^2}
$$

$$
= \alpha \sum_{i=1}^{X_i} X_i + \alpha' X + \beta \sum_{i=1}^{X_i} X_iX_j + \beta' X \sum_{i=1}^{X_i} X_i
$$

where

$$
\alpha = \frac{N-1}{n-1} - \frac{1}{n} + \frac{a^2(N-n)}{6^2(n-1)}
$$

$$
\alpha' = \frac{n-1}{n}
$$

$$
\beta = -\left( \frac{a^2(N-n)}{6^2(n-1)(N-2)} + \frac{1}{n} \right)
$$

$$
\beta' = -\frac{1}{n}
$$

215211
The matrix of the above quadratic form can be written as

\[
\begin{bmatrix}
\alpha & \beta & \beta & \ldots & \beta' & \ldots & \ldots & \beta \\
\beta & \alpha & \beta & \ldots & \beta' & \ldots & \ldots & \beta \\
\beta & \beta & \alpha & \ldots & \beta' & \ldots & \ldots & \beta \\
\vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \ddots & \vdots \\
\beta' & \beta' & \ldots & \alpha' & \ldots & \ldots & \ldots & \beta' \\
\vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\
\beta' & \beta' & \ldots & \ldots & \ldots & \ldots & \alpha & \beta \\
\end{bmatrix}
\]

Now consider any principal minor of order $r$ which does not involve the terms $\alpha'$ and $\beta'$. The value of this principal minor is noted as

\[
\begin{bmatrix}
\alpha+(r-1)\beta \\
\alpha-\beta \\
\end{bmatrix}^{r-1}
\]

If this value is greater than zero then the quadratic form will be positive definite and we have the result that the estimator $t_2$ is necessary better estimator of order $r$ than the estimator $t_1$. 
\[
\left[ \alpha + (r-1)\beta \right] \left[ \alpha - \beta \right]^{r-1}
\]

\[
= \left\{ \frac{N-1}{n-1} - \frac{1}{n} + \frac{\alpha^2 (N-n)}{6^2 (n-1)} + (r-1) \left[ -\frac{\alpha^2 (N-n)}{6^2 (n-1)(N-2)} - \frac{1}{n} \right] \right\}
\]

\[
\left[ \frac{N-1}{n-1} - \frac{1}{n} + \frac{\alpha^2 (N-n)}{6^2 (n-1)} + \frac{\alpha^2 (N-n)}{6^2 (n-1)(N-2)} + \frac{1}{n} \right]^{r-1}
\]

\[
= \left[ \frac{N-1}{n-1} - \frac{r}{n} + \frac{\alpha^2 (N-n)(N-r-1)}{6^2 (n-1)(N-2)} \right] \left[ \frac{N-1}{n-1} + \frac{\alpha^2 (N-n)(N-1)}{6^2 (n-1)(N-2)} \right]^{r-1}
\]

Here it is noted that \( r < N \). So the above expression is always positive.

Now if the \( r \)th order principal minor involves the terms \( \alpha' \) and \( \beta' \), also then the value of that principal minor is given by following expression

\[
(\alpha - \beta)^{r-2} \left[ \alpha' (\alpha + (r-2)\beta) - (r-1) \beta' \right]^2
\]

If the value of this expression is positive then the quadratic form is positive definite and we have the result that the estimator \( t_2 \) is necessary better of order \( r \) than the estimator \( t_1 \).
\[(\alpha - \beta)^{r-2} \left[ \alpha' (\alpha + (r-2)\beta) - (r-1)\beta' \right] \]

\[= \left[ \frac{N-1}{n^{-1}} - \frac{1}{n} + \frac{a^2(N-n)}{6^2(n-1)} \right]^{r-2} \left\{ \frac{n-1}{n^{-1}} - \frac{1}{n} + \frac{a^2(N-n)}{6^2(n-1)} \right\} \]

\[= \left[ \frac{N-1}{n^{-1}} - \frac{1}{n} + \frac{a^2(N-n)}{6^2(n-1)} \right]^{r-2} \left[ \frac{(N-r)}{n^{-1}} + \frac{a^2(N-n)(N-r)}{6^2n(N-2)} \right] \]

\[> 0.\]

Thus we have proved that the matrix is positive, and we have the result that the estimator \( t_2 \) is a necessary better estimator of order \( r \) than the estimator \( t_1 \).

**Theorem 3.6.2.** The estimator \( t_2 \) is a necessary better estimator of order \( N \) than the estimator \( t_3 \).

**Proof:**

\[
\min \mathcal{E} \delta_1 \left[ \frac{\text{Var}(t_3) - \text{Var}(t_2)}{6^2} \right] 
\]

\[= \sum_{i=1}^{N} \frac{X_i^2}{\sum_{i=1}^{N} \frac{1}{(N-1)\xi_i} - \frac{X}{nX_1}} \]

\[\]
From Cauchy-Schwartz inequality we have

\[
\left[ \sum_{i=1}^{N} \frac{1}{(n-1)\Sigma X_i} \right] \left[ \sum_{i=1}^{N} \frac{\Sigma X_i}{(n-1)X} \right] \geq 1.
\]

i.e.

\[
\sum_{i=1}^{N} \frac{X}{(n-1)\Sigma X_i} \geq \frac{(N-1)X}{(n-1)X + X_i(N-n)}
\]

Thus

\[
\frac{(N-1)X}{(n-1)X + X_i(N-n)} \geq \frac{X}{nX_i}
\]

\[
= \frac{n(N-1)XX_i - X[ (n-1)X + X_i(N-n) ]}{nX_i [ (n-1)X + (N-n)X_i ]}
\]

\[
= X \left[ n(N-1)X_i - (n-1)X - (N-n)X_i \right]
\]

\[
= X \left[ NX_i(n-1) - (n-1)X \right]
\]

\[
= (n-1)X[ NX_i - X ]
\]

\[\geq 0 \text{ if } NX_i > X, i=1,2,\ldots,N.\]