PROBABILITY PROPORTIONAL TO SIZE SAMPLING

By

ALI HUSSEIN BAGOUR

Bachelor of Science in Statistics
King Abdul Aziz University
Jeddah, Saudi Arabia
1990

Master of Science in Experimental Statistics
New Mexico State University
Las Cruces, New Mexico
1997

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
July, 2004

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to Dr. William D. Warde for offering his support, expertise, and help throughout my Doctor of Philosophy program. Many thanks must also be given to my committee members, Dr. Larry Claypool, Dr. Mark E. Payton, and Dr. Nohpill Park, for reviewing my work and providing many helpful comments.

Special thanks go to all the faculty members of the Department of Statistics for helping me to accomplish my educational goals.

Finally, I would like to finish by thanking my family and friends for their emotional support.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

Table Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

1.1 Introduction

It should be clear that, whether we use a sample for a descriptive purpose or for an analytical purpose, the sample should be selected by using proper statistical methods in order to obtain results with the desired precision. One key issue in the selection of a sample is to allocate a certain non-zero probability to each unit of the population. This method of obtaining a sample, generally referred to as probability sampling, is usually classified as either equal probability sampling or unequal probability sampling. We are interested in unequal probability sampling.

1.2 Unequal Probability Sampling

As with the usual equal probability sampling designs, we can select a sample in unequal probability sampling by using either sampling with replacement (wr) or sampling without replacement (wor). In the former case, the sampling scheme is called unequal probability sampling with replacement (upswr) or probability proportional to size sampling with replacement (ppswr) and in the latter case the scheme is called unequal probability sampling without replacement (upswor) or probability proportional to size sampling without replacement (πpswor). Taking selection probabilities, in the case of ppswr, or inclusion probabilities, in the case of πpswor, proportional to some well-chosen auxiliary variable is one way of controlling the estimation error through the sampling design. A well-chosen auxiliary variable is roughly proportional to the variable of interest. Probability proportional to size sampling has the advantage that the most important units have a large chance of being selected by the selection procedure. Horvitz and Thompson (1952) used the data that are taken from a survey conducted by the Statistical Laboratory of Iowa State College to show that substantial reduction in the variance can be obtained by using unequal probability sampling. The data are given in Table 1.1. The total number of households for 20 blocks in Ames, Iowa needs to be estimated. The variable of interest in this example is the number of households, Y_i. The well-chosen auxiliary variable X_i in this example is the estimated number of households which was obtained by a team of observers who drove through the 20 blocks and made estimates of the number of households on each block.

1.2.1 Unequal Probability Sampling With Replacement (PPSWR Sampling)

The idea of probability proportional to size sampling was first given by Neyman(1934). Hansen and Hurwitz (1943) developed the general theory of probability proportional to size with replacement. One unit was selected at each of the n draws. They allocated the selection probability to the *ith* unit of the population given by $p_i = \frac{Z_i}{Z}$, where Z_i is the measure of size (auxiliary variable) for the *ith* population unit and $Z = \sum_{i=1}^{N} Z_i$.

An unbiased estimate for population total Y in unequal probability sampling designs with replacement was

$$y'_{HH} = \frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{p_i} , \tag{1.1}$$

where n was the size of the sample and p_i was the probability of the selection. Since it was based on sampling with replacement, Hartley and Rao (1962) called it multinomial sampling. There are different forms for the variance of this estimator. Cochran (1953) proved that

$$Var(y'_{HH}) = \frac{1}{n}\left(\sum_{i=1}^{N}\frac{Y_i^2}{p_i} - Y^2\right) , \tag{1.2}$$

where $Y = \sum_{i=1}^{N} Y_i$ and N was the size of the population.

Mukhopadhyay (2000) proved that

$$Var(y'_{HH}) = \frac{1}{2n}\sum_{\substack{i=1 \\ j\neq i}}^{N}\sum_{j=1}^{N} p_i p_j \left(\frac{Y_i}{p_i} - \frac{Y_j}{p_j}\right)^2 . \tag{1.3}$$

Beg and Hanif (1991) gave another form of variance, which is easily compared with the ratio estimator:

$$Var(y'_{HH}) = \frac{1}{n}\sum_{i=1}^{N}\frac{1}{p_i}(Y_i - p_i Y)^2 . \tag{1.4}$$

An unbiased variance estimator of $Var(y'_{HH})$ is given in Cochran (1953):

$$V\hat{a}r(y'_{HH}) = \frac{1}{n(n-1)}\sum_{i=1}^{n}\left(\frac{y_i}{p_i} - y'_{HH}\right)^2 . \tag{1.5}$$

Brewer and Hanif (1983) proved that

$$V\hat{a}r(y'_{HH}) = \frac{1}{2n^2(n-1)}\sum_{\substack{i=1 \\ j\neq i}}^{n}\sum_{j=1}^{n}\left(\frac{y_i}{p_i} - \frac{y_j}{p_j}\right)^2 . \tag{1.6}$$

Sampling statisticians use sampling with replacement frequently because:

1. Selection is simple
2. An unbiased variance estimator exists
3. It can be used for multistage sampling
4. Any sampling size can be used.

