National College of Business
Administration and Economics

NCBA&E

SAMPLING WITH UNEQUAL PROBABILITIES AND WITHOUT REPLACEMENT

By
MUHAMMAD QAISER SHAHBAZ

Doctorate of Philosophy
In
STATISTICS
May; 2003

National College of Business Administration and Economics.

SAMPLING WITH UNEQUAL PROBABILITIES AND WITHOUT REPLACEMENT

BY

MUHAMMAD QAISER SHAHBAZ

DOCTORATE OF PHILOSOPHY

IN

STATISTICS

May, 2003

National College of Business Administration and Economics

SAMPLING WITH UNEQUAL PROBABILITIES AND WITHOUT REPLACEMENT

BY

MUHAMMAD QAISER SHAHBAZ

A Dissertation Submitted to the
National College of Business Administration & Economics

In Partial Fulfillment of the
Requirements for the Degree of

DOCTORATE OF PHILOSOPHY
IN
STATISTICS

May, 2003

NATIONAL COLLEGE OF BUSINESS ADMINISTRATION AND ECONOMICS

SAMPLING WITH UNEQUAL PROBABILITIES AND WITHOUT REPLACEMENT

BY

MUHAMMAD QAISER SHAHBAZ

A dissertation submitted to School of Management Studies, in partial fulfillment of the requirements for the degree of

DOCTORATE OF PHILOSOPHY IN STATISTICS

Director Institute of Advanced Studies

Dissertation Committee:

Chairman

Member

Member

Declaration

This is to certify that the research work I am submitting has not already been submitted and shall not in future be submitted for obtaining similar degree from any other university.

Muhammad Qaiser Shahbaz

24 – 05 – 2003.

DEDICATIONS

To my Mother, the living legend,

All my Sisters

And

My Wife.

Acknowledgements

All and every praise be for ***Allah***, the most magnificent, the most beneficent. I pray to Him for His guidance and protection throughout my life. It is with His guidance and help that I have been able to complete this huge task.

First of all I would like to thank my supervisor, Prof. Dr. Muhammad Hanif, who has always been kind and helpful throughout this research work. I am also highly indebted to him as this work would not have been completed without his thought provoking ideas, sincere guidance and patience.

I also pay my special gratitude to Prof. Dr. Munir Ahmad for his guidance regarding my thesis that has definitely supported me to complete this task. His sublime personality is a mile stone for the students who think that sky is the limit.

I would also like to thank Prof. Dr. Muhammad Samiuddin for his valuable suggestions and enlightening comments about my research work. He is also a source of inspiration for me and for all the students who want to serve the subject of Statistics.

I am highly indebted to the prayers of my mother, beneath whose feet lay my heavens, who has always encouraged me in my endeavor life and has been a source of inspiration for me. I would not have been able to complete this work without blessings of Allah and prayers of my mother. Thanks are also due to all my sisters for their love and encouragements throughout the course of this thesis. I am also thankful to my wife for her help, encouragement and pleasurable acts during the course of my thesis. I would also like to thank my nephews and nieces for their love and criticism during my work.

Finally, I would like to thank all my colleagues and friends, who always helped and encouraged me whenever I had any problem regarding my research.

Contents

List of Tables

Summary

After describing the basic theory of survey sampling with reference to equal and unequal probability sampling, some selected selection procedures have been discussed, which can be used with Horvitz and Thompson estimator. Some of the popular estimators of population total (other than the Horvitz – Thompson estimator) have been discussed. The Model based sampling inference has been presented along with the famous model based estimators.

Some approximate formulae for variance of the Horvitz – Thompson estimator that use only the first order inclusion probabilities have been obtained. Some special cases of these approximations have also been given.

Three new selection procedures for use with Horvitz – Thompson estimator have been developed. These selection procedures are applicable for a sample of size two and are strictly without replacement. Some fundamental results related to inclusion probabilities and joint inclusion probabilities have been verified for these newly developed selection procedures. Empirical study for these new selection procedures have been carried out in order to see their performance for various types of populations. The regression analysis has been carried out in order to see the effect of coefficient of variation and correlation coefficient on the variance of these estimators. It has been found that these two coefficients have significant effect on the variance of Horvitz – Thompson estimator under the newly developed selection procedures.

A general procedure has been developed by introducing a constant in the revised probabilities of selection that helps in developing a number of other selection procedures. It has been found that the Yates – Grundy draw-by-draw (1953) and the Brewer (1963a) procedures are special cases of the general selection procedure. Empirical study has been conducted to obtain a suitable value of the constant for various sorts of populations.

A series of modified Murthy estimators has been developed by using the general Murthy (1957) estimator. These estimators have been developed by using various selection procedures in the general Murthy (1957) estimator. It has been found that the estimator used by Durbin (1953) for his rejective procedure is a special case of Murthy (1957) estimator under the Durbin (1967) draw-by-draw procedure. The unbiasedness of the new estimators has been verified and their design-based variances have been obtained. Empirical study has been carried out in order to see the performance of the new estimators.

The model based study of the modified Murthy estimator under the Durbin (1953) draw-by-draw procedure has been conducted and it is found that this estimator achieves the Godambe – Joshi (1965) lower bound to the variance of any estimator in unequal probability sampling.

Abbreviations

PPS	Probability Proportional to size sampling with replacement.
π PS	Probability Proportional to size sampling without replacement.
$\text{Var}\left(\hat{\theta}\right)$	Variance of Sampling Distribution of an Estimator $\hat{\theta}$.
Var	Sample Variance or Variance Estimator
d – b – d	Draw by Draw
MM1	Modified Murthy Estimator 1
MM2	Modified Murthy Estimator 2
MM3	Modified Murthy Estimator 3
RM	Modified Raj Estimator

Chapter 1

Introduction and Literature Survey

1.1 Introduction:

The theory of sampling has its origin way back in the history of mankind. People took a portion of a totality, generally called the *statistical population*, to decide about its nature. The selected part is generally referred to as a *sample*. The proper scientific tools used to obtain a sample from a population are referred to as the *sampling techniques* and a specific procedure of selecting a sample is called a *sampling design*. The collection of a sampling design, selection procedure and the estimation procedure is referred as the *sampling strategy*.

It should be clear that, whether we use a sample for the descriptive purpose or for the analytical purpose, the sample should be selected by using proper statistical methods in order to obtain the results with a desired precision. One key issue in the selection of a sample is to allocate certain non – zero probability to each and every unit of the population of being selected in the sample. This method of obtaining a sample is generally referred as the *Probability or the Random Sampling* otherwise it is referred as the *non – probability sampling* methods.

The probability sampling is usually classified as:

1. Equal Probability Sampling.
2. Unequal Probability Sampling.

1.1.1 The Probability of Selection:

The term *probability of selection* of a unit is defined as "*the probability, allocated to each and every unit of the population, of being selected at any specific draw*". In probability sampling we generally allocate a non – zero probability of selection to each and every unit of the population.

1.1.2 The Probability of Inclusion:

The term *probability of inclusion* of a population unit in the sample is defined as "*the total probability, assigned to a population unit, for being included in the sample in all draws.*" The probability of inclusion of ith unit in the sample is denoted by π_i.

1.1.3 The Design and Model Based Inference in Sampling:

The design based or randomization paradigm uses sample selection probabilities to provide the basis for its inference. For the most part, it is only the first and second order probabilities of the individual population units that are relevant. In the general case the first order inclusion probabilities varies from unit to unit and the second order probabilities from pair to pair. The properties of design – based estimators are defined in terms of their behavior over repeated sampling. In its pure form the design-based inference rests on what may be termed as the *Representative Principle*. Royall (1970) has argued that survey sampling is out of step with statistics as a whole. Statisticians working in other fields use their data to build models and analyzed them in those terms (using model based inference). Survey statisticians, on the other hand, have allowed themselves to be reduced into using an entirely irrelevant source of probability structure not related to the data themselves but only to the manner in which they have been collected. He suggests that in many instances a suitable model for inferential purposes was one for which the classical ratio estimator is optimal. This is one where the survey variable Y is linear homogeneous in an explanatory variable X and the variance function for Y is also linear homogeneous in X. Sometime it is assumed that the values Y_i's have been generated from a super-population model given as:

$$\left.\begin{aligned} &Y_i = \beta Z_i + \varepsilon_i \ \ with \quad E(\varepsilon_i) = 0 \\ &E(\varepsilon_i, \varepsilon_j) = \begin{cases} \sigma_i^2 & i = j \\ 0 & otherwise \end{cases} \\ &\sigma_i^2 = \sigma^2 Z_i^{2\gamma} \qquad \frac{1}{2} \le \gamma \le 1 \end{aligned}\right\} \tag{1.1.1}$$

The sampling methods that use model (1.1.1) as a super-population model for characteristics Y_i's are called the model based sampling methods. In model based sampling an estimator $\hat{Y}$ is said to be model unbiased if:

$$E_M\left(\hat{Y}\right)=E_M\left(Y\right) \tag{1.1.2}$$

where E_M is the expectation taken over all possible populations that can be drawn from model (1.1.1) and Y is total of a specific population. In design based sampling it is assumed that the sample has been drawn from a population by using an appropriate sampling design with probability of a specific sample denoted as $P(s)$. Further, in design based sampling it is assumed that the probability of inclusion of a specific unit, say i, in the sample is $\sum_{S \ni i} P(s)=\pi_i$ and that the joint probability of inclusion of ith and jth unit in the sample is $\sum_{S \ni i,j} P(s)=\pi_{ij}$. In design based sampling an estimator $\hat{Y}$ is defined as design unbiased if and only if:

$$E_D\left(\hat{Y}\right)=\sum_{s}\hat{Y}\,P(s)=Y \tag{1.1.3}$$

where summation is taken over all possible samples of size n that can be drawn from the population by using the sampling design used. It is sometimes useful to use the quantity $E_D\,E_M\left(\hat{Y}-Y\right)^2$ to decide about the performance of an estimator in unequal probability sampling. The quantity $E_D\,E_M\left(\hat{Y}-Y\right)^2$ is generally called the anticipated variance of an estimator $\hat{Y}$ under the super population model (1.1.1). In brief, "in design – based inference, the expectation (and hence also the biases, variances etc) are defined over all possible samples. In model – based inference, they are defined over all possible realizations of the assumed model". The useful references of model (1.1.1) are Cochran (1953), Brewer (1963b), Godambe and Joshi (1965), Rao (1966), Brewer and Hanif (1969a), Royall (1970), Royall and Herson (1973a,1973b), Foreman and Brewer (1971), Vos (1974), Cassel, Sarndal and Wretman (1976), T. G. Rao (1977), Brewer (1979), Hanif and Brewer (1980), Isaki and Fuller (1982), Brewer and Sarndal (1983), Hansen, Madow, and Tepping (1983), Kalton (1983), Sarndal and Wright (1984), Chaudhuri and Vos (1986), Brewer, Hanif and Tam (1988), Smith (1991), Sarndal, Swensson and Wretman (1992) and many others.

1.1.4 Unequal Probability Sampling:

The term *unequal probability sampling* is defined by Marriot (1990) in his *Dictionary of Statistical Terms* as *"A method of selection in which the units are selected with probability proportionate to a given measure related to the characteristics under study is called unequal probability sampling (UPS) or commonly known as the probability proportional to size (PPS) sampling."*

Like the usual *equal probability sampling* designs we can select a sample in *unequal probability sampling* by using either sampling with replacement (*wr*) or by using sampling without replacement (*wor*). In the former case the sampling scheme is called the *unequal probability sampling with replacement* (*upswr*) or the *probability proportional to size sampling with replacement* (*ppswr*) and in the latter case the scheme is called the *unequal probability sampling without replacement or the probability proportional to size sampling without replacement* (*πpswor*). The unequal probability sampling without replacement has attracted a number of survey statisticians towards itself due to its complexities. The importance and complexity of this sort of sampling can be judged from the following words of Prof. Amato Herzel (1986)

> *It is well known that the problems related to sampling without replacement with unequal probabilities have attracted and continue to attract the attention of many scholars and statisticians especially in recent years. It also seems that the argument in question is just of passing interest but the response to an authentic need arising from organizational and administrative necessities mainly in multistage sample surveys and also that the subject present both theoretical and practical.*

1.2 Unequal Probability Sampling With Replacement (PPS Sampling):

The unequal probability sampling with replacement is also called the multinomial sampling, suggested by Hartley and Rao (1962) from the fact that each and every unit has a specified probability of selection and a specific population unit can be selected more then once in a sample.

Basic theory of unequal probability sampling or probability proportional to size sampling with replacement was developed by Hansen and Hurwitz (1943). Prior to that, there had been a substantial development in sampling theory and practice but all these had been used on the assumption that the probability of selection within each stratum would be equal. Hansen and Hurwitz demonstrated the technique of unequal probability sampling via two – stage sampling. They selected the first stage units in usual way. The second stage units were selected with probabilities proportional to a size measure.

This first suggestion for use of unequal probability sampling thus can be associated with the technique of multi – stage sampling with probability proportional to size. Unequal probability sampling can, however, be used in a single stage design and need not necessarily be with probability exactly proportional to size, though some sort of size measure is almost always used as a starting point for assigning selection probabilities.

Smith (1994) had been reported as saying "*In their seminal paper 'On the theory of sampling from a finite population' Hansen and Hurwitz (1943), we were presented not just with a theoretical extension of work of Neyman but with a complete new product.*"

Hansen and Hurwitz (1943) proposed the idea of probability proportional to size sampling with replacement (*ppswr*). One unit was selected at each of the n draws. They allocated the selection probability to i-th unit of the population given by $P_i = \frac{Z_i}{Z}$ were Z_i is the measure of size for i-th population unit and $Z = \sum_{i=1}^{N} Z_i$.

Using the above notations Hansen and Hurwitz proposed the following estimator for population total Y for use with unequal probability sampling with replacement:

$$y'_{HH} = \frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{p_i}. \tag{1.2.1}$$

The Variance of (1.2.1) is given by:

$$\text{Var}\,(y'_{HH}) = \frac{1}{n}\left[\sum_{i=1}^{N}\frac{Y_i}{P_i} - Y^2\right] \tag{1.2.2}$$

$$= \frac{1}{n}\sum_{i=1}^{N}P_i\left(\frac{Y_i}{P_i} - Y\right)^2 \tag{1.2.3}$$

$$= \frac{1}{2n}\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}P_iP_j\left(\frac{Y_i}{P_i} - \frac{Y_j}{P_j}\right)^2 \quad \text{[Raj (1954)]} \tag{1.2.4}$$

$$= \frac{1}{n}\sum_{i=1}^{N}\frac{1}{P_i}(Y_i - P_iY_i)^2 \quad \text{[Beg and Hanif (1991)]} \tag{1.2.5}$$

Possible unbiased variance estimators of $Var(y'_{HH})$ are:

$$\text{var}(y'_{HH}) = \frac{1}{n(n-1)}\sum_{i=1}^{n}\left(\frac{y_i}{p_i} - y'_{HH}\right)^2 \tag{1.2.6}$$

and

$$\text{var}(y'_{HH}) = \frac{1}{2n(n-1)}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}\left(\frac{y_i}{p_i} - \frac{y_j}{p_j}\right)^2 \quad \text{[Raj (1954)]} \tag{1.2.7}$$

1.3 Unequal Probability Sampling Without Replacement (πpswor) Using the Horvitz and Thompson Estimator

The concept of unequal probability sampling without replacement was first used by Madow (1949) when he used the systematic sampling with probability proportional to size and avoid the chances of any unit being selected more than once. Madow does not provide any theoretical framework for unequal probability sampling without replacement. Narian (1951) provided a selection procedure but without any theoretical framework.

Horvitz and Thompson (1952) were the first to provide a complete theoretical framework for unequal probability sampling without replacement when they suggested the following estimator of population total for use with unequal probability sampling without replacement:

$$y'_{HT} = \sum_{i \in S} \frac{Y_i}{\pi_i} \tag{1.3.1}$$

The Horvitz and Thompson estimator enjoys following properties:

(i) It is the only unbiased estimator of the class in which same weight is attached to a particular population unit whenever it is selected (Horvitz and Thompson – 1952).

(ii) It is admissible in the class of all homogeneous linear unbiased estimators of population total Y; that is, there does not exist any member of that class which has a smaller variance than y'_{HT} for all Y_I. (Roy and Chakravarti – 1960).

(iii) If the Y_i *are all exactly proportional to the corresponding* π_i and the sample size is fixed, the variance of y'_{HT} is zero. This is the property usually associated with the ratio estimator and will be referred to as the *ratio estimator property* (Brewer – 1963a).

(iv) Under the model (1.1.1) the expected variance of the Horvitz and Thompson estimator achieves lower bound of the expected variance for any sample design – unbiased estimator given by Godambe and Joshi (1965) as:

$$E_M E_D \left(y'_{HT} - Y\right)^2 \geq \sum_{i=1}^{N} \sigma_i^2 \left(\frac{1}{\pi_i} - 1\right) \tag{1.3.2}$$

Horvitz and Thomson developed the following variance formula for y'_{HT} :

$$\operatorname{Var}\left(y'_{HT}\right) = \sum_{i=1}^{N} \frac{1-\pi_i}{\pi_i} Y_i^2 + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j} Y_i Y_j \tag{1.3.3}$$

An alternative expression for the variance of y'_{HT}, developed independently by Sen (1953) and by Yates and Grundy (1953), is given as:

$$\mathrm{Var}_{SYG}\left(y'_{HT}\right)=\frac{1}{2}\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}\left(\pi_i\pi_j-\pi_{ij}\right)\left(\frac{Y_i}{\pi_i}-\frac{Y_j}{\pi_j}\right)^2 \qquad (1.3.4)$$

Horvitz and Thompson derived following unbiased estimator of (1.3.3) :

$$var_{HT}\left(y'_{HT}\right)=\sum_{i=1}^{n}\frac{1-\pi_i}{\pi_i^2}y_i^2+\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}\frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j\pi_{ij}}y_iy_j \qquad (1.3.5)$$

The estimator of (1.3.4) proposed by Sen (1953) and independently by Yates and Grundy (1953) is given by:

$$var_{SYG}\left(y'_{HT}\right)=\frac{1}{2}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}\left(\frac{\pi_i\pi_j-\pi_{ij}}{\pi_{ij}}\right)\left(\frac{y_i}{\pi_i}-\frac{y_j}{\pi_j}\right)^2 \qquad (1.3.6)$$

The estimators given in (1.3.5) and (1.3.6) are unbiased under the condition $\pi_{ij}\neq 0$ for all $j\neq i$ and therefore are *conditionally unbiased estimators*. Both (1.3.5) and (1.3.6) can assume negative values, but (1.3.6) rarely seems to assume negative values. Expression (1.3.6) has also performed much better than (1.3.5) in a number of empirical comparisons, starting from those of Yates and Grundy's (1953). Brewer and Hanif (1969a), and Vijayan (1975) showed that for n = 2, it is the only possible non – negative variance estimator. Sen (1953) also compared the efficiency of (1.3.5) and (1.3.6) taking a population of five units and selecting all possible samples of two. He demonstrated that the expression (1.3.6) is positive for all samples, whereas (1.3.5) assume negative value for some samples. He further showed that for n = 2 and $\pi_{ij}>\pi_i\pi_j$ for all $i\neq j$, (1.3.5) is always positive when the selection is made without replacement and the Horvitz – Thompson estimator is used. Raj (1956a) proved further that the expression (1.3.6) is always positive. Rao (1963a) also proved that under the Midzuno (1951) and Yates – Grundy (1953) selection procedures for $\pi pswor$ that (1.3.6) was always positive.

Rao and Singh (1973) used Brewer's (1963a) selection procedure to compare (1.3.5) and (1.3.6) for the case n = 2, employing a wide variety of populations. Their empirical evidence also indicates that (1.3.6) is more stable than (1.3.5). A similar result was given by Brewer and Hanif (1969a), Lanke (1974) by using the Hajek's (1964) Method – I and by Shahbaz (2001).

The variance expression given in (1.3.3) and (1.3.4) and the variance estimators given in (1.3.5) and (1.3.6) require the calculation of joint inclusion probabilities (π_{ij}) and therefore they are very difficult to apply as the calculation of π_{ij} became cumbersome as the sample size increases. Attempts have been made to approximate the variance of Horvitz and Thompson estimator such that it does not involve the joint inclusion probabilities, the π_{ij}'s. A simple approximation to π_{ij} in terms of π_i's and π_j's for selection procedures that ensure $\pi_i = 2p_i$ is given by Brewer (1963a), Durbin (1967), Rao (1965) and Sampford (1967) as:

$$\pi_{ij} = \frac{\pi_i \pi_j}{2 + \sum_{k=1}^{N} \frac{\pi_k}{1-\pi_k}} \left[\frac{1}{1-\pi_i} + \frac{1}{1-\pi_j} \right] \qquad (1.3.7)$$

Brewer and Hanif (1983) gave two approximations to π_{ij}'s in terms of π_i's and π_j's. The first approximation given by Brewer and Hanif (1983) is:

$$\pi_{ij} = A\pi_i \pi_j + B(\pi_i + \pi_j) + C(\pi_i^2 + \pi_j^2) \qquad (1.3.8)$$

with $A = \frac{n^2}{n^2 - \sum_{j=1}^{N} \pi_j^2}$, $B = \frac{-n \sum_{j=1}^{N} \pi_j^2}{(N-2)\left[n^2 - \sum_{j=1}^{N} \pi_j^2\right]}$ and $C = \frac{n^2}{(N-2)\left[n^2 - \sum_{j=1}^{N} \pi_j^2\right]}$

Also A>1, B<0, $B = (-nA) \Big/ \left[(N-2)\sum_{i=1}^{N} \pi_i^2\right]$ and $C = A/(N-2)$. The second approximation given by Brewer and Hanif (1983) is much simpler than the first approximation. This approximation is given as:

$$\pi_{ij} = \frac{(n-1)\sum_{r=0}^{\infty} \pi_i^{2^r} \pi_j^{2^r}}{\prod_{i=0}^{r} \sum_{K=1}^{N} \pi_k^{2^i}} \qquad (1.3.9)$$

The approximation (1.3.9) performs reasonably well even when one or two values of π_i are close to unity, each term being less than half the preceding one. Brewer and Hanif further showed that the approximation (1.3.9) may not result in a feasible set of π_{ij} when two of π_i are close to unity. Herzel (1986) suggested another approximation for π_{ij}. This approximation is given as:

$$\pi_{ij} = \pi_i \pi_j - \frac{\pi_i(1-\pi_i)+\pi_j(1-\pi_j)}{N-2} + \frac{n-\sum_{k=1}^{N}\pi_k^2}{(N-1)(N-2)} + \ldots\ldots \quad (1.3.10)$$

The approximation (1.3.10) may produce negative values of π_{ij}'s. An example of this is N = 4, n = 2, π_i = 0.2, 0.25, 0.75 and 0.8, Hanif (1994). Hanif and Ahmad (2001) proposed another approximation to π_{ij}. This approximation is given as:

$$\pi_{ij} = \left(\frac{a_i + a_j}{2}\right)\pi_i \pi_j \quad (1.3.11)$$

where a_i and a_j are appropriately chosen. Hanif and Ahmad (2001) showed that using $a_i = a_j = \frac{(n-1)}{n-\pi_i}$ in (1.3.11) an approximate formula for variance of Horvitz and Thompson estimator can be given as:

$$\mathrm{Var}\left(y'_{HT}\right) \approx \sum_{i=1}^{N}\pi_i\left(1-\frac{n-1}{n-\pi_i}\pi_i\right)\left\{\frac{Y_i}{\pi_i}-\frac{Y}{n}\right\}^2 \quad (1.3.12)$$

The systematic sampling procedures have also been used to obtain an approximate expression for variance of Horvitz – Thompson estimator. Hartley and Rao (1962) derived the following expression under the random systematic method:

$$\mathrm{Var}\left(y'_{HT}\right) \approx \sum_{i=1}^{N}\pi_i\left(1-\frac{n-1}{n}\pi_i\right)\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2 - \frac{n-1}{n^2}\sum_{i=1}^{N}\left(2\pi_i^3 - \frac{\pi_i^2}{2}\sum_{j=1}^{N}\pi_j^2\right)\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2$$

$$+\frac{2(n-1)}{n^3}\left(\sum_{i=1}^{N}\pi_i Y_i - \frac{Y}{n}\sum_{j=1}^{N}\pi_j^2\right)^2 \quad (1.3.13)$$

The expression (1.3.13) is correct to order N^0. Rao (1963a) further showed that the asymptotic variance formula to order N^0 for a sample of size 2 is given as:

$$\mathrm{Var}\left(y'_{HT}\right)=\sum_{i=1}^{N}\pi_i\left(1-\frac{\pi_i}{2}\right)\left(\frac{Y_i}{\pi_i}-\frac{Y}{2}\right)^2-\frac{1}{2}\sum_{i=1}^{N}\left(\pi_i^3-\frac{\pi_i^2}{4}\sum_{j=1}^{N}\pi_j^2\right)\left(\frac{Y_i}{\pi_i}-\frac{Y}{2}\right)^2$$

$$+\lambda\left(\sum_{i=1}^{N}\pi_i Y_i-\frac{Y}{2}\sum_{j=1}^{N}\pi_j^2\right)^2 \qquad (1.3.14)$$

The value of λ in (1.3.14) is 3/32 for Narain's (1951) procedure, 1/8 for Carroll – Hartley (1964) rejective procedure and 1/4 for the Random Systematic procedure (1950). Rao (1965) further showed that λ = 0 for the Brewer (1963a) selection procedure. Since Rao – Sampford, Rao (1965) and Sampford (1967), procedure, Durbin (1967) draw-by-draw procedure and the Brewer (1963a) procedure are in same equivalence class therefore λ = 0 for all these procedures. Rao (1963a) further showed that the approximate formula to the order N^1 for a sample of size n is:

$$\mathrm{Var}\left(y'_{HT}\right)\approx\sum_{i=1}^{N}\pi_i\left(1-\frac{n-1}{n}\pi_i\right)\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2 \qquad (1.3.15)$$

Also an unbiased estimator of (1.3.15) to order N^1 is:

$$\mathrm{var}\left(y'_{HT}\right)=\frac{1}{2n^2(n-1)}\sum_{i=1}\sum_{\substack{j=1\\j\neq i}}^{n}\left[1-\left(p_i+p_j\right)+n\sum_{k=1}^{N}\pi_k^2\right]\left(\frac{y_i}{\pi_i}-\frac{y_j}{\pi_j}\right)^2 \qquad (1.3.16)$$

The estimators given in (1.3.5) and (1.3.6) are design based variance estimators of (1.3.3) and (1.3.4). Another approximate design based estimator for (1.3.4) proposed by Jessen (1978) is given as:

$$\mathrm{var}_J\left(y'_{HT}\right)=\frac{n-\sum_{i=1}^{N}\pi_i^2}{2n(n-1)}\sum_{i=1}\sum_{\substack{j=1\\j\neq i}}^{n}\left(\frac{y_i}{\pi_i}-\frac{y_j}{\pi_j}\right)^2 \qquad (1.3.17)$$

The variance estimators given in (1.3.5) and (1.3.6) are design based estimators. A model based variance estimator of $Var\left(y'_{HT}\right)$, proposed by Hanif and Brewer (1980), under the model (1.1.1) is given as:

$$\text{var}_{HB}\left(y'_{HT}\right)=\frac{n}{n-1}\left[1-\frac{\sum_{i=1}^{N}\pi_i^{2\gamma}}{\sum_{i=1}^{N}\pi_i^{2\gamma-1}}\right]\sum_{i=1}^{n}\left(\frac{y_i}{\pi_i}-\frac{y'_{HT}}{n}\right)^2 \qquad (1.3.18)$$

It is interesting to note that the estimator given in (1.3.18) reduces to the estimator given in (1.3.17) for $\gamma = 1$, therefore Jessen estimator is a special case of the Hanif and Brewer (1980) estimator, Hanif et. al. (1993). Another model based variance estimator proposed by Kumar, Gupta and Agarwal (1985) is given as:

$$var_{KGA}\left(y'_{HT}\right)=\frac{\sum_{i=1}^{N}P_i^{2\gamma-1}\left(1-nP_i\right)}{2(n-1)\sum_{i=1}^{N}P_i^{2\gamma-1}}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\left(\frac{y_i}{\pi_i}-\frac{y_j}{\pi_j}\right)^2 \qquad (1.3.19)$$

Hanif et. al. (1993) verified that the Kumar – Gupta – Agarwal estimator is identical to the Hanif – Brewer estimator. Hanif et. al. (1993) has also proposed following model based variance estimator under the model (1.1.1):

$$\text{var}_N\left(y'_{HT}\right)=\frac{1}{n-1}\sum_{i=1}^{N}P_i^{2\gamma-1}\left(1-nP_i\right)\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\left[\frac{\left(\frac{Y_i}{\pi_i}-\frac{Y_j}{\pi_j}\right)^2}{P_i^{2(\gamma-1)}+P_j^{2(\gamma-1)}}\right] \qquad (1.3.20)$$

The estimator (1.3.20) is model – unbiased and hence it is design – model unbiased. However, the estimator (1.3.20) is design biased but the bias decreases as the sample size increases, Hanif et. al. (1993).

Samiuddin and Asad (1981) proposed a sampling strategy for use with the estimator (1.3.1) following the lines of Rao, Hartley and Cochran (1962). The difference between the strategy of Samiuddin – Asad and that of Rao, Hartley and Cochran is that the former divide a population into (n+1) groups or blocks and selects a sample of n groups from these. Samiuddin and Asad showed that the joint probability of inclusion of ith and jth unit in the sample can be written as:

$$\pi_{ij}=\frac{\pi_i\,\pi_j\left(P_R+P_T-1\right)}{P_R\,P_T} \qquad (1.3.21)$$

where P_R and P_T are the probabilities of selection of block R and T respectively. Also π_i and π_j are probabilities of inclusion of ith and jth units respectively. Samiuddin and Asad obtained the following variance expression of Horvitz–Thompson estimator under (1.3.21):

$$V ar\left(y'_{HT}\right)=\sum_{R=1}^{n+1}\left\{\sum_{i\in R}\frac{Y_i^2}{\pi_i}-\frac{Y_R^2}{P_R}\right\}+\left\{\sum_{R=1}^{n+1}\frac{1-P_R}{P_R}Y_R^2+\sum_{\substack{R=1\\R\neq T}}^{n+1}\sum_{T=1}\frac{(1-P_R)(1-P_T)}{P_R\,P_T}Y_R\,Y_T\right\} \quad (1.3.22)$$

It should be noted that for above sampling scheme a number of $\pi_{ij}'s=0$, therefore a unbiased estimator of (1.3.22) does not exist and approximate estimators are suggested.

Hajek (1964) introduced the idea of poisson sampling in which a sequence of Bernoulli trials is conducted in order to ensure the inclusion of a unit in the sample. Hajek (1964) shows that the sample size in this method is not fixed. In poisson sampling there is a chance of an empty sample. Ogus and Clark (1971) modified the poisson sampling design in order to avoid the chances of an empty sample. Ogus and Clark (1971) showed that the chances of an empty sample can be avoided if we continue the poisson sampling method until a non – empty sample is not selected. Ogus and Clark (1971) called this sampling design the modified poisson sampling design. Brewer et. al. (1984) showed that the sampling variance of Horvitz – Thompson estimator under the poisson or modified poisson sampling can be put in a much simpler form as:

$$\text{Var}\left(y'_{HT}\right)=\sum_{i=1}^{N}(1-\pi_i)\frac{Y_i^2}{\pi_i} \quad (1.3.21)$$

An unbiased estimator of (1.3.21) is given as:

$$\text{var}\left(y'_{HT}\right)=\sum_{S}(1-\pi_i)\frac{y_i^2}{\pi_i} \quad (1.3.22)$$

Brewer et. al. (1972) and Brewer et. al. (1984) proposed another sampling design that can be used with the Horvitz – Thompson estimator. This sampling design is called the collocated sampling and this method has a much smaller sampling variance as compared to the poisson or modified poisson sampling design. Brewer et. al. (1984) showed that the sampling variance of Horvitz – Thompson estimator under this sampling design is given as:

$$\operatorname{Var}\left(y'_{HT}\right)=\sum_{i=1}^{N}\frac{Y_i^2}{\pi_i}-\frac{N}{N-1}\sum_{i=1}^{N}Y_i^2+\frac{1}{N-1}\left[Y^2-2\sum_{i=1}^{N}\frac{Y_i}{\pi_i}\sum_{j=1}^{i-1}Y_j\right] \quad (1.3.23)$$

1.4 Description of Selected Selection Procedures:

A number of selection procedures are available in the literature that can be used with the Horvitz – Thompson estimator in unequal probability sampling without replacement. These selection procedures have their own advantages and disadvantages. Some of these procedures impose rigorous restrictions on initial probabilities of selection whereas some of these methods require a number of iterations to evaluate probabilities of inclusion and the joint probabilities of inclusion. Some of the developed selection procedures are somewhat simpler in application as they produce a compact formula for evaluation of inclusion probability and joint inclusion probabilities but they are applicable to a sample of size 2 only, see for example Brewer and Hanif (1983) for description of some fifty selection procedures along with their classification on the basis of the sample selection method. Following selection procedures have been widely used in real life surveys as they produce compact formulae for evaluation of π_i and π_{ij}.

1.4.1 Sen – Midzuno Procedure:

This selection procedure is reported by Horvitz and Thompson (1952) and is applicable for a sample of any size. This selection procedure is stated as:

- Select first unit with probability q_i
- Select a sample of size n – 1 from remaining units with equal probability and without replacement.

The quantities π_i and π_{ij} for this selection procedure are given as:

$$\pi_i=q_i+\frac{n-1}{N-1}(1-q_i) \quad with \quad \sum_{i=1}^{N}q_i=1 \quad (1.4.1)$$

$$\pi_{ij}=\frac{n-1}{N-1}\left[\frac{N-n}{N-2}(q_i+q_j)+\frac{n-2}{N-2}\right] \quad (1.4.2)$$

where q_i are revised selection probabilities.

1.4.2 Yates – Grundy Draw–by–Draw Procedure:

This selection procedure was developed by Yates and Grundy (1953) and also reported by Durbin (1953a) and Hajek (1964). This selection procedure is stated as:

- Select first unit with probability proportional to size.
- Select second unit with probability proportional to size of remaining units.

The quantities π_i and π_{ij} for this selection procedure are given as:

$$\pi_i = p_i\left[1+\sum_{j=1}^{N}\frac{p_j}{1-p_j}-\frac{p_i}{1-p_i}\right] \tag{1.4.3}$$

$$\pi_{ij} = p_i p_j\left[\frac{1}{1-p_i}+\frac{1}{1-p_j}\right] \tag{1.4.4}$$

This procedure can be considered as one of the simplest procedure as it does not impose any sort of restrictions on initial probabilities of selection and final probabilities of inclusion.

1.4.3 Brewer Selection Procedure:

Brewer (1963a) developed this selection procedure for use with unequal probability sampling without replacement. The selection procedure of Brewer is strictly without replacement procedure. The selection procedure is given as:

- Select first unit with probability proportional to $\frac{p_i(1-p_i)}{(1-2p_i)}$
- Select the second unit with probability proportional to size of the remaining units.

The probability of inclusion π_i and joint probability of inclusion π_{ij} for this selection procedure is given as:

$$\pi_i = 2p_i \tag{1.4.5}$$

$$\pi_{ij} = \frac{2p_i p_j}{k}\left[\frac{1}{1-2p_i}+\frac{1}{1-2p_j}\right] \tag{1.4.6}$$

with $$k = 1+\sum_{j=1}^{N}\frac{p_j}{1-2p_j} \tag{1.4.7}$$

1.4.4 Durbin's Draw–by–Draw Procedure:

This selection procedure was proposed by Durbin (1967). This procedure uses the idea of revised probabilities. This procedure is a draw – by – draw procedure, as the procedure of Brewer, and is therefore strictly without replacement procedure. This procedure is stated as:

- Select first unit with probability proportional to size.
- Select second unit with probability proportional to $p_j\left[\frac{1}{1-2p_i}+\frac{1}{1-2p_j}\right]$

The probability of inclusion and joint probability of inclusion for this selection procedure is same as that for the Brewer (1963a) selection procedure.

1.4.5 Rao – Sampford Selection Procedure:

This selection procedure is a rejective selection procedure developed independently by Rao (1965b) and by Sampford (1967). This selection procedure is stated as:

- Select one element with a revised probability q_i and with replacement.
- Select one element with probability proportional to size.
- Repeat above two steps if same unit is selected twice.

This procedure also produces same probability of inclusion and joint probability of inclusion as the procedures of Brewer (1963a) and Durbin (1967).

1.4.6 Prabu – Ajgonkar Selection Procedure:

This selection procedure was developed by Deshpande, Prabu and Ajgonkar (1982) and uses sampling with replacement at successive draws. This selection procedure is stated as:

- Select first unit with probability proportional to size and with replacement.
- Select second element with probability $\frac{p_i(1-p_i)}{A(1-2p_i)}$ where A is a normalizing constant.
- If same unit is selected twice, then select one more element, from remainder of the population, with probability proportional to size.

The probability of inclusion for this procedure is same as the of Brewer (1963a) method. The joint probability of inclusion for this selection procedure is given as:

$$\pi_{ij} = \frac{p_i p_j}{A}\left[\frac{1}{1-2p_i} + \frac{1}{1-2p_j}\right] \qquad (1.4.8)$$

1.4.7 Durbin's Rejective Procedure:

This selection procedure was proposed by Durbin (1953) and is stated as:

- Select two units with probabilities p_i and with replacement.
- Repeat first step if same elements are selected until two distinct elements turn up.

The probability of inclusion and joint probability of inclusion for this selection procedure are given as:

$$\pi_i = \frac{2p_i(1-p_i)}{1-\sum_{i=1}^{N} p_i^2} \qquad (1.4.9)$$

$$\pi_{ij} = \frac{2p_i p_j}{1-\sum_{i=1}^{N} p_i^2} \qquad (1.4.10)$$

Durbin uses following estimator for estimation of population total under this selection procedure:

$$y_D = \frac{1}{2}\left[\frac{y_i}{p_i} + \frac{y_j}{p_j}\right] \qquad (1.4.11)$$

The estimator given in (1.4.11) is a biased estimator of population total, but the bias is generally negligible as compared to mean squared error.

1.4.8 Yates – Grundy Rejective Procedure:

This selection procedure was developed by Yates and Grundy (1953) and can be applicable to a sample of any size. For a sample of size 2 this procedure is given as:

- Select two units with probability proportional to size and with replacement.
- Repeat first step if same unit is selected twice.

For this selection procedure the probability of inclusion and joint probability of inclusion are given as:

$$\pi_i = \frac{2p_i(1-p_i)}{1-\sum_{j=1}^{N} p_j^2} \tag{1.4.12}$$

$$\pi_{ij} = \frac{2p_i p_j}{1-\sum_{j=1}^{N} p_j^2} \tag{1.4.13}$$

These selection procedures have been applied to a number of artificial and natural populations for evaluation of population variance of Horvitz – Thompson estimator in chapter 3 and 4.

1.5 Unequal Probability Sampling Without Replacement (πpswor) Using Special Estimators:

Das (1951), Raj (1956a), Murthy (1957) and Rao – Hartley – Cochran (1962) developed special estimators for use with unequal probability sampling without replacement.

1.5.1 Das's Estimator:

Das (1951) proposed the following estimator for use with unequal probability sampling without replacement:

$$t = \sum_{r=1}^{n} c_r t_r , \tag{1.5.1}$$

where c_r are arbitrary numbers with $\sum_{r=1}^{n} c_r = 1$ and

$$t_r = \frac{(1-p_1)(1-p_1-p_2)......(1-p_1-p_2-...-p_{r-1})}{(N-1)......(N-r+1)p_1......p_r} y_r \tag{1.5.2}$$

There is a defect in Das estimator that its variance estimator may produce negative value in some cases.

1.5.2 Raj's Estimator:

Raj (1956a) proposed a series of ordered estimators that can be used with unequal probability sampling without replacement. The estimator proposed by Raj (1956a) has the general form as:

$$t_{mean} = \frac{1}{n}\sum_{r=1}^{n} t_r \, , \tag{1.5.3}$$

where

$$t_1 = \frac{y_1}{p_1} \text{ and } t_r = \sum_{i=1}^{r-1} y_i + \frac{y_r}{p_r}\left(1 - \sum_{i=1}^{r-1} p_i\right) for\ r > 1 \tag{1.5.4}$$

The Raj's estimator t_{mean} for a sample of size two is:

$$t_{mean} = \frac{1}{2}\left[\frac{y_i}{p_i}(1+p_i) + \frac{y_j}{p_j}(1-p_i)\right] \tag{1.5.5}$$

The variance of (1.5.5) is:

$$\mathrm{V}ar(t_{mean}) = \frac{1}{8}\sum_{i=1}^{N}\sum_{\substack{j=1 \\ j\neq i}} P_i P_j (2 - P_i - P_j)\left(\frac{Y_i}{P_i} - \frac{Y_j}{P_j}\right)^2 \tag{1.5.6}$$

Pathak (1967a) derived a formula for (1.5.3) for a sample of any size. This variance expression is given as:

$$\mathrm{V}ar(t_{mean}) = \frac{1}{2n^2}\sum_{i=1}^{N}\sum_{\substack{j=1 \\ j\neq i}} P_i P_j \left[1 + \sum_{r=2}^{n} Q_{ij}(r-1)\right]\left(\frac{Y_i}{P_i} - \frac{Y_j}{P_j}\right)^2 \tag{1.5.7}$$

where $Q_{ij}(r-1)$ denotes the probability of non – inclusion of one or both of the units i and j in the first (r – 1) sample units. An unbiased variance estimator proposed by Raj (1956a) for a sample of any size is given as:

$$\mathrm{var}(t_{mean}) = \frac{1}{n(n-1)}\sum_{k=1}^{n}(t_k - \bar{t})^2 \tag{1.5.8}$$

where $\bar{t} = \frac{1}{n}\sum_{k=1}^{n} t_k$. An unbiased estimator for (1.5.6) for n = 2 is:

$$\mathrm{var}(t_{mean}) = \frac{(1-p_i)^2}{4}\left(\frac{y_i}{p_i} - \frac{y_j}{p_j}\right)^2 \tag{1.5.9}$$

The estimators (1.5.8) and (1.5.9) are non-negative for all p_i's.

1.5.3 Murthy's Estimator:

Murthy (1957) symmetries the Raj (1956a) estimator by using the Rao – Blackwell theorem to produce following unordered estimator of population total:

$$t_{symm} = \frac{1}{P(s)} \sum_{i=1}^{n} P(s|i) y_i \qquad (1.5.10)$$

where $P(s|i)$ is the probability of obtaining a sample "s" given that ith unit has been already selected and $P(s)$ is the probability of obtaining a sample "s". Murthy estimator for a sample of size two is given as:

$$t_{symm} = \frac{1}{2 - p_i - p_j} \left[\frac{y_i}{p_i}(1 - p_j) + \frac{y_j}{p_j}(1 - p_i) \right] \qquad (1.5.11)$$

The variance of (1.5.11) is:

$$Var(t_{symm}) = \frac{1}{2} \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}} \frac{P_i P_j (1 - P_i - P_j)}{2 - P_i - P_j} \cdot \left(\frac{Y_i}{P_i} - \frac{Y_j}{P_j} \right)^2 \qquad (1.5.12)$$

An unbiased variance estimator for n = 2 is:

$$\text{var}(t_{symm}) = \frac{(1 - p_i)(1 - p_j)}{2} \left(\frac{y_i}{p_i} - \frac{y_j}{p_j} \right)^2 \qquad (1.5.13)$$

Murthy showed that the estimator given in (1.5.11) always performs better than the estimator given in (1.5.5).

Pathak (1967a) derived the variance expression for Murthy (1957) estimator for a sample of size n. This expression is:

$$Var(t_{symm}) = \frac{1}{2} \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}} P_i P_j \left[1 - \sum_{S \ni i j} \frac{P(s|i) P(s|j)}{P(s)} \right] \left(\frac{Y_i}{P_i} - \frac{Y_j}{P_j} \right)^2 \qquad (1.5.14)$$

An unbiased variance estimator of (1.5.14) given by Pathak (1967b) is:

$$\text{var}(t_{symm}) = \frac{1}{2} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}} p_i p_j \left[\frac{p(s) p(s|ij) - p(s|i) p(s|j)}{\{p(s)\}^2} \right] \left(\frac{y_i}{p_i} - \frac{y_j}{p_j} \right)^2 \qquad (1.5.15)$$

where $p(s|ij)$ denotes the conditional probability of selecting a sample "s", given that units *i* and *j* were selected in that order at the first two draws.

A Generalized Murthy estimator has also been proposed which has the form:

$$y'_{GM} = \frac{1}{P(s)} \sum_{i \in s} P(s|i) y_i \tag{1.5.16}$$

where $P(s|i)$ and $P(s)$ are defined above.

Samiuddin et. el. (1992) has shown that the Generalized Murthy estimator y'_{GM} is both design and model unbiased. Further the estimator y'_{GM} achieves the Godambe – Joshi lower bound if and only if $\frac{P(s|i)}{P(s)} = \frac{1}{\pi_i}$, Samiuddin et. al. (1992).

1.5.4 Rao – Hartley – Cochran Estimator:

Rao – Hartley – Cochran (1962) proposed sampling strategy for use with unequal probability sampling and the estimator of population total. Their estimator is:

$$y'_{RHC} = \sum_{i=1}^{n} \frac{\pi_i \, y_{iT}}{p_{iT}} \tag{1.5.17}$$

where p_{iT} is the probability of T-th unit selected from the i-th group. Also $\pi_i = \sum_{T=1}^{N_i} p_{iT}$ *and* $\sum_{i=1}^{n} \pi_i = 1$. The Rao – Hartley – Cochran estimator can be used for any sample size. The variance of (1.5.17) is:

$$Var(y'_{RHC}) = \frac{n\left(\sum_{i=1}^{n} N_i^2 - N\right)}{N(N-1)} \cdot \left[\sum_{i=1}^{n} \sum_{T=1}^{N_i} \frac{Y_{iT}^2}{n P_{iT}} - \frac{Y^2}{n}\right] \tag{1.5.18}$$

Rao, Hartley and Cochran (1962) further showed that, since the population size can be written as N = n R + k, where 0<k<n and R is a positive integer, the variance given in (1.5.18) can be written as:

$$Var(y'_{RHC}) = \left[1 - \frac{n-1}{N-1} + \frac{k(n-k)}{N(N-1)}\right]\left[\sum_{i=1}^{n} \sum_{T=1}^{N_i} \frac{Y_{iT}^2}{n P_{iT}} - \frac{Y^2}{n}\right] \tag{1.5.19}$$

Further, if N is an exact multiple of n, then k = 0 and (1.5.19) becomes:

$$Var(y'_{RHC}) = \left[1 - \frac{n-1}{N-1}\right]\left[\sum_{i=1}^{n} \sum_{T=1}^{N_i} \frac{Y_{iT}^2}{n P_{iT}} - \frac{Y^2}{n}\right] \tag{1.5.20}$$

An unbiased variance estimator of (1.5.18) is:

$$\operatorname{var}\left(y'_{RHC}\right)=\frac{\left(\sum_{i=1}^{n}N_i^2-N\right)}{\left(N^2-\sum_{i=1}^{n}N_i^2\right)}\cdot\sum_{i=1}^{n}\pi_i\left(\frac{y_{it}}{p_{it}}-y'_{RHC}\right)^2 \qquad (1.5.21)$$

The unbiased variance estimators for (1.5.19) and (1.5.20) are:

$$var\left(y'_{RHC}\right)=\frac{N^2+k(n-k)-Rn}{N^2(n-1)-k(n-k)}\cdot\sum_{i=1}^{n}\pi_i\left(\frac{y_{it}}{p_{it}}-y'_{RHC}\right)^2 \qquad (1.5.22)$$

and $$var\left(y'_{RHC}\right)=\frac{1}{(n-1)}\left(1-\frac{n}{N}\right)\cdot\sum_{i=1}^{n}\pi_i\left(\frac{y_{it}}{p_{it}}-y'_{RHC}\right)^2 \qquad (1.5.23)$$

1.5.5 Basu's Estimator:

Basu (1971) suggested that it was natural to estimate the ratio:

$$\left[\sum_{i=1}^{N}Y_i-\sum_{i=1}^{n}y_i\right]\Big/\left[\sum_{i=1}^{N}P_i-\sum_{i=1}^{n}p_i\right] \qquad (1.5.24)$$

by some sort of average of the observed ratio. Two particular averages which he suggested were $\sum_{i=1}^{n}y_i\Big/\sum_{i=1}^{n}p_i$, which led to the conventional ratio estimator, and $\frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{p_i}$ which led to the predictive estimator:

$$y'_{BASU}=\sum_{i=1}^{n}y_i+\frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{p_i}\left[1-\sum_{j=1}^{n}p_j\right] \qquad (1.5.25)$$

He claimed that these two estimators had as much face validity as unordered forms of individual Raj estimator, and though they were not unbiased, they were far to calculate. His argument for face validity appears to be based on their being symmetric function of sample values and possessing the ratio estimator property.

1.6 The Model Based Inference in Sampling:

The model based sampling has been in application from early thirties when Smith, H. F. (1938) introduced the idea in agricultural surveys. Cochran (1953), Brewer (1963b), Royall and Herson (1973). Cassel, Sarndal and Wretman (1976) has also given the idea of model based sampling inference. The model (1.1.1) has been popularly used in the model based sampling. Godambe – Joshi obtained the lower bound to the variance of any estimator for which $\pi_i \infty \sigma_i$. The lower bound is given in (1.3.2). Since the emergence of model based sampling inference a number of estimators have been proposed for use in unequal probability sampling.

1.6.1 Horvitz – Thompson Ratio Estimator:

This estimator has been proposed by Brewer (1963b) and reproduced by Hajek (1971). The estimator is:

$$y'_{HTR} = \frac{\sum_{i \in S} Y_i \pi_i^{-1}}{\sum_{i \in S} Z_i \pi_i^{-1}} \cdot Z \tag{1.6.1}$$

This estimator has Horvitz – Thompson estimator in numerator and in denominator. The mean square error for estimator given in (1.6.1) has been derived by Brewer and Hanif (1983) and is given as:

$$MSE\left(y'_{HTR}\right) \doteq Var\left(y'_{HT}\right) + \frac{Y^2}{Z^2} Var\left(z'_{HT}\right) - 2\frac{Y}{Z} Cov\left(y'_{HT}, z'_{HT}\right) \tag{1.6.2}$$

where $\mathrm{Var}\left(y'_{HT}\right)$ is variance of Horvitz – Thompson estimator for Y, $\mathrm{Var}\left(z'_{HT}\right)$ is variance of Horvitz – Thompson estimator for Z and $\mathrm{Cov}\left(y'_{HT}, z'_{HT}\right)$ is covariance of Horvitz – Thompson estimator between variables Y and Z.

1.6.2 Brewer's Ratio Estimator:

Brewer (1979) proposed a model based estimator of population total for use with unequal probability sampling without replacement. The estimator proposed by Brewer is given as:

$$y'_B = \frac{\sum_{i \in S} Y_i \left(\pi_i^{-1} - 1\right)}{\sum_{i \in S} Z_i \left(\pi_i^{-1} - 1\right)} \sum_{i \notin S} Z_i \tag{1.6.3}$$

The estimator proposed by Brewer is appropriate for large samples and hence the asymptotic theory is appropriate. The estimator is also appropriate if sample is a significant proportion of the population provided that the finite population correction factor is allowed. The asymptotic variance of y_B' under the model (1.3.2), for large N and n, may achieve Godambe – Joshi lower bound given in (1.3.3).

Hanif (1994) has shown that when $\pi_i = \frac{nZ_i}{Z}$ the estimator (1.6.3) reduces to the Horvitz – Thompson estimator. Further, Hanif (1994) has shown that when $\pi_i = \frac{n}{N}$ the estimator (1.6.3) reduces to usual ratio estimator.

1.6.3 The Generalized Regression Estimator:

A generalized regression estimator proposed by Cassel, Sarandal and Wretman (1976) is given as:

$$y_{CSW}' = \sum_{i \in S} \left(Y_i - \hat{\beta}_{CSW} Z_i\right)\pi_i^{-1} + \hat{\beta}_{CSW} Z, \tag{1.6.5}$$

where

$$\hat{\beta}_{CSW} = \frac{\sum_{i \in S} Y_i Z_i \pi_i^{-2}}{\sum_{i \in S} Z_i^2 \pi_i^{-2}} \tag{1.6.6}$$

A more general class of regression estimators, also proposed by Cassel, Sarandal and Wretman, is given as:

$$y_{CSW}' = y_{HT}' + \hat{\beta}\left(Z - z_{HT}'\right) \tag{1.6.7}$$

where y_{HT}' and z_{HT}' are Horvitz – Thompson estimators for variables Y and Z respectively.

Samiuddin et. al. (1992) while discussing a general check on estimators and with other forms of regression estimates. Some of the numerical evidence by them shows that these are surprisingly close to actual Cassel, Sarandal and Wretman (1976) estimate.

1.7 Chronological Development:

The theory of unequal probability sampling was first introduced in early 1940's when Hansen and Hurwitz (1943) provided the basic theory for unequal probability sampling with replacement. The general theory of unequal probability sampling without replacement was developed by Horvitz and Thomson (1952). Since that time a large number of developments have been made in the field of unequal probability sampling with and without replacement. These developments can be classified as:

i. The early period up to 1950.
ii. From 1951 to 1960.
iii. From 1961 to 1970.
iv. From 1971 to 1980.
v. From 1981 to 1990.
vi. From 1991 onward.

1.7.1 Early Period up to 1950:

The theory of unequal probability sampling has its emergence since the time when Neyman (1934) produced his historical paper on theory of sampling. Hansen and Hurwitz (1943) first used the theory of unequal probability sampling in the early 1940's in two stage sampling. The procedure used by them was an unequal probability sampling with replacement procedure. They showed that the technique developed by them can bring a substantial decease in the sampling variance as compared with simple random sampling if an additional probability indicator bench marks variable, closely related with the variable under study, is used. After the development made by Hansen and Hurwitz (1943), Madow (1949) uses the idea of unequal probability sampling in *order systematic sampling* to avoid the danger of a unit being selected. This method has a draw back that π_{ij} for a number of samples under this selection procedure are zero. Goodman and Kish (1950) suggested *random systematic sampling* in unequal probability sampling without replacement.

1.7.2 From 1951 to 1960:

In this period a number of new selection procedures and estimators were developed in the field of unequal probability sampling.

Narain (1951) initiated the work in this period when he developed his selection procedure for use with unequal probability sampling. The selection procedure developed by him was a *whole sample* procedure. He also introduced the idea of *working probabilities* to obtain the values of π_i and π_{ij} but he could not provide rigorous theory for unequal probability sampling without replacement. Midzuno (1951), Lahiri (1951) and Ikeda and Midzuno (1952) continued the work of development when they developed two more selection procedures. There selection procedures enjoy the property that they make the classical ratio estimator unbiased. The main turning point in the theory of unequal probability sampling without replacement came in 1952 when Horvitz and Thompson (1952) provided basic theory of estimation and developed their own estimator to estimate the population total. Horvitz and Thompson (1952) derived the variance and variance estimator for their estimator. A draw back of the variance estimator developed by Horvitz and Thompson is that it can assume negative value for some of the samples. Yates and Grundy (1953) developed two selection procedures; one procedure was a draw – by – draw procedure whereas the other one was a rejective procedure. Yates and Grundy (1953) and Sen (1953) derived the variance expression and variance estimator of Horvitz and Thompson estimator. Sen (1953) compared the efficiency of two famous estimators of variance of Horvitz – Thompson estimator and showed that the Sen – Yates – Grundy estimator is more stable as compared with the Horvitz – Thompson variance estimator. Raj (1956a) carried out similar study by using two famous selection procedures. Durbin (1953a) developed another selection procedure by using idea of *rejective sampling*. He used the usual Hansen and Hurwitz estimator with his selection procedure and showed that the said estimator is biased under his selection procedure. Grundy (1954) developed a systematic procedure also.

The development continued in the following years Durbin (1953a), Durbin (1953b), Raj (1956b), Stevens (1958) are the other main contributions. The method developed by Durbin (1956b) is based on the criterion of grouping the population. The method of Raj (1956b) is based on the minimization of population variance of the Horvitz and Thompson estimator on the assumption that the values of Yi has been generated from the model

$$Y_i = \alpha + \beta Z_i \tag{1.7.1}$$

The method of Stevens (1958) combines both equal and unequal probability sampling. Stevens (1958) uses the method of dividing a population into groups and then selecting a group with probability proportional to the aggregate of group size. Finally he selected a simple random sample from the selected group to achieve the ultimate sample.

This period also witnessed the development of some famous estimators for use with unequal probability sampling without replacement to estimate the population total and mean. Das (1951) derived his estimator by using a draw – by – draw procedure. Raj (1956a) developed a series of unbiased estimators following the Yates – Grundy draw – by – draw procedure. The estimators derived by Raj were based on the ordered samples. Murthy (1957) uses the idea of sufficiency to improve Raj's (1956a) estimator. In this process Raj's (1956a) estimator is symmetrized. He developed new estimator which was more efficient than the Raj (1956a) estimator. Deming (1960) gave a systematic method to minimize the sampling variance of the estimator.

1.7.3 From 1961 to 1970:

This period may be considered as one of two decades in which unequal probability sampling flourished tremendously.

The work began in 1961 when Rao (1961) produced some estimator to find the sampling variance of Horvitz – Thompson estimator. Rao (1961) also developed a selection procedure for use with the Horvitz – Thompson estimator. Hartley and Rao (1962) developed an approximate expression for variance of Horvitz – Thompson estimator under the random systematic procedure for a sample of size n and for a sample of size 2. The approximation was true for order N^0. Rao (1963b) further obtained the approximation to order N^1. Rao (1963a) compared the stability of Horvitz – Thompson and Sen – Yates – Grundy variance estimator under Midzuno (1951) and Yates – Grundy (1953) selection procedures and proved that the later is always positive under these selection procedures. He also obtained a variance estimator for the resulting approximate variance expression. Hanurav (1962a) gave four selection procedures. All of these procedures were based on the draw – by – draw method. The chief feature of these procedures was that the sample size was not fixed in these procedures. Hanurav (1966 and 1967) developed three more selection procedures for use with the Horvitz – Thompson estimator. Rao – Hartley – Cochran (1962) developed a new estimator which was based on the random division of population into number of groups and then selecting one element from each group. The method of Rao – Hartley – Cochran uses the Hansen – Hurwitz estimator with a sample size one in each group. Chikkagoudar (1967) proposed his estimator that was an extension of the Rao – Hartley – Cochran estimator. Chikkagoundar suggested that the population should be divided into "k" groups of appropriate sizes and a sample of size n_i should be selected from each group using the Midzuno (1951) strategy. Brewer (1963a), Rao (1965), Durbin (1967) and Sampford (1967) developed their separate procedure. The common thing in all of these methods is that they all produce same value of inclusion probability and joint probabilities of inclusion for n = 2. The method of Sampford (1967) is an extension of the Rao (1965) method. Durbin (1967) produced another procedure which was based on the idea of grouped sampling. Fellegi (1963) introduced another method which uses both actual and working probabilities to obtain the probability of inclusion and joint probability of inclusion in the sample. The method developed by Fellegi is applicable for a sample of any size but he shows that the applicability is very cumbersome as the sample size increases. Carroll and Hartley (1964) proposed three selection procedures. One of these procedures was a draw – by – draw procedure; another one is the rejective procedure whereas the third one is the whole sample procedure. They also proposed that the revised probabilities should be chosen such that the inclusion probability should be equal to np_i. Vijayan (1968) generalizes the methods of Hanurav to develop his selection procedure for use with unequal probability sampling without replacement. Brewer and Hanif (1969a) produced a variance estimator optimization procedure. This procedure is developed under the consideration to optimize the estimator of variance of Horvitz and Thompson estimator. This method actually fixes the quantities π_{ij}'s in a sample such that the resulting values of π_i and π_{ij} optimizes the stability of Sen – Yates – Grundy variance estimator. Brewer and Hanif (1969a) further showed that the Sen – Yates – Grundy variance estimator is the only non – negative variance estimator for a sample of size 2. Hajek (1964a) developed five selection procedures. First of these methods was a *poisson sampling method*; that is the method was such that there was no fixed sample size and the binomial trials are carried out to ensure the inclusion of a unit in the sample. Hajek also showed that the joint probability of inclusion for any two units with this procedure is simply the product of respective inclusion probabilities of that set of units. The other four procedures were an approximation to the Carroll and Hartley (1964) rejective procedure. Main difference among four methods of Hajek and method of Carroll and Hartley was that Hajek used his own set of working probabilities. Hajek termed his four methods as Hajek's methods I to IV. Pathak (1967a , 1967b) derived the variance expressions for Raj and Murthy estimators for sample of any size. Connor (1966) derived an exact formula for joint probability of inclusion of any two units in the sample in unequal probability sampling. However the evaluation of this formula for any pair of units involves adding contributions from all possible combinations of units separating the two in the pair. This can become tedious for large N. Nevertheless Connor's formula does make the estimation of variance for random systematic procedure more amenable to computer programming than it is for other systematic procedures. Jessen (1969) proposed his four methods. Jessen used the idea of probability decrement to obtain the probabilities of inclusion and the joint probabilities of inclusion. The methods developed by Jessen are somewhat equal in their application and with a slight modification of one procedure Jessen developed the other one. Rao and Bayless (1969) conducted an empirical study to compare the stability of various variance estimators in unequal probability sampling for a sample of size 2. Bayless and Rao (1970) extended the study for a sample of size n = 3 and n = 4. They found that the Sen–Yates–Grundy variance estimator is stable whereas the Horvitz–Thompson variance estimator is not much stable in unequal probability sampling without replacement. Sankaranarayanan (1969) proposed an inclusion probability proportional to size sampling design by using the Lahiri's method of selection.

1.7.4 From 1971 to 1980:

Ogus and Clark (1971) gave the idea of *modified poisson sampling.* The method developed by Ogus and Clark (1971) ensures that there is no chance of an empty selection that was possible for simple *poisson sampling.* Basu (1971) wrote his historical paper, which lead many survey statisticians in development of new methods that can be used either with equal probability sampling and unequal probability sampling. Fuller (1971) proposed his method which selects the units in such a way that the quantities π_{ij}'s are approximately proportional to $\pi_i\pi_j$. Dodds and Fryer (1971) generalizes the results of Brewer (1963a) and Durbin (1967) when he developed a general selection procedure for use with unequal probability sampling. Dodds and Fryer (1971) actually used working probabilities for selection at both the draws. They also developed the parameterized Brewer (1963a) method but they did not study the said procedure completely. Mukhopadhyay (1972) gave another selection procedure which was applicable for any n but this method has a very serious draw back that it can not be applied with a considerable ease. Another reason for its deficiency is that Sinha (1973) proposed two selection procedures which superseded the method of Mukhopadhyay (1972) in that these methods were easy to apply. Sinha termed his two methods as *extension* and *reduction* procedures. The reason for these two names is that in extension procedure Sinha used the idea of full sample space that can constitute a feasible set of π_i, π_j and π_{ij}. In reduction procedure Sinha reduced the said sample space considerably. Das and Mohanty (1973) also proposed another selection procedure that can be used with unequal probability sampling without replacement and for any sample size. The method developed by Das and Mohanty can be termed as a whole sample procedure as it lists all possible samples of specified size and selects a random sample with equal probability. Chief feature of this procedure is its simplicity and simple evaluation of joint probabilities of inclusion; and hence evaluation of Sen – Yates – Grundy variance estimator. Rao and Singh (1973) studied the Horvitz – Thompson (1952) and Sen – Yates – Grundy (1953) variance estimators by using the Brewer (1963a) selection procedure under a variety of populations. They found that the former variance estimator assume negative values for some samples whereas the later does not assume negative value for any of the sample and hence is stable. Lanke (1974) carried out similar study by using the Hajek (1964b) Method – I and obtain the same conclusion. Hanif (1974) developed an asymptotic variance formula for use with unequal probability sampling without replacement. Vijayan (1975) showed that the Sen – Yates – Grundy (1953) variance estimator is the only non – negative variance estimator for a sample of size 2. Bebbington (1975) introduced the concept of *list sequential sampling* that uses equal probability of selection at first draw and varying probabilities at the subsequent draws to obtain a sample of any size. Brewer (1975) generalizes his selection procedure for any n. However the joint inclusion probabilities are very tedious to calculate as they require the recursive relation that involves the conditional second order probabilities in all subsequent draws given every one of the possible combination of draws in the preceding draws. Chaudhuri (1976) proposed another draw – by – draw procedure. This procedure is applicable to any sample size. Also, this procedure uses a standard selection procedure which is applicable to a sample of size two and which ensures that $\pi_i = 2P_i$. This procedure actually combines unequal probability sampling with equal probability sampling. Singh (1978) developed another selection procedure which selects a sample of size n in two steps. This procedure applies some cumbersome conditions on initial probabilities of selection. Choudhry (1979) proposed a selection procedure that is applicable to any sample size. This method uses Yates – Grundy draw – by – draw procedure for a sample of size (n – 1) and uses a set of working probabilities to select last unit from the sample; in order to make overall inclusion probability proportional to size. Chromy (1979) developed another selection procedure, which keeps history of the previous draws also. This procedure also uses the idea of conditional probabilities for any population unit to be included in the sample. Hanif and Brewer (1980) also derived the model based variance estimator for variance of Horvitz and Thompson estimator and showed that Jessen (1978) variance estimator is a special case of their estimator.

This period also saw substantial development in the model based sampling inference. Royall (1971) uses the linear regression model in sample survey designs to improve their efficiencies. Royall and Herson (1973a, 1973b) proposed the idea of robust estimation under a linear stochastic model. Brewer (1979) also uses the idea of robustness in sample survey designs. He also introduces a model based estimator of population total for use with unequal probability sampling without replacement. Cassel, Sarandal and Wretman (1976) gave some results for the generalized difference and regression estimators under the super population model. Sarandal (1978) compare the design – based and model – based sampling inference. Chaudhuri (1979) compare some sampling strategies under a super-population model. Also, Chaudhury and Arnab (1979a) compare the efficiency of various sampling strategies under a super-population model.

Some other notable references in this period are of Deshpande (1977), Agrwal and Goel (1977), Sunter (1977), Jessen (1978), Choudhry and Singh (1979), Agarwal (1979) and Singh and Srivastava (1980).

1.7.5 From 1981 to 1990:

Brewer, Early and Hanif (1984) derived the expression for variance and variance estimators for *poisson* and *modified poisson sampling.* and introduced the idea of *collocated sampling* and derive the variance expression of Horvitz – Thompson estimator under this sampling design. The main difference in *poisson sampling* and *collocated sampling* is that in the former the random variable used in the Bernoulli trials are considered as uniformly distributed whereas in the latter same random variable is treated as uniformly spaced. Chaudhuri (1981) and Chaudhuri (1985) developed two more selection procedures for use with unequal probability sampling without replacement. Deshpande (1982a) also developed two more methods in addition to his previous two selection procedures. The methods developed by Deshpande uses sampling with replacement at second draw. The probability of selection is taken proportional to the measure of size at first draw and proportional to a suitable measure at the subsequent draws. Samiuddin and Asad (1981) developed their selection procedure and estimator for use with unequal probability sampling without replacement and for any sample size on the lines of Rao – Hartley and Cochran (1962) method but they were unable to obtain an unbiased variance estimator. Agarwal, Kumar and Dey (1982) developed two more selection procedures along with the procedures of Gupta, Nigam and Kumar (1982) and Gupta, Nigam and Kumar (1984). Agarwal, Singh and Sing (1984) uses systematic sampling with varying probabilities to reduce the sampling variance of systematic sampling. Srivastava and Singh (1981) increase the number of selection procedures by developing their selection procedure along the selection procedure of Sengupta (1981). Hansen, Hurwitz and Tepping (1983) evaluate the model dependent and probability sampling inference in sample surveys. Brewer and Hanif (1983) tried to approximate the variance of Horvitz – Thompson estimator when they proposed two approximations for joint inclusion probabilities in terms of marginal probabilities of inclusion. Another attempt was made by Herzel (1986) when he gave his approximation for the joint probabilities of inclusion in terms of marginal probabilities of inclusion. Kumar, Gupta and Agarwal (1985) derived the model based variance estimator for variance of Horvitz – Thompson estimator and it was also established that their variance estimator is identical to the Hanif – Brewer (1980) variance estimator. Hanif, Beg and Khawaja (1990) has given a comprehensive bibliography of the selection procedures that can be used with the Horvitz – Thompson estimator. Their list contain about 120 selection procedures.

The development for selection procedures were continued in this decade some worth mentioning references are of Chao (1982), Kumar and Agarwal (1982), Nigam and Gupta (1984), Kumar, Gupta and Nigam (1987), Kumar, Srivastava and Agarwal (1986), Saxena, Singh and Srivastava (1986), Sunter (1986), Agarwal, Singh and Singh (1984), Gupta (1989), Gupta, Nigam and Kumar (1984), Kumar and Agarwal (1985), Kumar (1987), Agarwal and Singh (1986), Dey and Srivastava (1987), Kumar, Kuthuria and Agarwal (1984) and Gabbler and Schweigkoffer (1990).

1.7.6 From 1991 Onward:

Samiuddin and Kattan (1991) initiated the work in this decade by developing their selection procedure. Beg and Hanif (1991) compare sampling with and without replacement in case of unequal probability sampling. Further developments continued in this field with the work of Gupta, Kumar and Nigam (1992). Sahoo, Sahoo and Mohanty (1993) continued the quest of bringing refinements in the field of unequal probability sampling when they introduced their own selection procedure and increased the tally of selection procedures for use with unequal probability sampling without replacement. Samiuddin et. al. (1992) give a detailed summary of model and design based sampling inference. They have also proposed the generalized Murthy estimator and have shown that this estimator is both design and model unbiased and also achieves the Godambe – Joshi lower bound. Hanif et. al. (1993) have compared various variance estimators for Horvitz – Thompson estimator and proved that the estimators developed by Hanif and Brewer (1980) and by Kumar, Gupta and Agarwal (1985) are both model unbiased. Hanif et. al. (1994) have also given a general class of Murthy estimator. Tracy and Mukhopadhyay (1994) obtained a variance estimator for Horvitz – Thompson estimator under the Midzuno strategy. Gazali (1996, 1998) modified the Jessen methods 2 and 3 to produce two more selection procedures. The methods developed by Gazali have a larger applicability as compared with the Jessen's methods. Aires (2000) have compared conditional poisson and pareto sampling designs. Hanif and Ahmad (2001) tried to obtain a solution for joint probabilities of inclusion in terms of marginal probabilities of inclusion in order to simplify the computations for variance and variance estimator of Horvitz – Thompson estimator but a feasible solution has yet to be produced.

Chapter 2

First Order Inclusion Probabilities

2.1 Introduction:

In this chapter some approximate expressions for the variance of Horvitz – Thompson estimator has been derived by using only the first order inclusion probabilities, that is, the probabilities that depend on single unit only. The estimator of population total proposed by Horvitz and Thompson (1952) is given as:

$$y'_{HT} = \sum_{i \in S} \frac{y_i}{\pi_i} \tag{1.3.1}$$

The sampling variance of the estimator given in equation (1.3.1) has two popular forms given by Horvitz and Thompson (1952) and independently by Sen (1953) and Yates and Grundy (1953) as in (1.3.9) and (1.3.10).

Both the variance expressions given in equations (1.3.9) and (1.3.10) require the calculations of joint probabilities of inclusion, π_{ij}. These probabilities are very complicated to evaluate when sample size increases from 2. Many efforts have been made to approximate these joint inclusion probabilities via some suitable relation with the marginal inclusion probabilities. Some of the most famous references for these approximations have been given in the previous chapter. In this chapter some more approximations for joint inclusion probabilities in terms of the marginal inclusion probabilities have been obtained. These results have been given in the following pages.

2.2 An Alternative Expression for Variance of Horvitz – Thompson Estimator:

In this section an alternative expression for the variance of Horvitz-Thompson estimator has been derived. This expression is very useful to obtain a suitable approximation for the variance using only the first order inclusion probabilities.

Consider

$$\mathrm{Var}_{SYG}\left(y'_{HT}\right)=\frac{1}{2}\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}\left(\pi_i\,\pi_j-\pi_{ij}\right)\left(\frac{Y_i}{\pi_i}-\frac{Y_j}{\pi_j}\right)^2$$

$$=\frac{1}{2}\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}\left(\pi_i\,\pi_j-\pi_{ij}\right)\left[\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)-\left(\frac{Y_j}{\pi_j}-\frac{Y}{n}\right)\right]^2,\ \text{where}\ \ Y=\sum_{i=1}^{N}Y_i$$

$$=\frac{1}{2}\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}\left(\pi_i\,\pi_j-\pi_{ij}\right)\left[\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2+\left(\frac{Y_j}{\pi_j}-\frac{Y}{n}\right)^2-2\left(\frac{Y_i}{\pi_j}-\frac{Y}{n}\right)\left(\frac{Y_j}{\pi_j}-\frac{Y}{n}\right)\right]$$

$$=\frac{1}{2}\left[\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}\left(\pi_i\,\pi_j-\pi_{ij}\right)\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2+\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}\left(\pi_i\,\pi_j-\pi_{ij}\right)\right.$$

$$\left.\left(\frac{Y_j}{\pi_j}-\frac{Y}{n}\right)^2-2\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}\left(\pi_i\,\pi_j-\pi_{ij}\right)\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)\left(\frac{Y_j}{\pi_i}-\frac{Y}{n}\right)\right]$$

$$=\frac{1}{2}\left[\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}\left\{\pi_i\,\pi_j\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2-\pi_{ij}\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2\right\}\right.$$

$$+\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}\left\{\pi_i\,\pi_j\left(\frac{Y_j}{\pi_j}-\frac{Y}{n}\right)^2-\pi_{ij}\left(\frac{Y_j}{\pi_j}-\frac{Y}{n}\right)^2\right\}$$

$$\left.-2\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}\left(\pi_i\,\pi_j-\pi_{ij}\right)\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)\left(\frac{Y_j}{\pi_j}-\frac{Y}{n}\right)\right]$$

$$=\frac{1}{2}\left[\sum_{i=1}^{N}\pi_i\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2\sum_{\substack{j=1\\ j\neq i}}^{N}\pi_j-\sum_{i=1}^{N}\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2\sum_{\substack{j=1\\ j\neq i}}^{N}\pi_{ij}+\sum_{i=1}^{N}\pi_i\right.$$

$$\sum_{\substack{j=1\\j\neq i}}^{N}\pi_j\left(\frac{Y_j}{\pi_j}-\frac{Y}{n}\right)^2-\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}\pi_{ij}\left(\frac{Y_j}{\pi_j}-\frac{Y}{n}\right)^2$$

$$\left.-2\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}\left(\pi_i\pi_j-\pi_{ij}\right)\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)\left(\frac{Y_j}{\pi_j}-\frac{Y}{n}\right)\right]$$

$$=\frac{1}{2}\left[\sum_{i=1}^{N}\pi_i\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2\left(n-\pi_i\right)-\sum_{i=1}^{N}\left(n\pi_i-\pi_i\right)\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2\right.$$

$$\sum_{j=1}^{N}\pi_j\left(\frac{Y_j}{\pi_j}-\frac{Y}{n}\right)^2\left(n-\pi_j\right)-\sum_{j=1}^{N}\left(n\pi_j-\pi_j\right)\left(\frac{Y_j}{\pi_j}-\frac{Y}{n}\right)^2$$

$$\left.-2\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}\left(\pi_i\pi_j-\pi_{ij}\right)\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)\left(\frac{Y_j}{\pi_j}-\frac{Y}{n}\right)\right]$$

$$=\sum_{j=1}^{N}\pi_j\left(\frac{Y_j}{\pi_j}-\frac{Y}{n}\right)^2\left(n-\pi_j\right)-\sum_{j=1}^{N}\left(n\pi_j-\pi_j\right)\left(\frac{Y_j}{\pi_j}-\frac{Y}{n}\right)^2$$

$$-\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}\left(\pi_i\pi_j-\pi_{ij}\right)\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)\left(\frac{Y_j}{\pi_j}-\frac{Y}{n}\right)$$

$$=\sum_{i=1}^{N}n\pi_i\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2-\sum_{i=1}^{N}\pi_i^2\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2-\sum_{i=1}^{N}n\pi_i\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2$$

$$+\sum_{i=1}^{N}\pi_i\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2-\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}\left(\pi_i\pi_j-\pi_{ij}\right)\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)\left(\frac{Y_j}{\pi_j}-\frac{Y}{n}\right)$$

$$=\sum_{i=1}^{N}\pi_i\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2-\sum_{i=1}^{N}\pi_i^2\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2-\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}\left(\pi_i\pi_j-\pi_{ij}\right)\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)\left(\frac{Y_j}{\pi_j}-\frac{Y}{n}\right)$$

$$=\sum_{i=1}^{N}\pi_i\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2-\sum_{i=1}^{N}\pi_i^2\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2$$

$$+\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}\left(\pi_{ij}-\pi_i\pi_j\right)\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)\left(\frac{Y_j}{\pi_j}-\frac{Y}{n}\right) \qquad (2.2.1)$$

In equation (2.2.1) only the last term contains the joint probability of inclusion, π_{ij}. An approximate result for the variance of Horvitz – Thompson estimator may be obtained by manipulating only the last term appropriately. Some more approximations have been obtained as:

2.3 Approximations for π_{ij}

The first approximation that has been suggested is somewhat similar to one given by Hanif and Ahmad (2001). To obtain the approximation consider the third term of equation (2.2.1):

$$\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}\left(\pi_{ij}-\pi_i\pi_j\right)\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)\left(\frac{Y_j}{\pi_j}-\frac{Y}{n}\right) \qquad (2.3.1)$$

Now substituting $\pi_{ij}=a_i\pi_i\pi_j+a_j\pi_i\pi_j$ in (2.3.1):

$$=\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}\left(a_i\pi_i\pi_j+a_j\pi_i\pi_j-\pi_i\pi_j\right)\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)\left(\frac{Y_j}{\pi_j}-\frac{Y}{n}\right)$$

$$=\sum_{i=1}^{N}a_i\pi_i\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)\sum_{\substack{j=1\\ j\neq i}}^{N}\pi_j\left(\frac{Y_j}{\pi_j}-\frac{Y}{n}\right)+\sum_{j=1}^{N}a_j\pi_j\left(\frac{Y_j}{\pi_j}-\frac{Y}{n}\right)$$

$$\sum_{\substack{i=1\\ i\neq j}}^{N}\pi_i\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)-\sum_{i=1}^{N}\pi_i\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)\sum_{\substack{j=1\\ j\neq i}}^{N}\pi_j\left(\frac{Y_j}{\pi_j}-\frac{Y}{n}\right)$$

$$=\sum_{i=1}^{N}a_i\pi_i\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)\left[\sum_{j=1}^{N}\pi_j\left(\frac{Y_j}{\pi_j}-\frac{Y}{n}\right)-\pi_i\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)\right]$$

$$+\sum_{i=1}^{N}a_j\pi_j\left(\frac{Y_j}{\pi_j}-\frac{Y}{n}\right)\left[\sum_{j=1}^{N}\pi_i\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)-\pi_j\left(\frac{Y_j}{\pi_j}-\frac{Y}{n}\right)\right]$$

$$-\sum_{i=1}^{N}\pi_i\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)\left[\sum_{j=1}^{N}\pi_j\left(\frac{Y_j}{\pi_j}-\frac{Y}{n}\right)-\pi_i\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)\right]$$

$$= -\sum_{i=1}^{N} a_i\,\pi_i^2\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2 - \sum_{j=1}^{N} a_j\,\pi_j^2\left(\frac{Y_j}{\pi_j}-\frac{Y}{n}\right)^2 + \sum_{i=1}^{N}\pi_i^2\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2$$

$$= -2\sum_{i=1}^{N} a_i\,\pi_i^2\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2 + \sum_{i=1}^{N}\pi_i^2\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2 \qquad (2.3.2)$$

Substituting (2.3.2) in (2.2.1):

$$\mathrm{Var}_{SYG}\left(y'_{HT}\right) \approx \sum_{i=1}^{N}\pi_i\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2 - 2\sum_{i=1}^{N} a_i\,\pi_i^2\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2$$

$$\approx \sum_{i=1}^{N}\left(\pi_i - 2a_i\,\pi_i^2\right)\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2$$

$$\approx \sum_{i=1}^{N}\pi_i\left(1-2a_i\,\pi_i\right)\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2 \qquad (2.3.3)$$

Now, the values of a_i can be chosen accordingly to obtain a reasonable approximation for the variance of Horvitz – Thompson estimator.

The second approximation is relatively simpler than the first one. For this approximation again consider the (2.3.1)

$$\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}\left(\pi_{ij}-\pi_i\,\pi_j\right)\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)\left(\frac{Y_j}{\pi_j}-\frac{Y}{n}\right) \qquad (2.3.1)$$

Now, substituting $\pi_{ij} = a_i\,\pi_j + a_j\,\pi_i$ in (2.3.1):

$$= \sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}\left(a_i\,\pi_j + a_j\,\pi_i - \pi_i\,\pi_j\right)\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)\left(\frac{Y_j}{\pi_j}-\frac{Y}{n}\right)$$

$$= \sum_{i=1}^{N} a_i\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)\sum_{\substack{j=1\\ j\neq i}}^{N}\pi_j\left(\frac{Y_j}{\pi_j}-\frac{Y}{n}\right) + \sum_{j=1}^{N} a_j\left(\frac{Y_j}{\pi_j}-\frac{Y}{n}\right)$$

$$\sum_{\substack{i=1\\ i\neq j}}^{N}\pi_i\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right) - \sum_{i=1}^{N}\pi_i\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)\sum_{\substack{j=1\\ j\neq i}}^{N}\pi_j\left(\frac{Y_j}{\pi_j}-\frac{Y}{n}\right)$$

$$= \sum_{i=1}^{N} a_i\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)\left[\sum_{j=1}^{N}\pi_j\left(\frac{Y_j}{\pi_j}-\frac{Y}{n}\right)-\pi_i\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)\right]$$

$$+\sum_{i=1}^{N} a_j\left(\frac{Y_j}{\pi_j}-\frac{Y}{n}\right)\left[\sum_{j=1}^{N}\pi_i\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)-\pi_j\left(\frac{Y_j}{\pi_j}-\frac{Y}{n}\right)\right]$$

$$-\sum_{i=1}^{N} \pi_i\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)\left[\sum_{j=1}^{N}\pi_j\left(\frac{Y_j}{\pi_j}-\frac{Y}{n}\right)-\pi_i\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)\right]$$

$$= -\sum_{i=1}^{N} a_i\,\pi_i\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2-\sum_{j=1}^{N} a_j\,\pi_j\left(\frac{Y_j}{\pi_j}-\frac{Y}{n}\right)^2+\sum_{i=1}^{N}\pi_i^2\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2$$

$$= -2\sum_{i=1}^{N} a_i\,\pi_i\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2+\sum_{i=1}^{N}\pi_i^2\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2 \qquad (2.3.4)$$

Substituting (2.3.4) in equation (2.2.1) another approximation for variance of Horvitz – Thompson estimator is:

$$\mathrm{Var}_{SYG}\left(y'_{HT}\right)\approx \sum_{i=1}^{N}\pi_i\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2-\sum_{i=1}^{N}\pi_i^2\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2$$

$$-2\sum_{i=1}^{N} a_i\,\pi_i\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2+\sum_{i=1}^{N}\pi_i^2\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2$$

$$\approx \sum_{i=1}^{N}\pi_i\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2-2\sum_{i=1}^{N} a_i\,\pi_i\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2$$

$$\approx \sum_{i=1}^{N}\left(\pi_i-2a_i\,\pi_i\right)\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2$$

$$\approx \sum_{i=1}^{N}\pi_i\left(1-2a_i\right)\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2 \qquad (2.3.5)$$

This approximation seems to be reasonably simpler as compared to (2.3.3).

For the third approximation again consider (2.3.1):

$$\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}\left(\pi_{ij}-\pi_i\,\pi_j\right)\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)\left(\frac{Y_j}{\pi_j}-\frac{Y}{n}\right) \qquad (2.3.1)$$

Substituting $\pi_{ij} = a\ \pi_i\ \pi_j$ in (2.3.1):

$$= \sum_{i=1}^{N}\sum_{\substack{j=1 \\ j \neq i}} \left(a\,\pi_i\,\pi_j - \pi_i\,\pi_j\right)\left(\frac{Y_i}{\pi_i} - \frac{Y}{n}\right)\left(\frac{Y_j}{\pi_j} - \frac{Y}{n}\right)$$

$$= (a-1)\sum_{i=1}^{N} \pi_i \left(\frac{Y_i}{\pi_i} - \frac{Y}{n}\right)\sum_{\substack{j=1 \\ j \neq i}}^{N} \pi_j \left(\frac{Y_j}{\pi_j} - \frac{Y}{n}\right)$$

$$= \sum_{i=1}^{N} \pi_i\,(a-1)\left(\frac{Y_i}{\pi_i} - \frac{Y}{n}\right)\left[\sum_{j=1}^{N} \pi_j \left(\frac{Y_j}{\pi_j} - \frac{Y}{n}\right) - \pi_i \left(\frac{Y_i}{\pi_i} - \frac{Y}{n}\right)\right]$$

$$= -(a-1)\sum_{i=1}^{N} \pi_i^2 \left(\frac{Y_i}{\pi_i} - \frac{Y}{n}\right)^2 \qquad (2.3.6)$$

Substituting (2.3.6) in (2.2.1) following approximate expression for variance of Horvitz – Thompson estimator is obtained:

$$\mathrm{Var}_{\mathrm{SYG}}\left(y'_{\mathrm{HT}}\right) \approx \sum_{i=1}^{N} \pi_i \left(\frac{Y_i}{\pi_i} - \frac{Y}{n}\right)^2 - \sum_{i=1}^{N} \pi_i^2 \left(\frac{Y_i}{\pi_i} - \frac{Y}{n}\right)^2 - \sum_{i=1}^{N} \pi_i^2\,(a-1)\left(\frac{Y_i}{\pi_i} - \frac{Y}{n}\right)^2$$

$$\approx \sum_{i=1}^{N} \left[\pi_i - \pi_i^2 - (a-1)\pi_i^2\right]\left(\frac{Y_i}{\pi_i} - \frac{Y}{n}\right)^2$$

$$\approx \sum_{i=1}^{N} \left[\pi_i - \pi_i^2 - a\ \pi_i^2 + \pi_i^2\right]\left(\frac{Y_i}{\pi_i} - \frac{Y}{n}\right)^2$$

$$\approx \sum_{i=1}^{N} \pi_i\,(1 - a\,\pi_i)\left(\frac{Y_i}{\pi_i} - \frac{Y}{n}\right)^2 \qquad (2.3.7)$$

This is a special case of the one given by Hartley and Rao (1962) for $a = \frac{n-1}{n}$ and is as:

$$\mathrm{Var}_{\mathrm{SYG}}\left(y'_{\mathrm{HT}}\right) \approx \sum_{i=1}^{N} \pi_i \left(1 - \frac{n-1}{n}\pi_i\right)\left(\frac{Y_i}{\pi_i} - \frac{Y}{n}\right)^2 \qquad (2.3.8)$$

Another approximation has been obtained. For this approximation again consider (2.3.1):

$$\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}\left(\pi_{ij}-\pi_i\pi_j\right)\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)\left(\frac{Y_j}{\pi_j}-\frac{Y}{n}\right) \qquad (2.3.1)$$

Substituting $\pi_{ij}=a\pi_i+b\pi_j$ in (2.3.1):

$$=\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}\left(a\pi_i+b\pi_j-\pi_i\pi_j\right)\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)\left(\frac{Y_j}{\pi_j}-\frac{Y}{n}\right)$$

$$=-\sum_{i=1}^{N}(a+b)\pi_i\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2+\sum_{i=1}^{N}\pi_i^2\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2 \qquad (2.3.9)$$

Substituting (2.3.9) in (2.2.1) following approximation for variance of Horvitz – Thompson estimator is obtained.

$$\text{Var}_{\text{SYG}}\left(y'_{\text{HT}}\right)\approx\sum_{i=1}^{N}\pi_i\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2-\sum_{i=1}^{N}\pi_i^2\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2-\sum_{i=1}^{N}(a+b)\pi_i\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2$$

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

$$\approx\sum_{i=1}^{N}\pi_i\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2-\sum_{i=1}^{N}(a+b)\pi_i\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2$$

$$\approx(1-a-b)\sum_{i=1}^{N}\pi_i\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2 \qquad (2.3.10)$$

This approximation require suitable values of "a" and "b" for a close approximation of true variance of Horvitz – Thompson estimator. It is also interesting to note that expression given in (2.3.10) transforms to one given by Hartley and Rao (1962) for $a=b=\frac{n-1}{2n}$.