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

Horvitz and Thompson (1952) developed a general theory of sampling without replacement. Narain (1951) and Madow (1949) presented the idea of probability proportional to size, but without the mathematical background. Horvitz and Thompson (1952) gave their estimator for population total Y, which was

$$y'_{HT} = \sum_{i=1}^{n} \frac{y_i}{\pi_i}, \qquad (1.7)$$

where π_i is the probability of inclusion in the sample of the ith population unit. They developed their variance

$$Var_{HT}(y'_{HT}) = \sum_{i=1}^{N} \frac{1-\pi_i}{\pi_i} Y_i^2 + \sum_{\substack{i=1 \\ j \neq i}}^{N} \sum_{j=1}^{N} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j} Y_i Y_j, \qquad (1.8)$$

where π_{ij} is the probability that both unit i and unit j are included in the sample. This is applicable when the sample size is a random variable (Hanurav (1967)). This variance form will be denoted by HT form. Yates and Grundy (1953) and Sen (1953) independently developed the variance expression

$$Var_{SYG}(y'_{HT}) = \frac{1}{2}\sum_{\substack{i=1 \\ j\neq i}}^{N}\sum_{j=1}^{N}\left(\pi_i\pi_j - \pi_{ij}\right)\left(\frac{Y_i}{\pi_i} - \frac{Y_j}{\pi_j}\right)^2 . \tag{1.9}$$

This is applicable for fixed sample size (Brewer and Hanif (1983)). This variance form will be denoted by SYG form. An unbiased variance estimator for (1.8) was given by Horvitz and Thompson (1952)

$$\hat{Var}_{HT}(y'_{HT}) = \sum_{i=1}^{n}\frac{1-\pi_i}{\pi_i^2}y_i^2 + \sum_{\substack{i=1 \\ j\neq i}}^{n}\sum_{j=1}^{n}\frac{\pi_{ij} - \pi_i\pi_j}{\pi_i\pi_j\pi_{ij}}y_i y_j . \tag{1.10}$$

Concurrently Sen (1953) and Yates and Grundy (1953) also gave an unbiased variance estimator for (1.9). Their expression was

$$\hat{Var}_{SYG}(y'_{HT}) = \frac{1}{2}\sum_{\substack{i=1 \\ j\neq i}}^{n}\sum_{j=1}^{n}\left(\frac{\pi_i\pi_j - \pi_{ij}}{\pi_{ij}}\right)\left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j}\right)^2 . \tag{1.11}$$

Equation (1.10) can yield a negative variance, but (1.11) rarely seems to do so in practice (Raj (1956), Rao and Singh (1973), Lanke (1974), and Rao (1963)).

A number of selection procedures were developed on the basis of the estimator given by Horvitz and Thompson (1952). A list of these procedures can be found in Brewer and Hanif (1983), and Chaudhuri and Vos (1986), but most of these selection procedures are limited to n = 2 due to the difficulty of finding a compact form for π_{ij} as the sample size increases (Yates and Grundy (1953)). A sample of size two is the most important case in practice. For example, when stratified sampling is considered with a very large number of strata, it makes sense to select a sample of size two units from each stratum (Durbin (1967)). The two variance forms of the Horvitz and Thompson estimator given in (1.8) and (1.9) require the calculation of joint inclusion probabilities (π_{ij}) and therefore they are very difficult to apply because the calculation of π_{ij} becomes cumbersome as the sample size increases. Sarndal (1996) discussed important problems that are associated with π_{ij}'s. In summary, if the selection procedure requires that π_i is proportional to the measure of size and the sample size is fixed, it becomes tedious and often computationally difficult to calculate π_{ij}'s and check whether the conditions $\pi_{ij} < \pi_i \pi_j$ are satisfied in order to ensure that the estimated variance is positive. In addition, they are used with a cumbersome double summation in the variance of the Horvitz and Thompson estimator (1952). Also, Brewer and Donadio (2003) stated that if one or more of the $\frac{N(N-1)}{2}$ distinct values of π_{ij} are zero, then the estimated variance of the Horvitz and Thompson estimator (1952) is biased and if π_{ij} is very small compared to $\pi_i \pi_j$, then the estimated variance will be unstable. Attempts have been made to approximate the variance of the Horvitz and Thompson estimator such that it does not involve the joint inclusion probabilities, π_{ij}. A simple approximation to π_{ij} in terms of π_i and π_j for selection procedures that ensures that π_i is proportional to the known measure of size z_i is given by Brewer (1963a), Durbin (1967), Rao (1965) and Sampford (1967) as

$$\pi_{ij} = \frac{\pi_i \pi_j}{2 + \sum_{k=1}^{N} \frac{\pi_k}{1-\pi_k}} \left[\frac{1}{1-\pi_i} + \frac{1}{1-\pi_j} \right]. \tag{1.12}$$

Hajek (1981) proposed the following approximation for π_{ij}

$$\pi_{ij} \approx \pi_i \pi_j \left[1 - (1-\pi_i)(1-\pi_j) d^{-1} \right], \tag{1.13}$$

where $d = \sum_{i=1}^{N} \pi_i (1-\pi_i)$. Hajek (1964, 1981) showed the validity of this approximation when rejective sampling such as Durbin's rejective procedure (1953) is carried out in each stratum. Brewer and Hanif (1983) gave two approximations to π_{ij} in terms of π_i and π_j. The first approximation given by Brewer and Hanif (1983) is

$$\pi_{ij} = A\pi_i \pi_j + B\left(\pi_i + \pi_j\right) + C\left(\pi_i^2 + \pi_j^2\right) , \tag{1.14}$$

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

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

$$\pi_{ij} = \frac{(n-1)\sum_{r=0}^{\infty} \pi_i^{2^r} \pi_j^{2^r}}{\prod_{i=0}^{r} \sum_{K=1}^{N} \pi_k^{2^i}} . \tag{1.15}$$

Herzel (1986) suggested another approximation for π_{ij}. This approximation is given as

$$\pi_{ij} = \pi_i \pi_j - \frac{\pi_i (1-\pi_i) + \pi_j (1-\pi_j)}{N-2} + \frac{n - \sum_{k=1}^{N} \pi_k^2}{(N-1)(N-2)} . \tag{1.16}$$

Brewer (2002) proposed another approximation for π_{ij}. This approximation is given as

$$\pi_{ij} = \left(\frac{a_i + a_j}{2}\right) \pi_i \pi_j , \tag{1.17}$$

where a_i and a_j are appropriately chosen. Brewer (2002) showed that using

$a_i = a_j = \dfrac{(n-1)}{n-\pi_i}$ in (1.17) gives an approximate formula for the variance of the Horvitz

and Thompson estimator as

$$Var_{SYG}\left(y_{HT}^{\prime}\right) \approx \sum_{i=1}^{N} \pi_i \left(1-\frac{n-1}{n-\pi_i}\pi_i\right)\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2. \tag{1.18}$$

Hartley and Rao (1962), used a random systematic procedure to obtain an approximate expression for the variance of the Horvitz and Thompson estimator, which is given as

$$Var_{SYG}\left(y_{HT}^{\prime}\right) \approx \sum_{i=1}^{N} \pi_i \left(1-\frac{n-1}{n}\pi_i\right)\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2. \tag{1.19}$$

When the balanced sampling design is maximum entropy or is close to maximal entropy, it can be approximated by a conditional Poisson sampling. Thus a general form of variance approximation can be written as

$$Var_{approx}\left(y_{HT}^{\prime}\right) \approx \sum_{i=1}^{N} \frac{b_i}{\pi_i^2}(Y_i - Y_i^*)^2, \tag{1.20}$$

where $Y_i^* = \pi_i \dfrac{\sum_{k=1}^{N} b_k Y_k / \pi_k}{\sum_{k=1}^{N} b_k}$,

and the coefficients b_i take several forms (Deville and Tille, 2002) (see also Matei and Tille, 2003). Hajek (1981) proposed the most common value for b_i:

$$b_i = \frac{\pi_i(1-\pi_i)N}{N-1}. \tag{1.21}$$

Berger (2003) extended the Hajek (1964, 1981) variance approximation to accommodate the systematic sampling design. Deville and Tille (2002) proposed another approximation for b_i in (1.20) by solving the following equation system by the fixed-point technique (see Matei and Tille, 2003):

$$b_i - \frac{b_i^2}{\sum_{k=1}^{N} b_k} = \pi_i(1-\pi_i). \tag{1.22}$$

Starting from Hajek (1964), Brewer (2002) proposed the following approximation variance (see Matei and Tille, 2003):

$$Var(y_{HT}^{/}) \approx \sum_{i=1}^{N} \pi_i(1-\pi_i)(\frac{Y_i}{\pi_i} - \frac{\tilde{Y}}{n})^2, \tag{1.23}$$

where $\tilde{Y} = \sum_{i=1}^{N} a_i Y_i$,

and $a_i = \frac{n(1-\pi_i)}{\sum_{k=1}^{N} \pi_k(1-\pi_k)}$.

Also, Brewer and Donadio (2003) proposed another approximation for π_{ij}. This approximation is the same as the one given by Brewer (2002):

$$\pi_{ij} = \left(\frac{c_i + c_j}{2}\right)\pi_i \pi_j,$$

which was given in equation (1.17). They proposed the approximate variance

$$Var_{SYG}(y_{HT}^{/}) \approx \sum_{i=1}^{N} \pi_i \left(1 - c_i \pi_i\right)\left(\frac{Y_i}{\pi_i} - \frac{Y}{n}\right)^2, \tag{1.24}$$

$$\text{where } c_i = \frac{1 - \frac{1}{n} - \frac{1}{n^2}\sum_{k=1}^{N} c_k \pi_k^2}{1 - \frac{2}{n}\pi_i}, \tag{1.25}$$

or $c_i = c = \dfrac{n-1}{n - \frac{1}{n}\sum_{k=1}^{N} \pi_k^2}$. (1.26)

But all these approximations have limitations in one way or another.

1.2.3 Description of Selected Selection Procedures

The Horvitz and Thompson estimator (1952) requires evaluation of the probability of inclusion, π_i and the variance and the variance estimators of the Horvitz and Thompson estimator (1952) require evaluation of joint inclusion probabilities, π_{ij}. A selection procedure is needed to evaluate π_i and π_{ij}. A number of selection procedures available in the literature can be used with the Horvitz and 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 probabilities and joint inclusion probabilities but they are applicable to a sample of size 2 only. Brewer and Hanif (1983) described fifty selection procedures along with their classification on the basis of the sample selection method. The following selection procedures have been widely used in real life surveys, as they produce compact formulae for evaluation of π_i and π_{ij}.