Chapter 3

The New Selection Procedures

3.1 Introduction:

In this chapter some new selection procedures have been developed for use with the Horvitz and Thompson estimator. These selection procedures have been developed with the view of minimization of sampling variance of Horvitz and Thompson estimator. Some important results regarding the inclusion probabilities and joint inclusion probabilities have also been derived. The derivation of these procedures along with verification of results has been given in the following sections.

3.2 The New Selection Procedures:

In this section some new selection procedures have been derived that can be used with the Horvitz and Thompson estimator by obtaining the expressions for inclusion probability of ith unit in the sample and the joint inclusion probability for ith and jth unit in the sample.

3.2.1 The New Selection Procedure – I:

The New selection procedure – I is stated as:

- Select first unit with probability proportional to $\frac{p_i}{1-2p_i}$
- Select second unit with probability proportional to size of the remaining units.

The probability of inclusion for ith unit (π_i) in the sample for this selection procedure is obtained as under:

$$\pi_i = \frac{\frac{p_i}{1-2p_i}}{\sum_{i=1}^{N}\frac{p_i}{1-2p_i}} + \sum_{\substack{j=1\\ j\neq i}}^{N}\frac{\frac{p_j}{1-2p_j}}{\sum_{j=1}^{N}\frac{p_j}{1-2p_j}}\cdot\frac{p_i}{1-p_j}$$

$$= \frac{1}{\sum_{i=1}^{N} \frac{p_i}{1-2p_i}} \left[\frac{p_i}{1-2p_i} + \sum_{\substack{j=1 \\ j \neq i}}^{N} \frac{p_j \, p_i}{(1-p_j)(1-2p_j)} \right]$$

$$= \frac{p_i}{d} \left[\frac{1}{1-2p_i} + \sum_{j=1}^{N} \frac{p_j}{(1-p_j)(1-2p_j)} - \frac{p_i}{(1-p_i)(1-2p_i)} \right] \quad with \; d = \sum_{i=1}^{N} \frac{p_i}{1-2p_i}$$

$$= \frac{p_i}{d} \left[\frac{(1-p_i)-p_i}{(1-p_i)(1-2p_i)} + \sum_{j=1}^{N} \frac{p_j}{(1-p_j)(1-2p_j)} \right]$$

$$= \frac{p_i}{d} \left[\frac{1-2p_i}{(1-p_i)(1-2p_i)} + \sum_{j=1}^{N} \frac{p_j}{(1-p_j)(1-2p_j)} \right]$$

$$= \frac{p_i}{d} \left[\frac{1}{1-p_i} + \sum_{j=1}^{N} \frac{p_j}{(1-p_j)(1-2p_j)} \right] \qquad for \; p_i \; and \; p_j < \frac{1}{2} \qquad (3.2.1)$$

The probability of inclusion of both ith and jth unit in the sample (π_{ij}) for this selection procedure is obtained as:

$$\pi_{ij} = p_i \, p_{j|i} + p_j \, p_{i|j}$$

$$= \frac{\frac{p_i}{1-2p_i}}{\sum_{i=1}^{N} \frac{p_i}{1-2p_i}} \cdot \frac{p_j}{1-p_i} + \frac{\frac{p_j}{1-2p_j}}{\sum_{j=1}^{N} \frac{p_j}{1-2p_j}} \cdot \frac{p_i}{1-p_j}$$

$$= \frac{1}{\sum_{i=1}^{N} \frac{p_i}{1-2p_i}} \left[\frac{p_i \, p_j}{(1-p_i)(1-2p_i)} + \frac{p_i \, p_j}{(1-p_j)(1-2p_j)} \right]$$

$$= \frac{p_i \, p_j}{d} \left[\frac{1}{(1-p_i)(1-2p_i)} + \frac{1}{(1-p_j)(1-2p_j)} \right] \quad for \; p_i, \, p_j < \frac{1}{2} \qquad (3.2.2)$$

3.2.2 The New Selection Procedure – II:

New selection procedure – II is stated as:

- Select first unit with probability proportional to $\frac{2p_i(1-p_i)}{1-4p_i}$
- Select second unit with probability proportional to size of the remaining units.

The expressions for π_i and π_{ij} are derived as:

$$\pi_i = \frac{\dfrac{2p_i(1-p_i)}{1-4p_i}}{\sum_{i=1}^{N}\dfrac{2p_i(1-p_i)}{1-4p_i}} + \sum_{\substack{j=1\\ j\neq i}}^{N} \frac{\dfrac{2p_j(1-p_j)}{1-4p_j}}{\sum_{j=1}^{N}\dfrac{2p_j(1-p_j)}{1-4p_j}} \cdot \frac{p_i}{1-p_j}$$

$$= \frac{1}{\sum_{i=1}^{N}\dfrac{2p_i(1-p_i)}{1-4p_i}}\left[\frac{2p_i(1-p_i)}{1-4p_i} + \sum_{\substack{j=1\\ j\neq i}}^{N}\frac{2p_j(1-p_j)}{1-4p_j}\cdot\frac{p_i}{1-p_j}\right]$$

$$= \frac{1}{\sum_{i=1}^{N}\dfrac{2p_i(1-p_i)}{1-4p_i}}\left[\frac{2p_i(1-p_i)}{1-4p_i} + \sum_{\substack{j=1\\ j\neq i}}^{N}\frac{2p_i p_j}{1-4p_j}\right]$$

$$= \frac{p_i}{\sum_{i=1}^{N}\dfrac{2p_i(1-p_i)}{1-4p_i}}\left[\frac{2(1-p_i)}{1-4p_i} + 2\sum_{\substack{j=1\\ j\neq i}}^{N}\frac{p_j}{1-4p_j}\right]$$

$$= \frac{p_i}{\sum_{i=1}^{N}\dfrac{2p_i(1-p_i)}{1-4p_i}}\left[\frac{2(1-p_i)}{1-4p_i} + 2\sum_{j=1}^{N}\frac{p_j}{1-4p_j} - \frac{2p_i}{1-4p_i}\right]$$

$$= \frac{p_i}{\sum_{i=1}^{N}\dfrac{2p_i(1-p_i)}{1-4p_i}}\left[\frac{2(1-2p_i)}{1-4p_i} + 2\sum_{j=1}^{N}\frac{p_j}{1-4p_j}\right] \qquad (3.2.3)$$

Now

$$\sum_{i=1}^{N}\frac{2p_i(1-p_i)}{1-4p_i} = \sum_{i=1}^{N}\frac{p_i(2-2p_i)}{1-4p_i}$$

$$= \frac{1}{2}\sum_{i=1}^{N}\frac{p_i(4-4p_i)}{1-4p_i}$$

$$= \frac{1}{2}\sum_{i=1}^{N}\frac{3p_i + p_i(1-4p_i)}{1-4p_i}$$

$$= \frac{1}{2}\left[3\sum_{i=1}^{N}\frac{p_i}{1-4p_i} + \sum_{i=1}^{N}p_i\right]$$

$$= \frac{1}{2}\left[1+3\sum_{i=1}^{N}\frac{p_i}{1-4p_i}\right]$$

$$= \frac{1}{2}b \quad with\ b=1+3\sum_{i=1}^{N}\frac{p_i}{1-4p_i} \tag{3.2.4}$$

Substituting this value in (3.2.3)

$$\pi_i = \frac{2p_i}{b}\left[\frac{2(1-2p_i)}{1-4p_i}+2\sum_{j=1}^{N}\frac{p_j}{1-4p_j}\right]$$

$$= \frac{2p_i}{b}\left[\frac{2-4p_i}{1-4p_i}+2\sum_{j=1}^{N}\frac{p_j}{1-4p_j}\right]$$

$$= \frac{2p_i}{b}\left[\frac{(1-4p_i)}{1-4p_i}+\frac{1}{1-4p_i}+2\sum_{j=1}^{N}\frac{p_j}{1-4p_j}\right]$$

$$= \frac{2p_i}{b}\left[1+\frac{1}{1-4p_i}+2\sum_{j=1}^{N}\frac{p_j}{1-4p_j}\right] \quad for\ p_i,\ p_j<\frac{1}{4} \tag{3.2.5}$$

Again

$$\pi_{ij} = p_i p_{j|i} + p_j p_{i|j}$$

$$= \frac{\dfrac{2p_i(1-p_i)}{1-4p_i}}{\sum_{i=1}^{N}\dfrac{2p_i(1-p_i)}{1-4p_i}}\cdot\frac{p_j}{1-p_i}+\frac{\dfrac{2p_j(1-p_j)}{1-4p_j}}{\sum_{j=1}^{N}\dfrac{2p_j(1-p_j)}{1-4p_j}}\cdot\frac{p_i}{1-p_j}$$

$$= \frac{1}{\sum_{i=1}^{N}\dfrac{2p_i(1-p_i)}{1-4p_i}}\left[\frac{2p_i(1-p_i)}{1-4p_i}\cdot\frac{p_j}{1-p_i}+\frac{2p_j(1-p_j)}{1-4p_j}\cdot\frac{p_i}{1-p_j}\right]$$

Substituting the value of $\sum_{i=1}^{N}\frac{2p_i(1-p_i)}{1-4p_i}$ from (3.2.4)

$$\pi_{ij} = \frac{1}{\frac{b}{2}}\left[\frac{2p_i p_j}{1-4p_i}+\frac{2p_i p_j}{1-4p_j}\right]$$

$$= \frac{4p_i p_j}{b}\left[\frac{1}{1-4p_i}+\frac{1}{1-4p_j}\right] \quad for\ p_i, p_j<\frac{1}{4} \tag{3.2.6}$$

3.2.3 The New Selection Procedure – III :

The third New procedure is a mixture of Brewer (1963) and Durbin (1967) selection procedures. This procedure is stated as:

- Select first unit i with probability proportional to $\frac{p_i(1-p_i)}{(1-2p_i)}$
- Select second unit j with probability proportional to

$$p_j\left[\frac{1}{1-2p_i}+\frac{1}{1-2p_j}\right] \text{ from the remaining units.}$$

The expressions for π_i and π_{ij} are derived as:

$$\pi_i = \frac{\frac{p_i(1-p_i)}{1-2p_i}}{\sum_{i=1}^{N}\frac{p_i(1-p_i)}{1-2p_i}} + \sum_{\substack{j=1\\j\neq i}}^{N}\frac{\frac{p_j(1-p_j)}{1-2p_j}}{\sum_{i=1}^{N}\frac{p_j(1-p_j)}{1-2p_j}}\cdot\frac{p_i\left[\frac{1}{1-2p_i}+\frac{1}{1-2p_j}\right]}{\sum_{\substack{j=1\\j\neq i}}^{N}p_i\left[\frac{1}{1-2p_i}+\frac{1}{1-2p_j}\right]} \tag{3.2.7}$$

Now

$$\sum_{i=1}^{N}\frac{p_i(1-p_i)}{1-2p_i} = \frac{1}{2}\sum_{i=1}^{N}\frac{2p_i(1-p_i)}{1-2p_i}$$

$$= \frac{1}{2}\sum_{i=1}^{N}\frac{p_i+p_i(1-2p_i)}{1-2p_i}$$

$$= \frac{1}{2}\left[1+\sum_{i=1}^{N}\frac{p_i}{1-2p_i}\right]$$

$$= \frac{1}{2}k \quad \text{where} \quad k=\left[1+\sum_{i=1}^{N}\frac{p_i}{1-2p_i}\right]$$

Again

$$\sum_{\substack{j=1\\j\neq i}}^{N}p_j\left[\frac{1}{1-2p_i}+\frac{1}{1-2p_j}\right] = \sum_{\substack{j=1\\j\neq i}}^{N}\left[\frac{p_j}{1-2p_i}+\frac{p_j}{1-2p_j}\right]$$

$$= \sum_{\substack{j=1\\j\neq i}}^{N}\frac{p_j}{1-2p_i}+\sum_{\substack{j=1\\j\neq i}}^{N}\frac{p_j}{1-2p_j}$$

$$= \frac{1-p_i}{1-2p_i}+\sum_{j=1}^{N}\frac{p_j}{1-2p_j}-\frac{p_i}{1-2p_i}$$

$$= \left[1+\sum_{i=1}^{N}\frac{p_i}{1-2p_i}\right] = k$$

Substituting the values of $\sum_{i=1}^{N}\frac{p_i(1-p_i)}{1-2p_i}$ and $\sum_{\substack{j=1\\j\neq i}}^{N} p_j\left[\frac{1}{1-2p_i}+\frac{1}{1-2p_j}\right]$ in (3.2.7)

$$\pi_i = \frac{\frac{p_i(1-p_i)}{1-2p_i}}{\frac{1}{2}k} + \sum_{\substack{j=1\\j\neq i}}^{N}\frac{\frac{p_j(1-p_j)}{1-2p_j}}{\frac{1}{2}k}\cdot\frac{p_i\left[\frac{1}{1-2p_i}+\frac{1}{1-2p_j}\right]}{k}$$

$$= \frac{\frac{2p_i(1-p_i)}{1-2p_i}}{k} + \sum_{\substack{j=1\\j\neq i}}^{N}\frac{\frac{2p_j(1-p_j)}{1-2p_j}}{k}\cdot\frac{p_i\left[\frac{1}{1-2p_i}+\frac{1}{1-2p_j}\right]}{k}$$

$$= \frac{p_i}{k}\left[\frac{2(1-p_i)}{1-2p_i}+\frac{1}{k}\sum_{\substack{j=1\\j\neq i}}^{N}\frac{2p_j(1-p_j)}{1-2p_j}\cdot\left\{\frac{1}{1-2p_i}+\frac{1}{1-2p_j}\right\}\right]$$

$$= \frac{p_i}{k}\left[\frac{(2-2p_i)}{1-2p_i}+\frac{1}{k}\sum_{\substack{j=1\\j\neq i}}^{N}\frac{p_j(2-2p_j)}{1-2p_j}\cdot\left\{\frac{1}{1-2p_i}+\frac{1}{1-2p_j}\right\}\right]$$

$$= \frac{p_i}{k}\left[\frac{(1+1-2p_i)}{1-2p_i}+\frac{1}{k}\sum_{\substack{j=1\\j\neq i}}^{N}\frac{p_j(1+1-2p_j)}{1-2p_j}\cdot\left\{\frac{1}{1-2p_i}+\frac{1}{1-2p_j}\right\}\right]$$

$$= \frac{p_i}{k}\left[\frac{1+(1-2p_i)}{1-2p_i}+\frac{1}{k}\sum_{\substack{j=1\\j\neq i}}^{N}\frac{p_j+p_j(1-2p_j)}{1-2p_j}\cdot\left\{\frac{1}{1-2p_i}+\frac{1}{1-2p_j}\right\}\right]$$

$$= \frac{p_i}{k}\left[1+\frac{1}{1-2p_i}+\frac{1}{k}\sum_{\substack{j=1\\j\neq i}}^{N}\left(p_j+\frac{p_j}{1-2p_j}\right)\cdot\left\{\frac{1}{1-2p_i}+\frac{1}{1-2p_j}\right\}\right]$$

$$= \frac{p_i}{k}\left[1+\frac{1}{1-2p_i}+\frac{1}{k}\sum_{\substack{j=1\\j\neq i}}^{N}\left\{\frac{p_j}{1-2p_i}+\frac{p_j}{1-2p_j}+\frac{p_j}{(1-2p_i)(1-2p_j)}+\frac{p_j}{(1-2p_j)^2}\right\}\right]$$

$$= \frac{p_i}{k}\left[1+\frac{1}{1-2p_i}+\frac{1}{k}\left\{\sum_{\substack{j=1\\j\neq i}}^{N}\frac{p_j}{1-2p_i}+\sum_{\substack{j=1\\j\neq i}}^{N}\frac{p_j}{1-2p_j}+\sum_{\substack{j=1\\j\neq i}}^{N}\frac{p_j}{(1-2p_i)(1-2p_j)}+\sum_{\substack{j=1\\j\neq i}}^{N}\frac{p_j}{(1-2p_j)^2}\right\}\right]$$

$$= \frac{p_i}{k}\left[1+\frac{1}{1-2p_i}+\frac{1}{k}\left\{k+\frac{1}{1-2p_i}\sum_{j=1}^{N}\frac{p_j}{1-2p_j}+\sum_{j=1}^{N}\frac{p_j}{(1-2p_j)^2}-\frac{2p_i}{(1-2p_i)^2}\right\}\right]$$

$$= \frac{p_i}{k}\left[1+\frac{1}{1-2p_i}+1+\frac{1}{k}\left\{\sum_{j=1}^{N}\frac{p_j}{(1-2p_j)^2}+\frac{1}{1-2p_i}\sum_{j=1}^{N}\frac{p_j}{1-2p_j}-\frac{2p_i}{(1-2p_i)^2}\right\}\right]$$

$$= \frac{p_i}{k}\left[2+\frac{1}{1-2p_i}+\frac{1}{k}\left\{\sum_{j=1}^{N}\frac{p_j}{(1-2p_j)^2}+\frac{1}{1-2p_i}\sum_{j=1}^{N}\frac{p_j}{1-2p_j}-\frac{2p_i}{(1-2p_i)^2}\right\}\right] \qquad (3.2.8)$$

$$= \frac{p_i}{k}\left[2+\frac{1}{1-2p_i}+\frac{1}{k}\left\{\sum_{j=1}^{N}\frac{p_i}{(1-2p_j)^2}+\frac{k-1}{1-2p_i}-\frac{2p_i}{(1-2p_i)^2}\right\}\right]$$

$$= \frac{p_i}{k}\left[4\left(\frac{1-p_i}{1-2p_i}\right)+\frac{1}{k}\left\{\sum_{j=1}^{N}\frac{p_j}{(1-2p_j)^2}-\frac{1}{(1-2p_i)^2}\right\}\right] \qquad (3.2.9)$$

Again

$\pi_{ij} = p_i p_{j|i} + p_j p_{i|j}$

$$= \frac{\frac{p_i(1-p_i)}{1-2p_i}}{\sum_{i=1}^{N}\frac{p_i(1-p_i)}{1-2p_i}}\cdot\frac{p_j\left[\frac{1}{1-2p_i}+\frac{1}{1-2p_j}\right]}{\sum_{\substack{j=1\\j\neq i}}^{N}p_j\left[\frac{1}{1-2p_i}+\frac{1}{1-2p_j}\right]}+\frac{\frac{p_j(1-p_j)}{1-2p_j}}{\sum_{i=1}^{N}\frac{p_j(1-p_j)}{1-2p_j}}\cdot\frac{p_i\left[\frac{1}{1-2p_i}+\frac{1}{1-2p_j}\right]}{\sum_{\substack{i=1\\i\neq j}}^{N}p_i\left[\frac{1}{1-2p_i}+\frac{1}{1-2p_j}\right]}$$

$$= \frac{\frac{p_i(1-p_i)}{1-2p_i}}{k/2}\cdot\frac{p_j\left[\frac{1}{1-2p_i}+\frac{1}{1-2p_j}\right]}{k}+\frac{\frac{p_j(1-p_j)}{1-2p_j}}{k/2}\cdot\frac{p_i\left[\frac{1}{1-2p_i}+\frac{1}{1-2p_j}\right]}{k}$$

$$= \frac{p_ip_j}{k^2}\left[\frac{1}{1-2p_i}+\frac{1}{1-2p_j}\right]\cdot\left[\frac{2-2p_i}{1-2p_i}+\frac{2-2p_j}{1-2p_j}\right]$$

$$= \frac{p_ip_j}{k^2}\left[\frac{1}{1-2p_i}+\frac{1}{1-2p_j}\right]\cdot\left[\frac{1+1-2p_i}{1-2p_i}+\frac{1+1-2p_j}{1-2p_j}\right]$$

$$= \frac{p_ip_j}{k^2}\left[\frac{1}{1-2p_i}+\frac{1}{1-2p_j}\right]\cdot\left[2+\frac{1}{1-2p_i}+\frac{1}{1-2p_j}\right] \qquad (3.2.10)$$

3.3 Some Useful Results for The New Selection Procedures:

In this section some important results regarding the inclusion probabilities and joint inclusion probabilities have been verified for selection procedures developed in the previous section. These results play a fundamental role in deciding whether a specific expression can be used as an inclusion probability or joint inclusion probability. These results have been verified in following sub – sections.

3.3.1 The New Selection Procedure – I:

In this subsection some important results regarding the inclusion probabilities and joint inclusion probabilities have been verified for the New selection procedure – I.

Result – 1: $\sum_{i=1}^{N}\pi_i = n$ for this selection procedure.

Proof: To prove this result, consider (3.2.1) as:

$$\pi_i = \frac{p_i}{d}\left[\frac{1}{1-p_i}+\sum_{j=1}^{N}\frac{p_j}{(1-p_j)(1-2p_j)}\right] \tag{3.2.1}$$

Summing both sides:

$$\sum_{i=1}^{N}\pi_i = \sum_{i=1}^{N}\left[\frac{p_i}{d}\left\{\frac{1}{1-p_i}+\sum_{j=1}^{N}\frac{p_j}{(1-p_j)(1-2p_j)}\right\}\right]$$

$$= \frac{1}{d}\sum_{i=1}^{N}p_i\left[\frac{1}{1-p_i}+\sum_{j=1}^{N}\frac{p_j}{(1-p_j)(1-2p_j)}\right]$$

$$= \frac{1}{d}\left[\sum_{i=1}^{N}\frac{p_i}{1-p_i}\left(1+\frac{1}{1-2p_i}\right)\right]$$

$$= \frac{1}{d}\left[\sum_{i=1}^{N}\frac{p_i}{1-p_i}\left(\frac{2-2p_i}{1-2p_i}\right)\right]$$

$$= \frac{2}{d}\sum_{i=1}^{N}\frac{p_i}{1-2p_i}=2. \tag{3.3.1}$$

Result – 2: The quantity π_{ij}, obtained under this selection procedure, satisfies the relation $\sum_{\substack{j=1 \\ j \neq i}}^{N} \pi_{ij} = (n-1)\pi_i$.

Proof: Consider (3.2.2) as:

$$\pi_{ij} = \frac{p_i p_j}{d}\left[\frac{1}{(1-p_i)(1-2p_i)} + \frac{1}{(1-p_j)(1-2p_j)}\right] \tag{3.2.2}$$

Summing π_{ij} when $(j \neq i)$:

$$\sum_{\substack{j=1 \\ j \neq i}}^{N} \pi_{ij} = \sum_{\substack{j=1 \\ j \neq i}}^{N}\left[\frac{p_i p_j}{d}\left\{\frac{1}{(1-p_i)(1-2p_i)} + \frac{1}{(1-p_j)(1-2p_j)}\right\}\right]$$

$$= \frac{p_i}{d}\left[\frac{1}{(1-p_i)(1-2p_i)}\sum_{\substack{j=1 \\ j \neq i}}^{N} p_j + \sum_{\substack{j=1 \\ j \neq i}}^{N}\frac{p_j}{(1-p_j)(1-2p_j)}\right]$$

$$= \frac{p_i}{d}\left[\frac{(1-p_i)}{(1-p_i)(1-2p_i)} + \sum_{j=1}^{N}\frac{p_j}{(1-p_j)(1-2p_j)} - \frac{p_i}{(1-p_i)(1-2p_i)}\right]$$

On Simplification:

$$\sum_{\substack{j=1 \\ j \neq i}}^{N} \pi_{ij} = \frac{p_i}{d}\left[\frac{1}{(1-p_i)} + \sum_{j=1}^{N}\frac{p_j}{(1-p_j)(1-2p_j)}\right] = \pi_i \tag{3.3.2}$$

Result – 3: The quantity π_{ij}, obtained under this selection procedure, satisfies the relation $\sum_{i=1}^{N}\sum_{\substack{j=1 \\ j \neq i}} \pi_{ij} = n(n-1)$ where n is the sample size.

Proof: Consider (3.2.2) as:

$$\pi_{ij} = \frac{p_i p_j}{d}\left[\frac{1}{(1-p_i)(1-2p_i)} + \frac{1}{(1-p_j)(1-2p_j)}\right] \tag{3.2.2}$$

Applying double summation on both sides:

$$\sum_{i=1}^{N}\sum_{\substack{j=1 \\ j \neq i}} \pi_{ij} = \sum_{i=1}^{N}\sum_{\substack{j=1 \\ j \neq i}}\left[\frac{p_i p_j}{d}\left\{\frac{1}{(1-p_i)(1-2p_i)} + \frac{1}{(1-p_i)(1-2p_i)}\right\}\right]$$

$$= \sum_{i=1}^{N}\left[\sum_{\substack{j=1\\j\neq i}}^{N}\left\{\frac{p_i\,p_j}{d}\left(\frac{1}{(1-p_i)(1-2p_i)}+\frac{1}{(1-p_j)(1-2p_j)}\right)\right\}\right] \qquad (3.3.3)$$

Also $\sum_{\substack{j=1\\j\neq i}}^{N}\pi_{ij} = \sum_{\substack{j=1\\j\neq i}}^{N}\left[\frac{p_i\,p_j}{d}\left\{\frac{1}{(1-p_i)(1-2p_i)}+\frac{1}{(1-p_j)(1-2p_j)}\right\}\right]=\pi_i$ (3.3.2)

Substituting (3.3.2) in (3.3.3)

$$\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}\pi_{ij} = \sum_{i=1}^{N}\pi_i = 2 \qquad (3.3.4)$$

Since n = 2, therefore equation (3.3.4) can be written as $\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}\pi_{ij}=n(n-1)$. Hence the result.

Result 4: The Sen – Yates – Grundy variance estimator is always non-negative under this selection procedure:

Proof: The Sen – Yates – Grundy variance estimator is given in (1.3.6). The quantities π_i and π_{ij} under the new selection procedure – I are given in (3.2.1) and (3.2.2). Now, the Sen – Yates – Grundy variance estimator is non-negative if:

$$\pi_i\,\pi_j-\pi_{ij}\geq 0$$

Now

$$\pi_i\,\pi_j-\pi_{ij} = \frac{p_i}{d}\left[\frac{1}{1-p_i}+\sum_{h=1}^{N}\frac{p_h}{(1-p_h)(1-2p_h)}\right]\cdot\frac{p_j}{d}\left[\frac{1}{1-p_j}+\sum_{h=1}^{N}\frac{p_h}{(1-p_h)(1-2p_h)}\right]$$
$$-\frac{p_i\,p_j}{d}\left[\frac{1}{(1-p_i)(1-2p_i)}+\frac{1}{(1-p_j)(1-2p_j)}\right]$$

Writing $A=\sum_{h=1}^{N}\frac{p_h}{(1-p_h)(1-2p_h)}$ above equation becomes

$$\pi_i\,\pi_j-\pi_{ij} = \frac{p_i\,p_j}{d}\left[\frac{1}{d}\left\{\frac{1}{1-p_i}+A\right\}\left\{\frac{1}{1-p_j}+A\right\}-\left\{\frac{1}{(1-p_i)(1-2p_i)}+\frac{1}{(1-p_j)(1-2p_j)}\right\}\right]$$
$$= \frac{p_i\,p_j}{d}\left[\frac{1}{d}\left\{\frac{1}{(1-p_i)(1-p_j)}+\frac{A}{(1-p_i)}+\frac{A}{(1-p_j)}+A^2\right\}\right.$$

$$-\left\{\frac{2}{(1-2p_i)}-\frac{1}{(1-p_i)}+\frac{2}{(1-2p_j)}-\frac{1}{(1-p_j)}\right\}\Bigg]$$

$$=\frac{p_i\, p_j}{d}\left[\frac{1}{d}\left\{A^2+\frac{(A+d)(2-p_i-p_j)+1}{(1-p_i)(1-p_j)}\right\}-\frac{4(1-p_i-p_j)}{(1-2p_i)(1-2p_j)}\right] \qquad (3.3.5)$$

Now since the first term within the main brackets of equation (3.3.5) is always greater than or equal to the second term therefore (3.3.5) is always non-negative for all values of p_i and p_j, making Sen – Yates – Grundy variance estimator non-negative for all samples. Further, it has been numerically checked that the relation $\pi_i\,\pi_j-\pi_{ij}\geq 0$ is true for all values of π_i, π_j and π_{ij} for this selection procedure.

3.3.2 The New Selection Procedure – II:

Result – 1: $\sum_{i=1}^{N}\pi_i=n$ for this selection procedure.

Proof: To prove this result, again consider the value of π_i from (3.2.5) as:

$$\pi_i=\frac{2p_i}{b}\left[1+\frac{1}{1-4p_i}+2\sum_{j=1}^{N}\frac{p_j}{1-4p_j}\right] \qquad (3.2.5)$$

Summing both sides:

$$\sum_{i=1}^{N}\pi_i=\sum_{i=1}^{N}\left[\frac{2p_i}{b}\left\{1+\frac{1}{1-4p_i}+2\sum_{j=1}^{N}\frac{p_j}{1-4p_j}\right\}\right]$$

$$=\frac{2}{b}\sum_{i=1}^{N}p_i\left[1+\frac{1}{1-4p_i}+2\sum_{j=1}^{N}\frac{p_j}{1-4p_j}\right]$$

$$=\frac{2}{b}\left[1+\sum_{i=1}^{N}\frac{p_i}{1-4p_i}+2\sum_{i=1}^{N}\frac{p_i}{1-4p_i}\right]$$

$$=\frac{2}{b}\left[1+3\sum_{i=1}^{N}\frac{p_i}{1-4p_i}\right]=2 \qquad (3.3.6)$$

Result – 2: The quantity π_{ij} , obtained under this selection procedure, satisfies the relation $\sum_{\substack{j=1\\ j\neq i}}^{N}\pi_{ij}=(n-1)\pi_i$.

Proof: To prove this result considers (3.2.6) as:

$$\pi_{ij} = \frac{4p_i p_j}{b}\left[\frac{1}{1-4p_i}+\frac{1}{1-4p_j}\right] \tag{3.2.6}$$

Summing π_{ij} when $(j \neq i)$:

$$\sum_{\substack{j=1\\ j\neq i}}^{N} \pi_{ij} = \sum_{\substack{j=1\\ j\neq i}}^{N}\left[\frac{4p_i p_j}{b}\left\{\frac{1}{1-4p_i}+\frac{1}{1-4p_j}\right\}\right]$$

$$= \frac{4p_i}{b}\left[\frac{1}{1-4p_i}\sum_{\substack{j=1\\ j\neq i}}^{N} p_j + \sum_{\substack{j=1\\ j\neq i}}^{N}\frac{p_j}{1-4p_j}\right]$$

$$= \frac{4p_i}{b}\left[\frac{1-p_i}{1-4p_i}+\sum_{j=1}^{N}\frac{p_j}{1-4p_j}-\frac{p_i}{1-4p_i}\right]$$

On Simplification:

$$\sum_{\substack{j=1\\ j\neq i}}^{N} \pi_{ij} = \frac{2p_i}{b}\left[1+\frac{1}{1-4p_i}+2\sum_{j=1}^{N}\frac{p_j}{1-4p_j}\right] = \pi_i \tag{3.3.7}$$

Result – 3: The quantity π_{ij}, obtained under this selection procedure, satisfies the relation $\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}} \pi_{ij} = n(n-1)$ where n is the sample size.

Proof: Consider (3.2.6) as:

$$\pi_{ij} = \frac{4p_i p_j}{b}\left[\frac{1}{1-4p_i}+\frac{1}{1-4p_j}\right] \tag{3.2.6}$$

Applying double sum on both sides:

$$\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}} \pi_{ij} = \sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}\left[\frac{4p_i p_j}{b}\left\{\frac{1}{1-4p_i}+\frac{1}{1-4p_j}\right\}\right]$$

$$= \sum_{i=1}^{N}\left[\sum_{\substack{j=1\\ j\neq i}}^{N}\left\{\frac{4p_i p_j}{b}\left(\frac{1}{1-4p_i}+\frac{1}{1-4p_j}\right)\right\}\right] \tag{3.3.8}$$

Also from (3.3.7):

$$\sum_{\substack{j=1\\ j\neq i}}^{N}\pi_{ij}=\sum_{\substack{j=1\\ j\neq i}}^{N}\left[\frac{4p_i p_j}{b}\left\{\frac{1}{1-4p_i}+\frac{1}{1-4p_i}\right\}\right]=\pi_i \qquad (3.3.7)$$

Substituting (3.3.7) in (3.3.8):

$$\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}\pi_{ij}=\sum_{i=1}^{N}\pi_i=2 \qquad (3.3.9)$$

Since n = 2, therefore equation (3.3.9) can be written as $\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}\pi_{ij}=n(n-1)$. Hence the result.

Result 4: The Sen – Yates – Grundy variance estimator is always non-negative under this selection procedure:

Proof: Consider again (1.3.6). The quantities π_i and π_{ij} under the new selection procedure – II are given in (3.2.5) and (3.2.6). Now, the Sen – Yates – Grundy variance estimator is non-negative if:

$$\pi_i\,\pi_j-\pi_{ij}\geq 0$$

Now

$$\pi_i\,\pi_j-\pi_{ij}=\frac{2p_i}{b}\left[1+\frac{1}{1-4p_i}+2\sum_{h=1}^{N}\frac{p_h}{1-4p_h}\right]\cdot\frac{2p_j}{b}\left[1+\frac{1}{1-4p_j}+2\sum_{h=1}^{N}\frac{p_h}{1-4p_h}\right]-\frac{4p_i p_j}{b}\left[\frac{1}{1-4p_i}+\frac{1}{1-4p_j}\right]$$

Writing $B=2\sum_{h=1}^{N}\frac{p_h}{1-4p_h}$, above equation can be written as:

$$\pi_i\,\pi_j-\pi_{ij}=\frac{4p_i p_j}{b}\left[\frac{1}{b}\left\{1+\frac{1}{1-4p_i}+B\right\}\left\{1+\frac{1}{1-4p_j}+B\right\}-\left\{\frac{1}{1-4p_i}+\frac{1}{1-4p_j}\right\}\right]$$

$$=\frac{4p_i p_j}{b}\left[\frac{1}{b}\left\{(B+1)^2+\frac{2(B+1)(1-2p_i-2p_i)+1}{(1-4p_i)(1-4p_j)}\right\}-\frac{2(1-2p_i-2p_j)}{(1-4p_i)(1-4p_j)}\right] \qquad (3.3.10)$$

Now since the first term within the main brackets of equation (3.3.10) is always greater than or equal to the second term therefore (3.3.10) is always non-negative for all values of p_i and p_j, making Sen – Yates – Grundy variance estimator non-negative for all samples. Further, it has been numerically checked that the relation $\pi_i\,\pi_j-\pi_{ij}\geq 0$ is true for all values of π_i, π_j and π_{ij} for this selection procedure.

3.3.3 The New Selection Procedure – III:

Result – 1: $\sum_{i=1}^{N} \pi_i = n$ for this selection procedure.

Proof: Consider the value of π_i from (3.2.8) as:

$$\pi_i = \frac{p_i}{k}\left[2+\frac{1}{1-2p_i}+\frac{1}{k}\left\{\sum_{j=1}^{N}\frac{p_j}{(1-2p_j)^2}+\frac{1}{1-2p_i}\sum_{j=1}^{N}\frac{p_j}{1-2p_j}-\frac{2p_i}{(1-2p_i)^2}\right\}\right] \quad (3.2.8)$$

Summing both sides:

$$\sum_{i=i}^{N}\pi_i = \sum_{i=1}^{N}\frac{p_i}{k}\left[2+\frac{1}{1-2p_i}+\frac{1}{k}\left\{\sum_{j=1}^{N}\frac{p_j}{(1-2p_j)^2}+\frac{1}{1-2p_i}\sum_{j=1}^{N}\frac{p_j}{1-2p_j}-\frac{2p_i}{(1-2p_i)^2}\right\}\right]$$

$$= \frac{1}{k}\left[2\sum_{i=1}^{N}p_i+\sum_{i=1}^{N}\frac{p_i}{1-2p_i}+\frac{1}{k}\left\{\sum_{i=1}^{N}\frac{p_i}{(1-2p_i)^2}+\left(\sum_{i=1}^{N}\frac{p_i}{1-2p_i}\right)^2-2\sum_{i=1}^{N}\frac{p_i^2}{(1-2p_i)^2}\right\}\right]$$

$$= \frac{1}{k}\left[2+\sum_{i=1}^{N}\frac{p_i}{1-2p_i}+\frac{1}{k}\left\{\sum_{i=1}^{N}\frac{p_i}{1-2p_i}*k\right\}\right]$$

$$= \frac{1}{k}\left[2+\sum_{i=1}^{N}\frac{p_i}{1-2p_i}+\sum_{i=1}^{N}\frac{p_i}{1-2p_i}\right]$$

$$= \frac{2}{k}\left[1+\sum_{i=1}^{N}\frac{p_i}{1-2p_i}\right]=2 \quad (3.3.11)$$

Result – 2: The quantity π_{ij}, obtained under this selection procedure, satisfies the relation $\sum_{\substack{j=1\\ j\neq i}}^{N}\pi_{ij} = (n-1)\pi_i$.

Proof: To prove this result considers (3.2.10) as:

$$\pi_{ij} = \frac{p_i p_j}{k^2}\left[\frac{1}{1-2p_i}+\frac{1}{1-2p_j}\right]\cdot\left[2+\frac{1}{1-2p_i}+\frac{1}{1-2p_j}\right] \quad (3.2.10)$$

$$= \frac{2p_ip_j}{k^2}\left[\frac{1}{1-2p_i}+\frac{1}{1-2p_j}\right]+\frac{p_ip_j}{k^2}\left[\frac{1}{1-2p_i}+\frac{1}{1-2p_j}\right]^2$$

Summing π_{ij} when $(j\neq i)$:

$$\sum_{\substack{j=1\\j\neq i}}^{N}\pi_{ij} = \sum_{\substack{j=1\\j\neq i}}^{N}\left[\frac{2p_ip_j}{k^2}\left\{\frac{1}{1-2p_i}+\frac{1}{1-2p_j}\right\}+\frac{p_ip_j}{k^2}\left\{\frac{1}{1-2p_i}+\frac{1}{1-2p_j}\right\}^2\right]$$

$$= \sum_{\substack{j=1\\j\neq i}}^{N}\left[\frac{2p_ip_j}{k^2}\left\{\frac{1}{1-2p_i}+\frac{1}{1-2p_j}\right\}\right]+\sum_{\substack{j=1\\j\neq i}}^{N}\left[\frac{p_ip_j}{k^2}\left\{\frac{1}{1-2p_i}+\frac{1}{1-2p_j}\right\}^2\right]$$

$$= \frac{2p_i}{k}+\frac{p_i}{k^2}\sum_{\substack{j=1\\j\neq i}}^{N}\left[\frac{p_j}{(1-2p_i)^2}+\frac{p_j}{(1-2p_j)^2}+\frac{2p_j}{(1-2p_i)(1-2p_j)}\right]$$

$$= \frac{2p_i}{k}+\frac{p_i}{k^2}\left[\frac{k}{(1-2p_i)}+\sum_{j=1}^{N}\frac{p_j}{(1-2p_j)^2}+\frac{1}{1-2p_i}\sum_{j=1}^{N}\frac{p_j}{(1-2p_j)}-\frac{2p_i}{(1-2p_i)^2}\right]$$

$$= \frac{p_i}{k}\left[2+\frac{1}{1-2p_i}+\frac{1}{k}\left\{\sum_{j=1}^{N}\frac{p_j}{(1-2p_j)^2}+\frac{1}{1-2p_i}\sum_{j=1}^{N}\frac{p_j}{(1-2p_j)}-\frac{2p_i}{(1-2p_i)^2}\right\}\right] \quad (3.3.12)$$

Comparing (3.3.12) with (3.2.8) we have $\sum_{\substack{j=1\\j\neq i}}^{N}\pi_{ij}=\pi_i$. Since n is 2, therefore we have $\sum_{\substack{j=1\\j\neq i}}^{N}\pi_{ij}=(n-1)\pi_i$.

Result – 3: The quantity π_{ij}, obtained under this selection procedure, satisfies the relation $\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}\pi_{ij}=n(n-1)$ where n is the sample size.

Proof: Consider (3.2.10) as:

$$\pi_{ij} = \frac{p_ip_j}{k^2}\left[\frac{1}{1-2p_i}+\frac{1}{1-2p_j}\right]\cdot\left[2+\frac{1}{1-2p_i}+\frac{1}{1-2p_j}\right] \quad (3.2.10)$$

Applying the double summation we have:

$$\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}\pi_{ij} = \sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}\left[\frac{p_ip_j}{k^2}\left\{\frac{1}{1-2p_i}+\frac{1}{1-2p_j}\right\}\cdot\left\{2+\frac{1}{1-2p_i}+\frac{1}{1-2p_j}\right\}\right]$$

$$= \sum_{i=1}^{N}\left[\sum_{\substack{j=1\\j\neq i}}^{N}\left\{\frac{p_ip_j}{k^2}\left(\frac{1}{1-2p_i}+\frac{1}{1-2p_j}\right)\cdot\left(2+\frac{1}{1-2p_i}+\frac{1}{1-2p_j}\right)\right\}\right] \quad (3.3.13)$$

Also from (3.2.12):

$$\sum_{\substack{j=1\\j\neq i}}^{N}\pi_{ij} = \sum_{\substack{j=1\\j\neq i}}^{N}\left[\frac{p_i p_j}{k^2}\left\{\frac{1}{1-2p_i}+\frac{1}{1-2p_j}\right\}\cdot\left\{2+\frac{1}{1-2p_i}+\frac{1}{1-2p_j}\right\}\right] = \pi_i \qquad (3.3.12)$$

Substituting (3.3.12) in (3.3.13):

$$\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}\pi_{ij} = \sum_{i=1}^{N}\pi_i = 2 \qquad (3.3.14)$$

Since n = 2, therefore equation (3.3.14) can be written as $\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}\pi_{ij} = n(n-1)$.

Hence the result.

Results 4: The Sen – Yates – Grundy variance estimator is always non-negative under this selection procedure:

Proof: Consider again (1.3.6). The quantities π_i and π_{ij} under the new selection procedure – III are given in (3.2.8) and (3.2.10). Now, the Sen – Yates – Grundy variance estimator is non-negative if:

$$\pi_i\,\pi_j - \pi_{ij} \geq 0$$

or

$$\frac{p_i}{k}\left[2+\frac{1}{1-2p_i}+\frac{1}{k}\left\{\sum_{h=1}^{N}\frac{p_h}{(1-2p_h)^2}+\frac{1}{1-2p_i}\sum_{h=1}^{N}\frac{p_h}{1-2p_h}-\frac{2p_i}{(1-2p_i)^2}\right\}\right]$$

$$\cdot\frac{p_i}{k}\left[2+\frac{1}{1-2p_i}+\frac{1}{k}\left\{\sum_{h=1}^{N}\frac{p_h}{(1-2p_h)^2}+\frac{1}{1-2p_i}\sum_{h=1}^{N}\frac{p_h}{1-2p_h}-\frac{2p_i}{(1-2p_i)^2}\right\}\right]$$

$$-\frac{p_i p_j}{k^2}\left[\frac{1}{1-2p_i}+\frac{1}{1-2p_j}\right]*\left[2+\frac{1}{1-2p_i}+\frac{1}{1-2p_j}\right]\geq 0$$

Above equation, after simplification becomes:

$$\frac{4(1-p_i)(1-p_j)}{(1-2p_i)(1-2p_j)}\left[\frac{1-p_i}{1-2p_i}+\frac{1-p_j}{1-2p_j}\right]+\left(\frac{3-4p_i}{1-2p_i}\right)\left[\frac{C(1-2p_j)^2+D(1-2p_j)-2p_j}{k(1-2p_j)^2}\right]$$

$$+\left(\frac{3-4p_j}{1-2p_j}\right)\left[\frac{C(1-2p_i)^2+D(1-2p_i)-2p_i}{k(1-2p_i)^2}\right]+\left[\frac{3p_i(1-p_j)+3p_j(1-p_i)+p_i p_j}{(1-2p_i)(1-2p_j)}\right]$$

$$+\left[\frac{\left\{C(1-2p_i)^2+D(1-2p_i)-2p_i\right\}\cdot\left\{C(1-2p_j)^2+D(1-2p_j)-2p_j\right\}}{k^2(1-2p_i)^2(1-2p_j)^2}\right]$$

$$+\left[\frac{(1-p_i)(1-p_j)}{(1-2p_i)(1-2p_j)}\right]\geq 0 \qquad (3.2.15)$$

where $C=\sum_{h=1}^{N}\frac{p_h}{(1-2p_h)^2}$ *and* $D=\sum_{h=1}^{N}\frac{p_h}{(1-2p_h)}$. Equation (3.2.15) is true for all values of p_i and p_j, making Sen – Yates – Grundy variance estimator non-negative for all samples. Further, it has been numerically checked that the relation $\pi_i\,\pi_j-\pi_{ij}\geq 0$ is true for all values of π_i, π_j and π_{ij} for this selection procedure.

3.4 Empirical Study:

In this section the empirical study has been carried out by using the new selection procedures and some famous selection procedures given in chapter 1. Fifty natural populations have been used for this empirical study. The empirical study has been carried out by evaluating the variance of Horvitz – Thompson estimator for various selection procedures. Relative efficiency of various sampling designs has also been compared with the Brewer procedure, that is, variance of each procedure divided by the variance of Brewer procedure. After obtaining the variances the rank of a selection procedure is obtained with respect to its sampling variance, that is the procedure with smallest variance in a particular population has been assigned a rank of 1, the procedure with second smallest variance has been assigned a rank of 2 and so on. The ranking has been done following the idea of Jeffreys (1961) which says that "Rank correlations are effectively used to detect the monotone relation between two quantities". Now, this is obvious because a perfect monotone relation between two quantities is reduced to a perfect linear relation in terms of ranks. Departure and extent of departure from this is reflected in the magnitude of rank correlation. These ranks have been used along with the ranks of coefficient of variation and ranks of the correlation coefficient in deciding the overall performance of a selection procedure. The results of these analyses have been given in the following tables.

Table 3.1: Variances of Various Selection Procedures along with the Relative Efficiency (within brackets) of Brewer Procedure for Selected Natural Populations.

Procedure	Pop. – 1	Pop. – 2	Pop. – 3	Pop. – 4	Pop. – 5
Simple Random Sampling	12149.53 (115.23)	4946942.50 (32.41)	875830.81 (30.46)	4436.00 (6.23)	9541.78 (66.79)
Hansen – Hurwitz	14951.35 (93.64)	1607816.88 (99.71)	275095.13 (96.97)	309.58 (89.20)	7116.15 (89.56)
Midzuno – Sen	8567.60 (163.41)	753536.94 (212.74)	179187.86 (148.86)	1740.59 (15.86)	6103.93 (104.41)
Yates Grundy (d – b – d)	13383.07 (104.61)	1520483.75 (105.43)	246963.45 (108.01)	333.75 (82.74)	6132.15 (103.93)
Brewer	14000.22 (100.00)	1603085.00 (100.00)	266747.50 (100.00)	276.14 (100.00)	6373.32 (100.00)
Durbin (Rejective)	14014.51 (99.90)	1776858.88 (90.22)	281093.94 (94.90)	277.66 (99.45)	6416.62 (99.33)
Yates – Grundy (Rejective)	12865.85 (115.58)	1462906.00 (120.56)	233172.10 (124.55)	400.38 (63.07)	5980.32 (114.42)
New – I	14570.52 (96.09)	1685487.25 (95.11)	286405.00 (93.14)	249.27 (110.78)	6611.32 (96.40)
New – II	15670.78 (89.34)	1976598.75 (81.10)	347040.09 (76.86)	296.85 (93.02)	7469.12 (85.33)
New – III	14648.59 (95.57)	1710862.13 (93.70)	290845.50 (91.71)	244.67 (112.86)	6679.08 (95.42)

Table 3.1 (Continued)

Procedure	Pop. – 6	Pop. – 7	Pop. – 8	Pop. – 9	Pop. – 10
Simple Random Sampling	1796015.13 (2.86)	340089.34 (3.41)	720850.25 (64.30)	283385.69 (46.71)	35074512 (166.65)
Hansen – Hurwitz	52975.27 (96.83)	14721.07 (78.82)	499017.84 (92.88)	142668.53 (92.77)	61680976 (94.76)
Midzuno – Sen	281445.13 (18.23)	62507.72 (18.56)	552353.25 (83.91)	195738.14 (67.62)	10142310 (576.30)
Yates Grundy (d – b – d)	54243.95 (94.57)	10489.04 (110.62)	465838.94 (99.50)	135870.34 (97.42)	53669716 (108.91)
Brewer	51296.67 (100.00)	11603.40 (100.00)	463496.94 (100.00)	132359.27 (100.00)	58450276 (100.00)
Durbin (Rejective)	56572.57 (90.67)	12579.37 (92.24)	461990.38 (100.33)	132129.33 (100.17)	60436296 (96.71)
Yates – Grundy (Rejective)	62518.13 (75.58)	13098.16 (63.86)	468282.70 (104.22)	139031.90 (100.10)	49735460 (128.03)
New – I	56387.64 (90.97)	13964.13 (83.09)	461842.81 (100.36)	129430.63 (102.26)	62755576 (93.14)
New – II	133736.31 (38.36)	n. c. (n. c.)	459708.00 (100.82)	124073.31 (106.68)	72097800 (81.07)
New – III	59806.44 (85.77)	15325.12 (75.71)	461634.78 (100.40)	129011.20 (102.60)	63612676 (91.88)

n. c. = Not Calculated.

Table 3.1 (Continued)

Procedure	Pop. – 11	Pop. – 12	Pop. – 13	Pop. – 14	Pop. – 15
Simple Random Sampling	116657888 (25.75)	33290.53 (25.55)	113744360 (15.93)	9171896.00 (4.80)	5704569.00 (3.10)
Hansen – Hurwitz	33459728 (89.78)	9055.63 (93.92)	21936420 (82.58)	469457.03 (93.78)	190441.38 (92.99)
Midzuno – Sen	37576684 (79.94)	18249.91 (46.60)	30949710 (58.53)	2242999.50 (19.63)	1505043.13 (11.77)
Yates Grundy (d – b – d)	26741372 (112.34)	8830.26 (96.32)	14681897 (123.38)	417418.81 (105.47)	181002.19 (97.84)
Brewer	30039964 (100.00)	8504.88 (100.00)	18114750 (100.00)	440262.16 (100.00)	177091.80 (100.00)
Durbin (Rejective)	30205920 (99.45)	8481.25 (100.28)	20543326 (88.18)	456182.31 (96.51)	180367.11 (98.18)
Yates – Grundy (Rejective)	24568876 (135.46)	9132.03 (96.90)	12836245 (174.51)	425429.30 (106.69)	190376.50 (96.46)
New – I	33028670 (90.95)	8229.51 (103.35)	20378062 (88.89)	485847.69 (90.62)	180120.25 (98.32)
New – II	45083272 (66.63)	7774.19 (109.40)	n. c. (n. c.)	883123.13 (49.85)	201684.97 (87.81)
New – III	33664724 (89.23)	8198.06 (103.74)	21799216 (83.10)	498932.25 (88.24)	181345.52 (97.65)

n. c. = Not Calculated.

Table 3.1 (Continued)

Procedure	Pop. – 16	Pop. – 17	Pop. – 18	Pop. – 19	Pop. – 20
Simple Random Sampling	146.60 (20.11)	552.85 (72.72)	174762048 (353.02)	660286848 (353.10)	2886.59 (42.12)
Hansen – Hurwitz	33.05 (89.19)	431.97 (93.07)	677828992 (91.02)	2464964096 (94.58)	1317.38 (92.30)
Midzuno – Sen	69.39 (42.48)	362.19 (111.00)	27048020 (2280.95)	536794880 (434.33)	1654.17 (73.50)
Yates Grundy (d – b – d)	30.87 (95.50)	388.89 (103.38)	508257984 (121.39)	2131433344 (109.38)	1248.41 (97.40)
Brewer	29.48 (100.00)	402.02 (100.00)	616952000 (100.00)	2331453952 (100.00)	1215.89 (100.00)
Durbin (Rejective)	29.48 (99.98)	415.00 (96.87)	706549120 (87.32)	2452795392 (95.05)	1208.62 (100.60)
Yates – Grundy (Rejective)	32.36 (95.97)	380.09 (115.72)	432262848 (178.89)	1968792576 (131.54)	1276.32 (99.29)
New – I	28.68 (102.77)	414.75 (96.93)	693443648 (88.97)	2503432192 (93.13)	1190.16 (102.16)
New – II	28.46 (103.57)	462.03 (87.01)	n. c. (n. c.)	2919367680 (79.86)	1132.15 (107.40)
New – III	28.53 (103.32)	418.21 (96.13)	747877056 (82.49)	2551881472 (91.36)	1187.17 (102.42)

n. c. = Not Calculated.

Table 3.1 (Continued)

Procedure	Pop. – 21	Pop. – 22	Pop. – 23	Pop. – 24	Pop. – 25
Simple Random Sampling	51550.80 (8.86)	30627.20 (75.27)	2119.58 (68.74)	209.44 (652.98)	33002236 (19.71)
Hansen – Hurwitz	5149.11 (88.72)	24445.05 (94.31)	1544.81 (94.32)	1553.35 (88.04)	6976035 (93.22)
Midzuno – Sen	16686.42 (27.38)	18290.16 (126.04)	1207.00 (120.72)	225.46 (606.58)	12700635 (51.20)
Yates Grundy (d – b – d)	4553.31 (100.32)	22444.18 (102.71)	1410.09 (103.33)	1204.55 (113.53)	6204300 (104.82)
Brewer	4568.07 (100.00)	23053.29 (100.00)	1457.04 (100.00)	1367.58 (100.00)	6503349 (100.00)
Durbin (Rejective)	4524.93 (100.95)	23682.82 (97.34)	1488.73 (97.87)	1619.84 (84.43)	6660894 (97.63)
Yates – Grundy (Rejective)	4639.45 (103.03)	21953.56 (112.91)	1372.55 (114.10)	1100.55 (150.01)	6021660 (117.13)
New – I	4693.16 (97.33)	23647.68 (97.49)	1502.63 (96.97)	1459.49 (93.70)	6818343 (95.38)
New – II	5314.35 (85.96)	24867.00 (92.71)	1595.51 (91.32)	440.45 (310.49)	7788024 (83.50)
New – III	4731.29 (96.55)	23766.82 (97.00)	1511.44 (96.40)	1526.57 (89.58)	6895054 (94.32)

Table 3.1 (Continued)

Procedure	Pop. – 26	Pop. – 27	Pop. – 28	Pop. – 29	Pop. – 30
Simple Random Sampling	42895136 (20.80)	19825144 (67.08)	52686164 (566.09)	625482.69 (48.25)	983823.19 (37.67)
Hansen – Hurwitz	9738723 (91.62)	14469214 (91.91)	303246912 (98.35)	323717.97 (93.23)	391832.13 (94.58)
Midzuno – Sen	17535510 (50.88)	14055314 (94.61)	52643748 (566.55)	280560.22 (107.57)	646497.50 (57.32)
Yates Grundy (d – b – d)	8816018 (101.21)	13042031 (101.96)	282992928 (105.39)	282880.78 (106.69)	376569.28 (98.41)
Brewer	8922493 (100.00)	13298241 (100.00)	298253152 (100.00)	301799.25 (100.00)	370586.00 (100.00)
Durbin (Rejective)	9225706 (96.71)	13342208 (99.67)	326419424 (91.37)	303496.25 (99.44)	372031.84 (99.61)
Yates – Grundy (Rejective)	8854914 (109.64)	12856317 (110.98)	270043968 (122.48)	268174.70 (121.23)	382230.70 (101.65)
New – I	9066038 (98.42)	13539721 (98.22)	311582496 (95.72)	319729.00 (94.39)	365347.06 (101.43)
New – II	10794841 (82.66)	14173989 (93.82)	340107680 (87.69)	359089.44 (84.05)	357507.16 (103.66)
New – III	9093912 (98.12)	13585954 (97.88)	315275616 (94.60)	322796.81 (93.50)	364711.06 (101.61)

Table 3.1 (Continued)

Procedure	Pop. – 31	Pop. – 32	Pop. – 33	Pop. – 34	Pop. – 35
Simple Random Sampling	74491.10 (134.07)	710508.44 (88.46)	103107392 (41.58)	47647188 (646.46)	2582.05 (51.94)
Hansen – Hurwitz	105404.14 (94.75)	680122.13 (92.42)	44782804 (95.74)	323955264 (95.08)	1405.82 (95.40)
Midzuno – Sen	82039.90 (121.73)	661409.31 (95.03)	63940276 (67.06)	23977166 (1284.64)	1326.13 (101.13)
Yates Grundy (d – b – d)	98536.13 (101.35)	630053.25 (99.76)	43200928 (99.25)	285696896 (107.81)	1304.92 (102.78)
Brewer	99867.50 (100.00)	628548.56 (100.00)	42875176 (100.00)	308020000 (100.00)	1341.15 (100.00)
Durbin (Rejective)	100083.27 (99.78)	626755.63 (100.29)	43313968 (98.99)	310292864 (99.27)	1366.61 (98.14)
Yates – Grundy (Rejective)	97379.14 (107.83)	631461.30 (105.52)	43566616 (103.03)	266287760 (123.27)	1275.95 (111.67)
New – I	101085.67 (98.79)	627351.88 (100.19)	42636828 (100.56)	328738016 (93.70)	1376.70 (97.42)
New – II	103231.97 (96.74)	625298.63 (100.52)	42442840 (101.02)	364320960 (84.55)	1446.16 (92.74)
New – III	101230.53 (98.65)	627172.31 (100.22)	42611912 (100.62)	331335232 (92.96)	1382.08 (97.04)

Table 3.1 (Continued)

Procedure	Pop. – 36	Pop. – 37	Pop. – 38	Pop. – 39	Pop. – 40
Simple Random Sampling	566143.38 (6.02)	52956032 (7.99)	964.42 (137.23)	47313.47 (50.21)	770759.94 (59.22)
Hansen – Hurwitz	36479.28 (93.48)	4487930 (94.32)	1396.58 (94.76)	25019.09 (94.95)	482237.16 (94.65)
Midzuno – Sen	160931.47 (21.19)	10701504 (39.55)	1027.80 (128.76)	28879.18 (82.26)	345700.31 (132.03)
Yates Grundy (d – b – d)	34884.00 (97.76)	4021857 (105.25)	1302.01 (101.65)	23854.06 (99.58)	435082.28 (104.91)
Brewer	34102.43 (100.00)	4232811 (100.00)	1323.44 (100.00)	23755.04 (100.00)	456431.59 (100.00)
Durbin (Rejective)	35126.89 (97.08)	4397221 (96.26)	1323.73 (99.98)	24029.42 (98.86)	463041.13 (98.57)
Yates – Grundy (Rejective)	37032.77 (94.33)	3932415 (114.43)	1283.29 (75.63)	23965.81 (104.24)	417700.60 (116.62)
New – I	35172.03 (96.96)	4526397 (93.51)	1343.28 (98.52)	23710.00 (100.19)	476670.38 (95.75)
New – II	45789.69 (74.48)	5586715 (75.77)	1376.95 (96.11)	23652.76 (100.43)	517744.63 (88.16)
New – III	35608.39 (95.77)	4612076 (91.78)	1345.52 (98.36)	23707.10 (100.20)	479786.72 (95.13)

Table 3.1 (Continued)

Procedure	Pop. – 41	Pop. – 42	Pop. – 43	Pop. – 44	Pop. – 45
Simple Random Sampling	211.56 (83.86)	12693.78 (4.90)	10816.00 (109.32)	164.89 (7.21)	5305.21 (24.97)
Hansen – Hurwitz	199.48 (88.94)	750.29 (82.90)	12943.82 (91.35)	12.57 (94.61)	1377.74 (96.15)
Midzuno – Sen	187.31 (94.72)	2908.04 (21.39)	10732.41 (110.17)	47.20 (25.21)	1759.56 (75.29)
Yates Grundy (d – b – d)	177.17 (100.14)	880.51 (70.64)	10990.32 (107.58)	11.19 (106.37)	1270.41 (104.28)
Brewer	177.41 (100.00)	621.98 (100.00)	11823.61 (100.00)	11.90 (100.00)	1324.73 (100.00)
Durbin (Rejective)	177.54 (99.93)	743.99 (83.60)	12726.49 (92.91)	11.89 (100.03)	1384.64 (95.67)
Yates – Grundy (Rejective)	177.22 (106.72)	1086.12 (49.49)	10434.65 (133.20)	10.78 (114.12)	1237.41 (115.29)
New – I	177.82 (99.77)	536.93 (115.84)	12436.23 (95.07)	12.77 (93.16)	1386.21 (95.56)
New – II	180.33 (98.38)	1507.57 (41.26)	15763.95 (75.00)	14.67 (81.10)	1563.59 (84.72)
New – III	177.94 (99.70)	478.66 (129.94)	12759.32 (92.67)	12.88 (92.35)	1399.87 (94.63)

Table 3.1 (Continued)

Procedure	Pop. – 46	Pop. – 47	Pop. – 48	Pop. – 49	Pop. – 50
Simple Random Sampling	9137.78 (144.85)	268.47 (4.39)	3146908.00 (2.31)	1550119.88 (10.74)	1538.13 (10.51)
Hansen – Hurwitz	14674.05 (90.20)	12.60 (93.55)	81172.90 (89.44)	212971.75 (78.21)	173.11 (93.35)
Midzuno – Sen	2889.85 (458.03)	76.35 (15.44)	942010.75 (7.71)	330243.59 (50.43)	580.77 (27.82)
Yates Grundy (d – b – d)	11164.49 (118.56)	11.20 (105.20)	87418.94 (83.05)	130808.48 (127.33)	166.55 (97.03)
Brewer	13236.35 (100.00)	11.79 (100.00)	72601.00 (100.00)	166556.06 (100.00)	161.59 (100.00)
Durbin (Rejective)	13647.58 (96.99)	11.86 (99.39)	73507.27 (98.77)	169749.23 (98.12)	161.61 (99.99)
Yates – Grundy (Rejective)	9674.77 (158.75)	11.37 (94.07)	104570.30 (68.24)	113376.70 (169.60)	173.53 (16.21)
New – I	15011.58 (88.17)	12.99 (90.73)	65900.09 (110.17)	198599.81 (83.87)	159.87 (101.08)
New – II	21251.52 (62.28)	17.09 (68.95)	65602.80 (110.67)	408949.97 (40.73)	163.24 (98.99)
New – III	15583.45 (84.94)	13.24 (89.04)	65050.55 (111.61)	211150.94 (78.88)	159.82 (101.11)

Table 3.2 : Ranks of Sampling Variances of Various Selection procedures along with the ranks of C.V. (X) and ρ_{XY}

Pop. No.	CV	ρ_{XY}	SRS	HH	Sen-Mid	YG Dbd	Brew.	Dur. Rej.	YG. Rej.	New I	New II	New III
1.	20	7	2	9	1	4	5	6	3	7	10	8
2.	46	49	10	5	1	3	4	8	2	6	9	7
3.	38	35	10	5	1	3	4	6	2	7	9	8
4.	10	41	10	6	9	7	3	4	8	2	5	1
5.	9	10	10	8	2	3	4	5	1	6	9	7
6.	45	50	10	2	9	3	1	5	7	4	8	6
7.	44	44	9	6	8	1	2	3	4	5	-	7
8.	4	9	10	8	9	6	5	4	7	3	1	2
9.	12	21	10	8	9	6	5	4	7	3	1	2
10.	33	23	2	7	1	4	5	6	3	8	10	9
11.	39	25	10	6	8	2	3	4	1	5	9	7
12.	15	28	10	7	9	6	5	4	8	3	1	2
13.	48	29	9	7	8	2	3	5	1	4	-	6

Table 3.2 (Continued)

Pop. No.	CV	ρ_{XY}	SRS	HH	Sen-Mid	YG Dbd	Brew.	Dur. Rej.	YG Rej.	New I	New II	New III
14.	34	45	10	5	9	1	3	4	2	6	8	7
15.	37	40	10	7	9	4	1	3	6	2	8	5
16.	19	27	10	8	9	6	4	4	7	3	1	2
17.	25	12	10	8	1	3	4	6	2	5	9	7
18.	40	32	2	6	1	4	5	8	3	7	-	9
19.	43	13	2	7	1	4	5	6	3	8	10	9
20.	22	16	10	8	9	6	5	4	7	3	1	2
21.	29	33	10	7	9	2	3	1	4	5	8	6
22.	23	11	10	8	1	3	4	6	2	5	9	7
23.	18	14	10	8	1	3	4	5	2	6	9	7
24.	50	4	1	9	2	5	6	10	4	7	3	8
25.	31	30	10	7	9	2	3	4	1	5	8	6
26.	36	26	10	7	9	1	3	6	2	4	8	5

Table 3.2 (Continued)

Pop. No.	CV	ρ_{XY}	SRS	HH	Sen-Mid	YG Dbd	Brew.	Dur. Rej.	YG Rej.	New I	New II	New III
27.	21	8	10	9	7	2	3	4	1	5	8	6
28.	35	2	2	6	1	4	5	9	3	7	10	8
29.	16	20	10	8	2	3	4	5	1	6	9	7
30.	17	22	10	8	9	6	4	5	7	3	1	2
31.	13	1	1	10	2	4	5	6	3	7	9	8
32.	5	3	10	9	8	6	5	2	7	4	1	3
33.	11	24	10	8	9	5	4	6	7	3	1	2
34.	14	37	2	7	1	4	5	6	3	8	10	9
35.	27	18	10	8	3	2	4	5	1	6	9	7
36.	30	42	10	6	9	2	1	3	7	4	8	5
37.	41	38	10	5	9	2	3	4	1	6	8	7
38.	1	19	1	10	2	4	5	6	3	7	9	8
39.	24	15	10	8	9	5	4	7	6	3	1	2

Table 3.2 (Continued)

Pop. No.	CV	ρ_{XY}	SRS	HH	Sen-Mid	YG Dbd	Brew.	Dur. Rej.	YG Rej.	New I	New II	New III
40.	28	17	10	8	1	3	4	5	2	6	9	7
41.	3	5	10	9	8	1	3	4	2	5	7	6
42.	47	48	10	5	9	6	3	4	7	2	8	1
43.	49	47	3	9	2	4	5	7	1	6	10	8
44.	7	39	10	5	9	2	4	3	1	6	8	7
45.	26	34	10	4	9	2	3	5	1	6	8	7
46.	8	6	2	7	1	4	5	6	3	8	10	9
47.	6	46	10	5	9	1	3	4	2	6	8	7
48.	32	43	10	6	9	7	4	5	8	3	2	1
49.	42	31	10	7	8	2	3	4	1	5	9	6
50.	2	36	10	7	9	6	3	4	8	2	5	1

Table 3.3: Frequency Table of Ranks of Various Selection Procedures along with the Average Rank

Ranks	SRS	HH	Sen-Mid	YG Dbd	Brew	Dur. rej	YG Rej	New I	New II	New III
1	3	0	13	5	3	1	12	0	9	4
2	7	1	6	11	1	1	10	4	1	8
3	1	0	1	9	15	4	9	9	1	1
4	0	1	0	11	15	16	3	5	0	0
5	0	7	0	3	15	10	0	9	2	3
6	0	7	0	9	1	12	2	12	0	7
7	0	12	1	2	0	2	10	7	1	15
8	0	14	6	0	0	2	4	4	13	7
9	2	6	23	0	0	1	0	0	13	5
10	37	2	0	0	0	1	0	0	7	0
Average Ranks	8.16	7.06	5.80	3.62	3.82	5.00	3.70	5.06	6.85	5.72

Table 3.4: Average Ranks of Various Selection Procedures with ranks of Coefficient of Variation.

CV	SRS	HH	Sen-Mid	YG Dbd	Brew	Dur. Rej	YG Rej	New I	New II	New III
1 – 10	8.30	7.40	6.60	4.00	4.00	4.20	4.20	4.90	6.30	5.10
11 – 20	7.50	8.10	5.20	4.70	4.50	5.10	4.80	4.90	5.20	4.90
21 – 30	10.00	7.40	5.80	3.00	3.50	4.60	3.30	4.80	7.00	5.60
31 – 40	7.60	6.20	5.70	3.20	3.60	5.50	3.10	5.40	8.00	6.50
41 – 50	7.40	6.20	5.70	3.20	3.50	5.60	3.10	5.30	8.13	6.50

Table 3.5: Average Ranks of Various Selection Procedures with ranks of Correlation Coefficient.

ρ_{XY}	SRS	HH	Sen-Mid	YG Dbd	Brew	Dur. Rej	YG Rej	New I	New II	New III
1 – 10	5.80	8.40	4.10	3.90	4.60	5.60	3.40	5.90	6.80	6.50
11 – 20	8.30	8.10	3.00	3.60	4.30	5.50	2.90	5.50	7.50	6.30
21 – 30	9.10	7.30	8.00	4.00	3.90	4.80	4.40	4.10	4.44	4.30
31 – 40	8.40	6.00	6.50	3.10	3.40	4.40	3.00	5.40	8.11	6.50
41 – 50	9.20	5.50	7.40	3.50	2.90	4.70	4.80	4.40	7.33	5.00

Table 3.6: Average Ranks of various Procedures for populations having Negative Kurtosis and Negative or Positive Skewness.

	Freq.	SRS	HH	Sen-Mid	YG Dbd	Brew	Dur. Rej	YG Rej	New I	New II	New III
Positive S(X)	28	7.43	7.39	5.04	3.68	3.93	4.96	3.61	5.39	7.33	6.11
Negative S(X)	3	10.00	7.00	6.67	3.67	4.33	4.00	3.00	5.00	6.00	5.33
Combined S(X)	31	7.68	7.35	5.19	3.68	3.97	4.87	3.55	5.35	7.20	6.03

Table 3.7: Average Ranks of various Procedures for populations having Positive Kurtosis and Negative or Positive Skewness.

	Freq.	SRS	HH	Sen-Mid	YG Dbd	Brew	Dur. Rej	YG Rej	New I	New II	New III
Positive S(X)	16	8.75	6.44	6.38	3.06	3.50	5.38	3.25	4.94	7.07	5.88
Negative S(X)	3	10.00	7.33	9.00	6.00	4.00	4.33	7.67	2.67	2.33	1.67
Combined S(X)	19	8.95	6.58	6.79	3.53	3.58	5.21	3.95	4.58	6.32	5.21

Table 3.8: Correlation Coefficient between Rank of Variance of a Selection Procedure, Rank of the Correlation Coefficient and Rank of the Coefficient of Variation.

Estimator	Correlation Coefficient between Rank of Variance and Rank of C.V. (X).	Correlation Coefficient between Rank of Variance and Rank of ρ_{xy}
SRS	-0.0856	0.3145
HH	-0.3987	-0.7136
Sen-Mid	-0.0912	0.3998
YG (dbd)	-0.2560	-0.1374
Brewer	-0.2407	-0.5632
Dur. (Rej.)	0.2891	-0.2655
YG (Rej.)	-0.2279	0.1235
New I	0.1225	-0.2517
New II	0.3055	0.0927
New III	0.2738	-0.1432

Table 3.9: Regression Summary for Ranks of Various Selection Procedures for model

$$Rank\ (\Pr oc.) = \beta_0 + \beta_1 \left[Rank\ (CV)\right] + \beta_2 \left[Rank\ (\rho)\right] + \varepsilon$$

Coeffi-Cient.	SRS	HH	Sen-Mid	YG Dbd	Brew	Dur. Rej	YG Rej	New I	New II	New III
β_0	7.318	9.294	4.368	4.950	5.251	5.131	2.210	5.634	5.393	5.580
p-Value	0.000	0.000	0.000	0.000	0.000	0.000	0.008	0.000	0.000	0.000
β_1	-0.058	-0.017	-0.068	-0.035	-0.003	0.052	-0.042	0.027	0.071	0.064
p-Value	0.093	0.169	0.055	0.085	0.777	0.001	0.110	0.112	0.044	0.011
β_2	0.092	-0.071	0.124	-0.008	-0.049	-0.055	0.076	-0.047	-0.008	-0.056
p-Value	0.008	0.000	0.001	0.687	0.000	0.001	0.005	0.007	0.814	0.024
F	4.032	26.309	6.632	2.232	12.069	9.608	4.535	4.168	2.305	4.545
p-Value	0.024	0.000	0.003	0.119	0.000	0.000	0.016	0.022	0.112	0.016

3.5 Conclusions:

The empirical study of various selection procedures has been carried out along with the new selection procedures. The sampling variance of Horvitz – Thompson estimator has been calculated along with the relative efficiency of various selection procedures. These results are given in tables 3.1 and 3.2 respectively. From the analysis of these tables it can be seen that the new selection procedures II and III performs reasonably well in various populations. It has also been found that the new selection procedure III has a smaller variance in comparison with the new selection procedures I and II.

An overview of table–3.1 shows that the new selection procedure-I performs better than all other selection procedures in only a single population numbered 41. A detailed analysis of this population shows that this population has negative skewness for the Y variable. Further this population is platy – kurtic for both variables X and Y. Hence for such a population the new selection procedure-I is much more suitable as compared to the other populations.

The new selection procedure-II also performs better than all other procedure for 8% of the populations. These populations are numbered 9, 12, 16 and 20. Further study of these populations show some interesting results. Population 12 shows somewhat other results. This population has relatively high coefficient of variation for variable Y and a low correlation for variable X. Also, this population does not have a high coefficient of correlation. This is another sort of population where we can use the new selection procedure to draw a sample using unequal probability sampling without replacement. Population 16 shows the results which have quite resemblances with population 12. The results of population 20 are more interesting. This population has more than 100% coefficient of variation for variable Y, low coefficient of variation for variable X and relatively smaller value of coefficient of correlation. Thus a population having larger coefficient of variation for Y variable and low coefficient of correlation can produce efficient results with the new selection procedure.

Further analysis of table – 3.1 shows that in 8% of the populations the new selection procedure-III perform better than all other procedures included in the study. These populations are numbered as 4, 42, 48 and 50. Further analysis of these populations shows somewhat interesting results. From population 4 we can see that it has relatively low coefficient of variation for both variables and has a large value of coefficient of correlation. One recommendation for use of the new selection procedure – III is the populations having low value of coefficient of variation and larger value of correlation coefficient. Population 42 has more than 100% coefficient of variation for both of the variables and has high coefficient of correlation between variables X and Y, yet this population is producing smallest variance with the new selection procedure – III. Similar results hold for other two populations in which the new selection procedure – III is performing better than all other procedures.

The results of table – 3.1 further shows that the new selection procedures collectively perform better than the Brewer's (1963a) procedure in 28% populations. Table 3.3 contains the frequency of rank of various selection procedures along with the average rank of that procedure. From this table it can be seen that the Yates-Grundy rejective (1953) procedure clearly outperform all other procedures as its average rank is 3.08. This selection procedure is closely followed by the Yates-Grundy draw-by-draw (1953) and the Brewer (1963a) procedure. Also the Yates-Grundy rejective procedure performs better than all other procedures in 21 out of 50 populations, that is this procedure outperforms all other procedures under study in 42% of the populations. Table – 3.4 and table – 3.5 shows the average rank of various selection procedures along with the group ranks of coefficient of variation and correlation coefficient. A thorough study of table – 3.4 shows that the Yates-Grundy rejective (1953) procedure outperforms other procedures for all ranges of ranks except the highest where the Yates – Grundy draw-by-draw procedure is performing better. Table – 3.5 constitute the average ranks of various selection procedures along with the average ranks of correlation coefficients. A study of this table shows that the Yates-Grundy rejective (1953) procedure perform better than all other procedures for almost all ranks of correlation coefficient. Brewer's procedure performs better for high correlation coefficient. Also from the results of table 3.2, that contains ranks of various selection procedures the following table can be easily constructed that shows the comparative study of the new selection procedures:

Table 3.10: Comparative Study of New Selection procedures

Rank	New Method I	New Method II	New Method III
1	36	10	4
2	4	1	45
3	10	36	1

From above table it can be seen that the New Selection procedure – I clearly outperform other two new procedures.

Table 3.6 and Table 3.7 constitute the average rank of various selection procedures on the basis of Skewness and Kurtosis of the populations. From these two tables it can be readily seen that for populations having negative kurtosis and positive skewness the Yates – Grundy (1953) rejective procedure outperform other procedures involved in the study and is closely followed by the Yates – Grundy (1953) draw-by-draw and the Brewer (1963a) selection procedure. For populations having negative kurtosis and negative skewness same picture is obtained. For populations having positive kurtosis and positive skewness the Yates – Grundy (1953) draw-by-draw procedure outperform other procedures involved in the study and is followed by the Yates – Grundy (1953) rejective and Brewer (1963a) selection procedures. Further, for populations having positive kurtosis and negative skewness New selection procedure – III clearly outperform other procedures involved in the study and is closely followed by New selection procedures – I and II. In general for populations having negative kurtosis Yates – Grundy (1953) rejective procedure outperform other procedures and is followed by Yates – Grundy (1953) draw-by-draw and Brewer (1963a) selection procedures. Finally, for populations having positive kurtosis Yates – Grundy (1953) draw-by-draw procedure is best and is followed by Brewer (1963a) and Yates – Grundy (1953) rejective procedure.

Table – 3.8 shows the correlation coefficient between rank of variance of various selection procedures, rank of correlation coefficient and rank of coefficient of variation. From this table it can be seen that the rank of Hansen – Hurwitz, Yates – Grundy draw-by-draw, Brewer, Durbin rejective, New – I and New – III has negative correlation with rank of the correlation coefficient, indicating that these selection procedures will have a decrease in the variance with an increase in the correlation coefficient. Other procedures have positive correlation, indicating that the variance of these selection procedures will increase with an increase in correlation coefficient. This table further shows that the New selection procedures will perform better in populations having smaller coefficient of variation as the correlation coefficients between rank of variance of these selection procedures and coefficient of variation are positive.

Table 3.9 constitutes the values of regression coefficients and the F-statistic for the regression model:

$$Rank\ (\Pr oc.) = \beta_0 + \beta_1 [Rank\ (CV)] + \beta_2 [Rank\ (\rho)] + \varepsilon$$

This table shows very interesting results. From this table it can be seen that the regression is significant for all the selection procedures except the Yates-Grundy draw-by-draw (1953) procedure and the New selection procedure – II. Further, the coefficients of the regression model are very interesting. The β_1 coefficients for all the selection procedures are also given in this table. These coefficients show the partial effect of the coefficient of variation on the expected rank of an estimator. From these coefficients it can be seen that for majority of selection procedures this coefficient is negative and is insignificant. This coefficient is significant only for the Durbin rejective procedure and two new procedures which are procedure number II and III. This shows that the coefficient of variation has a significant effect on the expected rank of these three selection procedures. The β_2 row of this table shows the partial effect of correlation coefficient on the expected rank of a selection procedure. A close look of this row shows that the correlation coefficient has a significant effect in determination of expected rank for majority of selection procedures as this coefficient is significant for eight out of ten selection procedures under study. This coefficient is insignificant only for Yates-Grundy draw-by-draw and New procedure – II. Further, this coefficient is negative for seven out of ten procedures indicating that the expected rank of these selection procedures will decrease with an increase in the correlation coefficient. This coefficient is positive for Simple Random Sampling, Sen-Midzuno (1951) procedure and the Yates-Grundy rejective procedures indicating that with an increase in the correlation coefficient the expected rank of these selection procedures will increase.

Chapter 4

The General Selection Procedure

4.1 Introduction:

In this chapter a general selection procedure has been developed for use with the Horvitz and Thompson estimator. Some important results regarding the inclusion probabilities and joint inclusion probabilities have also been verified. The derivation of this procedure along with the verification of results has been given in the following sections.

4.2 The General Selection Procedure:

The general selection procedure is stated as:

- Select first unit with probability proportional to $\frac{a\,p_i(1-p_i)}{1-2a\,p_i}$
- Select second unit with probability proportional to size of the remaining units.

The expression for π_i is derived as:

$$\pi_i = \frac{\frac{a\,p_i(1-p_i)}{1-2a\,p_i}}{\sum_{i=1}^{N}\frac{a\,p_i(1-p_i)}{1-2a\,p_i}} + \sum_{\substack{j=1\\ j\neq i}}^{N}\frac{\frac{a\,p_j(1-p_j)}{1-2a\,p_j}}{\sum_{j=1}^{N}\frac{a\,p_j(1-p_j)}{1-2a\,p_j}}\cdot\frac{p_i}{1-p_j}$$

$$= \frac{1}{\sum_{i=1}^{N}\frac{a\,p_i(1-p_i)}{1-2a\,p_i}}\left[\frac{a\,p_i(1-p_i)}{1-2a\,p_i} + \sum_{\substack{j=1\\ j\neq i}}^{N}\frac{a\,p_j(1-p_j)}{1-2a\,p_j}\cdot\frac{p_i}{1-p_j}\right]$$

$$= \frac{1}{\sum_{i=1}^{N}\frac{a\,p_i(1-p_i)}{1-2a\,p_i}}\left[\frac{a\,p_i(1-p_i)}{1-2a\,p_i} + \sum_{\substack{j=1\\ j\neq i}}^{N}\frac{a\,p_i\,p_j}{1-2a\,p_j}\right]$$

$$= \frac{p_i}{\sum_{i=1}^{N} \frac{a\,p_i\,(1-p_i)}{1-2a\,p_i}} \left[\frac{a(1-p_i)}{1-2a\,p_i} + a \sum_{\substack{j=1 \\ j \neq i}}^{N} \frac{p_j}{1-2a\,p_j} \right]$$

$$= \frac{p_i}{\sum_{i=1}^{N} \frac{a\,p_i\,(1-p_i)}{1-2a\,p_i}} \left[\frac{a(1-p_i)}{1-2a\,p_i} + a \sum_{j=1}^{N} \frac{p_j}{1-2a\,p_j} - \frac{a\,p_i}{1-2a\,p_i} \right]$$

$$= \frac{p_i}{\sum_{i=1}^{N} \frac{a\,p_i\,(1-p_i)}{1-2a\,p_i}} \left[\frac{a(1-2\,p_i)}{1-2a\,p_i} + a \sum_{j=1}^{N} \frac{p_j}{1-2a\,p_j} \right] \qquad (4.2.1)$$

Now

$$\sum_{i=1}^{N} \frac{a\,p_i\,(1-p_i)}{1-2a\,p_i} = \sum_{i=1}^{N} \frac{p_i\,(a-a\,p_i)}{1-2a\,p_i}$$

$$= \frac{1}{2} \sum_{i=1}^{N} \frac{p_i\,(2a-2a\,p_i)}{1-2a\,p_i}$$

$$= \frac{1}{2} \left[\sum_{i=1}^{N} \frac{(2a-1)p_i}{1-2a\,p_i} + \sum_{i=1}^{N} \frac{p_i\,(1-2a\,p_i)}{1-2a\,p_i} \right]$$

$$= \frac{1}{2} \left[1 + (2a-1) \sum_{i=1}^{N} \frac{p_i}{1-2a\,p_i} \right]$$

$$= \frac{1}{2} c \qquad with\ \ c = 1 + (2a-1) \sum_{i=1}^{N} \frac{p_i}{1-2a\,p_i} \qquad (4.2.2)$$

From (4.2.1) and (4.2.2):

$$\pi_i = \frac{2\,p_i}{c} \left[\frac{a(1-2\,p_i)}{1-2a\,p_i} + a \sum_{j=1}^{N} \frac{p_j}{1-2a\,p_j} \right]$$

$$= \frac{2\,p_i}{c} \left[\frac{a-2a\,p_i}{1-2a\,p_i} + a \sum_{j=1}^{N} \frac{p_j}{1-2a\,p_j} \right]$$

$$= \frac{2\,p_i}{c} \left[\frac{(a-1)+(1-2a\,p_i)}{1-2a\,p_i} + a \sum_{j=1}^{N} \frac{p_j}{1-2a\,p_j} \right]$$

$$= \frac{2p_i}{c}\left[1+\frac{(a-1)}{1-2a\,p_i}+a\sum_{j=1}^{N}\frac{p_j}{1-2a\,p_j}\right] \; for \; p_i<\frac{1}{2a} \qquad (4.2.3)$$

The expression of π_{ij} for this selection procedure is derived as:

$$\pi_{ij} = p_i p_{j|i} + p_j p_{i|j}$$

$$= \frac{\frac{a\,p_i(1-p_i)}{1-2a\,p_i}}{\sum_{i=1}^{N}\frac{a\,p_i(1-p_i)}{1-2a\,p_i}}\cdot\frac{p_j}{1-p_i}+\frac{\frac{a\,p_j(1-p_j)}{1-2a\,p_j}}{\sum_{j=1}^{N}\frac{a\,p_j(1-p_j)}{1-2a\,p_j}}\cdot\frac{p_i}{1-p_j}$$

$$= \frac{1}{\sum_{i=1}^{N}\frac{a\,p_i(1-p_i)}{1-2a\,p_i}}\left[\frac{a\,p_i(1-p_i)}{1-2a\,p_i}\cdot\frac{p_j}{1-p_i}+\frac{a\,p_j(1-p_j)}{1-2a\,p_j}\cdot\frac{p_i}{1-p_j}\right]$$

Substituting the value of $\sum_{i=1}^{N}\frac{a\,p_i(1-p_i)}{1-2a\,p_i}$ from (4.2.2)

$$\pi_{ij} = \frac{1}{\frac{c}{2}}\left[\frac{a\,p_i\,p_j}{1-2a\,p_i}+\frac{a\,p_i\,p_j}{1-2a\,p_j}\right]$$

$$= \frac{2a\,p_i\,p_j}{c}\left[\frac{1}{1-2a\,p_i}+\frac{1}{1-2a\,p_j}\right] \qquad (4.2.4)$$

4.3 Some Useful Results for General Selection Procedure:

Result – 1 : $\sum_{i=1}^{N}\pi_i = n$ for this selection procedure.

Proof : To prove this result again consider the value of π_i:

$$\pi_i = \frac{2p_i}{c}\left[1+\frac{(a-1)}{1-2a\,p_i}+a\sum_{j=1}^{N}\frac{p_j}{1-2a\,p_j}\right] \qquad (4.2.3)$$

Summing both sides:

$$\sum_{i=1}^{N}\pi_i = \sum_{i=1}^{N}\left[\frac{2p_i}{c}\left\{1+\frac{(a-1)}{1-2a\,p_i}+a\sum_{j=1}^{N}\frac{p_j}{1-2a\,p_j}\right\}\right]$$

$$= \frac{2}{c}\sum_{i=1}^{N} p_i \left[1+\frac{(a-1)}{1-2a\,p_i}+a\sum_{j=1}^{N}\frac{p_j}{1-2a\,p_j}\right]$$

$$= \frac{2}{c}\left[1+(2a-1)\sum_{j=1}^{N}\frac{p_j}{1-2a\,p_j}\right]=2 \tag{4.3.1}$$

Result – 2 : The quantity π_{ij} , obtained under this selection procedure, satisfies the relation $\sum_{\substack{j=1\\ j\neq i}}^{N}\pi_{ij}=(n-1)\pi_i$.

Proof : Consider (4.2.4):

$$\pi_{ij} = \frac{2a\,p_i\,p_j}{c}\left[\frac{1}{1-2a\,p_i}+\frac{1}{1-2a\,p_j}\right]$$

Summing π_{ij} when $(j\neq i)$:

$$\sum_{\substack{j=1\\ j\neq i}}^{N}\pi_{ij} = \sum_{\substack{j=1\\ j\neq i}}^{N}\left[\frac{2a\,p_i\,p_j}{c}\left\{\frac{1}{1-2a\,p_i}+\frac{1}{1-2a\,p_j}\right\}\right]$$

$$= \frac{2a\,p_i}{c}\left[\frac{(1-p_i)}{1-2a\,p_i}+\sum_{j=1}^{N}\frac{p_j}{1-2a\,p_j}-\frac{p_i}{1-2a\,p_i}\right]$$

On Simplification

$$\sum_{\substack{j=1\\ j\neq i}}^{N}\pi_{ij} = \frac{2p_i}{c}\left[\frac{(a-1)+(1-2a\,p_i)}{1-2a\,p_i}+a\sum_{j=1}^{N}\frac{p_j}{1-2a\,p_j}\right]$$

$$= \frac{2p_i}{c}\left[1+\frac{(a-1)}{1-2a\,p_i}+a\sum_{j=1}^{N}\frac{p_j}{1-2a\,p_j}\right]=\pi_i \tag{4.3.2}$$

Result – 3 : The quantity π_{ij} , obtained under this selection procedure, satisfies the relation $\sum_{i=1}\sum_{\substack{j=1\\ j\neq i}}^{N}\pi_{ij}=n(n-1)$ where n is the sample size.