1.2.3.1 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 the 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:

$$\pi_i = q_i + \frac{n-1}{N-1}(1-q_i) \quad with \ \sum_{i=1}^{N} q_i = 1 \tag{1.27}$$

$$\pi_{ij} = \frac{n-1}{N-1}\left[\frac{N-n}{N-2}(q_i + q_j) + \frac{n-2}{N-2}\right] \tag{1.28}$$

where q_i are revised selection probabilities.

1.2.3.2 Yates-Grundy Draw-by-Draw Procedure

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

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

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

$$\pi_i = p_i\left[1 + \sum_{j=1}^{N} \frac{p_j}{1-p_j} - \frac{p_i}{1-p_i}\right] \tag{1.29}$$

$$\pi_{ij} = p_i p_j \left[\frac{1}{1-p_i} + \frac{1}{1-p_j} \right]. \tag{1.30}$$

This procedure is one of the simplest procedures as it does not impose any restriction on initial probabilities of selection and final probabilities of inclusion.

1.2.3.3 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 the 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:

$$\pi_i = 2p_i \tag{1.31}$$

$$\pi_{ij} = \frac{2p_i p_j}{k} \left[\frac{1}{1-2p_i} + \frac{1}{1-2p_j} \right] \tag{1.32}$$

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

1.2.3.4 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 is the procedure of Brewer, and is therefore strictly a without replacement procedure. This procedure is stated as:

- Select the first unit with probability proportional to size.
- Select the 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 are the same as those for the Brewer (1963a) selection procedure.

1.2.3.5 Rao-Sampford Selection Procedure

This selection procedure is a rejective selection procedure developed independently by Rao (1965) 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 the previous two steps if the 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).

1.2.3.6 Prabhu-Ajgonkar Selection Procedure

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

- Select the first unit with probability proportional to size and with replacement.
- Select the second element with probability $\frac{p_i(1-p_i)}{A(1-2p_i)}$ where A is a normalizing constant.
- If the 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 the same as that of the Brewer (1963a) method. The joint probability of inclusion for this selection procedure is given as:

$$\pi_{ij} = \frac{p_i p_j}{A}\left[\frac{1}{1-2p_i} + \frac{1}{1-2p_j}\right], \tag{1.33}$$

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

1.2.3.7 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 the first step if the 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:

$$\pi_i = \frac{2p_i(1-p_i)}{1-\sum_{i=1}^{N} p_i^2} \tag{1.34}$$

$$\pi_{ij} = \frac{2p_i p_j}{1-\sum_{i=1}^{N} p_i^2}. \tag{1.35}$$

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

$$y_D = \frac{1}{2}\left[\frac{y_i}{p_i} + \frac{y_j}{p_j}\right]. \tag{1.36}$$

The estimator given in (1.36) is a biased estimator of the population total, but the bias is generally negligible as compared to the mean squared error.

1.2.3.8 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 size of 2 this procedure is given as:

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

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

$$\pi_i = \frac{2p_i(1-p_i)}{1-\sum_{j=1}^{N} p_j^2} \tag{1.37}$$

$$\pi_{ij} = \frac{2p_i p_j}{1-\sum_{j=1}^{N} p_j^2}. \tag{1.38}$$

1.2.4 Unequal Probability Sampling Without Replacement Using Special Estimators

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

1.2.4.1 Das's Estimator

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

$$t = \sum_{r=1}^{n} c_r t_r \ , \tag{1.39}$$

where c_r are such that $\sum_{r=1}^{n} c_r = 1$, but for simplicity Das chose $c_r = \frac{1}{n}$ and

$$t_r = \frac{(1-p_1)(1-p_1-p_2)......(1-p_1-p_2-...-p_{r-1})}{(N-1)......(N-r+1)p_1......p_r} y_r \ . \tag{1.40}$$

This estimator takes into account the order of selection. There is a flaw in Das's estimator in that its variance estimator may produce a negative value in some cases (Brewer and Hanif (1983)).

1.2.4.2 Raj's Estimator

Raj (1956) developed a series of estimators based on the order of selection given by Das (1951). His t_{mean} estimators improved the sampling error; moreover, any sample size could be used and no π_{ij} was needed. The estimator proposed by Raj (1956) has the general form:

$$t_{mean} = \frac{1}{n}\sum_{r=1}^{n} t_r , \tag{1.41}$$

where

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

For a sample size of 2, he defined his estimate to be

$$t_{mean} = \frac{1}{2}\left[\frac{y_1}{p_1}(1+p_1) + \frac{y_2}{p_2}(1-p_1)\right], \tag{1.42}$$

with variance

$$Var(t_{mean}) = \frac{1}{8}\sum_{\substack{i=1 \\ j \neq i}}^{N}\sum_{j=1}^{N} p_i p_j (2 - p_i - p_j)\left(\frac{Y_i}{p_i} - \frac{Y_j}{p_j}\right)^2 . \tag{1.43}$$

An unbiased estimator for that variance is

$$V\hat{a}r(t_{mean}) = \frac{(1-p_1)^2}{4}\left(\frac{y_1}{p_1} - \frac{y_2}{p_2}\right)^2 . \tag{1.44}$$

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

$$Var(t_{mean}) = \frac{1}{2n^2}\sum_{\substack{i=1 \\ j \neq i}}^{N}\sum_{j=1}^{N} p_i p_j \left[1 + \sum_{r=2}^{n} Q_{ij}(r-1)\right]\left(\frac{Y_i}{p_i} - \frac{Y_j}{p_j}\right)^2 , \tag{1.45}$$

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 (1956) for a sample of any size is given as

$$V\hat{a}r(t_{mean}) = \frac{1}{n(n-1)}\sum_{k=1}^{n}(t_k - \bar{t})^2 , \tag{1.46}$$

where $\bar{t} = \frac{1}{n}\sum_{k=1}^{n} t_k$. Both the variance estimators are non-negative for all p_i's.