Proof : Take

$$\pi_{ij} = \frac{2a\,p_i\,p_j}{c}\left[\frac{1}{1-2a\,p_i}+\frac{1}{1-2a\,p_j}\right] \tag{4.2.4}$$

Applying the double summation:

$$\sum_{\substack{i=1}}^{N}\sum_{\substack{j=1\\j\neq i}}\pi_{ij} = \sum_{\substack{i=1}}^{N}\sum_{\substack{j=1\\j\neq i}}\left[\frac{2a\,p_i\,p_j}{c}\left\{\frac{1}{1-2a\,p_i}+\frac{1}{1-2a\,p_j}\right\}\right]$$

$$=\sum_{i=1}^{N}\left[\sum_{\substack{j=1\\j\neq i}}^{N}\left\{\frac{2a\,p_i\,p_j}{c}\left(\frac{1}{1-2a\,p_i}+\frac{1}{1-2a\,p_j}\right)\right\}\right] \tag{4.3.3}$$

Also from (4.3.2)

$$\sum_{\substack{j=1\\j\neq i}}^{N}\pi_{ij} = \sum_{\substack{j=1\\j\neq i}}^{N}\left[\frac{2a\,p_i\,p_j}{c}\left\{\frac{1}{1-2a\,p_i}+\frac{1}{1-2a\,p_j}\right\}\right]=\pi_i \tag{4.3.2}$$

Substituting (4.3.2) in (4.3.3) :

$$\sum_{\substack{i=1}}^{N}\sum_{\substack{j=1\\j\neq i}}\pi_{ij} = \sum_{i=1}^{N}\pi_i = 2 \tag{4.3.4}$$

Since n = 2, therefore equation (4.3.4) can be written as $\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}\pi_{ij}=n(n-1)$.

Hence the result.

Results 4: The Sen – Yates – Grundy variance estimator is always non-negative under this selection procedure:

Proof: Consider again (1.3.6). The quantities π_i and π_{ij} under the new selection procedure – III are given in (4.2.2) and (4.2.4). Now, the Sen – Yates – Grundy variance estimator is non-negative if:

$$\pi_i\,\pi_j-\pi_{ij}\geq 0$$

or $$\frac{2p_i}{c}\left[1+\frac{(a-1)}{1-2a\,p_i}+a\sum_{h=1}^{N}\frac{p_h}{1-2a\,p_h}\right]\cdot\frac{2p_j}{c}\left[1+\frac{(a-1)}{1-2a\,p_j}+a\sum_{h=1}^{N}\frac{p_h}{1-2a\,p_h}\right]$$

$$-\frac{2a\,p_i\,p_j}{c}\left[\frac{1}{1-2a\,p_i}+\frac{1}{1-2a\,p_j}\right]\geq 0$$

Writing $E=a\sum_{h=1}^{N}\frac{p_h}{(1-2a\,p_h)}$, above equation becomes:

$$\frac{2p_i\,p_j}{c}\left[\frac{2}{c}\left\{1+\frac{(a-1)}{(1-2a\,p_i)}+E\right\}\left\{1+\frac{(a-1)}{(1-2a\,p_j)}+E\right\}-a\left\{\frac{1}{(1-2a\,p_i)}+\frac{1}{(1-2a\,p_j)}\right\}\right]\geq 0$$

$$= \frac{2 p_i p_j}{c}\left[\frac{2}{c}\left\{(E+1)^2 + \frac{2(E+1)(a-1)(1-a p_i - a p_j)+(a-1)^2}{(1-2a p_i)(1-2a p_j)}\right\} - \frac{2a(1-a p_i - a p_j)}{(1-2a p_i)(1-2a p_j)}\right] \geq 0 \qquad (4.3.4)$$

Equation (4.3.4) is true for all values of p_i and p_j, making Sen – Yates – Grundy variance estimator non-negative for all samples. Further, it has been numerically checked that the relation $\pi_i \pi_j - \pi_{ij} \geq 0$ is true for all values of π_i, π_j and π_{ij} for this selection procedure.

4.4 Two Special Cases of the General Selection Procedure:

In this section two special cases of the general selection procedure have been derived for selected values of "a" by using the quantities π_i and π_{ij} of the general selection procedure. These special cases have been derived in the following sub-sections.

4.4.1 Special Case – 1:

The values of π_i and π_{ij} obtained under the general selection procedure transform to that of Yates and Grundy (1953) draw-by-draw selection procedure for $a = 0.5$.

Consider:

$$\pi_i = \frac{2 p_i}{c}\left[1 + \frac{(a-1)}{1-2a p_i} + a\sum_{j=1}^{N}\frac{p_j}{1-2a p_j}\right]$$

put a = 0.5

$$\pi_i = \frac{2 p_i}{c}\left[1 + \frac{0.5-1}{1-2(0.5)p_i} + 0.5\sum_{j=1}^{N}\frac{p_j}{1-2(0.5)p_j}\right]$$

$$= \frac{2 p_i}{c}\left[1 - \frac{0.5}{1-p_i} + 0.5\sum_{j=1}^{N}\frac{p_j}{1-p_j}\right] \qquad (4.4.1)$$

Since c will be 1 if we put $a = 0.5$, therefore (4.4.1) is:

$$\pi_i = 2 p_i\left[1 - \frac{0.5}{1-p_i} + 0.5\sum_{j=1}^{N}\frac{p_j}{1-p_j}\right]$$

$$= p_i\left[1+\frac{1-p_i-1}{1-p_i}+\sum_{j=1}^{N}\frac{p_j}{1-p_j}\right]$$

$$= p_i\left[1-\frac{p_i}{1-p_i}+\sum_{j=1}^{N}\frac{p_j}{1-p_j}\right] \qquad (4.4.2)$$

Again consider the quantity π_{ij}:

$$\pi_{ij} = \frac{2a\,p_i\,p_j}{c}\left[\frac{1}{1-2a\,p_i}+\frac{1}{1-2a\,p_j}\right]$$

putting a = 0.5 and on simplification with c = 1:

$$\pi_{ij} = p_i\,p_j\left[\frac{1}{1-p_i}+\frac{1}{1-p_j}\right] \qquad (4.4.3)$$

4.4.2 Special Case – 2:

The values of π_i and π_{ij} obtained under the general selection procedure transform to that of Brewer (1963) draw-by-draw selection procedure for a = 1.0. Consider:

$$\pi_i = \frac{2p_i}{c}\left[1+\frac{(a-1)}{1-2a\,p_i}+a\sum_{j=1}^{N}\frac{p_j}{1-2a\,p_j}\right] \qquad (4.2.4)$$

put a = 1

$$\pi_i = \frac{2p_i}{c}\left[1+\frac{1-1}{1-2p_i}+\sum_{j=1}^{N}\frac{p_j}{1-2p_j}\right]$$

$$= \frac{2p_i}{c}\left[1+\sum_{j=1}^{N}\frac{p_j}{1-2p_j}\right] \qquad (4.4.4)$$

Since $c=1+\sum_{j=1}^{N}\frac{p_j}{1-2p_j}$ for a = 1, therefore (4.4.4) is:

$$\pi_i = \frac{2p_i}{1+\sum_{j=1}^{N}\frac{p_j}{1-p_j}}\left[1+\sum_{j=1}^{N}\frac{p_j}{1-2p_j}\right]$$

$$= 2p_i \qquad (4.4.5)$$

Again consider the quantity π_{ij}:

$$\pi_{ij} = \frac{2a\,p_i\,p_j}{c}\left[\frac{1}{1-2a\,p_i} + \frac{1}{1-2a\,p_j}\right]$$

Substituting a = 1 and $c = 1 + \sum_{j=1}^{N} \frac{p_j}{1-2\,p_j}$

$$\pi_{ij} = \frac{2\,p_i\,p_j}{k}\left[\frac{1}{1-2\,p_i} + \frac{1}{1-2\,p_j}\right] \quad with\ k = 1 + \sum_{j=1}^{N} \frac{p_j}{1-2\,p_j} \qquad (4.4.6)$$

4.5 Empirical Study:

In this section the empirical study has been carried out for the general selection procedure by using fifty natural populations selected from standard text on sampling. The empirical study has been carried out by using various values of the constant "*a*" in the range on 0.5 to 1.5 with an increment of 0.1. The ranking has also been done in order to decide the nature of populations where the general selection procedure can be applied with a specified value of "*a*". The results of these analyses have been given in the following tables.

Table 4.1: Sampling Variance of Horvitz – Thompson estimator for Various values of "a" using selected Natural Populations.

Values of Constant "a"	Pop. – 1	Pop. – 2	Pop. – 3	Pop. – 4	Pop. – 5
a = 0.5	13383.07	1520483.75	246963.45	333.75	6132.15
a = 0.6	13497.68	1534528.25	250341.36	321.11	6171.69
a = 0.7	13616.46	1549658.38	253977.75	308.90	6215.10
a = 0.8	13739.66	1566014.38	257901.52	297.23	6262.82
a = 0.9	13867.49	1583758.00	262145.56	286.26	6315.37
a = 1.0	14000.22	1603085.00	266747.50	276.14	6373.32
a = 1.1	14138.12	1624226.63	271751.16	267.08	6437.35
a = 1.2	14281.50	1647462.00	277207.19	259.33	6508.22
a = 1.3	14430.67	1673133.50	283174.53	253.16	6586.83
a = 1.4	14585.99	1701662.50	289722.84	248.95	6674.24
a = 1.5	14747.84	1733575.75	296933.41	247.11	6771.64

Table 4.1 (Continued)

Values of Constant "a"	Pop. – 6	Pop. – 7	Pop. – 8	Pop. – 9	Pop. – 10
$a = 0.5$	54243.95	10489.04	465838.94	135870.34	53669716
$a = 0.6$	53025.25	10210.17	465359.03	135197.66	54549584
$a = 0.7$	52050.27	10098.45	464884.16	134510.59	55465080
$a = 0.8$	51381.81	10222.15	464414.78	133808.81	56418620
$a = 0.9$	51098.03	10678.13	463951.97	133091.88	57412752
$a = 1.0$	51296.66	11603.40	463496.97	132359.27	58450276
$a = 1.1$	52100.54	13184.52	463051.06	131610.63	59534312
$a = 1.2$	53664.91	15636.17	462614.97	130845.49	60668280
$a = 1.3$	56187.07	18991.09	462191.06	130063.41	61855988
$a = 1.4$	59919.01	21723.39	461780.44	129263.82	63101636
$a = 1.5$	65184.17	9770.08	461384.81	128446.45	64409928

Table 4.1 (Continued)

Values of Constant "a"	Pop. – 11	Pop. – 12	Pop. – 13	Pop. – 14	Pop. – 15
a = 0.5	26741372	8830.26	14681897	417418.81	181002.19
a = 0.6	27291148	8767.15	15196228	418652.88	179699.84
a = 0.7	27889582	8703.08	15780313	421258.88	178632.94
a = 0.8	28542522	8638.01	16448373	425500.47	177825.02
a = 0.9	29256668	8571.95	17218672	431701.28	177301.70
a = 1.0	30039964	8504.88	18114750	440262.09	177091.88
a = 1.1	30901648	8436.76	19166550	451682.13	177227.36
a = 1.2	31852838	8367.59	20409810	466589.47	177743.89
a = 1.3	32906718	8297.34	21877178	485781.94	178681.78
a = 1.4	34079228	8226.01	23552600	510283.00	180086.02
a = 1.5	35389744	8153.56	25144146	541420.38	182007.31

Table 4.1 (Continued)

Values of Constant "a"	Pop. – 16	Pop. – 17	Pop. – 18	Pop. – 19	Pop. – 20
$a = 0.5$	30.87	388.89	508257984	2131433344	1248.41
$a = 0.6$	30.57	391.09	526722016	2168050944	1242.30
$a = 0.7$	30.29	393.48	546602048	2206245376	1236.01
$a = 0.8$	30.01	396.08	568096832	2246127104	1229.52
$a = 0.9$	29.74	398.92	591447424	2287819008	1222.81
$a = 1.0$	29.48	402.02	616952000	2331453952	1215.89
$a = 1.1$	29.23	405.43	644981248	2377181184	1208.74
$a = 1.2$	29.00	409.19	676000448	2425169152	1201.36
$a = 1.3$	28.79	413.35	710589568	2475605504	1193.71
$a = 1.4$	28.61	417.99	749441408	2528702208	1185.81
$a = 1.5$	28.45	423.17	793241024	2584696064	1177.63

Table 4.1 (Continued)

Values of Constant "a"	Pop. – 21	Pop. – 22	Pop. – 23	Pop. – 24	Pop. – 25
a = 0.5	4553.31	22444.18	1410.09	1204.55	6204300
a = 0.6	4545.39	22555.23	1418.63	1230.76	6253134
a = 0.7	4542.34	22671.29	1427.57	1259.55	6306956
a = 0.8	4544.69	22792.72	1436.93	1291.46	6366241
a = 0.9	4553.04	22919.91	1446.75	1327.18	6431506
a = 1.0	4568.07	23053.29	1457.04	1367.58	6503351
a = 1.1	4590.53	23193.33	1467.85	1413.74	6582434
a = 1.2	4621.28	23340.58	1479.20	1466.77	6669514
a = 1.3	4661.27	23495.64	1491.15	1526.66	6765439
a = 1.4	4711.60	23659.17	1503.72	1586.23	6871190
a = 1.5	4773.51	23831.88	1516.98	1590.18	6987877

Table 4.1 (Continued)

Values of Constant "a"	Pop. – 26	Pop. – 27	Pop. – 28	Pop. – 29	Pop. – 30
$a = 0.5$	8816018	13042031	282992928	282880.78	376569.28
$a = 0.6$	8821950	13086805	285837152	286286.84	375399.41
$a = 0.7$	8834315	13134573	288778656	289870.41	374216.50
$a = 0.8$	8854238	13185566	291823648	293642.16	373020.03
$a = 0.9$	8883080	13240030	294979296	297614.16	371809.94
$a = 1.0$	8922493	13298242	298253152	301799.28	370586.00
$a = 1.1$	8974521	13360508	301654080	306211.41	369347.84
$a = 1.2$	9041691	13427179	305191616	310865.47	368094.84
$a = 1.3$	9127172	13498605	308876672	315777.63	366827.03
$a = 1.4$	9234999	13575222	312721568	320965.78	365544.22
$a = 1.5$	9370339	13657478	316740352	326448.91	364245.59

Table 4.1 (Continued)

Values of Constant "a"	Pop. – 31	Pop. – 32	Pop. – 33	Pop. – 34	Pop. – 35
a = 0.5	98536.13	630053.25	43200928	285696896	1304.92
a = 0.6	98787.30	629759.88	43131948	289909472	1311.52
a = 0.7	99045.73	629462.69	43064748	294242880	1318.42
a = 0.8	99311.65	629161.69	42999424	298702048	1325.65
a = 0.9	99585.52	628856.94	42936160	303292704	1333.22
a = 1.0	99867.48	628548.56	42875184	308019968	1341.15
a = 1.1	100158.16	628236.38	42816632	312890272	1349.47
a = 1.2	100457.72	627920.88	42760800	317909472	1358.19
a = 1.3	100766.70	627602.06	42707856	323084352	1367.34
a = 1.4	101085.45	627280.06	42658088	328422048	1376.96
a = 1.5	101414.48	626955.13	42611776	333930176	1387.06

Table 4.1 (Continued)

Values of Constant "a"	Pop. – 36	Pop. – 37	Pop. – 38	Pop. – 39	Pop. – 40
$a = 0.5$	34884.00	4021856.50	1302.01	23854.06	435082.28
$a = 0.6$	34584.60	4052012.25	1306.06	23832.94	438996.38
$a = 0.7$	34345.88	4087455.75	1310.23	23812.38	443077.50
$a = 0.8$	34177.74	4128862.50	1314.51	23792.49	447336.06
$a = 0.9$	34091.97	4177011.00	1318.91	23773.36	451783.34
$a = 1.0$	34102.43	4232810.50	1323.44	23755.04	456431.72
$a = 1.1$	34225.77	4297319.00	1328.09	23737.70	461294.97
$a = 1.2$	34481.86	4371787.50	1332.89	23721.47	466387.72
$a = 1.3$	34894.75	4457698.00	1337.83	23706.49	471726.59
$a = 1.4$	35493.79	4556808.50	1342.92	23692.88	477328.94
$a = 1.5$	36314.87	4671240.00	1348.16	23680.87	483214.72

Table 4.1 (Continued)

Values of Constant "a"	Pop. – 41	Pop. – 42	Pop. – 43	Pop. – 44	Pop. – 45
$a = 0.5$	177.17	880.51	10990.32	11.19	1270.41
$a = 0.6$	177.18	832.63	11129.63	11.30	1279.24
$a = 0.7$	177.22	782.37	11280.97	11.43	1288.99
$a = 0.8$	177.26	729.96	11445.84	11.57	1299.75
$a = 0.9$	177.33	676.00	11626.02	11.73	1311.63
$a = 1.0$	177.41	621.98	11823.62	11.90	1324.73
$a = 1.1$	177.52	571.31	12041.13	12.08	1339.18
$a = 1.2$	177.66	532.08	12281.61	12.29	1355.15
$a = 1.3$	177.82	524.98	12548.74	12.51	1372.79
$a = 1.4$	178.02	607.95	12847.09	12.75	1392.30
$a = 1.5$	178.26	860.12	13182.36	13.01	1413.90

Table 4.1 (Continued)

Values of Constant "a"	Pop. – 46	Pop. – 47	Pop. – 48	Pop. – 49	Pop. – 50
a = 0.5	11164.49	11.20	87418.94	130808.48	166.55
a = 0.6	11523.22	11.25	84203.74	136057.69	165.36
a = 0.7	11906.97	11.33	81093.89	142112.92	164.27
a = 0.8	12318.34	11.44	78108.34	149104.73	163.27
a = 0.9	12760.31	11.59	75268.91	157189.39	162.37
a = 1.0	13236.35	11.79	72601.00	166556.06	161.59
a = 1.1	13750.44	12.02	70134.16	177434.09	160.94
a = 1.2	14307.23	12.31	67903.05	190103.67	160.43
a = 1.3	14912.19	12.66	65948.48	204907.66	160.07
a = 1.4	15571.81	13.06	64318.77	222266.80	159.88
a = 1.5	16293.81	13.53	63071.63	242695.02	159.87

Table 4.2 : Ranks of various values of "a" along with the ranks of C.V. (X) and ρ_{XY}

Pop No.	CV	ρ_{XY}	Values of Constant "*a*"										
			0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5
1.	20	7	1	2	3	4	5	6	7	8	9	10	11
2.	46	49	1	2	3	4	5	6	7	8	9	10	11
3.	38	35	1	2	3	4	5	6	7	8	9	10	11
4.	10	41	11	10	9	8	7	6	5	4	3	2	1
5.	9	10	1	2	3	4	5	6	7	8	9	10	11
6.	45	50	8	6	4	3	1	2	5	7	9	10	11
7.	44	44	5	3	2	4	6	7	8	9	10	11	1
8.	4	9	11	10	9	8	7	6	5	4	3	2	1
9.	12	21	11	10	9	8	7	6	5	4	3	2	1
10.	33	23	1	2	3	4	5	6	7	8	9	10	11
11.	39	25	1	2	3	4	5	6	7	8	9	10	11
12.	15	28	11	10	9	8	7	6	5	4	3	2	1
13.	48	29	1	2	3	4	5	6	7	8	9	10	11

Table 4.2 (Continued)

Pop No.	CV	ρ_{XY}	Values of Constant "*a*"										
			0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5
14.	34	45	1	2	3	4	5	6	7	8	9	10	11
15.	37	40	10	8	6	5	3	1	2	4	7	9	11
16.	19	27	11	10	9	8	7	6	5	4	3	2	1
17.	25	12	1	2	3	4	5	6	7	8	9	10	11
18.	40	32	1	2	3	4	5	6	7	8	9	10	11
19.	43	13	1	2	3	4	5	6	7	8	9	10	11
20.	22	16	11	10	9	8	7	6	5	4	3	2	1
21.	29	33	5	3	1	2	4	6	7	8	9	10	11
22.	23	11	1	2	3	4	5	6	7	8	9	10	11
23.	18	14	1	2	3	4	5	6	7	8	9	10	11
24.	50	4	1	2	3	4	5	6	7	8	9	10	11
25.	31	30	1	2	3	4	5	6	7	8	9	10	11
26.	36	26	1	2	3	4	5	6	7	8	9	10	11

Table 4.2 (Continued)

Pop No.	CV	ρ_{XY}	Values of Constant "*a*"										
			0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5
27.	21	8	1	2	3	4	5	6	7	8	9	10	11
28.	35	2	1	2	3	4	5	6	7	8	9	10	11
29.	16	20	1	2	3	4	5	6	7	8	9	10	11
30.	17	22	11	10	9	8	7	6	5	4	3	2	1
31.	13	1	1	2	3	4	5	6	7	8	9	10	11
32.	5	3	11	10	9	8	7	6	5	4	3	2	1
33.	11	24	11	10	9	8	7	6	5	4	3	2	1
34.	14	37	1	2	3	4	5	6	7	8	9	10	11
35.	27	18	1	2	3	4	5	6	7	8	9	10	11
36.	30	42	8	7	5	3	1	2	4	6	9	10	11
37.	41	38	1	2	3	4	5	6	7	8	9	10	11
38.	1	19	1	2	3	4	5	6	7	8	9	10	11
39.	24	15	11	10	9	8	7	6	5	4	3	2	1

Table 4.2 (Continued)

Pop No.	CV	ρ_{XY}	Values of Constant "*a*"										
			0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5
40.	28	17	1	2	3	4	5	6	7	8	9	10	11
41.	3	5	1	2	3	4	5	6	7	8	9	10	11
42.	47	48	11	9	8	7	6	5	3	2	1	4	10
43.	49	47	1	2	3	4	5	6	7	8	9	10	11
44.	7	39	1	2	3	4	5	6	7	8	9	10	11
45.	26	34	1	2	3	4	5	6	7	8	9	10	11
46.	8	6	1	2	3	4	5	6	7	8	9	10	11
47.	6	46	1	2	3	4	5	6	7	8	9	10	11
48.	32	43	11	10	9	8	7	6	5	4	3	2	1
49.	42	31	1	2	3	4	5	6	7	8	9	10	11
50.	2	36	11	10	9	8	7	6	5	4	3	2	1

Table 4.3: Frequency Table of Ranks of General Selection Procedures for various "a" along with the Average Rank

Ranks	Values of Constant "a"										
	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5
1	32	-	1	-	2	1	-	-	1	-	13
2	-	32	1	1	-	2	1	1	-	12	-
3	-	2	32	1	1	-	1	-	12	-	-
4	-	-	1	33	1	-	1	13	-	1	-
5	2	-	1	1	32	1	13	-	-	-	-
6	-	1	1	-	2	45	-	1	-	-	-
7	-	1	-	1	12	1	33	1	1	-	-
8	2	1	1	12	-	-	1	33	-	-	-
9	-	1	12	-	-	-	-	1	35	1	-
10	1	12	-	-	-	-	-	-	1	35	1
11	13	-	-	-	-	-	-	-	-	1	36
Average Ranks	4.22	4.40	4.60	4.96	5.30	5.74	6.26	6.80	7.38	7.96	8.38

Table 4.4: Average Ranks of Various Values of "*a*" with ranks of Coefficient of Variation.

CV	Values of Constant "*a*"										
	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5
1 – 10	5.0	5.2	5.4	5.6	5.8	6.0	6.2	6.4	6.6	6.8	7.0
11 – 20	6.0	6.0	6.0	6.0	6.0	6.0	6.0	6.0	6.0	6.0	6.0
21 – 30	4.1	4.2	4.2	4.5	4.9	5.6	6.3	7.0	7.8	8.4	9.0
31 – 40	2.9	3.4	3.9	4.5	5.0	5.5	6.3	7.2	8.2	9.1	10.0
41 – 50	3.1	3.2	3.5	4.2	4.8	5.6	6.5	7.4	8.3	9.5	9.9

Table 4.5: Average Ranks of Various Values of "*a*" with ranks of Correlation Coefficient.

ρ_{XY}	Values of Constant "*a*"										
	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5
1 – 10	3.0	3.6	4.2	4.8	5.4	6.0	6.6	7.2	7.8	8.4	9.0
11 – 20	3.0	3.6	4.2	4.8	5.4	6.0	6.6	7.2	7.8	8.4	9.0
21 – 30	6.0	6.0	6.0	6.0	6.0	6.0	6.0	6.0	6.0	6.0	6.0
31 – 40	3.3	3.5	3.7	4.3	4.9	5.5	6.3	7.2	8.2	9.1	10.0
41 – 50	5.8	5.3	4.9	4.9	4.8	5.2	5.8	6.4	7.1	7.9	7.9

Table 4.6: Average Ranks of Values of "*a*" for populations having Negative Kurtosis and Negative or Positive Skewness.

		Values of Constant "*a*"										
	Freq.	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5
Positive S(X)	28	3.86	4.14	4.39	4.79	5.18	5.68	6.29	6.93	7.64	8.25	8.86
Negative S(X)	3	4.33	4.67	5.00	5.33	5.67	6.00	6.33	6.67	7.00	7.33	7.67
Combined S(X)	31	3.91	4.19	4.45	4.84	5.23	5.71	6.29	6.90	7.58	8.16	8.74

Table 4.7: Average Ranks of Values of "*a*" for populations having Positive Kurtosis and Negative or Positive Skewness.

		Values of Constant "*a*"										
	Freq.	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5
Positive S(X)	16	3.56	3.75	4.06	4.63	5.13	5.75	6.44	7.13	7.81	8.69	9.06
Negative S(X)	3	11.00	10.00	9.00	8.00	7.00	6.00	5.00	4.00	3.00	2.00	1.00
Combined S(X)	19	4.73	4.74	4.84	5.16	5.43	5.79	6.21	6.64	7.05	7.63	7.79

Table 4.8: Regression Summary for Ranks of Various Values of "*a*" for model

$$Rank\left(a.\right)=\beta_o+\beta_1\left[Rank\left(CV\right)\right]+\beta_2\left[Rank\left(\rho\right)\right]+\varepsilon$$

Coeffi-Cient.	Values of Constant "*a*"										
	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5
β_0	4.608	5.178	5.661	5.926	6.274	6.443	6.458	6.447	6.393	6.225	6.385
p-Value	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
β_1	-0.108	-0.093	-0.076	-0.052	-0.031	-0.073	0.019	0.045	0.073	0.104	0.126
p-Value	0.021	0.011	0.002	0.006	0.023	0.508	0.154	0.024	0.012	0.004	0.006
β_2	0.093	0.063	0.034	0.014	-0.077	-0.020	-0.026	-0.032	-0.034	-0.036	-0.048
p-Value	0.047	0.082	0.206	0.431	0.557	0.068	0.045	0.111	0.227	0.299	0.283
F	3.586	3.830	4.065	4.215	4.074	2.854	2.370	3.045	3.451	4.548	4.081
p-Value	0.036	0.029	0.024	0.021	0.023	0.068	0.105	0.057	0.040	0.016	0.023

4.6 Conclusions:

The results of empirical study of the general selection procedure have been given in previous section. These results constitute variance of Horvitz – Thompson estimator for various values of the constant "a", the ranks of sampling variance in various populations for various values of "a", the average ranks in correspondence with ranks of coefficient of variation, the average ranks in correspondence of the correlation coefficient, the frequency of ranks of various values of "a" and the regression summary. These results have been given in table – 4.1 to table – 4.6. The conclusions drawn from these results have been given in following pages.

Table – 4.1 constitute the sampling variance of Horvitz – Thompson estimator for various values of "a". From analysis of this table it can be easily seen that in some populations the sampling variance of Horvitz – Thompson estimator increases with an increase in the value of "a". These populations are numbered 1, 2, 3, 5, 10, 11, 13, 14, 17, 18, 19, 22 to 29, 31, 34, 35, 37, 38, 40, 41, 43 to 47 and 49. Further, study of these populations show that they exhibit somewhat similar trend. All the above mentioned populations, except 44, have positive skewness in the measure of size. Also the coefficient of skewness is positive in most of the populations for the basic variable of study. Further, the coefficient of kurtosis is negative for basic variable of study, Y, in most of the above mentioned populations. From these it is suggested that a smaller value of "a" should be used when a population has positive skewness in both measure of size and basic variable of study and negative kurtosis for variable Y. The analysis of coefficient of variation and correlation coefficient for above mentioned populations shows that these populations have larger coefficient of variation along with relatively high correlation coefficient. One recommendation for the use of smaller value of "a" in general selection procedure is that the population must have smaller coefficient of variation along with a relatively low correlation coefficient.

Also, from the analysis of this table it is easily seen that the sampling variance of Horvitz – Thompson estimator starts decreasing with an increase in value of "a". These populations are numbered 4, 8, 9, 12, 16, 20, 30, 31, 32, 33, 39, 42, 48 and 50. Further, study of these populations reveals some interesting results. It can be seen that the nature of skewness in measure of size has a mixed trend. Also the coefficient of skewness for basic variable of study has a mixed trend. It can also be seen that the larger value of "a" produces a smaller variance when the measure of size and basic variable of study has positive skewness and basic variable of study has negative kurtosis. The analysis of coefficient of variation and correlation coefficient reveals that these populations has haphazard trend for both coefficient of variation and correlation coefficient. One recommendation for the use of a larger value of "a" in the general selection procedure is that the direction of coefficient of variation and correlation must be opposite or correlation between these two must be negative.

Some populations in table – 4.1 show very interesting results. These populations are numbered 6, 7, 15 and 21. In these populations larger as well as smaller value of "a" increases the sampling variance of Horvitz – Thompson estimator. In these populations a middle value, around 1.0, of "a" produces smaller variance as compared to the smaller or larger values of "a". These populations have one thing in common, that is all these populations have positive skewness for both measure of size and basic variable of study. Also the kurtosis for both variables is positive in first two populations and negative in last two populations. From these it is concluded that the moderate value of "a" will produce a smaller variance if both the variable have positive skewness and same kurtosis. Further, the analysis of coefficient of variation and correlation coefficient shows that all these populations have larger correlation coefficient.

Table – 4.2 shows the ranks of sampling variance of Horvitz – Thomson estimator for various values of "a". From this table it can be seen that in 11 populations the rank decreases with increase in "a", in 35 cases rank increases with increase in "a" whereas in 4 cases there is no relation between the value of "a" and the rank. Table 4.3 shows the frequency of ranks for various values of "a". This table shows a very interesting picture. From this table it can be seen that the frequency of higher ranks increases with an increase in the value of constant "a". The frequency of rank 1 is maximum for a = 0.5 and the maximum frequency moves to the higher ranks with an increase in the value of the constant "a". This table also shows the average rank of various values of "*a*". From this table it can be seen that the average rank increases with an increase in the value of "*a*". Also from table 4.3 following table can be easily constructed that shows performance of various values of the constant "*a*":

Table 4.9: Frequency of Best Performance of various values of the Constant "*a*"

	Values of Constant "*a*"										
	0.5	**0.6**	**0.7**	**0.8**	**0.9**	**1.0**	**1.1**	**1.2**	**1.3**	**1.4**	**1.5**
Frequency of Best	32	-	1	-	2	1	-	-	1	-	13

From table 4.9 it can be readily seen that a = 0.5 clearly dominate other values involved in the study.

Table 4.4 shows the average ranks along with the group of ranks of coefficient of variations. From the analysis of table – 4.4 it is readily seen that a smaller value of "*a*" is suitable if the coefficient of variation has a smaller value. Further, analysis of this table shows that a value of "*a*" between 0.5 and 1.0 will produce a smaller variance for almost all coefficients of variation.

Table – 4.5 shows the average rank for various values of "*a*" along with the group ranks of correlation coefficients. A view of this table shows that the value of "*a*" between 0.5 and 1.0 will produce smaller sampling variance in comparison with other values. Also a = 0.5 outperforms all other values larger than it for small coefficient of correlation.

Table 4.6 and Table 4.7 constitute the average rank of various values of "*a*" on the basis of skewness and Kurtosis of the populations. From these two tables it can be readily seen that the smaller values of "*a*" have smaller average ranks for all situations. Further the average rank of populations having negative kurtosis is smaller than that of having positive kurtosis.

Table – 4.8 constitutes the regression summary for the model:

$$Rank\ (\text{Pr}\ oc.) = \beta_0 + \beta_1 \left[Rank\ (CV)\right] + \beta_2 \left[Rank\ (\rho)\right] + \varepsilon$$

This table constitutes the regression coefficients and the F-statistic for testing significance of the regression along with the p-values. The regression coefficients β_1 are also given in this table. These coefficients have a negative sign for the values of "a" from 0.5 to 1.0. This indicates that the expected rank of Horvitz – Thompson estimator will decrease with an increase in the rank of coefficient of variation for a = 0.5 to a = 1.0. Further, the regression coefficients β_1 have a positive sign for the values of "a" from 1.1 to onward. This indicates that the expected rank of the Horvitz – Thompson estimator will increase with an increase in the coefficient of variation for a = 1.1 to a = 1.5. From this discussion it can be concluded that the smaller values of "a" should be used in the General Selection Procedure for populations having larger coefficient of variation and larger values of "a" should be used for populations having smaller coefficient of variation for variable X.

Further, table – 4.8 constitutes the regression coefficients β_2. These coefficients show the partial effect of the correlation coefficient on the expected rank of the variance of Horvitz – Thompson estimator for various choices of the constant "a". From this table it can be seen that this coefficient has a negative sign for the values of "a" from 0.9 to onward. This indicates that the expected rank of the variance of Horvitz – Thompson estimator will decrease with an increase in the rank of correlation coefficient for larger values of the constant "a". Further, the expected rank will increase with an increase in the rank of correlation coefficient for smaller values of "a" as the regression coefficient β_2 has a positive sign for the values of "a" from 0.5 to 0.8. From this discussion it can be concluded that the larger value of "a" will be useful in the General Selection Procedure for the populations having larger correlation coefficient.

The overall conclusion that can be drawn from the study of table – 4.6 is that a larger value of "a" should be used for populations having larger correlation coefficient and smaller value of coefficient of variation and vice versa.

Chapter 5

Modified Murthy Estimators

5.1 Introduction:

In this chapter a series of new estimators has been developed following the idea of Murthy (1957). These estimators have been developed by using different selection procedures in the general Murthy estimator given as:

$$t_{symm} = \frac{1}{P(S)} \sum_{i=1}^{n} P(S|i) y_i \qquad (1.5.10)$$

Murthy (1957) used the Yates – Grundy draw – by – draw procedure in estimator (1.5.10) to obtain following unbiased estimator of population total for a sample of size 2:

$$t_{symm} = \frac{\left[\frac{y_i}{p_i}(1-p_j) + \frac{y_j}{p_j}(1-p_i)\right]}{(2-p_i-p_j)} \qquad (1.5.11)$$

Three new estimators have been developed by using various selection procedures in the estimator given in equation (1.5.10). A fourth estimator has been proposed following the idea of Raj (1956a).

5.2 Modified Murthy Estimators:

In this section and the following sub – sections some new estimators has been developed that can be used with unequal probability sample without replacement to estimate the population total. These estimators have been developed by using Brewer, Durbin and New selection procedure – III in the general Murthy estimator given in equation (1.7.10).

5.2.1 Modified Murthy Estimator – I:

This estimator has been developed by using the Brewer (1963a) selection procedure in the general Murthy estimator given as:

$$t_{symm} = \frac{1}{P(S)} \sum_{i=1}^{n} P(S|i) y_i \qquad (1.5.10)$$

Above estimator for a sample of size 2 is given as:

$$t_{symm} = \frac{P(S|i) y_i + P(S|j) y_j}{P(S)} \qquad (5.2.1)$$

Now, for Brewer selection procedure:

$$P(S|i) = \frac{p_j}{1 - p_i} \ , \quad P(S|j) = \frac{p_i}{1 - p_j}$$

and $$P(S) = \frac{2 p_i p_j}{k} \left[\frac{1}{1 - 2 p_i} + \frac{1}{1 - 2 p_j} \right] = \frac{4 p_i p_j (1 - p_i - p_j)}{k (1 - 2 p_i)(1 - 2 p_j)}$$

where $k = 1 + \sum_{i=1}^{N} \frac{p_i}{1 - 2 p_i}$

Substituting these values in equation (5.2.1):

$$t_{MM1} = \frac{y_i \frac{p_j}{1 - p_i} + y_j \frac{p_i}{1 - p_j}}{\frac{4 p_i p_j (1 - p_i - p_j)}{k (1 - 2 p_i)(1 - 2 p_j)}}$$

$$= \frac{p_i p_j \left[\frac{y_i}{p_i (1 - p_i)} + \frac{y_j}{p_j (1 - p_j)} \right]}{\frac{4 p_i p_j (1 - p_i - p_j)}{k (1 - 2 p_i)(1 - 2 p_j)}}$$

$$= \frac{\left[\frac{y_i}{p_i} (1 - p_j) + \frac{y_j}{p_j} (1 - p_i) \right]}{(1 - p_i)(1 - p_j)} \cdot \frac{k (1 - 2 p_i)(1 - 2 p_j)}{4 (1 - p_i - p_j)}$$

$$= \frac{k (1 - 2 p_i)(1 - 2 p_j) \left[\frac{y_i}{p_i} (1 - p_j) + \frac{y_j}{p_j} (1 - p_i) \right]}{4 (1 - p_i)(1 - p_j)(1 - p_i - p_j)} \qquad (5.2.2)$$

This estimator is a slight modification of the usual Murthy estimator given in equation (1.5.11).

5.2.2 Modified Murthy Estimator – II:

The modified Murthy estimator – II has been developed by using the Durbin (1967) selection procedure in the general Murthy estimator given in equation (5.2.1). Now, for Durbin (1967) selection procedure:

$$P(S|i) = \frac{p_j}{k}\left[\frac{1}{1-2p_i} + \frac{1}{1-2p_j}\right] = \frac{2p_j(1-p_i-p_j)}{k(1-2p_i)(1-2p_j)}$$

$$P(S|j) = \frac{p_i}{k}\left[\frac{1}{1-2p_i} + \frac{1}{1-2p_j}\right] = \frac{2p_i(1-p_i-p_j)}{k(1-2p_i)(1-2p_j)}$$

and $$P(S) = \frac{2p_i p_j}{k}\left[\frac{1}{1-2p_i} + \frac{1}{1-2p_j}\right] = \frac{4p_i p_j(1-p_i-p_j)}{k(1-2p_i)(1-2p_j)}$$

Now, substituting above values in equation (5.2.1) the estimator of population total for use with unequal probability sampling without replacement is:

$$t_{MM2} = \frac{\left[y_i \frac{2p_j(1-p_i-p_j)}{k(1-2p_i)(1-2p_j)} + y_j \frac{2p_i(1-p_i-p_j)}{k(1-2p_i)(1-2p_j)}\right]}{\frac{4p_i p_j(1-p_i-p_j)}{k(1-2p_i)(1-2p_j)}}$$

$$= \frac{k(1-2p_i)(1-2p_j)}{4p_i p_j(1-p_i-p_j)}\left[\frac{2y_i p_j(1-p_i-p_j) + 2y_j p_i(1-p_i-p_j)}{k(1-2p_i)(1-2p_j)}\right]$$

$$= \frac{2y_i p_j(1-p_i-p_j) + 2y_j p_i(1-p_i-p_j)}{4p_i p_j(1-p_i-p_j)}$$

$$= \frac{2p_i p_j(1-p_i-p_j)\left[\frac{y_i}{p_i} + \frac{y_j}{p_j}\right]}{4p_i p_j(1-p_i-p_j)}$$

$$= \frac{1}{2}\left[\frac{y_i}{p_i} + \frac{y_j}{p_j}\right] \tag{5.2.3}$$

This estimator is same as used by Durbin (1967) for his rejective selection procedure.

5.2.3 Modified Murthy Estimator – III:

This estimator has been developed by using the new selection procedure – III in the general Murthy estimator given in equation (5.2.1). For new selection procedure – III, the conditional probabilities and the probability of sample are given as:

$$P(S|i)=\frac{p_j}{k}\left[\frac{1}{1-2p_i}+\frac{1}{1-2p_j}\right]=\frac{2p_j(1-p_i-p_j)}{k(1-2p_i)(1-2p_j)} \tag{5.2.4}$$

$$P(S|j)=\frac{p_i}{k}\left[\frac{1}{1-2p_i}+\frac{1}{1-2p_j}\right]=\frac{2p_i(1-p_i-p_j)}{k(1-2p_i)(1-2p_j)} \tag{5.2.5}$$

and $$P(S)=\frac{p_i p_j}{k^2}\left[\frac{1}{1-2p_i}+\frac{1}{1-2p_j}\right]\left[2+\frac{1}{1-2p_i}+\frac{1}{1-2p_j}\right]$$

$$=\frac{4p_i p_j(1-p_i-p_j)\{(1-p_i-p_j)+(1-2p_i)(1-2p_j)\}}{k^2(1-2p_i)^2(1-2p_j)^2} \tag{5.2.6}$$

Now, substituting the values from equation (5.2.4), (5.2.5) and (5.2.6) in equation (5.2.1) we have:

$$t_{MM3}=\frac{\left[y_i\frac{2p_j(1-p_i-p_j)}{k(1-2p_i)(1-2p_j)}+y_j\frac{2p_i(1-p_i-p_j)}{k(1-2p_i)(1-2p_j)}\right]}{\frac{4p_i p_j(1-p_i-p_j)\{(1-p_i-p_j)+(1-2p_i)(1-2p_j)\}}{k^2(1-2p_i)^2(1-2p_j)^2}}$$

$$=\left[\frac{2y_i p_j(1-p_i-p_j)+2y_j p_i(1-p_i-p_j)}{k(1-2p_i)(1-2p_j)}\right]$$

$$\cdot\frac{k^2(1-2p_i)^2(1-2p_j)^2}{4p_i p_j(1-p_i-p_j)[(1-p_i-p_j)+(1-2p_i)(1-2p_j)]}$$

$$=\frac{2kp_i p_j(1-p_i-p_j)(1-2p_i)(1-2p_j)\left[\frac{y_i}{p_i}+\frac{y_j}{p_j}\right]}{4p_i p_j(1-p_i-p_j)[(1-p_i-p_j)+(1-2p_i)(1-2p_j)]}$$

$$=\frac{k(1-2p_i)(1-2p_j)}{2[(1-p_i-p_j)+(1-2p_i)(1-2p_j)]}\left[\frac{y_i}{p_i}+\frac{y_j}{p_j}\right] \tag{5.2.7}$$

This estimator is a slight modification of the estimator given in equation (5.2.3).

5.2.4 Modified Raj Estimator:

This estimator has been developed following the idea of Raj (1956a). Raj (1956a) proposed a series of estimators based on the ordered samples. The estimators proposed by Raj (1956a) has a general shape as:

$$t_n = \sum_{i=1}^{n-1} y_i + \frac{y_n}{p_n}\left(1 - \sum_{i=1}^{n-1} p_i\right) \tag{1.7.4}$$

Raj (1956a) proposed that the average of above series of estimators can be used as an unbiased estimator of population total. Raj's estimator for n = 2 is given as:

$$t_2 = y_1 + \frac{y_2}{p_2}(1 - p_1) \tag{5.2.8}$$

Parallel to above estimator a modified estimator of population total is given as:

$$t_{RM} = y_j + \frac{y_i}{p_i}(1 - p_i)\left(b - \frac{p_j}{1 - p_j}\right) \tag{5.2.9}$$

where $b = \sum_{i=1}^{N} \frac{p_i}{1 - p_i}$

5.3 Unbiasedness of The Modified Estimators:

In this section the unbiasedness of modified estimators, developed in the previous section, has been proved. The unbiasedness has been proved by using directly the estimator and the probability of sample. The unbiasedness of all three estimators have been proved in the following subsections.

5.3.1 The Modified Murthy Estimator – I:

The unbiasedness of the modified Murthy estimator – I has been proved in this subsection as under:

Now $t_{MM1} = \dfrac{k(1-2p_i)(1-2p_j)\left[\frac{y_i}{p_i}(1-p_j) + \frac{y_j}{p_j}(1-p_i)\right]}{4(1-p_i)(1-p_j)(1-p_i-p_j)}$

also $\quad P(S) = \frac{2p_i p_j}{k}\left[\frac{1}{1-2p_i} + \frac{1}{1-2p_j}\right] = \frac{4p_i p_j(1-p_i-p_j)}{k(1-2p_i)(1-2p_j)}$

Now

$$E\left(t_{MM1}\right)=\sum_{\substack{i=1\\j>i}}^{N}\sum_{j=1}^{N} t_{MM1}\,P(S)$$

$$=\frac{1}{2}\sum_{\substack{i=1\\j\neq i}}^{N}\sum_{j=1} t_{MM1}\,P(S)$$

$$=\frac{1}{2}\sum_{\substack{i=1\\j\neq i}}^{N}\sum_{j=1}\left[\frac{k\left(1-2P_i\right)\left(1-2P_j\right)\left[\frac{Y_i}{P_i}\left(1-P_j\right)+\frac{Y_j}{p_j}\left(1-P_i\right)\right]}{4\left(1-P_i\right)\left(1-P_j\right)\left(1-P_i-P_j\right)}\cdot\frac{4P_iP_j\left(1-P_i-P_j\right)}{k\left(1-2P_i\right)\left(1-2P_j\right)}\right]$$

$$=\frac{1}{2}\sum_{\substack{i=1\\j\neq i}}^{N}\sum_{j=1}\frac{P_iP_j\left[\frac{Y_i}{P_i}\left(1-P_j\right)+\frac{Y_j}{P_j}\left(1-P_i\right)\right]}{\left(1-P_i\right)\left(1-P_j\right)}$$

$$=\frac{1}{2}\sum_{\substack{i=1\\j\neq i}}^{N}\sum_{j=1}P_iP_j\left[\frac{Y_i}{P_i\left(1-P_i\right)}+\frac{Y_j}{P_j\left(1-P_j\right)}\right]$$

$$=\frac{1}{2}\sum_{\substack{i=1\\j\neq i}}^{N}\sum_{j=1}\left[\frac{Y_i}{1-P_i}P_j+\frac{Y_j}{1-P_j}P_i\right]$$

$$=\frac{1}{2}\left[\sum_{i=1}^{N}\frac{Y_i}{1-P_i}\sum_{\substack{j=1\\j\neq i}}^{N}P_j+\sum_{j=1}^{N}\frac{Y_j}{1-P_j}\sum_{\substack{i=1\\i\neq j}}^{N}P_i\right]$$

$$=\frac{1}{2}\left[\sum_{i=1}^{N}\frac{Y_i}{1-P_i}\left(1-P_i\right)+\sum_{j=1}^{N}\frac{Y_j}{1-P_j}\left(1-P_j\right)\right]$$

$$=\frac{1}{2}\left[\sum_{i=1}^{N}Y_i+\sum_{j=1}^{N}Y_j\right]$$

$$=\frac{1}{2}\left[Y+Y\right]$$

$$=\frac{1}{2}*2Y\ \Rightarrow\ E\left(t_{MM1}\right)=Y$$

From above result it can be seen that the modified Murthy estimator – I is unbiased for population total.

5.3.2 The Modified Murthy Estimator – II:

In this subsection the unbiasedness of modified Murthy estimator – II has been proved.

Now $t_{MM2} = \frac{1}{2}\left[\frac{y_i}{p_i} + \frac{y_j}{p_i}\right]$

also $P(S) = \frac{2\,p_i\,p_j}{k}\left[\frac{1}{1-2\,p_i} + \frac{1}{1-2\,p_j}\right] = \frac{4\,p_i\,p_j\left(1-p_i-p_j\right)}{k\left(1-2\,p_i\right)\left(1-2\,p_j\right)}$

So

$$E\left(t_{MM2}\right) = \sum_{i=1}^{N}\sum_{\substack{j=1\\ j>i}}^{N} t_{MM2}\,P(S)$$

$$= \frac{1}{2}\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}} t_{MM2}\,P(S)$$

$$= \frac{1}{2}\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}\left[\frac{1}{2}\left\{\frac{Y_i}{P_i}+\frac{Y_j}{P_j}\right\}\cdot\frac{4\,P_i\,P_j\left(1-P_i-P_j\right)}{k\left(1-2P_i\right)\left(1-2P_j\right)}\right]$$

$$= \frac{1}{2}\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}\left[\frac{1}{2}\left\{\frac{Y_i}{P_i}+\frac{Y_j}{P_j}\right\}\cdot\frac{2\,P_i\,P_j\left(2-2\,P_i-2\,P_j\right)}{k\left(1-2\,P_i\right)\left(1-2\,P_j\right)}\right]$$

$$= \frac{1}{2}\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}\left[\left\{\frac{Y_i}{P_i}+\frac{Y_i}{P_j}\right\}\cdot\frac{P_i\,P_j\left\{\left(1-2\,P_i\right)+\left(1-2\,P_j\right)\right\}}{k\left(1-2\,P_i\right)\left(1-2\,P_j\right)}\right]$$

$$= \frac{1}{2}\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}\left[\left\{Y_i\,P_i+Y_j\,P_i\right\}\cdot\frac{\left\{\left(1-2\,P_i\right)+\left(1-2\,P_j\right)\right\}}{k\left(1-2\,P_i\right)\left(1-2\,P_j\right)}\right]$$

$$= \frac{1}{2}\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}\left[\left\{Y_i\,P_j+Y_j\,P_i\right\}\cdot\left\{\frac{\left(1-2\,P_i\right)}{k\left(1-2\,P_i\right)\left(1-2\,P_j\right)}+\frac{\left(1-2\,P_j\right)}{k\left(1-2\,P_i\right)\left(1-2\,P_j\right)}\right\}\right]$$

$$= \frac{1}{2}\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}\left[\frac{Y_i\,P_j}{k\left(1-2\,P_i\right)}+\frac{Y_i\,P_j}{k\left(1-2\,P_j\right)}+\frac{Y_j\,P_i}{k\left(1-2\,P_i\right)}+\frac{Y_j\,P_i}{k\left(1-2\,P_j\right)}\right]$$

$$= \frac{1}{2}\left[\sum_{i=1}^{N}\frac{Y_i}{k\left(1-2P_i\right)}\sum_{\substack{j=1\\ j\neq i}}^{N}P_j+\sum_{j=1}^{N}\frac{P_j}{k\left(1-2P_j\right)}\sum_{\substack{i=1\\ i\neq j}}^{N}Y_i\right.$$

$$\left.+\sum_{i=1}^{N}\frac{P_i}{k\left(1-2P_i\right)}\sum_{\substack{j=1\\ j\neq i}}^{N}Y_j+\sum_{j=1}^{N}\frac{Y_j}{k\left(1-2P_j\right)}\sum_{\substack{i=1\\ i\neq j}}^{N}P_i\right]$$

$$= \frac{1}{2}\left[\sum_{i=1}^{N}\frac{Y_i}{k\left(1-2P_i\right)}\left(1-P_i\right)+\sum_{j=1}^{N}\frac{P_j}{k\left(1-2P_j\right)}\left(Y-Y_j\right)\right.$$

$$\left.+\sum_{i=1}^{N}\frac{P_i}{k\left(1-2P_i\right)}\left(Y-Y_i\right)+\sum_{j=1}^{N}\frac{Y_j}{k\left(1-2P_j\right)}\left(1-P_j\right)\right]$$

$$= \frac{1}{2k}\left[\sum_{i=1}^{N}\frac{Y_i}{\left(1-2P_i\right)}\left(1-P_i\right)+\sum_{j=1}^{N}\frac{P_j}{\left(1-2P_j\right)}\left(Y-Y_j\right)\right.$$

$$\left.+\sum_{i=1}^{N}\frac{P_i}{\left(1-2P_i\right)}\left(Y-Y_i\right)+\sum_{j=1}^{N}\frac{Y_j}{\left(1-2P_j\right)}\left(1-P_j\right)\right]$$

$$= \frac{1}{2k}\left[\sum_{i=1}^{N}\frac{Y_i}{\left(1-2P_i\right)}-\sum_{i=1}^{N}\frac{Y_iP_i}{\left(1-2P_i\right)}+Y\sum_{j=1}^{N}\frac{P_j}{\left(1-2P_j\right)}-\sum_{j=1}^{N}\frac{Y_jP_j}{\left(1-2P_j\right)}\right.$$

$$\left.+Y\sum_{i=1}^{N}\frac{P_i}{\left(1-2P_i\right)}-\sum_{i=1}^{N}\frac{Y_iP_i}{\left(1-2P_i\right)}+\sum_{j=1}^{N}\frac{Y_j}{\left(1-2P_j\right)}-\sum_{j=1}^{N}\frac{Y_jP_j}{\left(1-2P_j\right)}\right]$$

$$= \frac{1}{2k}\left[2Y\sum_{i=1}^{N}\frac{P_i}{\left(1-2P_i\right)}+2\sum_{i=1}^{N}\frac{Y_i}{\left(1-2P_i\right)}-4\sum_{i=1}^{N}\frac{Y_iP_i}{\left(1-2P_i\right)}\right]$$

$$= \frac{1}{2k}\left[2Y\sum_{i=1}^{N}\frac{P_i}{\left(1-2P_i\right)}+2\sum_{i=1}^{N}\frac{Y_i}{\left(1-2P_i\right)}\left(1-2P_i\right)\right]$$

$$= \frac{1}{2k}\left[2Y\sum_{i=1}^{N}\frac{P_i}{\left(1-2P_i\right)}+2\sum_{i=1}^{N}Y_i\right]$$

$$= \frac{1}{2k}\left[2Y\sum_{i=1}^{N}\frac{P_i}{\left(1-2P_i\right)}+2Y\right]$$

$$= \frac{1}{2k}\left[2Y\left\{1+\sum_{i=1}^{N}\frac{P_i}{1-2P_i}\right\}\right]$$

$$= \frac{1}{2k}[2kY] \Rightarrow E(t_{MM2}) = Y$$

which is unbiased. It is worth to quote here that the same estimator, when used by Durbin (1967) with his rejective selection procedure, turned out to be a biased estimator of population total Y.

5.3.3 The Modified Murthy Estimator – III:

In this subsection the unbiasedness of modified Murthy estimator – III has been proved. From subsection 5.2.3:

$$t_{MM3} = \frac{k(1-2p_i)(1-2p_j)}{2[(1-p_i-p_j)+(1-2p_i)(1-2p_j)]}\left[\frac{y_i}{p_i}+\frac{y_j}{p_j}\right]$$

and
$$P(S) = \frac{4p_i p_j(1-p_i-p_j)\{(1-p_i-p_j)+(1-2p_i)(1-2p_j)\}}{k^2(1-2p_i)^2(1-2p_j)^2}$$

Now, the unbiasedness of t_{MM3} has been proved as under:

$$E(t_{MM3}) = \sum_{i-1}^{N}\sum_{\substack{j-1\\ j>i}}^{N} t_{MM3}\,P(S)$$

$$= \frac{1}{2}\sum_{i-1}^{N}\sum_{\substack{j-1\\ j\neq i}} t_{MM3}\,P(S)$$

$$= \frac{1}{2}\sum_{i-1}^{N}\sum_{\substack{j=1\\ j\neq i}}\left[\left\{\frac{k(1-2P_i)(1-2P_j)}{2[(1-P_i-P_j)+(1-2P_i)(1-2P_j)]}\left(\frac{Y_i}{P_i}+\frac{Y_j}{P_j}\right)\right\}\right.$$

$$\left.\cdot\left\{\frac{4P_iP_j(1-P_i-P_j)\{(1-P_i-P_j)+(1-2P_i)(1-2P_j)\}}{k^2(1-2P_i)^2(1-2P_j)^2}\right\}\right]$$

$$= \frac{1}{2}\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}\left[\left\{\frac{Y_i}{P_i}+\frac{Y_j}{P_j}\right\}\cdot\left\{\frac{2P_iP_j(1-P_i-P_j)}{k(1-2P_i)(1-2P_j)}\right\}\right]$$

$$= \frac{1}{2}\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}\left[\left\{\frac{Y_i}{P_i}+\frac{Y_j}{P_j}\right\}\cdot\frac{P_i P_j\left(2-2P_i-2P_j\right)}{k\left(1-2P_i\right)\left(1-2P_j\right)}\right]$$

$$= \frac{1}{2}\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}\left[\left\{\frac{Y_i P_j+Y_j P_i}{P_i P_j}\right\}\cdot\frac{P_i P_j\left\{\left(1-2P_i\right)+\left(1-2P_j\right)\right\}}{k\left(1-2P_i\right)\left(1-2P_j\right)}\right]$$

$$= \frac{1}{2}\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}\left[\left\{Y_i P_j+Y_j P_i\right\}\cdot\left\{\frac{\left(1-2P_i\right)}{k\left(1-2P_i\right)\left(1-2P_j\right)}+\frac{\left(1-2P_j\right)}{k\left(1-2P_i\right)\left(1-2P_j\right)}\right\}\right]$$

$$= \frac{1}{2}\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}\left[\left\{Y_i P_j+Y_j P_i\right\}\cdot\left\{\frac{1}{k\left(1-2P_j\right)}+\frac{1}{k\left(1-2P_i\right)}\right\}\right]$$

$$= \frac{1}{2}\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}\left[\frac{Y_i P_j}{k\left(1-2P_i\right)}+\frac{Y_i P_j}{k\left(1-2P_j\right)}+\frac{Y_j P_i}{k\left(1-2P_i\right)}+\frac{Y_j P_i}{k\left(1-2P_j\right)}\right]$$

$$= \frac{1}{2}\left[\sum_{i=1}^{N}\frac{Y_i}{k\left(1-2P_i\right)}\sum_{\substack{j=1\\j\neq i}}^{N}P_j+\sum_{j=1}^{N}\frac{P_j}{k\left(1-2P_j\right)}\sum_{\substack{i=1\\i\neq j}}^{N}Y_i\right.$$

$$\left.+\sum_{i=1}^{N}\frac{P_i}{k\left(1-2P_i\right)}\sum_{\substack{j=1\\j\neq i}}^{N}Y_j+\sum_{j=1}^{N}\frac{Y_j}{k\left(1-2P_j\right)}\sum_{\substack{i=1\\i\neq j}}^{N}P_i\right]$$

$$= \frac{1}{2k}\left[\sum_{i=1}^{N}\frac{Y_i}{\left(1-2P_i\right)}\left(1-P_i\right)+\sum_{j=1}^{N}\frac{P_j}{\left(1-2P_j\right)}\left(Y-Y_j\right)\right.$$

$$\left.+\sum_{i=1}^{N}\frac{P_i}{\left(1-2P_i\right)}\left(Y-Y_i\right)+\sum_{j=1}^{N}\frac{Y_j}{\left(1-2P_j\right)}\left(1-P_j\right)\right]$$

$$= \frac{1}{2k}\left[\sum_{i=1}^{N}\frac{Y_i}{\left(1-2P_i\right)}-\sum_{i=1}^{N}\frac{Y_i P_i}{\left(1-2P_i\right)}+Y\sum_{j=1}^{N}\frac{P_j}{\left(1-2P_j\right)}-\sum_{j=1}^{N}\frac{Y_j P_j}{\left(1-2P_j\right)}\right.$$

$$\left.+Y\sum_{i=1}^{N}\frac{P_i}{\left(1-2P_i\right)}-\sum_{i=1}^{N}\frac{Y_i P_i}{\left(1-2P_i\right)}+\sum_{j=1}^{N}\frac{Y_j}{\left(1-2P_j\right)}-\sum_{j=1}^{N}\frac{Y_j P_j}{\left(1-2P_j\right)}\right]$$

$$= \frac{1}{2k}\left[2Y\sum_{i=1}^{N}\frac{P_i}{\left(1-2P_i\right)}+2\sum_{i=1}^{N}\frac{Y_i}{\left(1-2P_i\right)}-4\sum_{i=1}^{N}\frac{Y_i P_i}{\left(1-2P_i\right)}\right]$$

$$= \frac{1}{2k}\left[2Y\sum_{i=1}^{N}\frac{P_i}{(1-2P_i)}+2\sum_{i=1}^{N}\frac{Y_i}{(1-2P_i)}(1-2P_i)\right]$$

$$= \frac{1}{2k}\left[2Y\sum_{i=1}^{N}\frac{P_i}{(1-2P_i)}+2Y\right]$$

$$= \frac{1}{2k}\left[2Y\left\{1+\sum_{i=1}^{N}\frac{P_i}{1-2P_i}\right\}\right]$$

$$= \frac{1}{2k}[2kY] \;\Rightarrow\; E(t_{MM3}) = Y$$

which proves the unbiasedness of modified Murthy estimator – III.

5.3.4 The Modified Raj Estimator:

In this sub – section the unbiasedness of the modified Raj estimator has been proved. To prove that unbiasedness consider the estimator given in equation (5.11) as:

$$t_{RM} = y_j + \frac{y_i}{p_i}(1-p_i)\left(b-\frac{p_j}{1-p_j}\right)$$

Since above is an ordered estimator, therefore for this estimator:

$$P(S)=\frac{p_i\,p_j}{1-p_j}$$

Now to prove unbiasedness of t_{RM} we proceed as under:

$$E(t_{RM}) = \sum_{\substack{i=1}}^{N}\sum_{\substack{j=1\\ j\neq i}} t_{RM}\,P(S)$$

$$= \sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}\left[Y_j+\frac{Y_i}{P_i}(1-P_i)\left(b-\frac{P_j}{1-P_j}\right)\right]\frac{P_i\,P_j}{1-P_j}$$

$$= \sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}\left[Y_j+Y_i\left(\frac{1-P_i}{P_i}\right)\left(b-\frac{P_j}{1-P_j}\right)\right]\frac{P_i\,P_j}{1-P_j}$$

$$= \sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}Y_j\frac{P_i\,P_j}{1-P_i}+\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}\left[Y_i\left(\frac{1-P_i}{P_i}\right)\left(b-\frac{P_j}{1-P_j}\right)\right]\frac{P_i\,P_j}{1-P_j}$$

$$= \sum_{\substack{i=1 \\ j \neq i}}^{N} \sum_{j=1}^{N} Y_j \frac{P_i P_j}{1-P_i} + \sum_{\substack{i=1 \\ j \neq i}}^{N} \sum_{j=1}^{N} Y_i \left(b - \frac{P_j}{1-P_j} \right) P_j$$

$$= \sum_{\substack{i=1 \\ j \neq i}}^{N} \sum_{j=1}^{N} Y_j \frac{P_i P_j}{1-P_i} + \sum_{\substack{i=1 \\ j \neq i}}^{N} \sum_{j=1}^{N} b Y_i P_j - \sum_{\substack{i=1 \\ j \neq i}}^{N} \sum_{j=1}^{N} Y_i \frac{P_j^2}{1-P_j}$$

$$= \sum_{j=1}^{N} Y_j P_j \sum_{\substack{i=1 \\ i \neq j}}^{N} \frac{P_i}{1-P_i} + \sum_{i=1}^{N} b Y_i \sum_{\substack{j=1 \\ j \neq i}}^{N} P_j - \sum_{i=1}^{N} Y_i \sum_{\substack{j=1 \\ j \neq i}}^{N} \frac{P_j^2}{1-P_j}$$

$$= \sum_{j=1}^{N} Y_j P_j \left[\sum_{i=1}^{N} \frac{P_i}{1-P_i} - \frac{P_j}{1-P_j} \right] + b \sum_{i=1}^{N} Y_i (1-P_i) - \sum_{i=1}^{N} Y_i \left[\sum_{j=1}^{N} \frac{P_j^2}{1-P_j} - \frac{P_i^2}{1-P_i} \right]$$

$$= \sum_{j=1}^{N} Y_j P_j \left[b - \frac{P_j}{1-P_j} \right] + b \sum_{i=1}^{N} Y_i (1-P_i) - \sum_{i=1}^{N} Y_i \left[\sum_{j=1}^{N} \frac{P_j^2}{1-P_j} - \frac{P_i^2}{1-P_i} \right]$$

$$= b \sum_{i=1}^{N} Y_i - \sum_{i=1}^{N} Y_i \sum_{i=1}^{N} \frac{P_i^2}{1-P_i}$$

$$= \sum_{i=1}^{N} Y_i \sum_{i=1}^{N} \frac{P_i}{1-P_i} - \sum_{i=1}^{N} Y_i \sum_{i=1}^{N} \frac{P_i^2}{1-P_i}$$

$$= \sum_{i=1}^{N} Y_i \left[\sum_{i=1}^{N} \frac{P_i}{1-P_i} - \sum_{i=1}^{N} \frac{P_i^2}{1-P_i} \right]$$

$$= \sum_{i=1}^{N} Y_i \left[\sum_{i=1}^{N} \frac{P_i}{1-P_i} (1-P_i) \right]$$

$$= \sum_{i=1}^{N} Y_i \sum_{i=1}^{N} P_i$$

$$= \sum_{i=1}^{N} Y_i \quad \Rightarrow \quad E(t_{RM}) = Y$$

Chapter 6

Design Based Study

6.1 Introduction:

A series of new estimator have been developed in chapter 5 of this thesis. In this chapter the design based study of these estimators has been carried out. The design based variance of all four estimators has been obtained along with the empirical study of these estimators with some of the famous estimators available in the literature. These studies have been given in the following sections and subsections.

6.2 Design Based Variances:

6.2.1 Modified Murthy Estimator – I:

In this subsection variance of the modified Murthy estimator – I has been developed. Now design based variance of this estimator is given as:

$$Var(t_{MM1}) = E_D(t_{MM1}^2) - [E_D(t_{MM1})]^2$$

$$= \sum_{\substack{i=1 \\ j>i}}^{N}\sum_{j=1} t_{MM1}^2 P(S) - Y^2 \qquad \because E_D(t_{MM1}) = Y$$

$$= \frac{1}{2}\sum_{\substack{i=1 \\ j\neq i}}^{N}\sum_{j=1} t_{MM1}^2 P(S) - Y^2$$

$$= \frac{1}{2}\sum_{\substack{i=1 \\ j\neq i}}^{N}\sum_{j=1}\left[\frac{k(1-2P_i)(1-2P_j)\left\{\frac{Y_i}{P_i}(1-P_j)+\frac{Y_j}{P_j}(1-P_i)\right\}}{4(1-P_i)(1-P_j)(1-P_i-P_j)}\right]^2$$

$$\cdot\frac{4P_iP_j(1-P_i-P_j)}{k(1-2P_i)(1-2P_j)} - \left[\sum_{i=1}^{N} Y_i\right]^2$$

$$= \frac{1}{2}\sum_{\substack{i=1\\ j\neq i}}^{N}\sum_{j=1}\left[\frac{k^2(1-2P_i)^2(1-2P_j)^2\left\{\frac{Y_i}{P_i}(1-P_j)+\frac{Y_j}{P_j}(1-P_i)\right\}^2}{4^2(1-P_i)^2(1-P_j)^2(1-P_i-P_j)^2}\right]$$

$$\cdot\frac{4P_i\,P_j(1-P_i-P_j)}{k(1-2P_i)(1-2P_j)}-\sum_{i=1}^{N}Y_i^2-\sum_{\substack{i=1\\ j\neq i}}^{N}\sum_{j=1}Y_i\,Y_j$$

$$= \frac{1}{2}\sum_{\substack{i=1\\ j\neq i}}^{N}\sum_{j=1}\left[\frac{k\,P_i\,P_j(1-2P_i)(1-2P_j)\left\{\frac{Y_i}{P_i}(1-P_j)+\frac{Y_j}{P_j}(1-P_i)\right\}^2}{4(1-P_i)^2(1-P_j)^2(1-P_i-P_j)}\right]-\sum_{i=1}^{N}Y_i^2-\sum_{\substack{i=1\\ j\neq i}}^{N}\sum_{j=1}Y_i\,Y$$

$$= \frac{1}{2}\sum_{\substack{i=1\\ j\neq i}}^{N}\sum_{j=1}\left[\frac{k\,P_i\,P_j(1-2P_i)(1-2P_j)}{4(1-P_i-P_j)}\left\{\frac{Y_i}{P_i(1-P_i)}+\frac{Y_j}{P_j(1-P_j)}\right\}^2\right]-\sum_{i-1}^{N}Y_i^2-\sum_{\substack{i=1\\ j\neq i}}^{N}\sum_{j=1}Y_i\,Y_j$$

$$= \frac{1}{2}\left[\sum_{\substack{i=1\\ j\neq i}}^{N}\sum_{j=1}\frac{k\,P_i\,P_j(1-2P_i)(1-2P_j)}{4(1-P_i-P_j)}\left\{\frac{Y_i}{P_i(1-P_i)}+\frac{Y_j}{P_j(1-P_j)}\right\}^2-2\sum_{i=1}^{N}Y_i^2-2\sum_{\substack{i-1\\ j\neq i}}^{N}\sum_{j=1}Y_i\,Y_j\right]$$

$$= \frac{1}{2}\left[\sum_{\substack{i=1\\ j\neq i}}^{N}\sum_{j=1}\frac{k\,P_i\,P_j(1-2P_i)(1-2P_j)}{4(1-P_i-P_j)}\left\{\frac{Y_i^2}{P_i^2(1-P_i)^2}+\frac{Y_j^2}{P_j^2(1-P_{ji})^2}\right.\right.$$

$$\left.\left.+2\frac{Y_i\,Y_j}{P_i\,P_j(1-P_i)(1-P_j)}\right\}-\sum_{i=1}^{N}Y_i^2-\sum_{j=1}^{N}Y_j^2-2\sum_{\substack{i=1\\ j\neq i}}^{N}\sum_{j=1}Y_i\,Y_j\right]$$

$$= \frac{1}{2}\left[\sum_{\substack{i=1\\ j\neq i}}^{N}\sum_{j=1}\frac{k\,P_i\,P_j(1-2P_i)(1-2P_j)}{4(1-P_i-P_j)}\left\{\frac{Y_i^2}{P_i^2(1-P_i)^2}+\frac{Y_j^2}{P_j^2(1-P_{ji})^2}\right.\right.$$

$$\left.\left.+2\frac{Y_i\,Y_j}{P_i\,P_j(1-P_i)(1-P_j)}\right\}-\sum_{\substack{i=1\\ j\neq i}}^{N}\sum_{j=1}\frac{Y_i^2}{1-P_i}P_j-\sum_{\substack{i=1\\ j\neq i}}^{N}\sum_{j=1}\frac{Y_j^2}{1-P_j}P_i-2\sum_{\substack{i-1\\ j\neq i}}^{N}\sum_{j=1}Y_i\,Y_j\right]$$

$$= \frac{1}{2}\sum_{\substack{i=1 \\ j\neq i}}^{N}\sum_{j=1} \frac{P_i P_j}{4(1-P_i-P_j)}\left[\frac{Y_i^2 .k(1-2P_i)(1-2P_j)}{P_i^2(1-P_i)^2} + \frac{Y_j^2 .k(1-2P_i)(1-2P_j)}{P_j^2(1-P_j)^2}\right.$$

$$+2\frac{Y_i Y_j .k(1-2P_i)(1-2P_j)}{P_i P_j(1-P_i)(1-P_j)} - \frac{Y_i^2 .4(1-P_i-P_j)}{P_i(1-P_i)} - \frac{Y_j^2 .4(1-P_i-P_j)}{P_j(1-P_j)}$$

$$\left. -2\frac{Y_i Y_j .4(1-P_i-P_j)}{P_i P_j}\right]$$

$$= \frac{1}{2}\sum_{\substack{i=1 \\ j\neq i}}^{N}\sum_{j-1} \frac{P_i P_j}{4(1-P_i-P_j)}\left[\frac{Y_i^2 .k(1-2P_i)(1-2P_j)}{P_i^2(1-P_i)^2} - \frac{Y_i^2 .4(1-P_i-P_j)}{P_i(1-P_i)}\right.$$

$$+\frac{Y_j^2 .k(1-2P_i)(1-2P_j)}{P_j^2(1-P_j)^2} - \frac{Y_j^2 .4(1-P_i-P_j)}{P_j(1-P_j)} - 2\frac{Y_i Y_j .4(1-P_i-P_j)}{P_i P_j}$$

$$\left. +2\frac{Y_i Y_j .k(1-2P_i)(1-2P_j)}{P_i P_j(1-P_i)(1-P_j)}\right]$$

$$= \frac{1}{2}\sum_{\substack{i=1 \\ j\neq i}}^{N}\sum_{j-1} \frac{P_i P_i}{4(1-P_i-P_j)}\left[\frac{Y_i^2}{P_i^2}\left\{\frac{k(1-2P_i)(1-2P_j)}{(1-P_i)^2} - \frac{4P_i(1-P_i-P_j)}{(1-P_i)}\right\}\right.$$

$$+\frac{Y_j^2}{P_j^2}\left\{\frac{k(1-2P_i)(1-2P_j)}{(1-P_j)^2} - \frac{4P_j(1-P_i-P_j)}{(1-P_j)}\right\}$$

$$\left. -2\frac{Y_i Y_i}{P_i P_j}\left\{4(1-P_i-P_i) - \frac{k(1-2P_i)(1-2P_j)}{(1-P_i)(1-P_j)}\right\}\right]$$

$$\Rightarrow \mathrm{V}ar(t_{MM1}) = \frac{1}{2}\sum_{\substack{i=1 \\ j\neq i}}^{N}\sum_{j=1} \frac{P_i P_i}{4(1-P_i-P_j)}\left[\frac{Y_i^2}{P_i^2}A_{ij} + \frac{Y_j^2}{P_j^2}B_{ij} - 2\frac{Y_i Y_j}{P_i P_j}C_{ij}\right] \tag{6.2.1}$$

where $$A_{ij} = \frac{k(1-2P_i)(1-2P_j)}{(1-P_i)^2} - \frac{4P_i(1-P_i-P_j)}{(1-P_i)} \tag{6.2.2}$$

$$B_{ij} = \frac{k(1-2P_i)(1-2P_j)}{(1-P_j)^2} - \frac{4P_j(1-P_i-P_j)}{(1-P_j)} \tag{6.2.3}$$

$$C_{ij} = 4(1-P_i-P_j) - \frac{k(1-2P_i)(1-2P_j)}{(1-P_i)(1-P_j)} \tag{6.2.4}$$

6.2.2 Modified Murthy Estimator – II:

The design based variance of modified Murthy estimator – II has been obtained in this subsection. To obtain the variance let us proceed as under:

$$\operatorname{Var}(t_{MM2}) = E_D\left(t_{MM2}^2\right) - \left[E_D\left(t_{MM2}\right)\right]^2$$

$$= \sum_{i=1}^{N}\sum_{\substack{j=1\\ j>i}}^{N} t_{MM2}^2\, P(S) - Y^2 \qquad \because\ E_D\left(t_{MM2}\right) = Y$$

$$= \frac{1}{2}\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N} t_{MM2}^2\, P(S) - Y^2$$

$$= \frac{1}{2}\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N}\left[\frac{1}{2}\left\{\frac{Y_i}{P_i}+\frac{Y_j}{P_j}\right\}\right]^2 \cdot \frac{4P_iP_j\left(1-P_i-P_j\right)}{k\left(1-2P_i\right)\left(1-2P_j\right)} - \left[\sum_{i=1}^{N} Y_i\right]^2$$

$$= \frac{1}{2}\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N}\left[\frac{1}{4}\left\{\frac{Y_i}{P_i}+\frac{Y_j}{P_j}\right\}^2\right] \cdot \frac{4P_iP_j\left(1-P_i-P_j\right)}{k\left(1-2P_i\right)\left(1-2P_j\right)} - \sum_{i=1}^{N} Y_i^2 - \sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N} Y_iY_j$$

$$= \frac{1}{2}\left[\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N}\frac{P_iP_j\left(1-P_i-P_j\right)}{k\left(1-2P_i\right)\left(1-2P_j\right)}\left\{\frac{Y_i}{P_i}+\frac{Y_j}{P_j}\right\}^2 - 2\sum_{i=1}^{N} Y_i^2 - 2\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N} Y_iY_j\right]$$

$$= \frac{1}{2}\left[\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N}\frac{P_iP_j\left(1-P_i-P_j\right)}{k\left(1-2P_i\right)\left(1-2P_j\right)}\left\{\frac{Y_i^2}{P_i^2}+\frac{Y_j^2}{P_j^2}+2\frac{Y_iY_j}{P_iP_j}\right\} - \sum_{i=1}^{N} Y_i^2 - \sum_{j=1}^{N} Y_j^2 - 2\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N} Y_iY_j\right]$$

$$= \frac{1}{2}\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N}\left[\frac{P_iP_j\left(1-P_i-P_j\right)}{k\left(1-2P_i\right)\left(1-2P_j\right)}\left\{\frac{Y_i^2}{P_i^2}+\frac{Y_j^2}{P_j^2}+2\frac{Y_iY_j}{P_iP_j}\right\} - \frac{Y_i^2}{1-P_i}P_i - \frac{Y_j^2}{1-P_j}P_i - 2Y_iY_j\right]$$

$$= \frac{1}{2}\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N} P_iP_j\left[\frac{Y_i^2}{P_i^2}\cdot\frac{\left(1-P_i-P_j\right)}{k\left(1-2P_i\right)\left(1-2P_j\right)} + \frac{Y_j^2}{P_j^2}\cdot\frac{\left(1-P_i-P_j\right)}{k\left(1-2P_i\right)\left(1-2P_j\right)}\right.$$

$$\left. + 2\frac{Y_iY_j}{P_iP_j}\cdot\frac{\left(1-P_i-P_j\right)}{k\left(1-2P_i\right)\left(1-2P_j\right)} - \frac{Y_i^2}{P_i\left(1-P_i\right)} - \frac{Y_j^2}{P_j\left(1-P_j\right)} - 2\frac{Y_iY_j}{P_iP_j}\right]$$

$$= \frac{1}{2}\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N} P_iP_j\left[\frac{Y_i^2}{P_i^2}\cdot\frac{\left(1-P_i-P_j\right)}{k\left(1-2P_i\right)\left(1-2P_j\right)} - \frac{Y_i^2}{P_i\left(1-P_i\right)} + \frac{Y_j^2}{P_j^2}\cdot\frac{\left(1-P_i-P_j\right)}{k\left(1-2P_i\right)\left(1-2P_j\right)}\right.$$

$$-\frac{Y_j^2}{P_j(1-P_j)}-2\frac{Y_i Y_j}{P_i P_j}+2\frac{Y_i Y_j}{P_i P_j}\cdot\frac{(1-P_i-P_j)}{k(1-2P_i)(1-2P_j)}\Bigg]$$

$$=\frac{1}{2}\sum_{\substack{i=1\,j=1\\j\neq i}}^{N}\sum P_i P_j\left[\frac{Y_i^2}{P_i^2}\left\{\frac{(1-P_i-P_j)}{k(1-2P_i)(1-2P_j)}-\frac{P_i}{(1-P_i)}\right\}+\frac{Y_j^2}{P_j^2}\right.$$

$$\left.\cdot\left\{\frac{(1-P_i-P_j)}{k(1-2P_i)(1-2P_j)}-\frac{P_j}{(1-P_j)}\right\}-2\frac{Y_i Y_j}{P_i P_j}\left\{1-\frac{(1-P_i-P_j)}{k(1-2P_i)(1-2P_j)}\right\}\right]$$

$$=\frac{1}{2}\sum_{\substack{i=1\,j=1\\j\neq i}}^{N}\sum P_i P_j\left[\frac{Y_i^2}{P_i^2}\left\{1-\frac{(1-P_i-P_j)}{k(1-2P_i)(1-2P_j)}\right\}+\frac{Y_j^2}{P_j^2}\right.$$

$$\left.\cdot\left\{1-\frac{(1-P_i-P_j)}{k(1-2P_i)(1-2P_j)}\right\}-2\frac{Y_i Y_j}{P_i P_j}\left\{1-\frac{(1-P_i-P_j)}{k(1-2P_i)(1-2P_j)}\right\}\right]$$

$$=\frac{1}{2}\sum_{\substack{i=1\,j=1\\j\neq i}}^{N}\sum P_i P_j\left[\left\{1-\frac{(1-P_i-P_j)}{k(1-2P_i)(1-2P_j)}\right\}\cdot\left\{\frac{Y_i^2}{P_i^2}+\frac{Y_j^2}{P_j^2}-2\frac{Y_i Y_j}{P_i P_j}\right\}\right]$$

$$\Rightarrow \mathrm{V}ar(t_{MM2})=\frac{1}{2}\sum_{\substack{i=1\,j=1\\j\neq i}}^{N}\sum P_i P_j\left[1-\frac{(1-P_i-P_j)}{k(1-2P_i)(1-2P_j)}\right]\cdot\left\{\frac{Y_i}{P_i}-\frac{Y_j}{P_j}\right\}^2 \qquad (6.2.5)$$

6.2.3 The Modified Murthy Estimator – III:

In this section the design based variance of modified Murthy estimator – III has been obtained. To obtain this variance consider modified Murthy estimator – III from equation (5.2.7) and let us proceed as under:

$$\mathrm{V}ar(t_{MM3})=E_D(t_{MM3}^2)-[E_D(t_{MM3})]^2$$

$$=\sum_{\substack{i=1\,j=1\\j>i}}^{N}\sum t_{MM3}^2\,P(S)-Y^2 \qquad \because\ E_D(t_{MM3})=Y$$

$$=\frac{1}{2}\sum_{\substack{i=1\,j=1\\j\neq i}}^{N}\sum t_{MM3}^2\,P(S)-Y^2$$

$$=\frac{1}{2}\sum_{\substack{i=1\,j=1\\j\neq i}}^{N}\sum\left[\frac{k(1-2P_i)(1-2P_j)}{2(1-P_j)(1-2P_i)+2(1-P_i)(1-2P_j)}\left\{\frac{Y_i}{P_i}+\frac{Y_j}{P_j}\right\}\right]^2$$

$$\cdot\frac{2P_iP_j(1-P_i-P_j)[2(1-P_j)(1-2P_i)+2(1-P_i)(1-2P_j)]}{k^2(1-2P_i)^2(1-2P_j)^2}-Y^2$$

$$=\frac{1}{2}\sum_{\substack{i=1\,j=1\\j\neq i}}^{N}\sum\left[\frac{2P_iP_j(1-P_i-P_j)}{2(1-P_j)(1-2P_i)+2(1-P_i)(1-2P_j)}\left\{\frac{Y_i}{P_i}+\frac{Y_j}{P_j}\right\}^2\right]-\left[\sum_{i=1}^{N}Y_i\right]^2$$

$$=\frac{1}{2}\sum_{\substack{i=1\,j=1\\j\neq i}}^{N}\sum\left[\frac{2P_iP_j(1-P_i-P_j)}{2(1-P_i-P_j)+2(1-2P_i)(1-2P_j)}\left\{\frac{Y_i}{P_i}+\frac{Y_j}{P_j}\right\}^2\right]-\left[\sum_{i=1}^{N}Y_i\right]^2$$

$$=\frac{1}{2}\sum_{\substack{i=1\,j=1\\j\neq i}}^{N}\sum\left[\frac{P_iP_j(1-P_i-P_j)}{(1-P_i-P_j)+(1-2P_i)(1-2P_j)}\left\{\frac{Y_i^2}{P_i^2}+\frac{Y_j^2}{P_j^2}+2\frac{Y_iY_j}{P_iP_j}\right\}\right]-\sum_{i=}^{N}Y_i^2-\sum_{\substack{i=1\,j=1\\j\neq i}}^{N}\sum Y_iY_j$$

$$=\frac{1}{2}\left[\sum_{\substack{i=1\,j=1\\j\neq i}}^{N}\sum\frac{P_iP_j(1-P_i-P_j)}{(1-P_i-P_j)+(1-2P_i)(1-2P_j)}\left\{\frac{Y_i^2}{P_i^2}+\frac{Y_j^2}{P_j^2}+2\frac{Y_iY_j}{P_iP_j}\right\}\right.$$

$$\left.-\sum_{\substack{i=1\,j=1\\j\neq i}}^{N}\sum\frac{Y_i^2}{1-P_i}P_j-\sum_{\substack{i=1\,j=1\\j\neq i}}^{N}\sum\frac{Y_j^2}{1-P_j}P_i-\sum_{\substack{i=1\,j=1\\j\neq i}}^{N}\sum Y_iY_j\right]$$

$$=\frac{1}{2}\sum_{\substack{i=1\,j=1\\j\neq i}}^{N}\sum P_iP_j\left[\frac{(1-P_i-P_j)}{(1-P_i-P_j)+(1-2P_i)(1-2P_j)}\left\{\frac{Y_i^2}{P_i^2}+\frac{Y_j^2}{P_j^2}+2\frac{Y_iY_j}{P_iP_j}\right\}\right.$$

$$\left.-\frac{Y_i^2}{P_i(1-P_i)}-\frac{Y_j^2}{P_j(1-P_j)}-2\frac{Y_iY_j}{P_iP_j}\right]$$

$$=\frac{1}{2}\sum_{\substack{i=1\,j=1\\j\neq i}}^{N}\sum\frac{P_iP_j(1-P_i-P_j)}{(1-P_i-P_j)+(1-2P_i)(1-2P_j)}\left[\frac{Y_i^2}{P_i^2}+\frac{Y_j^2}{P_j^2}+2\frac{Y_iY_j}{P_iP_j}-\frac{Y_i^2}{P_i(1-P_i)}\right.$$

$$\cdot\frac{(1-P_i-P_j)+(1-2P_i)(1-2P_j)}{(1-P_i-P_j)}-\frac{Y_j^2}{P_j(1-P_j)}\cdot\frac{(1-P_i-P_j)+(1-2P_i)(1-2P_j)}{(1-P_i-P_j)}$$

$$\left.-2\frac{Y_iY_i}{P_iP_j}\cdot\frac{(1-P_i-P_j)+(1-2P_i)(1-2P_j)}{(1-P_i-P_j)}\right]$$

$$=\frac{1}{2}\sum_{\substack{i=1\,j=1\\j\neq i}}^{N}\sum\frac{P_iP_j(1-P_i-P_j)}{(1-P_i-P_j)+(1-2P_i)(1-2P_j)}\left[\frac{Y_i^2}{P_i^2}\right.$$

$$\cdot\left\{\frac{(1-P_i)(1-P_i-P_j)-P_i(1-P_i-P_j)-P_i(1-2P_i)(1-2P_j)}{(1-P_i)(1-P_i-P_j)}\right\}+\frac{Y_j^2}{P_j^2}$$

$$\cdot\left\{\frac{(1-P_j)(1-P_i-P_j)-P_j(1-P_i-P_j)-P_j(1-2P_i)(1-2P_j)}{(1-P_j)(1-P_i-P_j)}\right\}-2\frac{Y_iY_j}{P_iP_j}$$

$$\cdot\left\{\frac{(1-P_i-P_j)+(1-2P_i)(1-2P_j)-(1-P_i-P_j)}{(1-P_i-P_j)}\right\}\Bigg]\Bigg]$$

$$=\frac{1}{2}\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}\frac{P_iP_j(1-P_i-P_j)}{(1-P_i-P_j)+(1-2P_i)(1-2P_j)}\left[\frac{Y_i^2}{P_i^2}\cdot\left\{\frac{(1-P_i-P_j)(1-2P_i)-P_i(1-2P_i)(1-2P_j)}{(1-P_i)(1-P_i-P_i)}\right\}\right.$$

$$\left.+\frac{Y_j^2}{P_j^2}\left\{\frac{(1-P_i-P_j)(1-2P_j)-P_j(1-2P_i)(1-2P_j)}{(1-P_j)(1-P_i-P_j)}\right\}-2\frac{Y_iY_j}{P_iP_j}\left\{\frac{(1-2P_i)(1-2P_j)}{(1-P_i-P_j)}\right\}\right]$$

$$=\frac{1}{2}\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}\frac{P_iP_j(1-P_i-P_j)}{(1-P_i-P_j)+(1-2P_i)(1-2P_j)}\left[\frac{Y_i^2}{P_i^2}\cdot\left\{\frac{(1-2P_i)[(1-P_i-P_j)-P_i(1-2P_j)]}{(1-P_i)(1-P_i-P_j)}\right\}\right.$$

$$\left.+\frac{Y_i^2}{P_j^2}\left\{\frac{(1-2P_j)[(1-P_i-P_j)-P_j(1-2P_i)]}{(1-P_j)(1-P_i-P_j)}\right\}-2\frac{Y_iY_i}{P_iP_j}\left\{\frac{(1-2P_i)(1-2P_j)}{(1-P_i-P_i)}\right\}\right]$$

$$=\frac{1}{2}\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}\frac{P_iP_j(1-P_i-P_j)}{(1-P_i-P_j)+(1-2P_i)(1-2P_j)}\left[\frac{Y_i^2}{P_i^2}\cdot\left\{\frac{(1-2P_i)[1-P_i-P_j-P_i+2P_iP_j]}{(1-P_i)(1-P_i-P_j)}\right\}\right.$$

$$\left.+\frac{Y_j^2}{P_j^2}\cdot\left\{\frac{(1-2P_j)[1-P_i-P_j-P_j+2P_iP_j]}{(1-P_j)(1-P_i-P_j)}\right\}-2\frac{Y_iY_j}{P_iP_j}\left\{\frac{(1-2P_i)(1-2P_j)}{(1-P_i-P_j)}\right\}\right]$$

$$=\frac{1}{2}\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}\frac{P_iP_j(1-P_i-P_j)}{(1-P_i-P_j)+(1-2P_i)(1-2P_j)}\left[\frac{Y_i^2}{P_i^2}\cdot\left\{\frac{(1-2P_i)[(1-2P_i)-P_j+2P_iP_j]}{(1-P_i)(1-P_i-P_j)}\right\}\right.$$

$$\left.+\frac{Y_j^2}{P_j^2}\cdot\left\{\frac{(1-2P_j)[(1-2P_j)-P_i+2P_iP_j]}{(1-P_j)(1-P_i-P_j)}\right\}-2\frac{Y_iY_j}{P_iP_j}\cdot\left\{\frac{(1-2P_i)(1-2P_j)}{(1-P_i-P_j)}\right\}\right]$$

$$=\frac{1}{2}\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}\frac{P_iP_j(1-P_i-P_j)}{(1-P_i-P_j)+(1-2P_i)(1-2P_j)}\left[\frac{Y_i^2}{P_i^2}\cdot\left\{\frac{(1-2P_i)[1(1-2P_i)-P_j(1-2P_i)]}{(1-P_i)(1-P_i-P_j)}\right\}\right.$$

$$\left.+\frac{Y_j^2}{P_i^2}\cdot\left\{\frac{(1-2P_j)[1(1-2P_j)-P_i(1-2P_j)]}{(1-P_j)(1-P_i-P_j)}\right\}-2\frac{Y_iY_j}{P_iP_j}\left\{\frac{(1-2P_i)(1-2P_j)}{(1-P_i-P_j)}\right\}\right]$$

$$= \frac{1}{2}\sum_{\substack{i=1 \\ j\neq i}}^{N}\sum_{j=1} \frac{P_i P_j (1-P_i-P_j)}{(1-P_i-P_j)+(1-2P_i)(1-2P_j)} \left[\frac{Y_i^2}{P_i^2} \cdot \left\{ \frac{(1-2P_i)^2(1-P_j)}{(1-P_i)(1-P_i-P_j)} \right\} \right.$$

$$\left. + \frac{Y_j^2}{P_j^2} \cdot \left\{ \frac{(1-2P_j)^2(1-P_i)}{(1-P_j)(1-P_i-P_j)} \right\} - 2\frac{Y_i Y_j}{P_i P_j} \cdot \left\{ \frac{(1-2P_i)(1-2P_j)}{(1-P_i-P_j)} \right\} \right]$$

$$= \frac{1}{2}\sum_{\substack{i=1 \\ j\neq i}}^{N}\sum_{j=1} \frac{P_i P_j}{(1-P_i-P_j)+(1-2P_i)(1-2P_j)} \left[\frac{Y_i^2}{P_i^2} \cdot \left\{ \frac{(1-2P_i)^2(1-P_j)}{(1-P_i)} \right\} \right.$$

$$\left. + \frac{Y_j^2}{P_j^2} \cdot \left\{ \frac{(1-2P_j)^2(1-P_i)}{(1-P_j)} \right\} - 2\frac{Y_i Y_j}{P_i P_j} \{(1-2P_i)(1-2P_j)\} \right]$$

$$Var(t_{MM3}) = \frac{1}{2}\sum_{\substack{i=1 \\ j\neq i}}^{N}\sum_{j=1} \frac{P_i P_j}{(1-P_i-P_j)+(1-2P_i)(1-2P_j)} \left[\frac{Y_i^2}{P_i^2} \cdot D_{ij} + \frac{Y_j^2}{P_j^2} \cdot E_{ij} - 2\frac{Y_i Y_j}{P_i P_j} \cdot F_{ij} \right] \quad (6.2.6)$$

where $$D_{ij} = \frac{(1-2P_i)^2(1-P_j)}{(1-P_j)} \quad (6.2.7)$$

$$E_{ij} = \frac{(1-2P_j)^2(1-P_i)}{(1-P_i)} \quad (6.2.8)$$

$$F_{ij} = (1-2P_i)(1-2P_j) \quad (6.2.9)$$

6.2.4 The Modified Raj Estimator:

In this sub-section variance of modified Raj estimator has been obtained. To obtain this variance consider the modified Raj estimator given in equation (5.2.9) given as:

$$t_{RM} = y_j + \frac{y_i}{p_i}(1-p_i)\left(k - \frac{p_j}{1-p_j}\right) \quad (5.2.9)$$