1.2.4.3 Murthy's Estimator

Murthy (1957) suggested that the Raj (1956) estimator could be improved by the process of unordering. The unordered estimator of the population total is

$$t_{symm} = \frac{1}{P(s)}\sum_{i=1}^{n} P(s|i) y_i \ , \tag{1.47}$$

where $P(s|i)$ is the probability of obtaining a sample "s" given that the *ith* unit has been selected and $P(s)$ is the probability of obtaining a sample "s". The Murthy estimator for a sample size of 2 is given as

$$t_{symm} = \frac{1}{2-p_1-p_2}\left[\frac{y_1}{p_1}(1-p_2)+\frac{y_2}{p_2}(1-p_1)\right] . \tag{1.48}$$

The variance of (1.48) is

$$Var(t_{symm}) = \frac{1}{2}\sum_{i=1}^{N}\sum_{\substack{j=1 \\ j\neq i}} \frac{p_i p_j (1-p_i-p_j)}{2-p_i-p_j} \cdot \left(\frac{Y_i}{p_i}-\frac{Y_j}{p_j}\right)^2 . \tag{1.49}$$

An unbiased variance estimator for n = 2 is

$$\hat{Var}(t_{symm}) = \frac{(1-p_1)(1-p_2)}{4}\left(\frac{y_1}{p_1}-\frac{y_2}{p_2}\right)^2 . \tag{1.50}$$

Murthy showed that the estimator given in (1.48) always performs better than the estimator given in (1.42). Pathak (1967a) derived the variance expression for the Murthy (1957) estimator for a sample of size n. This expression is

$$Var\left(t_{symm}\right)=\frac{1}{2}\sum_{\substack{i=1\\ j\neq i}}^{N}\sum_{j=1} p_i\, p_j\left[1-\sum_{S\ni ij}\frac{P(s|i)P(s|j)}{P(s)}\right]\left(\frac{Y_i}{p_i}-\frac{Y_j}{p_j}\right)^2 . \tag{1.51}$$

An unbiased variance estimator of (1.51) given by Pathak (1967b) is

$$V\hat{a}r\left(t_{symm}\right)=\frac{1}{2}\sum_{\substack{i=1\\ j\neq i}}^{n}\sum_{j=1} p_i\, p_j\left[\frac{P(s)P(s|ij)-P(s|i)P(s|j)}{\{P(s)\}^2}\right]\left(\frac{y_i}{p_i}-\frac{y_j}{p_j}\right)^2 , \tag{1.52}$$

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

A Generalized Murthy estimator which has also been proposed, equation (1.47), has the form

$$y'_{GM}=\frac{1}{P(s)}\sum_{i\in s}P(s|i)\,y_i , \tag{1.53}$$

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

Samiuddin et al. (1992) have shown that the Generalized Murthy estimator y'_{GM} is both design and model unbiased. Further the estimator y'_{GM} achieves the Godambe-Joshi (1965) lower bound if and only if $\frac{P(s|i)}{P(s)}=\frac{1}{\pi_i}$. Bayless and Rao (1970) and Rao and Bayless (1969) did empirical studies for sample size 2, 3, and 4, taking a number of populations and concluded that the performance of Murthy's estimator was reasonably good, due to the Rao-Blackwell Theorem.

1.2.4.4 Rao-Hartley-Cochran Estimator

Rao et al. (1962) proposed a sampling strategy for use with unequal probability sampling and the estimator of population total. The population units are divided randomly into n groups, where the group sizes are predetermined. Then one unit is selected from each group. Their estimator is

$$y'_{RHC} = \sum_{i=1}^{n} \frac{\pi_i \, y_{iT}}{p_{iT}}, \tag{1.54}$$

where p_{iT} is the probability of the Tth unit being selected from the ith 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 (1.54) is

$$Var\left(y'_{RHC}\right) = \frac{n\left(\sum_{i=1}^{n} N_i^2 - N\right)}{N(N-1)} \cdot \left[\sum_{i=1}^{n}\sum_{T=1}^{N_i} \frac{Y_{iT}^2}{n\, p_{iT}} - \frac{Y^2}{n}\right]. \tag{1.55}$$

Rao et al. (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 (1.55) can be written as

$$Var\left(y'_{RHC}\right) = \left[1 - \frac{n-1}{N-1} + \frac{k(n-k)}{N(N-1)}\right]\left[\sum_{i=1}^{n}\sum_{T=1}^{N_i} \frac{Y_{iT}^2}{n\, p_{iT}} - \frac{Y^2}{n}\right]. \tag{1.56}$$

Further, if N is an exact multiple of n, then k = 0 and (1.56) becomes

$$Var\left(y'_{RHC}\right) = \left[1 - \frac{n-1}{N-1}\right]\left[\sum_{i=1}^{n}\sum_{T=1}^{N_i} \frac{Y_{iT}^2}{n\, p_{iT}} - \frac{Y^2}{n}\right]. \tag{1.57}$$

An unbiased variance estimator of (1.55) is

$$\hat{Var}\left(y'_{RHC}\right)=\frac{\left(\sum_{i=1}^{n} N_i^2 - N\right)}{\left(N^2-\sum_{i=1}^{n} N_i^2\right)}\cdot\sum_{i=1}^{n}\pi_i\left(\frac{y_{it}}{p_{it}}-y'_{RHC}\right)^2. \tag{1.58}$$

The unbiased variance estimators for (1.56) and (1.57) are

$$\hat{Var}\left(y'_{RHC}\right)=\frac{N^2+k(n-k)-Nn}{N^2(n-1)-k(n-k)}\cdot\sum_{i=1}^{n}\pi_i\left(\frac{y_{it}}{p_{it}}-y'_{RHC}\right)^2 \tag{1.59}$$

and

$$\hat{Var}\left(y'_{RHC}\right)=\frac{1}{(n-1)}\left(1-\frac{n}{N}\right)\cdot\sum_{i=1}^{n}\pi_i\left(\frac{y_{it}}{p_{it}}-y'_{RHC}\right)^2. \tag{1.60}$$

This thesis has two objectives. The first objective is to find approximate expressions for each form of the variance of the Horvitz and Thompson estimator, form (1.8) and form (1.9), using four approximations for π_{ij} as linear combination of π_i and π_j. The second objective is to develop a general selection procedure which can be used on the basis of the estimator given by Horvitz and Thompson (1952).

In Chapter 2 of this thesis, the derivation of the four proposed approximation expressions for the variance of the Horvitz and Thompson estimator are given using the variance form of Sen (1953) and Yates and Gundy (1953). Also, the evaluation of the proposed variance approximations is given using an empirical study. In Chapter 3 the derivation of the four proposed approximation expressions for the variance of the Horvitz and Thompson estimator are given using the variance form of Horvitz and Thompson (1952). Also, the evaluation of the proposed variance approximations is given using an empirical study. The development and special cases of the proposed general selection procedure are given in Chapter 4. Also, the verification of the important properties regarding the probability of inclusion and joint probability of inclusion are given in Chapter 4. Finally, an empirical study to specify the proposed general selection procedure is given in Chapter 4. Conclusions and future work are given in Chapter 5.

Table 1.1

Data of the Statistical Laboratory of Iowa State College

Block No.	Y_i	X_i
1	19	18
2	9	9
3	17	14
4	14	12
5	21	24
6	22	25
7	27	23
8	35	24
9	20	17
10	15	14
11	18	18
12	37	40
13	12	12
14	47	30
15	27	27
16	25	26
17	25	21
18	13	9
19	19	19
20	12	12

CHAPTER 2

THE PROPOSED APPROXIMATE EXPRESSIONS FOR VARIANCE OF THE HORVITZ AND THOMPSON ESTIMATOR USING THE VARIANCE FORM OF SEN AND YATES-GRUNDY

2.1 Introduction

In this chapter some approximate expressions for the variance of the Horvitz and Thompson estimator have been derived using the variance form of Sen (1953) and Yates and Grundy (1953). These approximate expressions require only the marginal inclusion probabilities π_i's. The estimator of the population total proposed by Horvitz and Thompson (1952) is given as

$$y'_{HT} = \sum_{i=1}^{n} \frac{y_i}{\pi_i} . \tag{2.1}$$

The variance of the estimator given in (2.1) has the popular form given concurrently by Sen (1953) and Yates and Grundy (1953) as in (1.9). This variance form is valid only if the sample size is fixed (Brewer and Hanif (1983)).

The variance expression given in (1.9) requires 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 by some suitable relation with the marginal inclusion probabilities.

Some of the most famous references for these approximations were given in chapter one. In this chapter, four approximations for joint inclusion probabilities π_{ij}'s as linear combinations of the marginal inclusion probabilities π_i and π_j have been obtained and used to derive four approximation expressions for the variance of the Horvitz and Thompson estimator using the variance form of Sen (1953) and Yates and Grundy (1953).

2.2 An Alternative Expression for Variance of the Horvitz and Thompson Estimator Using the Variance Form of Sen and Yates-Grundy

Brewer and Donadio (2003) derived an alternative expression for the variance of the Horvitz and Thompson estimator using the variance form of Sen (1953) and Yates and Grundy (1953). This expression was used to obtain a suitable approximation for the variance using only the marginal inclusion probabilities. To derive this expression let us recall formula (1.9).

$$Var_{SYG}\left(y'_{HT}\right) = \frac{1}{2}\sum_{i=1}^{N}\sum_{\substack{j=1 \\ j\neq i}}\left(\pi_i \pi_j - \pi_{ij}\right)\left(\frac{Y_i}{\pi_i}-\frac{Y_j}{\pi_j}\right)^2 \tag{1.9}$$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$+\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}(\pi_{ij}-\pi_i\pi_j)\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)\left(\frac{Y_j}{\pi_j}-\frac{Y}{n}\right). \qquad (2.2)$$

In equation (2.2) only the last term contains the joint probability of inclusion, π_{ij}. An approximate result for the variance of the Horvitz and Thompson estimator may be obtained by manipulating only the last term appropriately. The alternative expression for the variance form of Sen (1953) and Yates and Grundy (1953) (equation 2.2) was used to derive the variance approximations for the variance of the Horvitz and Thompson estimator.