Now, design based variance of this estimator is obtained as under:

$$Var(t_{RM}) = E_D(t_{RM}^2) - [E_D(t_{RM})]^2$$

$$= \sum_{\substack{i=1 \\ j\neq i}}^{N}\sum_{j=1} t_{RM}^2 P(S) - Y^2 \qquad \because E_D(t_{RM}) = Y$$

$$= \sum_{\substack{i=1\,j=1\\ j\neq i}}^{N}\sum \left[Y_j + \frac{Y_i}{P_i}(1-P_i)\left(b - \frac{P_j}{1-P_j}\right)\right]^2 \frac{P_i P_j}{1-P_i} - Y^2$$

$$= \sum_{\substack{i=1\,j=1\\ j\neq i}}^{N}\sum \left[Y_j + \frac{Y_i}{P_i}(1-P_i)\left(b - \frac{P_j}{1-P_j}\right)\right]^2 \frac{P_i P_j}{1-P_i} - \left[\sum_{i=1}^{N} Y_i\right]^2$$

$$= \sum_{\substack{i=1\,j=1\\ j\neq i}}^{N}\sum \left[Y_j + \frac{Y_i}{P_i}\left(b - bP_i - \frac{P_j}{1-P_j} + \frac{P_i P_j}{1-P_j}\right)\right]^2 \frac{P_i P_j}{1-P_i} - \sum_{i=1}^{N} Y_i^2 - \sum_{\substack{i=1\,j=1\\ j\neq i}}^{N}\sum Y_i Y_j$$

$$= \sum_{\substack{i=1\,j=1\\ j\neq i}}^{N}\sum P_i P_j \left[\left\{Y_j + \frac{Y_i}{P_i}\left(\sum_{i=1}^{N}\frac{P_i}{1-P_i}(1-P_i) - \frac{P_j - P_i P_j}{1-P_j}\right)\right\}^2 \frac{1}{1-P_i} - \frac{Y_i^2}{P_i(1-P_i)} - \frac{Y_i Y_j}{P_i P_j}\right]$$

$$= \sum_{\substack{i=1\,j=1\\ j\neq i}}^{N}\sum P_i P_j \left[\left\{Y_j + \frac{Y_i}{P_i}\left(\sum_{i=1}^{N} P_i - \frac{P_j(1-P_i)}{1-P_j}\right)\right\}^2 \frac{1}{1-P_i} - \frac{Y_i^2}{P_i(1-P_i)} - \frac{Y_i Y_j}{P_i P_j}\right]$$

$$= \sum_{\substack{i=1\,j=1\\ j\neq i}}^{N}\sum P_i P_j \left[\left\{Y_j + \frac{Y_i}{P_i}\left(1 - \frac{P_j(1-P_i)}{1-P_j}\right)\right\}^2 \frac{1}{1-P_i} - \frac{Y_i^2}{P_i(1-P_i)} - \frac{Y_i Y_j}{P_i P_j}\right]$$

$$= \frac{1}{2}\sum_{\substack{i=1\,j=1\\ j\neq i}}^{N}\sum P_i P_j \left[2\left\{Y_j + \frac{Y_i}{P_i}\left(1 - \frac{P_j(1-P_i)}{1-P_j}\right)\right\}^2 \frac{1}{1-P_i} - 2\frac{Y_i^2}{P_i(1-P_i)} - 2\frac{Y_i Y_j}{P_i P_j}\right]$$

$$= \frac{1}{2}\sum_{\substack{i=1\,j=1\\ j\neq i}}^{N}\sum P_i P_j \left[2\left\{Y_j^2 + \frac{Y_i^2}{P_i^2}\left(1 - \frac{P_j(1-P_i)}{1-P_j}\right)^2 + 2\frac{Y_i Y_j}{P_i}\left(1 - \frac{P_j(1-P_i)}{1-P_j}\right)\right\}\right.$$

$$\left.\cdot\frac{1}{1-P_i} - \frac{Y_i^2}{P_i(1-P_i)} - \frac{Y_j^2}{P_j(1-P_j)} - 2\frac{Y_i Y_j}{P_i P_j}\right]$$

$$= \frac{1}{2}\sum_{\substack{i=1\,j=1\\ j\neq i}}^{N}\sum P_i P_j \left[2\left\{Y_j^2 + \frac{Y_i^2}{P_i^2} + \frac{Y_i^2}{P_i^2}\cdot\frac{P_j^2(1-P_i)^2}{(1-P_j)^2} - 2\cdot\frac{Y_i^2}{P_i^2}\cdot\frac{P_j(1-P_i)}{(1-P_j)} + 2\frac{Y_i Y_j}{P_i}\right.\right.$$

$$\left.\left. - 2\frac{Y_i Y_j}{P_i}\cdot\frac{P_j(1-P_i)}{(1-P_j)}\right\}\cdot\frac{1}{1-P_i} - \frac{Y_i^2}{P_i(1-P_i)} - \frac{Y_j^2}{P_j(1-P_j)} - 2\frac{Y_i Y_j}{P_i P_j}\right]$$

$$= \frac{1}{2}\sum_{\substack{i=1}}^{N}\sum_{\substack{j=1\\ j\neq i}} P_i\, P_j \left[2\left\{ \frac{Y_j^2}{1-P_i} + \frac{Y_i^2}{P_i^2(1-P_i)} + \frac{Y_i^2}{P_i^2}\cdot\frac{P_j^2(1-P_i)}{(1-P_j)^2} - 2\frac{Y_i^2}{P_i^2}\cdot\frac{P_j}{(1-P_j)} \right.\right.$$

$$\left.\left. + 2\frac{Y_i\,Y_j}{P_i(1-P_i)} - 2\frac{Y_i\,Y_j}{P_j}\cdot\frac{P_j}{1-P_j} \right\} - \frac{Y_i^2}{P_i(1-P_i)} - \frac{Y_j^2}{P_j(1-P_j)} - 2\frac{Y_i\,Y_j}{P_i\,P_j} \right]$$

$$= \frac{1}{2}\sum_{\substack{i=1}}^{N}\sum_{\substack{j=1\\ j\neq i}} P_i\, P_j \left[\frac{Y_i^2}{P_i^2(1-P_i)} + \frac{Y_i^2}{P_i^2}\cdot\frac{P_j^2(1-P_i)}{(1-P_j)^2} - 2\frac{Y_i^2}{P_i^2}\cdot\frac{P_j}{1-P_j} + \frac{Y_i^2}{1-P_j} - \frac{Y_i^2}{P_i(1-P_i)} \right.$$

$$- \frac{Y_j^2}{P_j(1-P_j)} - 2\frac{Y_i\,Y_j}{P_i\,P_j} + \frac{Y_j^2}{P_j^2(1-P_j)} + \frac{Y_j^2}{P_j^2}\cdot\frac{P_i^2(1-P_j)}{(1-P_i)^2} - 2\frac{Y_j^2}{P_j^2}\cdot\frac{P_i}{1-P_i} + \frac{Y_j^2}{1-P_i}$$

$$\left. - 2\frac{Y_i\,Y_j}{P_j}\cdot\frac{P_i}{1-P_i} + 2\frac{Y_i\,Y_j}{P_i(1-P_i)} - 2\frac{Y_i\,Y_j}{P_i}\cdot\frac{P_j}{1-P_j} + 2\frac{Y_i\,Y_j}{P_j(1-P_j)} \right]$$

$$= \frac{1}{2}\sum_{\substack{i=1}}^{N}\sum_{\substack{j=1\\ j\neq i}} P_i\, P_j \left[\frac{Y_i^2}{P_i^2}\left\{ \frac{1}{1-P_i} + \frac{P_j^2(1-P_i)}{(1-P_j)^2} - 2\frac{P_j}{1-P_j} + \frac{P_i^2}{1-P_j} - \frac{P_i}{1-P_i} \right\} + \frac{Y_j^2}{P_j^2} \right.$$

$$\cdot\left\{ \frac{1}{1-P_j} + \frac{P_i^2(1-P_j)}{(1-P_i)^2} - 2\frac{P_i}{1-P_i} + \frac{P_j^2}{1-P_i} - \frac{P_j}{1-P_j} \right\} - 2\frac{Y_i\,Y_j}{P_i\,P_j}\cdot$$

$$\left.\left\{ 1 - \frac{P_j}{1-P_i} + \frac{P_j^2}{1-P_j} - \frac{P_i}{1-P_j} + \frac{P_i^2}{1-P_i} \right\} \right]$$

$$= \frac{1}{2}\sum_{\substack{i=1}}^{N}\sum_{\substack{j=1\\ j\neq i}} P_i\, P_j \left[\frac{Y_i^2}{P_i^2}\left\{ 1 + \frac{P_j^2(1-P_i)}{(1-P_j)^2} + \frac{P_i^2-2P_j}{1-P_j} \right\} + \frac{Y_j^2}{P_j^2}\left\{ 1 + \frac{P_i^2(1-P_j)}{(1-P_i)^2} \right.\right.$$

$$\left.\left. + \frac{P_j^2-2P_i}{1-P_i} \right\} - 2\frac{Y_i\,Y_j}{P_i\,P_j}\left\{ 1 - \frac{P_j}{1-P_i} + \frac{P_j^2}{1-P_j} - \frac{P_i}{1-P_j} + \frac{P_i^2}{1-P_i} \right\} \right]$$

$$Var(t_{RM}) = \frac{1}{2}\sum_{\substack{i=1}}^{N}\sum_{\substack{j=1\\ j\neq i}} P_i\, P_j \left[\frac{Y_i^2}{P_i^2}G_{ij} + \frac{Y_j^2}{P_j^2}H_{ij} - 2\frac{Y_i\,Y_j}{P_i\,P_j}M_{ij} \right] \qquad (6.2.10)$$

where $$G_{ij} = 1 + \frac{P_j^2(1-P_i)}{(1-P_j)^2} + \frac{P_i^2-2P_j}{1-P_j} \qquad (6.2.11)$$

$$H_{ij} = 1 + \frac{P_i^2(1-P_j)}{(1-P_i)^2} + \frac{P_j^2-2P_i}{1-P_i} \qquad (6.2.12)$$

$$M_{ij} = 1 - \frac{P_j}{1-P_i} + \frac{P_j^2}{1-P_j} - \frac{P_i}{1-P_j} + \frac{P_i^2}{1-P_i} \qquad (6.2.13)$$

6.3 Empirical Study:

In this section the empirical study has been carried out to decide about the performance of new estimators. The sampling variance of various estimators has been calculated, by using the exact variance formulae, in this empirical study to decide about their performance. Ranking has also been done to have a better insight about performance of various estimators. The results of this empirical study have been given in following pages.

Table 6.1: Sampling Variance of Various Estimators using selected Natural Populations.

Estimator	Pop. – 1	Pop. – 2	Pop. – 3	Pop. – 4	Pop. – 5
Simple Random Sampling	12149.53	4946942.00	875830.69	4436.00	9541.78
Hansen – Hurwitz	14951.35	1607816.88	275095.13	309.58	7116.15
Horvitz – Thompson (YG dbd)	13383.07	1520483.75	246963.45	333.75	6132.15
Horvitz – Thompson (Brewer)	14000.22	1603085.00	266747.50	276.14	6373.32
Rao – Hartley – Cochran	14164.44	1523194.88	260616.44	275.18	6325.47
Raj (Ordered)	14091.58	1524429.88	259980.36	278.59	6416.57
Murthy (Unordered)	14035.80	1517847.63	258876.52	275.10	6339.61
Modified Murthy – I (t_{MM1})	15520.62	1661465.63	308736.72	252.06	7059.93
Modified Murthy – II (t_{MM2})	14000.22	1603084.88	266747.50	276.14	6373.32
Modified Murthy – III (t_{MM3})	15473.10	1748660.88	315597.03	250.11	7097.06
Raj Modified	28904.40	3273470.25	578977.00	430.63	12481.94

Table 6.1 (Continued)

Estimator	Pop. – 6	Pop. – 7	Pop. – 8	Pop. – 9	Pop. – 10
Simple Random Sampling	1796014.88	340089.34	720850.19	283385.69	35074512
Hansen – Hurwitz	52975.27	14721.07	499017.84	142668.53	61680976
Horvitz – Thompson (YG dbd)	54243.95	10489.04	465838.94	135870.34	53669716
Horvitz – Thompson (Brewer)	51296.67	11603.40	463496.94	132359.27	58450276
Rao – Hartley – Cochran	49443.59	13085.40	465750.00	133157.30	58434608
Raj (Ordered)	48824.30	12303.71	466823.94	133291.91	58025412
Murthy (Unordered)	48342.14	11702.98	464600.19	132622.94	57757348
Modified Murthy – I (t_{MM1})	81201.60	73559.35	462425.72	125830.42	68508664
Modified Murthy – II (t_{MM2})	51296.66	11603.40	463496.91	132359.25	58450276
Modified Murthy – III (t_{MM3})	83349.91	55840.13	461328.50	125580.77	69103216
Raj Modified	129905.50	53528.73	863973.88	236540.81	127728224

Table 6.1 (Continued)

Estimator	Pop. – 11	Pop. – 12	Pop. – 13	Pop. – 14	Pop. – 15
Simple Random Sampling	116657888	33290.53	113744368	9171896.00	5704569.00
Hansen – Hurwitz	33459728	9055.63	21936420	469457.03	190441.38
Horvitz – Thompson (YG dbd)	26741372	8830.26	14681897	417418.81	181002.19
Horvitz – Thompson (Brewer)	30039964	8504.88	18114750	440262.16	177091.80
Rao – Hartley – Cochran	31698690	8579.01	20249004	438159.91	180418.14
Raj (Ordered)	30696964	8557.67	18817320	433918.16	177205.83
Murthy (Unordered)	30411432	8528.08	18169780	430489.16	176051.25
Modified Murthy – I (t_{MM1})	41422784	7911.39	35983216	641929.06	192122.86
Modified Murthy – II (t_{MM2})	30039962	8504.88	18114750	440262.16	177091.80
Modified Murthy – III (t_{MM3})	40342760	7891.28	32965098	637801.00	193667.75
Raj Modified	70589960	15060.20	48578436	1057900.13	354665.00

Table 6.1 (Continued)

Estimator	Pop. – 16	Pop. – 17	Pop. – 18	Pop. – 19	Pop. – 20
Simple Random Sampling	146.60	552.85	174762048	660286848	2886.59
Hansen – Hurwitz	33.05	431.97	677828992	2464964096	1317.38
Horvitz – Thompson (YG dbd)	30.87	388.89	508257984	2131433344	1248.41
Horvitz – Thompson (Brewer)	29.48	402.02	616952000	2331453952	1215.89
Rao – Hartley – Cochran	30.04	392.70	602514688	2335229184	1248.04
Raj (Ordered)	29.90	396.57	597693568	2294652928	1234.53
Murthy (Unordered)	29.55	393.14	585524800	2280688640	1228.40
Modified Murthy – I (t_{MM1})	28.21	420.83	881214976	2707175168	1174.35
Modified Murthy – II (t_{MM2})	29.48	402.02	616952000	2331453952	1215.89
Modified Murthy – III (t_{MM3})	28.10	429.69	899908160	2753691136	1164.74
Raj Modified	51.08	775.55	1466860928	5027871232	2215.32

Table 6.1 (Continued)

Estimator	Pop. – 21	Pop. – 22	Pop. – 23	Pop. – 24	Pop. – 25
Simple Random Sampling	51550.81	30627.20	2119.58	209.44	33002238
Hansen – Hurwitz	5149.11	24445.05	1544.81	1553.35	6976035
Horvitz – Thompson (YG dbd)	4553.31	22444.18	1410.09	1204.55	6204300
Horvitz – Thompson (Brewer)	4568.07	23053.29	1457.04	1367.58	6503349
Rao – Hartley – Cochran	4805.84	22815.38	1441.82	1471.59	6510966
Raj (Ordered)	4684.54	22853.97	1445.18	1377.47	6449763
Murthy (Unordered)	4634.86	22730.95	1437.94	1344.68	6403505
Modified Murthy – I (t_{MM1})	5237.20	23868.57	1532.03	2093.34	7242271
Modified Murthy – II (t_{MM2})	4568.07	23053.29	1457.04	1367.58	6503349
Modified Murthy – III (t_{MM3})	5175.78	24227.21	1553.50	1918.31	7343009
Raj Modified	9262.33	45075.14	2881.68	3306.84	13262048

Table 6.1 (Continued)

Estimator	Pop. – 26	Pop. – 27	Pop. – 28	Pop. – 29	Pop. – 30
Simple Random Sampling	42895136	19825144	52686160	625482.63	983823.13
Hansen – Hurwitz	9738723	14469214	303246912	323717.97	391832.13
Horvitz – Thompson (YG dbd)	8816018	13042031	282992928	282880.78	376569.28
Horvitz – Thompson (Brewer)	8922493	13298241	298253152	301799.25	370586.00
Rao – Hartley – Cochran	9089475	13504600	283030464	302136.78	371209.38
Raj (Ordered)	8889974	13383291	285806816	302384.94	371006.19
Murthy (Unordered)	8793231	13290966	284498176	300846.41	369816.72
Modified Murthy – I (t_{MM1})	9480750	13960476	302430816	347341.31	357149.44
Modified Murthy – II (t_{MM2})	8922493	13298241	298253152	301799.25	370586.00
Modified Murthy – III (t_{MM3})	9503254	13960058	317287456	348246.56	357766.81
Raj Modified	16971576	25660730	594504704	638116.25	679523.06

Table 6.1 (Continued)

Estimator	Pop. – 31	Pop. – 32	Pop. – 33	Pop. – 34	Pop. – 35
Simple Random Sampling	74491.11	710508.50	103107384	47647192	2582.05
Hansen – Hurwitz	105404.14	680122.13	44782804	323955264	1405.82
Horvitz – Thompson (YG dbd)	98536.13	630053.25	43200928	285696896	1304.92
Horvitz – Thompson (Brewer)	99867.50	628548.56	42875176	308020000	1341.15
Rao – Hartley – Cochran	99856.55	634780.63	42425816	306904992	1331.83
Raj (Ordered)	100050.18	633407.31	42707620	307791776	1332.04
Murthy (Unordered)	99759.47	629949.88	42605616	306929472	1327.55
Modified Murthy – I (t_{MM1})	102582.31	628720.88	41872868	355647840	1401.77
Modified Murthy – II (t_{MM2})	99867.50	628548.44	42875176	308020000	1341.15
Modified Murthy – III (t_{MM3})	102678.01	627323.06	42129608	356810208	1415.90
Raj Modified	194354.52	1169550.00	80347616	668498496	2666.64

Table 6.1 (Continued)

Estimator	Pop. – 36	Pop. – 37	Pop. – 38	Pop. – 39	Pop. – 40
Simple Random Sampling	566143.38	52956028	964.42	47313.47	770760.00
Hansen – Hurwitz	36479.28	4487930	1396.58	25019.09	482237.16
Horvitz – Thompson (YG dbd)	34884.00	4021857	1302.01	23854.06	435082.28
Horvitz – Thompson (Brewer)	34102.43	4232811	1323.44	23755.04	456431.59
Rao – Hartley – Cochran	33673.18	4251723	1323.07	23702.30	456856.25
Raj (Ordered)	33717.23	4168227	1326.90	23699.21	455342.41
Murthy (Unordered)	33465.47	4138452	1323.24	23619.03	453588.00
Modified Murthy – I (t_{MM1})	39025.23	4965894	1368.31	23357.24	502383.94
Modified Murthy – II (t_{MM2})	34102.43	4232811	1323.44	23755.04	456431.59
Modified Murthy – III (t_{MM3})	39621.65	5077701	1370.77	23506.50	505061.22
Raj Modified	69934.57	9156076	2566.99	44647.03	943546.44

Table 6.1 (Continued)

Estimator	Pop. – 41	Pop. – 42	Pop. – 43	Pop. – 44	Pop. – 45
Simple Random Sampling	211.56	12693.78	10816.00	164.89	5305.21
Hansen – Hurwitz	199.48	750.29	597.89	12.57	1377.74
Horvitz – Thompson (YG dbd)	177.17	880.51	561.20	11.19	1270.41
Horvitz – Thompson (Brewer)	177.41	621.98	567.77	11.90	1324.73
Rao – Hartley – Cochran	177.32	666.93	531.46	11.91	1285.89
Raj (Ordered)	179.54	625.51	505.01	11.93	1296.28
Murthy (Unordered)	177.32	597.57	480.07	11.90	1290.58
Modified Murthy – I (t_{MM1})	178.98	349.49	4529.47	14.29	1433.92
Modified Murthy – II (t_{MM2})	177.41	621.98	567.77	11.90	1324.73
Modified Murthy – III (t_{MM3})	179.02	364.42	2335.98	14.29	1469.22
Raj Modified	321.28	642.25	1699.17	26.59	2709.35

Table 6.1 (Continued)

Estimator	Pop. – 46	Pop. – 47	Pop. – 48	Pop. – 49	Pop. – 50
Simple Random Sampling	9137.78	268.47	3146908.00	1550120.00	1538.13
Hansen – Hurwitz	14674.05	12.60	81172.90	212971.75	173.11
Horvitz – Thompson (YG dbd)	11164.49	11.20	87418.94	130808.48	166.55
Horvitz – Thompson (Brewer)	13236.35	11.79	72601.00	166556.06	161.59
Rao – Hartley – Cochran	13043.60	11.76	75761.38	198773.64	161.57
Raj (Ordered)	13216.87	11.79	73671.80	184318.23	162.30
Murthy (Unordered)	13048.93	11.74	72815.06	179193.66	161.58
Modified Murthy – I (t_{MM1})	18464.50	15.81	63353.90	356402.66	159.78
Modified Murthy – II (t_{MM2})	13236.35	11.79	72601.00	166556.06	161.59
Modified Murthy – III (t_{MM3})	18636.40	15.88	63821.39	329139.00	161.76
Raj Modified	31383.90	28.29	116564.49	509500.75	316.93

Table 6.2 : Ranks of Various Estimator along with the ranks of C.V. (X) and ρ_{XY}

Pop No	*CV*	ρ_{XY}	SRS	HH	HT (YG)	HT Bre	RHC	Raj	Mur.	MM 1	MM 2	MM 3	Raj Mod.
1.	20	7	1	8	2	3	7	6	5	10	3	9	11
2.	46	49	11	7	2	6	3	4	1	8	5	9	10
3.	38	35	11	7	1	5	4	3	2	8	5	9	10
4.	10	41	11	8	9	5	4	7	3	2	5	1	10
5.	9	10	10	9	1	4	2	6	3	7	4	8	11
6.	45	50	11	6	7	5	3	2	1	8	4	9	10
7.	44	44	11	7	1	2	6	5	4	10	2	9	8
8.	4	9	10	9	7	4	6	8	5	2	3	1	11
9.	12	21	11	9	8	4	6	7	5	2	3	1	10
10.	33	23	1	8	2	6	5	4	3	9	6	10	11
11.	39	25	11	7	1	3	6	5	4	9	2	8	10
12.	15	28	11	9	8	3	7	6	5	2	3	1	10
13.	48	29	11	7	1	2	6	5	4	9	2	8	10
14.	34	45	11	7	1	5	4	3	2	9	5	8	10
15.	37	40	11	7	6	2	5	4	1	8	2	9	10
16.	19	27	11	9	8	3	7	6	5	2	3	1	10

Table 6.2 (Continued)

Pop No	*CV*	ρ_{XY}	SRS	HH	HT (YG)	HT Bre	RHC	Raj	Mur.	MM 1	MM 2	MM 3	Raj Mod.
17.	25	12	10	9	1	5	2	4	3	7	5	8	11
18.	40	32	1	8	2	6	5	4	3	9	6	10	11
19.	43	13	1	8	2	5	7	4	3	9	5	10	11
20.	22	16	11	9	8	3	7	6	5	2	3	1	10
21.	29	33	11	7	1	2	6	5	4	9	2	8	10
22.	23	11	10	9	1	5	3	4	2	7	5	8	11
23.	18	14	10	8	1	5	3	4	2	7	5	9	11
24.	50	4	1	8	2	4	7	6	3	10	4	9	11
25.	31	30	11	7	1	4	6	3	2	8	4	9	10
26.	36	26	11	9	2	4	6	3	1	7	4	8	10
27.	21	8	10	9	1	3	6	5	2	8	3	7	11
28.	35	2	1	9	2	6	3	5	4	8	6	10	11
29.	16	20	10	7	1	3	5	6	2	8	3	9	11
30.	17	22	11	9	8	4	7	6	3	1	4	2	10
31.	13	1	1	10	2	5	4	7	3	8	5	9	11
32.	5	3	10	9	6	3	8	7	5	4	2	1	11
33.	11	24	11	9	8	6	3	5	4	1	6	2	10

Table 6.2 (Continued)

Pop No	*CV*	ρ_{XY}	SRS	HH	HT (YG)	HT Bre	RHC	Raj	Mur.	MM 1	MM 2	MM 3	Raj Mod.
34.	14	37	1	8	2	6	3	5	4	9	6	10	11
35.	27	18	10	8	1	5	3	4	2	7	5	9	11
36.	30	42	11	7	6	4	2	3	1	8	4	9	10
37.	41	38	11	7	1	4	6	3	2	8	4	9	10
38.	1	19	1	10	2	5	3	7	4	8	5	9	11
39.	24	15	11	9	8	6	5	4	3	1	6	2	10
40.	28	17	10	7	1	4	6	3	2	8	4	9	11
41.	3	5	10	9	1	4	2	8	2	6	4	7	11
42.	47	48	11	9	10	4	8	6	3	1	4	2	7
43.	49	47	11	7	4	5	3	2	1	10	5	9	8
44.	7	39	11	7	1	2	5	6	2	8	2	8	10
45.	26	34	11	7	1	5	2	4	3	8	5	9	10
46.	8	6	1	8	2	6	3	5	4	9	6	10	11
47.	6	46	11	7	1	4	3	4	2	8	4	9	10
48.	32	43	11	8	9	3	7	6	5	1	3	2	10
49.	42	31	11	7	1	2	6	5	4	9	2	8	10
50.	2	36	11	9	8	4	2	7	3	1	4	6	10

Table 6.3: Frequency Table of Ranks of Various Estimators along with the Average Rank

Ranks	SRS	HH	HT (YG)	HT Bre	RHC	Raj	Mur.	MM 1	MM 2	MM 3	Raj Mod.
1	10	-	21	-	-	-	6	6	-	7	-
2	-	-	12	6	6	2	13	6	8	5	-
3	-	-	-	9	12	7	13	-	9	-	-
4	-	-	1	14	4	12	10	1	13	-	-
5	-	-	-	13	6	10	8	-	13	-	-
6	-	1	3	8	12	11	-	1	7	1	-
7	-	18	2	-	8	6	-	6	-	2	1
8	-	11	8	-	2	2	-	16	-	10	2
9	-	18	2	-	-	-	-	10	-	19	-
10	11	2	1	-	-	-	-	4	-	6	26
11	29	-	-	-	-	-	-	-	-	-	21
Average Ranks	8.78	8.04	3.46	4.16	4.76	4.94	3.02	6.56	4.04	6.96	10.28

Table 6.4: Average Ranks of Various Estimators with ranks of Coefficient of Variation.

CV	SRS	HH	HT (YG)	HT Bre	RHC	Raj	Mur.	MM 1	MM 2	MM 3	Raj Mod.
1 – 10	8.6	8.5	3.8	4.1	3.8	6.5	3.3	5.5	3.9	6.0	10.6
11 – 20	7.8	8.6	4.8	4.2	5.2	5.8	3.8	5.0	4.1	5.3	10.5
21 – 30	10.5	8.1	2.9	4.2	4.2	4.2	2.7	6.5	4.2	7.0	10.5
31 – 40	8.0	7.7	2.7	4.4	5.1	4.0	2.7	7.6	4.3	8.3	10.3
41 – 50	9.0	7.3	3.1	3.9	5.5	4.2	2.6	8.2	3.7	8.2	9.5

Table 6.5: Average Ranks of Various Estimators with ranks of Correlation Coefficient.

Rho	SRS	HH	HT (YG)	HT Bre	RHC	Raj	Mur.	MM 1	MM 2	MM 3	Raj Mod.
1 – 10	5.5	8.8	2.6	4.2	4.8	6.3	3.6	7.2	4.0	7.1	5.5
11 – 20	8.4	8.4	2.6	4.6	4.4	4.6	2.8	6.4	4.6	7.4	8.4
21 – 30	10.0	8.3	4.7	3.9	5.9	5.0	3.6	5.0	3.7	5.0	10.0
31 – 40	9.0	7.4	2.4	3.8	4.4	4.6	2.8	7.7	3.8	8.6	9.0
41 – 50	11.0	7.3	5.0	4.3	4.3	4.2	2.3	6.5	4.1	6.7	11.0

Table 6.6: Average Ranks of various Estimators for populations having Negative Kurtosis and Negative or Positive Skewness.

	Freq.	SRS	HH	HT (YG)	HT Bre	RHC	Raj	Mur.	MM 1	MM 2	MM 3	Raj Mod.
Positive S(X)	28	7.86	8.11	3.25	4.39	4.43	4.89	3.04	6.86	4.32	7.32	10.50
Negative S(X)	3	10.33	7.67	3.00	3.00	5.33	6.67	3.00	6.00	2.67	6.00	10.67
Combined S(X)	31	8.10	8.06	3.23	4.26	4.52	5.06	3.03	6.77	4.16	7.19	10.52

Table 6.7: Average Ranks of various Estimators for populations having Positive Kurtosis and Negative or Positive Skewness.

	Freq.	SRS	HH	HT (YG)	HT Bre	RHC	Raj	Mur.	MM 1	MM 2	MM 3	Raj Mod.
Positive S(X)	16	9.69	7.81	3.06	4.06	5.13	4.44	2.88	7.13	3.88	7.25	9.88
Negative S(X)	3	11.00	9.00	8.00	3.67	5.33	6.33	3.67	1.33	3.67	3.00	10.00
Combined S(X)	19	9.89	8.00	3.84	4.00	5.16	4.74	3.00	6.21	3.84	6.58	9.89

Table 6.8: Regression Summary for Ranks of Various Estimators for model

$$Rank(Estim.) = \beta_0 + \beta_1 [Rank(CV)] + \beta_2 [Rank(\rho)] + \varepsilon$$

Coeffi- Cient.	SRS	HH	HT (YG)	HT Bre	RHC	Raj	Mur.	MM 1	MM 2	MM 3	Raj Mod.
β_0	6.397	9.468	3.378	4.390	4.349	6.903	4.042	4.898	4.274	5.348	10.949
p-Value	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
β_1	-0.045	-0.024	-0.062	0.0005	0.041	-0.056	-0.018	0.100	0.0009	0.088	-0.010
p-Value	0.237	0.004	0.057	0.973	0.036	0.000	0.162	0.001	0.951	0.010	0.116
β_2	0.139	-0.033	0.065	-0.009	-0.025	-0.021	-0.022	-0.035	-0.010	-0.024	-0.036
p-Value	0.001	0.000	0.045	0.485	0.197	0.101	0.089	0.242	0.475	0.461	0.000
F	6.792	21.385	2.913	0.281	2.471	17.133	4.053	5.716	0.286	3.616	24.013
p-Value	0.003	0.000	0.064	0.757	0.095	0.000	0.024	0.006	0.752	0.035	0.000

6.4 Conclusions:

The empirical study of various estimators has been given in the previous section. The results of this empirical study have been given in table – 6.1 through table 6.4. The analyses of these tables have been given in the following pages.

Table 6.1 constitutes the sampling variance of various estimators for fifty selected natural populations. From this table it can be readily seen that the modified Murthy estimator – I performs better than all other estimators in 12% of the populations. These populations are numbered as 30, 33, 39, 42, 48 and 50. Further analysis of these populations reveals that four of these populations have positive skewness for basic variable of study, that is Y. Also these populations are lepto–kurtic. One recommendation for the use of modified Murthy estimator – I for estimation are populations having positive skewness and kurtosis for basic variable of study. The results of correlation coefficient and coefficient of variation for these populations show that these populations have smaller value of coefficient of variation as well as the correlation coefficient, except population number 42, which has high value of coefficient of correlation as well as the coefficient of variation for measure of size. From this it is obvious that the modified Murthy estimator – I will perform better for populations that have smaller value of coefficient of correlation and coefficient of variation for measure of size and hence this estimator should be used in such populations. The modified Murthy estimator – II produces same sampling variance as the Horvitz – Thompson (1952) estimator under the Brewer (1963) procedure. Hence it is concluded that the modified Murthy estimator – II is equal in precision as the Horvitz – Thompson (1952) estimator under the Brewer (1963) selection procedure. The modified Murthy estimator – III performs better than all other estimators in 12% populations, numbered as 4, 8, 9, 12, 16 and 32. Further analysis of these populations reveals that all these populations have positive skewness in basic variable of study, whereas four of these populations have positive skewness for measure of size. Also five of these populations have negative kurtosis for measure of size. From this it is obvious that the modified Murthy estimator – III should be used in populations that have positive skewness along with negative kurtosis for measure of size. Also the analysis of coefficient of variation and correlation coefficient shows that these populations have smaller coefficient of variation for measure of size and small correlation coefficient between measure of size and basic variable of study. It is therefore suggested to use the modified Murthy estimator – II for estimation of population total when measure of size has smaller coefficient of variation, positive skewness and negative kurtosis. Finally, the modified Raj estimator does not perform better in these populations.

Table – 6.2 shows the ranks of various estimators, with reference to their sampling variance, in fifty populations under study. Table 6.3 constitutes the frequency of rank of an estimator along with its average rank. From the study of this table it can be seen that the Murthy estimator clearly outperform all other estimators in the comparison and is followed be the Horvitz – Thompson estimator under Yates – Grundy procedure and the modified Murthy estimator – II. Also from the study of this table it can be seen that the Murthy estimator does not perform better than all other estimators in a single population yet its performance is outstanding.

Table – 6.4 constitute average ranks of various estimators along with the group ranks of coefficient of variation. The analysis of this table shows that the Murthy (1957) estimator outperforms all other estimator. The New estimator – II closely followed the Murthy (1957) estimator for lower coefficient of variation. For moderate coefficient of variation Horvitz – Thompson estimator under Yates – Grundy (1953) draw – by – draw procedure closely follow the Murthy (1957) estimator. Other estimators do not perform much better.

Table – 6.5 shows the average ranks of various estimators along with the group ranks of correlation coefficient. From this table it is readily seen that the Murthy (1957) estimator outperform other estimators for almost all ranges of correlation coefficient and is closely followed by the Horvitz – Thompson (1952) estimator under Yates – Grundy draw – by – draw procedure. The New estimator – II performs better than the Horvitz – Thompson (1952) estimator for almost all the rank groups.

Table 6.6 and Table 6.7 constitute the average rank of various selection procedures on the basis of skewness and kurtosis of the populations.

From these tables it can be readily seen that for populations having negative kurtosis and positive skewness Murthy (1957) estimator clearly outperform other estimators involved in the study and is closely followed by Horvitz – Thompson (1952) estimator under Yates – Grundy (1953) draw-by-draw procedure and Horvitz – Thompson (1952) estimator under Brewer (1963a) selection procedure. For populations having negative kurtosis and negative skewness the Modified Murthy estimator – II is best and is followed by Murthy (1957) estimator, Horvitz – Thompson (1952) estimator under Yates – Grundy (1953) draw-by-draw procedure and Horvitz – Thompson (1952) estimator under Brewer (1963a) selection procedure. For populations having positive kurtosis and positive skewness Murthy (1957) estimator outperform other estimators involved in the study and is followed by Horvitz – Thompson (1952) estimator under Yates – Grundy (1953) draw-by-draw procedure and Modified Murthy estimator – II. Further, for populations having positive kurtosis and negative skewness Modified Murthy estimator – I is best and is followed by Modified Murthy estimator – III and Murthy (1957) estimator. In general for populations having negative kurtosis Murthy (1957) estimator clearly outperform other estimators involved in the study and is closely followed by Horvitz – Thompson (1952) estimator under Yates – Grundy (1953) draw-by-draw procedure and Horvitz – Thompson (1952) estimator under Brewer (1963a) selection procedure. Finally, for populations having positive kurtosis Murthy (1957) estimator outperform other estimators involved in the study and is closely followed by Modified Murthy estimator – II and Horvitz – Thompson (1952) estimator under Brewer (1963a) selection procedure.

Table 6.8 contains the regression summary for the model

$$Rank(Estim) = \beta_0 + \beta_1 \left[Rank(CV)\right] + \beta_2 \left[Rank(\rho)\right] + \varepsilon$$

From the study of this table it can be seen that the regression is significant for all the estimators under study except for the Horvitz – Thompson estimator (1952) under Yates – Grundy (1953) draw-by-draw procedure, the Horvitz – Thompson estimator under Brewer (1963a) procedure, the Rao, Hartley and Cochran (1962) and the New estimator – II. Further, the coefficients of these models show a very interesting picture. The coefficient of the CV(X) for Simple Random Sampling, Hansen – Hurwitz (1943) estimator, Horvitz – Thompson (1952) estimator under Yates – Grundy (1953) procedure, Raj (1956a) estimator, Murthy (1957) estimator and the Raj modified estimator is negative. This indicates that the average rank of these estimators will increase with a decrease in the rank of the CV(X) keeping rank of the correlation coefficient at a constant level. From this it can be concluded that these estimators will perform better for populations having larger coefficient of variation in the measure of size. For other estimators the coefficient for CV(X) is positive indicating that these estimators will perform better for populations having smaller coefficient of variation in the measure of size. Also the coefficient for rank of correlation coefficient is positive only for Simple Random Sampling and the Horvitz – Thompson (1952) estimator under Yates – Grundy (1953) draw-by-draw procedure. This indicates that these two methods will produce larger average rank for the variance in populations having larger rank for correlation coefficient. For rest of the procedures this coefficient is negative indicating that the average rank of these estimators will decrease with an increase in the rank of correlation coefficient. From this it is concluded that the Horvitz – Thompson (1952) estimator under Yates – Grundy (1953) draw-by-draw procedure will perform better for populations having small correlations. Overall it is concluded that the Modified Raj estimator will perform better in populations having larger coefficient of variation along with larger correlation coefficient. Other New estimators will perform better for populations having small coefficient of variation and large correlation coefficient.

Chapter 7

The Model Based Study

7.1 Introduction:

In this chapter the model based study of Modified Murthy Estimator – II, t_{MM2}, have been carried out under the linear stochastic model (1.1.1). The anticipated variance of this estimator has been obtained. Empirical study has been carried out to decide about the performance of various estimators.

7.2 Model Based Study of Modified Murthy Estimator – II:

In this section the model based study of modified Murthy estimator – II has been carried out. It has been shown that the modified Murthy estimator – II is model unbiased under the linear stochastic model given in (1.3.2) the anticipated variance of modified Murthy estimator – II has also been obtained by considering the super – population model given in (1.1.1). These studies have been given in following pages:

The New estimator – II is given as:

$$t_{MM2} = \frac{1}{2}\left[\frac{y_i}{p_i} + \frac{y_j}{p_j}\right] = \sum_{i \in S} \frac{y_i}{2p_i} \qquad (5.2.3)$$

The estimator (5.2.3) will be model unbiased if and only if:

$$E_M\left(t_{MM2} - Y\right) = 0$$

Substituting the value of t_{MM2} from (5.2.3) and the value of Y:

$$E_M\left(t_{MM2} - Y\right) = E_M\left(t_{MM2} - \beta Z - \sum_{i=1}^{N} \varepsilon_i\right)$$

$$= E_M\left(\sum_{i \in S} \frac{y_i}{2p_i} - \beta Z - \sum_{i=1}^{N} \varepsilon_i\right)$$

$$= E_M\left(\sum_{i \in S} \frac{\beta Z_i + \varepsilon_i}{2p_i} - \beta Z - \sum_{i=1}^{N} \varepsilon_i\right)$$

$$= \sum_{i \in S} \frac{\beta Z_i + E_M(\varepsilon_i)}{2 p_i} - \beta Z - \sum_{i=1}^{N} E_M(\varepsilon_i)$$

$$= \sum_{i \in S} \frac{\beta Z_i}{2 p_i} - \beta Z$$

$$= \sum_{i \in S} \frac{\beta Z_i}{2 Z_i / Z} - \beta Z$$

$$= \frac{1}{2} \sum_{i \in S} \beta Z - \beta Z$$

$= 0$ if and only if $\frac{1}{2} \sum_{i \in S} \beta Z = \beta Z$.