2.3 Approximations for π_{ij}

The first approximation that is suggested is somewhat similar to one given by Brewer (2002) and Brewer and Donadio (2003) (equation 1.17). To obtain the approximation, recall the third term of equation (2.2):

$$\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}\left(\pi_{ij}-\pi_i\pi_j\right)\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)\left(\frac{Y_j}{\pi_j}-\frac{Y}{n}\right). \tag{2.3}$$

Now substituting the first proposed approximation $\pi_{ij} = a_i\pi_i\pi_j + a_j\pi_i\pi_j$ in (2.3):

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

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

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

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

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

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

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

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

$$= -2 \sum_{i=1}^{N} a_i \pi_i^2 \left(\frac{Y_i}{\pi_i} - \frac{Y}{n} \right)^2 + \sum_{i=1}^{N} \pi_i^2 \left(\frac{Y_i}{\pi_i} - \frac{Y}{n} \right)^2 . \tag{2.4}$$

Substituting (2.4) in (2.2):

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

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

$$= \sum_{i=1}^{N} \pi_i \left(1 - 2 a_i \pi_i \right) \left(\frac{Y_i}{\pi_i} - \frac{Y}{n} \right)^2 . \tag{2.5}$$

The second approximation that is proposed is relatively simple compared to the first one. Substituting the second proposed approximation $\pi_{ij} = a_i \pi_j + a_j \pi_i$ in (2.3):

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

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

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

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

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

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

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

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

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

$$= -2\sum_{i=1}^{N} a_i\,\pi_i\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2 + \sum_{i=1}^{N}\pi_i^2\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2 . \tag{2.6}$$

Substituting (2.6) in equation (2.2) another approximation for the variance of the Horvitz and Thompson estimator is

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

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

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

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

$$= \sum_{i=1}^{N} \pi_i \left(1 - 2 a_i\right) \left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2. \qquad (2.7)$$

This approximation is simpler than (2.5).

The third approximation that is proposed was considered by Brewer and Donadio (2003) as a special case of their approximation $\pi_{ij} = \left(\frac{c_i + c_j}{2}\right) \pi_i \pi_j$ when $c_i = c_j = c$.

Substituting the third proposed approximation $\pi_{ij} = a\, \pi_i\, \pi_j$ in (2.3):

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

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

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

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

$$= -(a-1)\sum_{i=1}^{N} \pi_i^2 \left(\frac{Y_i}{\pi_i} - \frac{Y}{n}\right)^2 . \tag{2.8}$$

Substituting (2.8) in (2.2), the following approximate expression for the variance of the Horvitz and Thompson estimator is obtained:

$$Var_{SYG}\left(y_{HT}^{/}\right) \approx \sum_{i=1}^{N} \pi_i \left(\frac{Y_i}{\pi_i} - \frac{Y}{n}\right)^2 - \sum_{i=1}^{N} \pi_i^2 \left(\frac{Y_i}{\pi_i} - \frac{Y}{n}\right)^2 - \sum_{i=1}^{N} \pi_i^2 (a-1) \left(\frac{Y_i}{\pi_i} - \frac{Y}{n}\right)^2$$

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

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

$$= \sum_{i=1}^{N} \pi_i (1 - a\pi_i) \left(\frac{Y_i}{\pi_i} - \frac{Y}{n}\right)^2 . \tag{2.9}$$

This is a general case of the formula given by Hartley and Rao (1962) where they have $a = \dfrac{n-1}{n}$ and it is given as

$$Var_{SYG}\left(y_{HT}^{/}\right) \approx \sum_{i=1}^{N} \pi_i \left(1 - \frac{n-1}{n}\pi_i\right) \left(\frac{Y_i}{\pi_i} - \frac{Y}{n}\right)^2 . \tag{2.10}$$

This equation was given in (1.18).

A fourth approximation that is proposed is obtained as follows. Substituting the fourth proposed approximation $\pi_{ij} = a\pi_i + b\pi_j$ in (2.3):

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

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

$$= -\sum_{i=1}^{N}(a+b)\pi_i\left(\frac{Y_i}{\pi_i} - \frac{Y}{n}\right)^2 + \sum_{i=1}^{N}\pi_i^2\left(\frac{Y_i}{\pi_i} - \frac{Y}{n}\right)^2. \tag{2.11}$$

Substituting (2.11) in (2.2), the following approximation for the variance of the Horvitz and Thompson estimator is obtained:

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

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

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

$$= (1-a-b)\sum_{i=1}^{N}\pi_i\left(\frac{Y_i}{\pi_i} - \frac{Y}{n}\right)^2. \tag{2.12}$$

All four approximations require suitable values of a_i in (2.5) and (2.7), a in (2.9) and a and b in (2.12) for close approximations of the true variance of the Horvitz and Thompson estimator.

2.4 The Assisting of the Superpopulation Model

A superpopulation model is frequently used in the design of finite population to estimate the population total. The model described here is the same as that used by Smith (1938), in a survey of agricultural data. The model is described as:

$$Y_i = \beta\pi_i + \varepsilon_i \,, \tag{2.13}$$

where $E(\varepsilon_i) = E(\varepsilon_i, \varepsilon_j) = 0$ and $E(\varepsilon_i^2) = \sigma_i^2 = \sigma^2\pi_i^{2\gamma}$ and $\frac{1}{2} \le \gamma \le 1$.