Hence the modified Murthy estimator – II will be model unbiased if and only if $\frac{1}{2} \sum_{i \in S} \beta Z = \beta Z$.

The anticipated variance of New estimator – II is obtained as under:

Now:

$$E_D\, E_M\, (t_{MM2} - Y)^2 = E_D\, E_M \left[\frac{1}{2} \sum_{i \in S} \frac{Y_i}{P_i} - Y \right]^2$$

Substituting the value if Y_i from (1.1.1) and the value of Y:

$$E_D\, E_M\, (t_{MM2} - Y)^2 = E_D\, E_M \left[\frac{1}{2} \sum_{i \in S} \frac{\beta Z_i + \varepsilon_i}{P_i} - \beta Z - \sum_{i=1}^{N} \varepsilon_i \right]^2$$

$$= E_D\, E_M \left[\sum_{i \in S} \frac{\beta Z_i + \varepsilon_i}{2 P_i} - \beta Z - \sum_{i=1}^{N} \varepsilon_i \right]^2$$

$$= E_D\, E_M \left[\sum_{i \in S} \frac{\beta Z_i}{2 P_i} + \sum_{i \in S} \frac{\varepsilon_i}{2 P_i} - \beta Z - \sum_{i=1}^{N} \varepsilon_i \right]^2$$

Substituting $P_i = Z_i / Z$ in first term:

$$E_D\, E_M\, (t_{MM2} - Y)^2 = E_D\, E_M \left[\sum_{i \in S} \frac{\beta Z_i}{2 Z_i / Z} + \sum_{i \in S} \frac{\varepsilon_i}{2 P_i} - \beta Z - \sum_{i=1}^{N} \varepsilon_i \right]^2$$

$$= E_D\, E_M \left[\frac{1}{2} \sum_{i \in S} \beta Z + \sum_{i \in S} \frac{\varepsilon_i}{2 P_i} - \beta Z - \sum_{i=1}^{N} \varepsilon_i \right]^2$$

Since $\frac{1}{2}\sum_{i\in S}\beta Z = \beta Z$, therefore

$$E_D\, E_M\, (t_{MM2} - Y)^2 = E_D\, E_M \left[\sum_{i\in S} \frac{\varepsilon_i}{2P_i} - \sum_{i=1}^{N} \varepsilon_i \right]^2$$

Put $\pi_i = 2P_i$:

$$E_D\, E_M\, (t_{MM2} - Y)^2 = E_D\, E_M \left[\sum_{i\in S} \frac{\varepsilon_i}{\pi_i} - \sum_{i=1}^{N} \varepsilon_i \right]^2$$

$$= E_D\, E_M \left[\sum_{i\in S} \frac{\varepsilon_i}{\pi_i} - \sum_{i\in S} \varepsilon_i - \sum_{j\notin S} \varepsilon_j \right]^2$$

$$= E_D\, E_M \left[\sum_{i\in S} \varepsilon_i \left(\frac{1}{\pi_i} - 1 \right) - \sum_{j\notin S} \varepsilon_j \right]^2$$

$$= E_D\, E_M \left[\sum_{i\in S} \varepsilon_i^2 \left(\frac{1}{\pi_i} - 1 \right)^2 + \sum_{j\notin S} \varepsilon_j^2 - 2 \sum_{i\in S} \sum_{\substack{j\notin S \\ j\neq i}} \varepsilon_i\, \varepsilon_j \right]$$

$$= E_D \left[\sum_{i\in S} E_M \left(\varepsilon_i^2 \right) \left(\frac{1}{\pi_i} - 1 \right)^2 + \sum_{j\notin S} E_M \left(\varepsilon_j^2 \right) - 2 \sum_{i\in S} \sum_{\substack{j\in S \\ j\notin S}} E_M \left(\varepsilon_i\, \varepsilon_j \right) \right]$$

$$= E_D \left[\sum_{i\in S} \sigma_i^2 \left(\frac{1}{\pi_i} - 1 \right)^2 + \sum_{j\notin S} \sigma_j^2 \right]$$

$$= E_D \left[\sum_{i\in S} \sigma_i^2 \left(\frac{1}{\pi_i^2} + 1 - \frac{2}{\pi_i} \right) + \sum_{j\notin S} \sigma_j^2 \right]$$

$$= E_D \left[\sum_{i\in S} \left(\frac{\sigma_i^2}{\pi_i^2} + \sigma_i^2 - \frac{2\sigma_i^2}{\pi_i} \right) + \sum_{j\notin S} \sigma_j^2 \right]$$

$$= E_D \left[\sum_{i\in S} \frac{\sigma_i^2}{\pi_i^2} + \sum_{i\in S} \sigma_i^2 - \sum_{i\in S} \frac{2\sigma_i^2}{\pi_i} + \sum_{j\notin S} \sigma_j^2 \right]$$

$$= E_D \left[\sum_{i\in S} \frac{\sigma_i^2}{\pi_i^2} - \sum_{i\in S} \frac{2\sigma_i^2}{\pi_i} + \sum_{i=1}^{N} \sigma_i^2 \right]$$

$$= E_D\left[\sum_{i\in S}\frac{\sigma_i^2}{\pi_i^2} - 2\sum_{i\in S}\frac{\sigma_i^2}{\pi_i} + \sum_{i=1}^{N}\sigma_i^2\right]$$

$$= \sum_S\left[\sum_{i\in S}\frac{\sigma_i^2}{\pi_i^2} - 2\sum_{i\in S}\frac{\sigma_i^2}{\pi_i} + \sum_{i=1}^{N}\sigma_i^2\right]P(S)$$

$$= \sum_S\left[\sum_{i\in S}\frac{\sigma_i^2}{\pi_i^2}P(S) - 2\sum_{i\in S}\frac{\sigma_i^2}{\pi_i}P(S) + \sum_{i=1}^{N}\sigma_i^2 P(S)\right]$$

$$= \sum_{i=1}^{N}\sum_{S\ni i}\frac{\sigma_i^2}{\pi_i^2}P(S) - 2\sum_{i=1}^{N}\sum_{S\ni i}\frac{\sigma_i^2}{\pi_i}P(S) + \sum_{i=1}^{N}\sum_{S\ni i}\sigma_i^2 P(S)$$

$$= \sum_{i=1}^{N}\frac{\sigma_i^2}{\pi_i^2}\sum_{S\ni i}P(S) - 2\sum_{i=1}^{N}\frac{\sigma_i^2}{\pi_i}\sum_{S\ni i}P(S) + \sum_{i=1}^{N}\sigma_i^2\sum_{S}P(S)$$

Since $\sum_{S\ni i}P(S)=\pi_i$ and $\sum_{S}P(S)=1$, therefore:

$$E_D\,E_M\,(t_{MM2}-Y)^2 = \sum_{i=1}^{N}\frac{\sigma_i^2}{\pi_i^2}\pi_i - 2\sum_{i=1}^{N}\frac{\sigma_i^2}{\pi_i}\pi_i + \sum_{i=1}^{N}\sigma_i^2$$

$$= \sum_{i=1}^{N}\frac{\sigma_i^2}{\pi_i} - 2\sum_{i=1}^{N}\sigma_i^2 + \sum_{i=1}^{N}\sigma_i^2$$

$$= \sum_{i=1}^{N}\frac{\sigma_i^2}{\pi_i} - \sum_{i=1}^{N}\sigma_i^2$$

$$= \sum_{i=1}^{N}\sigma_i^2\left(\frac{1}{\pi_i}-1\right) \qquad (7.2.1)$$

From (7.2.1) it is obvious that the modified Murthy estimator – II achieves the Godambe – Joshi lower bound for variance of any estimator.

7.3 Empirical Study:

In this section the empirical study has been carried out using ten selected populations. The anticipated variance of Modified Murthy Estimator – II has been compared with the anticipated variance of Horvitz – Thompson estimator under various selection procedures. The results of this empirical study are given in following tables:

Table 7.1: Anticipated Variance of Modified Murthy Estimator – II along with the Anticipated Variance of Horvitz – Thompson Estimator.

Population Number – 1

Estimators	***Values of γ***					
	0.5	***0.6***	***0.7***	***0.8***	***0.9***	***1.0***
Modified Murthy – II (t_{MM2})	666.00	852.79	1098.27	1422.45	1852.61	2426.00
Horvitz – Thompson (Brewer)	666.00	852.79	1098.27	1422.45	1852.61	2426.00
Horvitz – Thompson (YG d-b-d)	663.20	849.93	1095.54	1420.16	1851.24	2426.33
Horvitz – Thompson (YG rej)	663.11	850.63	1097.51	1424.10	1858.19	2437.80

Population Number – 3

Estimator	***Values of γ***					
	0.5	***0.6***	***0.7***	***0.8***	***0.9***	***1.0***
Modified Murthy – II (t_{MM2})	14580.00	33532.30	78279.70	185725.30	448364.88	1102390
Horvitz – Thompson (Brewer)	14580.00	33532.30	78279.70	185725.30	448364.88	1102390
Horvitz – Thompson (YG d-b-d)	14302.21	32995.18	77303.58	184166.84	446660.94	1103764
Horvitz – Thompson (YG rej)	14170.92	32814.79	77215.04	184865.30	450820.09	1120684

Table 7.1 (Continued)

Population Number – 4

Estimator	Values of γ					
	0.5	***0.6***	***0.7***	***0.8***	***0.9***	***1.0***
Modified Murthy – II (t_{MM2})	2276.00	5090.29	11395.92	25538.59	57291.03	128653.50
Horvitz – Thompson (Brewer)	2276.00	5090.29	11395.92	25538.59	57291.03	128653.50
Horvitz – Thompson (YG d-b-d)	2271.64	5082.57	11383.27	25520.73	57274.83	128671.23
Horvitz – Thompson (YG rej)	2301.27	5151.36	11543.00	25891.65	58136.34	130672.69

Population Number – 5

Estimator	Values of γ					
	0.5	***0.6***	***0.7***	***0.8***	***0.9***	***1.0***
Modified Murthy – II (t_{MM2})	2352.00	5301.56	11957.61	26987.39	60947.38	137730
Horvitz – Thompson (Brewer)	2352.00	5301.56	11957.61	26987.39	60947.38	137730
Horvitz – Thompson (YG d-b-d)	2349.19	5296.56	11949.35	26975.65	60936.60	137741.55
Horvitz – Thompson (YG rej)	2380.57	5368.93	12116.32	27360.99	61826.23	139796.14

Table 7.1 (Continued)

Population Number – 8

Estimator	Values of γ					
	0.5	**0.6**	**0.7**	**0.8**	**0.9**	**1.0**
Modified Murthy – II (t_{MM2})	1260.00	2043.33	3314.40	5377.36	8726.29	14164.00
Horvitz – Thompson (Brewer)	1260.00	2043.33	3314.40	5377.36	8726.29	14164.00
Horvitz – Thompson (YG d-b-d)	1259.73	2042.99	3313.99	5376.94	8725.98	14164.11
Horvitz – Thompson (YG rej)	1265.55	2052.53	3329.63	5402.58	8768.02	14233.03

Population Number – 9

Estimator	Values of γ					
	0.5	**0.6**	**0.7**	**0.8**	**0.9**	**1.0**
Modified Murthy – II (t_{MM2})	539.00	735.38	1004.61	1374.20	1882.23	2581.50
Horvitz – Thompson (Brewer)	539.00	735.38	1004.61	1374.20	1882.23	2581.50
Horvitz – Thompson (YG d-b-d)	538.29	734.61	1003.83	1373.51	1881.81	2581.63
Horvitz – Thompson (YG rej)	540.38	737.68	1008.34	1380.11	1891.45	2595.68

Table 7.1 (Continued)

Population Number – 12

Estimator	*Values of γ*					
	0.5	***0.6***	***0.7***	***0.8***	***0.9***	***1.0***
Modified Murthy – II (t_{MM2})	930.60	1280.82	1770.69	2457.16	3420.84	4775.90
Horvitz – Thompson (Brewer)	930.60	1280.82	1770.69	2457.16	3420.84	4775.90
Horvitz – Thompson (YG d-b-d)	928.78	1278.93	1768.84	2455.57	3419.84	4776.11
Horvitz – Thompson (YG rej)	930.11	1281.44	1773.16	2462.62	3431.01	4793.46

Population Number – 13

Estimator	*Values of γ*					
	0.5	***0.6***	***0.7***	***0.8***	***0.9***	***1.0***
Modified Murthy – II (t_{MM2})	1875.06	2850.46	4740.83	8495.10	16134.27	32027.64
Horvitz – Thompson (Brewer)	1875.06	2850.46	4740.83	8495.10	16134.27	32027.64
Horvitz – Thompson (YG d-b-d)	1741.35	2686.62	4543.18	8282.86	16003.74	32303.43
Horvitz – Thompson (YG rej)	1692.24	2671.31	4629.99	8650.48	17107.52	35287.27

Table 7.1 (Continued)

Population Number – 16

Estimator	*Values of γ*					
	0.5	***0.6***	***0.7***	***0.8***	***0.9***	***1.0***
Modified Murthy – II (t_{MM2})	425.00	618.09	904.28	1330.66	1969.07	2929.50
Horvitz – Thompson (Brewer)	425.00	618.09	904.28	1330.66	1969.07	2929.50
Horvitz – Thompson (YG d-b-d)	421.75	614.36	900.30	1326.96	1966.77	2930.75
Horvitz – Thompson (YG rej)	423.83	618.60	908.27	1341.31	1991.85	2973.74

Population Number – 17

Estimator	*Values of γ*					
	0.5	***0.6***	***0.7***	***0.8***	***0.9***	***1.0***
Modified Murthy – II (t_{MM2})	510.00	767.90	1162.94	1771.70	2715.58	4188.00
Horvitz – Thompson (Brewer)	510.00	767.90	1162.94	1771.70	2715.58	4188.00
Horvitz – Thompson (YG d-b-d)	504.96	761.78	1156.03	1764.95	2711.29	4191.09
Horvitz – Thompson (YG rej)	507.06	766.75	1166.44	1785.46	2750.17	4263.00

Table 7.2: Relative Efficiency of Modified Murthy Estimator – II.

Population Number – 1

	Values of γ					
	0.5	**0.6**	**0.7**	**0.8**	**0.9**	**1.0**
RE (t_{MM2} to YG d-b-d)	99.58	99.66	99.75	99.84	99.93	100.01
RE (t_{MM2} to YG rej)	99.57	99.75	99.93	100.12	100.3	100.49

Population Number – 3

	Values of γ					
	0.5	**0.6**	**0.7**	**0.8**	**0.9**	**1.0**
RE (t_{MM2} to YG d-b-d)	98.09	98.4	98.75	99.16	99.62	100.12
RE (t_{MM2} to YG rej)	97.19	97.86	98.64	99.54	100.55	101.66

Table 7.2 (Continued)

Population Number – 4

	Values of γ					
	0.5	**0.6**	**0.7**	**0.8**	**0.9**	**1.0**
RE (t_{MM2} to YG d-b-d)	99.81	99.85	99.89	99.93	99.97	100.01
RE (t_{MM2} to YG rej)	101.11	101.2	101.29	101.38	101.48	101.57

Population Number – 5

	Values of γ					
	0.5	**0.6**	**0.7**	**0.8**	**0.9**	**1.0**
RE (t_{MM2} to YG d-b-d)	99.88	99.91	99.93	99.96	99.98	100.01
RE (t_{MM2} to YG rej)	101.21	101.27	101.33	101.38	101.44	101.5

Table 7.2 (Continued)

Population Number – 8

	Values of γ					
	0.5	***0.6***	***0.7***	***0.8***	***0.9***	***1.0***
RE (t_{MM2} to YG d-b-d)	99.98	99.98	99.99	99.99	100.01	100.01
RE (t_{MM2} to YG rej)	100.44	100.45	100.46	100.47	100.48	100.49

Population Number – 9

	Values of γ					
	0.5	***0.6***	***0.7***	***0.8***	***0.9***	***1.0***
RE (t_{MM2} to YG d-b-d)	99.87	99.89	99.92	99.95	99.98	100.01
RE (t_{MM2} to YG rej)	100.26	100.31	100.37	100.43	100.49	100.55

Table 7.2 (Continued)

Population Number – 12

	Values of γ					
	0.5	***0.6***	***0.7***	***0.8***	***0.9***	***1.0***
RE (t_{MM2} to YG d-b-d)	99.8	99.85	99.9	99.94	99.97	100
RE (t_{MM2} to YG rej)	99.95	100.05	100.14	100.22	100.3	100.37

Population Number – 13

	Values of γ					
	0.5	***0.6***	***0.7***	***0.8***	***0.9***	***1.0***
RE (t_{MM2} to YG d-b-d)	92.87	94.25	95.83	97.5	99.19	100.86
RE (t_{MM2} to YG rej)	90.25	93.72	97.66	101.83	106.03	110.18

Table 7.2 (Continued)

Population Number – 16

	Values of γ					
	0.5	**0.6**	**0.7**	**0.8**	**0.9**	**1.0**
RE (t_{MM2} to YG d-b-d)	99.24	99.4	99.56	99.72	99.88	100.04
RE (t_{MM2} to YG rej)	99.72	100.08	100.44	100.8	101.16	101.51

Population Number – 17

	Values of γ					
	0.5	**0.6**	**0.7**	**0.8**	**0.9**	**1.0**
RE (t_{MM2} to YG d-b-d)	99.01	99.2	99.41	99.62	99.84	100.07
RE (t_{MM2} to YG rej)	99.42	99.85	100.3	100.78	101.27	101.79

7.4 Conclusions:

The empirical study has been carried out by evaluating the anticipated variance of modified Murthy estimator – II along with the anticipated variance of Horvitz – Thompson estimator under various selection procedures. Three selection procedures have been selected to carry out this study. These selection procedures are the Brewer (1963a) selection procedure; which yield inclusion probabilities exactly proportional to measure of size, Yates – Grundy (1953) draw-by-draw procedure and Yates – Grundy (1953) rejective procedure. The procedures of Yates – Grundy do not yield inclusion probabilities exactly proportional to size. The results of this analysis are given in Table – 7.1 and Table 7.2. From these analyses it is clear that the Horvitz – Thompson estimator under the Yates – Grundy draw-by-draw procedure perform reasonably well as compared to the other procedures for almost all the selected populations. The New estimator – II has same anticipated variance as the Horvitz – Thompson estimator under the Brewer selection procedure. The Horvitz – Thompson estimator under Yates – Grundy rejective procedure has smaller anticipated variance as compared to the New estimator – II and Horvitz – Thompson estimator under the Brewer selection procedure for small values of γ. The performance of modified Murthy estimator – II is better than the Horvitz – Thompson estimator under Yates – Grundy rejective procedure for larger values of γ. Finally, the analysis yield that the modified Murthy estimator – II has same anticipated variance as the Horvitz – Thompson estimator under the Brewer selection procedure.

References

1. Agarwal, P and Kumar, P. (1982) "A note on the Rao's inclusion probabilities proportional to size sampling schemes" *Biometrical J.* 24(8), 807 – 810.
2. Agarwal, S. K., Kumar, P. and Dey, A. (1982) "On unequal probability sampling of two units without replacement", *J. Roy. Stat. Soc.* B, 44(1), 43 – 46.
3. Agarwal, S. K. and Goel, B. B. P. S. (1977) "A note on unequal probability sampling", *J. Ind. Soc. Agri. Stat.* 29(2), 98 – 100.
4. Agarwal, S. K. and Singh, P. (1986) "A sampling scheme with inclusion probability to size using PPS systematic sampling", *J. Ind. Soc. Agri. Stat.* 38, 153 – 160.
5. Agarwal, R., Singh, D. and Singh, P. (1984) "Systematic sampling using varying probabilities", *J. Ind. Soc. Agri. Stat.* 31(1), 99 – 109.
6. Agarwal, R., Singh, P. and Singh, D. (1984) "IPPs sampling scheme through grouping", *Biometrical J.* 26, 527 – 533.
7. Aires, N. (2000) "Comparison between conditional poisson sampling and pareto sampling designs" In ICES – II.
8. Basu, D. (1971) "An essay on the logical foundation of survey sampling – part one" *Foundation of Statistical Inference.* Holt, Rinehard and Winston, Edited by Godambe and Sprott.
9. Bayless, D. L. and Rao, J. N. K. (1970) "An empirical study of stabilities of estimators and variance estimators in unequal probability sampling for n = 3 and 4", *J. Amer. Stat. Assoc.* 65, 1645 – 1667.
10. Beg, M. A. and Hanif (1991) "Comparison of sampling with and without replacement using unequal probability sampling", *Bull. Inter. Stat. Inst.* 48(4), 118 – 119.
11. Brewer, K. R. W. (1963a) "A model of systematic sampling with unequal probabilities", *Aust. J. Stat.* 5, 5 – 13.
12. Brewer, K. R. W. (1963b) "Ratio estimation and finite population: some results deducible from assumption of an underlying stochastic process", *Aust. J. Stat.* 5, 93 – 105.

13. Brewer, K. R. W. (1967) "A note of Fellegi's method of sampling without replacement with probabilities proportional to size", *J. Amer. Stat. Assoc.*, 62, 79 – 85.
14. Brewer, K. R. W. (1975) "A simple procedure for sampling πpswor", *Aust. J. Stat.* 17(3), 166 – 172.
15. Brewer, K. R. W. (1979) "A class of robust sampling design for large scale surveys", *J. Amer. Stat. Assoc.* 74(4), 911 – 915.
16. Brewer, K. R. W. and Hanif, M. (1969a) "Sampling without replacement with probability of inclusion proportional to size – I: Methods using Horvitz – Thompson estimator", *Unpublished manuscript.*
17. Brewer, K. R. W. and Hanif, M. (1969b) "Sampling without replacement with probability of inclusion proportional to size – I: Methods using special estimators", *Unpublished manuscript.*
18. Brewer, K. R. W., Early, L. J. and Joyce, S. F. (1972) "Selecting several samples from a single population", Aust. J. Stat., 14, 231–239.
19. Brewer, K. R. W., Early, L. J. and Hanif, M. (1984) "Poisson, modified poisson and collocated sampling", *J. Stat. Plan. and Infer.* 10, 15 – 30.
20. Brewer, K. R. W. and Hanif, M. (1983) "*Sampling with Unequal Probabilities*", Lecture notes to Statistics, No. 15, Springer – Verlag.
21. Brewer, K. R. W. and Hanif, M. and Tam, S. M. (1988) "How nearly can a model based predictor and design based estimator reconciled?", *J. Amer. Stat. Assoc.* 83, 128 – 132.
22. Brewer, K. R. W. and Sarndal, C. E. (1983) "Six approaches to enumerative survey sampling (with discussion)", *Incomplete data in survey sampling*, 3(VIII), 363 – 413.
23. Brewer, K. R. W. and Undy, G. C. (1962) "Samples of two units drawn with unequal probabilities without replacement", *Aust. J. Stat.* 4, 89 – 100.
24. Carroll, J. L. and Hartley, H. O. (1964) "The systematic method of unequal probability sampling without replacement", *Abstract in Biometrics*, 20, 908 – 909.

25. Cassel, C. M., Sarndal, C. E. and Wretman, J. H. (1976) "Some results on generalized difference estimation and generalized regression estimation for finite populations", *Biometrika*, 63, 615–620.
26. Cassel, C. M., Sarndal, C. E. and Wretman, J. H. (1976) "*Foundation of inference in survey sampling*", John Wiley and Sons, Inc. New York.
27. Chao, M. T. (1982) "A general purpose unequal probability sampling plan", *Biometrika*, 69, 653 – 656.
28. Chaudhuri, A. (1971) "Some sampling schemes to use Horvitz – Thompson estimator in estimating finite population total", *Culc. Stat. Assoc. Bull.* 20, 37 – 66.
29. Chaudhuri, A. (1972) "A comparative study of Horvitz – Thompson and symmetrized Raj strategies", *Culc. Stat. Assoc.* Bull. 21, 21 – 44.
30. Chaudhuri, A. (1975a) "A simple method of sampling without replacement with inclusion probabilities exactly proportional to size" *Metrika*, 22, 147 – 152.
31. Chaudhuri, A. (1975b) "Some properties of estimators based on sampling schemes with varying probabilities", *Aust. J. Stat.* 17, 22–28.
32. Chaudhuri, A. (1976) "On the choice of sampling strategies", *Culc. Stat. Assoc. Bull.* 25, 119 – 128.
33. Chaudhuri, A. (1979) "A comparison among some sampling strategies under a model", *J. Ind. Soc. Agri. Stat.* 30(2), 17 – 22.
34. Chaudhuri, A. (1981) "Non – negative unbiased variance estimators", *Current Topics in Survey Sampling and Applied Statistics*, H. A. David (editor), Academic Press, New York.
35. Chaudhuri, A. (1985) "On optimal and related strategies for sampling on two occasions with varying probabilities", *J. Ind. Soc. Agri. Stat.* 37(1), 45 – 53.
36. Chaudhuri, A. (1996) "Estimation of total and variance estimation in a version of unequal probability sampling in three stages", *Pak. J. Stat.* 12(1), 1 – 7.
37. Chaudhuri, A. and Arnab, R. (1979a) "On the relative efficiencies of sampling strategies under super – population model", *Sankhya*. C, 41, 40 – 43.
38. Chaudhuri, A. and Vos, J. W. E. (1986) "*Unified Theory and Strategies of Survey Sampling*", North Holland series in Statistics and Probability.
39. Choudhry, G. H. and Singh, M. P. (1979) "Sampling with unequal probabilities and without replacement – a rejective method", *Survey Methodology*, 5(2), 162 – 177.

40. Chromy, J. R. (1979) "Sequential sample selection methods", *Unpublished manuscript*.

41. Cochran, W. G. (1953, 1963, 1977) "*Sampling Techniques*", John Wiley and Sons, Inc., New York.

42. Connor, W. S. (1966) "An exact formula for the probability that two specified sampling units will occur in a sample drawn with unequal probabilities", *J. Amer. Stat. Assoc.*, 61, 384 – 390.

43. Cumberland, W. G. and Royall, R. M. (1981) "Prediction model and unequal probability sampling", *J. Roy. Stat. Soc.* B, 43, 353 – 367.

44. Das, A. C. (1951) "On two phase sampling and sampling with varying probabilities", *Bull. Inter. Stat. Inst.* 33(2), 105 – 112.

45. Das, M. N. and Mohanty, S. (1973) "On PPS sampling without replacement ensuring selection probabilities exactly to size", *Aust. J. Stat.* 15, 87 – 94.

46. Deshpande, M. N. (1982) "A new fixed size sampling procedure with varying probabilities", *J. Ind. Soc. Agri. Stat.*, 34(2), 71 – 75.

47. Deshpande, M. N. (1984) "A new method of sampling with unequal probabilities", *J. Ind. Soc. Agri. Stat.*, 34(2), 23 – 26.

48. Deshpande, M. N., Prabhu – Ajgonkar, S. G. (1982) "An IPPS sampling scheme", *Statistica*, 36(4), 209 – 212.

49. Day, A. and Srivastava, A. K. (1987) "A simple procedure with inclusion probability proportional to size", *Survey Methodology*, 13, 85 – 92.

50. Dodds, D. J. and Freyer, J. G. (1971) "Some families of selection probabilities for sampling proportional to size", *J. Roy. Stat. Soc.*, B, 33, 263 – 272.

51. Durbin, J. (1953a) "Some results in sampling when the units are selected with unequal probabilities", *J. Roy. Stat. Soc*, B, 15, 262–269.

52. Durbin, J. (1953b) "Sampling with unequal probabilities of selection within strata. Unpublished to date. Seen by the author.

53. *Durbin*, J. (1967) "Design of multistage surveys for estimation of sampling errors", *Appl. Stat.*, 16, 152 – 164.

54. Fellegi, I. P. (1963) "Sampling with varying probabilities without replacement: rotating and non – rotating samples", *J. Amer. Stat. Assoc.* 58, 183 – 201.

55. Foreman, E. K. and Brewer, K. R. W. (1971) "The efficient use of supplementary information in standard sampling procedures", *J. Roy. Stat. Soc.* B, 33, 391 – 400.
56. Fuller, W. and Isaki, C. T. (1981) "Survey design under super population model", *Current Topics in Survey Sampling*, D. Kreweki, R. Plaket and J. N. K. Rao. Acedamic Press. 199 – 226.
57. Gabler, S. and Scheweigkoffer, R. (1990) "The existence of sampling designs with pre-assigned probabilities", *Matrika*. 37, 87 – 97.
58. Gazali, S. S. A. (1996) "Modification of Jessen method – 3", *Pak. J. Stat.* 12(3), 179 – 188.
59. Gazali, S. S. A. (1998) "Modification of Jessen method – 2", *Pak. J. Stat.* 14(1), 1 – 10.
60. Godambe, V. P. (1955) "A unified theory of sampling from finite populations", *J. Roy. Stat. Soc.* B, 17, 269 – 278.
61. Godambe, V. P. (1982) "Estimation in survey sampling: robustness and optimality", *J. Amer. Stat. Assoc.* 77, 393 – 406.
62. Godambe, V. P. and Joshi, V. M. (1965) "Admissibility and Bayes estimation in sampling finite population, I, II and III", *Ann. Math. Stat.* 36, 1707 – 1742.
63. Goodman, R. and Kish, L. (1950) "Controlled selection – a technique in probability sampling", *J. Amer. Stat. Assoc.* 45, 350 – 372.
64. Gupta, V. K. (1989) "A method of inclusion probability proportional to size selection", *Statistica*, XLIX (2), 277 – 282.
65. Hajek, J. (1964) "Asymptotic theory of rejective sampling with varying probabilities from a finite population", *Ann. Math. Stat.* 35, 1491–1523.
66. Hajek, J. (1971) "Comments on An essay on the logical foundation of survey sampling, part one, by Basu" in *Foundation of Statistical Inference*, Godambe, V. P. and Sprott, D. A. (editors) Holt, Rinheart and Winston, Toranto, 326.
67. Hajek, J. (1981) "*Sampling from a finite population*" Marcel Dekker Inc. New York.
68. Hanif, M. (1974) "An asymptotic variance formula for sampling without replacement with unequal probabilities", *Libyan J. Sci.* A, 47 – 51.
69. Hanif, M. and Brewer, K. R. W. (1980) "Sampling with unequal probabilities without replacement; a review", *Inter. Stat. Rev.* 48(3), 317 – 335.

70. Hanif, M. and Brewer, K. R. W. (1979 – 1980) "Generalization of the Horvitz and Thompson estimator", *Punj. Uni. J. Math.* 12–13, 123-34.
71. Hanif, M., Mukhopadhyay, P. and Bhattacharyya, S. (1993) "On estimating the variance of Horvitz – Thompson estimator", *Pak. J. Stat.*, 9(2), A, 123 – 136.
72. Hanif, M., Samiuddin, M. and Kattan, A. K. A. (1994) "Small sample performance of revised ratio estimator", *Pak. J. Stat.* 10(1), A, 123 –130.
73. Hanif, M. (1994) "*Design and model based sampling inference*", Unpublished Ph. D. Thesis.
74. Hanif, M. and Ahmad, M. (2001) "Approximate variance formula for variance of Horvitz – Thompson estimator using first order inclusion probabilities", *Presented at ISOSS – VII*, Lahore, 2001.
75. Hansen, M. H. and Hurwitz, W. N. (1943) "On the theory of sampling from a finite population", *Ann. Math. Stat.* 14, 333 – 362.
76. Hansen, M. H., Hurwitz, W. N. an Madow, W. G. (1953) "*Sample Survey Methods and Theory*", John Wiley and Sons Inc., New York.
77. Hansen, M. H., Madow, W. G. and Tepping, B. J. (1983) "An evaluation of model dependent and probability sampling inference in sample surveys, with discussion", *J. Amer. Stat. Assoc.* 78, 776 – 807.
78. Hanurav, T. V. (1962a) "Some Sampling schemes in probability sampling", *Sankhya*, A. 24, 428.
79. Hanurav, T. V. (1966) "An optimum πps – sampling", *Ann. Math. Stat.* 37, 1859.
80. Hanurav, T. V. (1967) "Optimum utilization of auxiliary information: πps sampling of two units from a stratum", *J. Roy. Stat. Soc.* B, 29, 374 – 391.
81. Hanurav, T. V. (1969) "Optimum utilization of auxiliary information: πps sampling of two units from a stratum – addenda and corrigenda", *J. Roy. Stat. Soc.* B, 31, 192 – 194.
82. Hartley, H. O. and Rao, J. N. K. (1962) "Sampling with unequal probabilities and without replacement", *Ann. Math. Stat.* 33, 350 – 374.
83. Herzel, A. (1986) "Sampling without replacement with unequal probabilities with pre-assigned joint inclusion probabilities of any order" *Metron*, 44, 1 – 4, 49 – 68.
84. Horvitz, D. G. and Thompson, D. J. (1952) "A generalization of sampling without replacement from a finite universe", *J. Amer. Stat. Assoc.* 47, 663 – 685.

85. Isaki, C. T. and Fuller, W. A. (1982) "Survey design under the regression super – population model", *J. Amer. Stat. Assoc.*, 77, 209 – 218.

86. Jeffreys, H. (1961) "Theory of Probability", 3rd Edition, Oxford University Press, Oxford.

87. Jessen, R. J. (1969) "Some methods of probability non – replacement sampling", *J. Amer. Stat. Assoc.* 64, 175 – 193.

88. Jessen, R. J. (1978) "*Statistical Survey Techniques*" John Wiley and Sons Inc., New York.

89. Kalton, G. (1983) "Model in the practice of survey sampling", *Inter. Stat. Review*, 51, 175 – 193.

90. Kish, L. (1965) "*Survey Sampling*", John Wiley and Sons Inc., New York.

91. Kumar, P. and Agarwal, S. K. (1982) "On the sample procedure of inclusion probability proportional to size sampling", *Biometrical J.* 24(8), 801 – 806.

92. Kumar, P. and Agarwal, S. K. (1985) "A general class of unequal probability sampling scheme for samples of two units from a finite population", *Cahiers de CERO*, 27(3 – 4) 229 – 233.

93. Kumar, P., Gupta, V. K. and Agarwal, S. K. (1985) "On inclusion probability proportional to size sampling scheme", *J. Stat. Plan. Infer.* 12, 127 – 131.

94. Kumar, P., Gupta, V. K. and Nigam, A. K. (1987) "On sample selection procedure using two auxiliary characteristics", *Statistica*, XLVII, 2, 219 – 226.

95. Kumar, P. Srivastava, A. K. and Agarwal, S. K. (1986) "A general class of probability sampling designs", *Statistica*, 26, 67 – 73.

96. Kumar, P. Kathuria, O. P. and Agarwal, S. K. (1984) "On sampling scheme with inclusion probability proportional to size", *J. Oper. Res. Stat. (Sr. Stat.)*, 15(4), 596 – 600.

97. Lahiri, D. B. (1974) "A method of sample selection providing unbiased ratio estimates", *Bull. Inter. Stat. Inst.* 33, 2, 133 – 140.

98. Lanke, J. (1974) "On non – negative variance estimators in survey sampling", *Sankhya*, C, 36(1), 33 – 42.

99. Madow, W. G. (1949) "On the theory of systematic sampling II", *Ann. Math. Stat.* 20, 333 – 354.

100. Marriot, F. H. C. (1990) "*A dictionary of Statistical Terms*", Longman.

101. Midzuno, H. (1952) "On the sampling system with probability proportionate to sum of size", *Ann. Inst. Math. Stat.* 3, 99 – 107.

102. Mukhopadhyay, P. (1972) "A sample scheme to realize a pre-assigned set of inclusion probabilities of first two order", *Culc. Stat. Inst. Bull.* 21, 87 – 122.

103. Murthy, M. N. (1957) "Ordered and unordered estimators in sampling without replacement", *Sankhya*, 18, 379 – 390.

104. Murthy, M. N. (1967, 1976) "*Sampling Theory and Methods*", Stat. Pub. Soc. Culcutta.

105. Narain, R. D. (1951) "On sampling without replacement with varying probabilities", *J. Ind. Soc, Agri. Stat.* 3, 169 – 175.

106. Neyman, J. (1934) "On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection", *J. Roy. Stat. Soc.*, 97, 558 – 606.

107. Nigam, A. K., Kumar, P. and Gupta, V. K. (1984) "Some methods of inclusion probability proportional to size sampling", *J. Roy. Stat. Soc.*, B, 46(3), 564 – 571.

108. Ogus, J. L. and Clark, D. F. (1971) "The annual survey of manufacturers: a report on Methodology", *Technical Paper # 24*, U. S. Bureau of Census, Washington.

109. Pathak, P. K. (1961) "Use of order statistics in sampling without replacement", *Sankhya*, A, 23, 409 – 414.

110. Pathak, P. K. (1967a) "Asymptotic efficiency of Des Raj's strategy – I", *Sankhya* A, 29, 283 – 298.

111. Pathak, P. K. (1967b) "Asymptotic efficiency of Des Raj's strategy–II", *Sankhya* A, 29, 299 – 304.

112. Raj, D. (1954) "Ratio estimation in sampling with equal and unequal probabilities", *Jour. Ind. Soc. Agri. Stat.*, 6 (2), 127 – 138.

113. Raj, D. (1956a) "Some estimators in sampling with varying probabilities without replacement", *J. Amer. Stat. Assoc.* 60, 278–284.

114. Raj, D. (1956b) "A note in determination of optimum probabilities in sampling without replacement", *Sankhya*, 17, 197 – 200.

115. Raj, D. (1968) "*Sampling Theory*", McGraw – Hill, New York.

116. Raj, D. (1972) "*The Design of Sample Surveys*", McGraw – Hill, New York.

117. Rao, J. N. K. (1961) "On the estimate of variance in unequal probability sampling", *Ann. Inst. Stat. Math.*, 13(1), 57 – 60.

118. Rao, J. N. K. (1963a) "On two systems of unequal probability sampling without replacement", *Ann. Inst. Stat. Math.*, 15, 67 – 72.

119. Rao, J. N. K. (1963b) "On three procedures of unequal probability sampling without replacement", *J. Amer. Stat. Assoc.*, 58, 192 – 215.

120. Rao, J. N. K. (1965) "On two simple schemes of unequal probability sampling without replacement", *J. Ind. Stat. Assoc.* 3, 173 – 180.

121. Rao, J. N. K. (1966) "On the relative efficiency of some estimators in pps sampling for multiple characteristics", *Sankhya*, A, 28, 61 – 70.

122. Rao, J. N. K. and Bayless, D. L. (1969) "An empirical study of the stabilities of estimators and variance estimators in unequal probability sampling of two units per stratum", *J. Amer. Stat. Assoc.*, 64, 540 – 549.

123. Rao, J. N. K., Hartley, H. O. and Cochran, W. G. (1962) "On a simple procedure of unequal probability sampling without replacement", *J. Roy. Stat. Soc.*, B, 24, 482 – 491.

124. Rao, J. N. K. and Singh, M. P. (1973) "On the choice of estimators in survey sampling", *Aust. Jour. Stat.*, 15, 95 – 104.

125. Rao, T. J. (1977) "Estimating the variance of the ratio estimator for the Midzuno – Sen sampling scheme", *Metrika*, 24, 203 – 208.

126. Royall, R. M. (1970) "On finite population sampling theory under certain regression models", *Biometrika*, 57, 377 – 387.

127. Royall, R, M. (1971) "Linear regression models in finite population sampling theory." *Foundations of Statistical Inference, Eds, V. P. Godambe and D. A. Sprott*, 259 – 274, Toronto, Holt.

128. Royall, R. M. and Herson, J. (1973a) "Robust estimation in finite populations – I", *J. Amer. Stat. Assoc.* 68, 880 – 889.

129. Royall, R. M. and Herson, J. (1973b) "Robust estimation in finite populations – II", *J. Amer. Stat. Assoc.* 68, 890 – 893.

130. Sahoo, I., Sahoo, L.N. and Mohanty, S. (1993) "An unequal probability sampling without replacement procedure", *J. Stat. Res.* 27(1), 9 – 14.

131. Samiuddin, M. and Asad, H. (1981) "A simple procedure of unequal probability sampling", *Biometrika*, 68(3), 728 – 731.

132. Samiuddin, M., Hanif, M. and Asad, H. (1978) "Some remarks on PPS sampling without replacement", *Presented at Fourth Australian Statistical Conference*, Canberra.

133. Samiuddin, M. and Kattan, K. A. K. (1991) "A procedure of unequal probability sampling", *Pak. J. Stat.* 7(3), 1 – 7.

134. Samiuddin, M., Kattan, A. K. A., Hanif, M. and Asad, H. (1992) "Some remarks on models, sampling schemes and estimators in unequal probability sampling", *Pak. J. Stat.*, 8(1), A, 1 – 18.

135. Sampford, M. R. (1962) "*Sampling Methods for Agriculture*", Oliver and Boyds, London.

136. Sampford, M. R. (1967) "On Sampling without replacement with unequal probabilities of selection", *Biometrika*, 54, 499 – 513.

137. Sankaranarayanan, K. (1969) "An IPPS sampling scheme using Lahiri's method of selection", *J. Ind. Soc. Agri. Stat.*, 21, 58 – 66.

138. Sarndal, C. E. (1978) "Design based and model based inference in survey sampling", *Scan. J. Stat.* 5(1), 27 – 52.

139. Sarndal, C. E. (1984) "Design – consistent and model – dependent estimation for small domains", J. Amer. Stat. Assoc. 79(3), 624 – 631.

140. Sarndal, C. E. and Wright, R. L. (1984) "Cosmetic form of estimators in survey sampling", *Scan. J. Stat.*, 11, 146 – 156.

141. Sarndal, C. W., Swensson, B. and Wretman, J. H. (1992) "*Model Assisted Survey Sampling*", Springer – Verlag.

142. Saxena, R. R., Singh, P. and Srivastava, A. K. (1986) "On unequal probability sampling plan", *Biometrika*, 73, 761 – 763.

143. Sen, A. R. (1953) "On the estimate of the variance in sampling with varying probabilities", *J. Ind. Soc. Agri. Stat.*, 5, 119 – 127.

144. Sengupta, S. (1981) "A sampling scheme to realize inclusion probabilities exactly proportional to size", *J. Ind. Soc. Agri. Stat.*, 33(2), 119 – 127.

145. Shahbaz, M. Q. (2001) "Probability proportional to size sampling: without replacement", *Unpublished M. Phil. thesis*, submitted at Government College, Lahore.

146. Shahbaz, M. Q. and Hanif, M. (2003) "A simple procedure for unequal probability sampling without replacement and a sample of size 2", *Pak. J. Stat.*, Vol. 19(1), 151 – 160.

147. Shahbaz, M. Q., Hanif, M. and Ahmad, M. (2001) "A new selection procedure for unequal probability sampling without replacement and a sample of size 2", *Proc. ICCS – VII*, Vol. 12, 291 – 296.

148. Shahbaz, M. Q., Hanif, M. and Ahmad, M. (2001) "Some estimators for unequal probability sampling", *Proc. ICCS – VII*, Vol. 12, 297 – 302.

149. Singh, P. (1978a) "A sampling scheme with inclusion probability proportional to size", *Sankhya*, C., 40, 122 – 128.

150. Singh, P. (1978b) "The selection of sample of two units with inclusion probability proportional to size", *Biometrika*, 65(2), 450 – 454.

151. Sihna, B. K. (1973) "On sampling scheme to realize pre-assigned sets of inclusion probabilities of first two orders", *Culc. Stat. Assoc. Bull.* 22, 89 – 110.

152. Som, R. K. (1973) "*A Manual of Sampling Techniques*", Heinemann Educational Book Ltd. U. K.

153. Smith, H. F. (1938) "An experimental law describing heterogeneity in the yields of agricultural crops", *J. Agri. Science*, 28, 1 – 23.

154. Smith, T. M. F. (1981) "Regression analysis for complex surveys. *In current topics in survey sampling*", 267 – 292. Academic Press, NY.

155. Srivastava, A. K. and Singh, D. (1981) "A sampling procedure with inclusion probability proportional to size", *Biometrika*, 68(3), 732 – 734.

156. Stevens, W. L. (1958) "Sampling without replacement with probability proportional to size", *J. Roy. Stat. Soc.* B, 20, 393 – 397.

157. Stuart, A. (1964) "Multistage sampling with preliminary random stratification of first stage units", *Rev. Inter. Stat. Inst.* 32, 193 – 220.

158. Sunter, A. B. (1977) "List sequential sampling with equal or unequal probabilities without replacement", *Inter. Stat. Rev.* 54(1), 33 – 50.