Useful references for model (3.11) are Brewer (1963b), Brewer (1979), Brewer and Donadio (2003), Cassel et al. (1976), Chaudhuri and Vos (1986), Cochran (1953), Foreman and Brewer (1971), Godambe and Joshi (1965), Hanif and Brewer (1980), Hansen et al. (1983), Kalton (1983), Rao (1966), Rao (1977), Royall (1970), Royall and Herson (1973a, 1973b), Sarndal and Wright (1984), Sarndal et al. (1992), and many others.

Under the superpopulation model, it will be shown in the following section that the expected value of the approximate variance formulas (formulas 2.5, 2.7, 2.9, and 2.12) that were derived after getting the approximations for π_{ij} will be approximately equal to the Godambe and Joshi (1965) lower bound for the expected variance, i.e.,

$$E_M[Var(y_{HT}^{/})] = \sum_{i=1}^{N} \sigma_i^2 \left(\frac{1}{\pi_i} - 1 \right). \tag{2.14}$$

The suitable values of a_i in formulas (2.5) and (2.7) and the suitable values of a and $a+b$ in formulas (2.9) and (2.12) will be obtained using the fact that the expected variance of the Horvitz and Thompson estimator is equal to the lower bound of the expected variance for any design-unbiased estimator under the superpopulation model (Godambe and Joshi (1965)). This approach has been used to obtain the values of c_i and c in formula (1.24) by Brewer and Donadio (2003).

2.4.1 The First Proposed Approximate Variance Formula

Recall formula (2.5) and model (2.13):

$$Var_{SYG}(y_{HT}^{/}) \approx \sum_{i=1}^{N} \pi_i \left(1-2\,a_i\,\pi_i\right)\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2 \qquad (2.5)$$

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

$$= \sum_{i=1}^{N} \pi_i (1-2a_i\pi_i)\left[\varepsilon_i\left(\frac{1}{\pi_i}-\frac{1}{n}\right)-\frac{1}{n}\sum_{j\neq i=1}^{N}\varepsilon_j\right]^2.$$

$$E_M\left(Var_{SYG}(y_{HT}^{/})\right) \approx \sum_{i=1}^{N} \pi_i (1-2a_i\pi_i)\left[\sigma_i^2\left(\frac{1}{\pi_i}-\frac{1}{n}\right)^2+\frac{1}{n^2}\sum_{j\neq i=1}^{N}\sigma_j^2\right]$$

$$= \sum_{i=1}^{N} (1-2a_i\pi_i)\left[\sigma_i^2\left(\frac{1}{\pi_i}-\frac{2}{n}\right)+\frac{\pi_i}{n^2}\sum_{j=1}^{N}\sigma_j^2\right]$$

$$= \sum_{i=1}^{N} \sigma_i^2\left[\left(\frac{1}{\pi_i}-\frac{1}{n}\right)-2a_i\left(1-\frac{2\pi_i}{n}\right)-\frac{2}{n^2}\sum_{j=1}^{N}a_j\pi_j^2\right]$$

$$= \sum_{i=1}^{N} \sigma_i^2\left[\left(\frac{1}{\pi_i}-1\right)-2a_i\left(1-\frac{2\pi_i}{n}\right)-\frac{2}{n^2}\sum_{j=1}^{N}a_j\pi_j^2-\frac{1-n}{n}\right]$$

$$= \sum_{i=1}^{N} \sigma_i^2\left(\frac{1}{\pi_i}-1\right)-\left[2\sum_{i=1}^{N}\sigma_i^2 a_i\left(1-\frac{2\pi_i}{n}\right)+\frac{2}{n^2}\sum_{i=1}^{N}\sigma_i^2\sum_{j=1}^{N}a_j\pi_j^2+\frac{1-n}{n}\sum_{i=1}^{N}\sigma_i^2\right]. \qquad (2.15)$$

Formula (2.15) shows that the expected value of the approximate variance formula (formula 2.5) is approximately equal to Godambe and Joshi's (1965) lower bound for the expected variance, i.e., the expected value of the approximate variance of the Horvitz and Thompson estimator is equal to Godambe and Joshi's (1965) lower bound for the expected variance minus some quantity, q_1. Now we will find the suitable values of a_i

in the first approximated variance of the Horvitz and Thompson estimator (formula 2.5) by equating q_1 to zero in order to make the expected value of the approximate variance formulas (formula 2.5) equal to Godambe and Joshi's (1965) lower bound. Two options will be considered to deal with q_1.

Option-1:

$$q_1 = \left[2\sum_{i=1}^{N} \sigma_i^2 a_i \left(1 - \frac{2\pi_i}{n}\right) + \frac{2}{n^2} \sum_{i=1}^{N} \sigma_i^2 \sum_{j=1}^{N} a_j \pi_j^2 + \frac{1-n}{n} \sum_{i=1}^{N} \sigma_i^2 \right] = 0$$

$$\sum_{i=1}^{N} \sigma_i^2 \left[-2a_i \left(1 - \frac{2\pi_i}{n}\right) - \frac{2}{n^2} \sum_{j=1}^{N} a_j \pi_j^2 - \frac{1-n}{n} \right] = 0.$$

This leads to the following equation:

$$a_i = \frac{\frac{n-1}{n} - \frac{2}{n^2} \sum_{j=1}^{N} a_j \pi_j^2}{2\left(1 - \frac{2\pi_i}{n}\right)}. \tag{2.16}$$

Equation (2.16) can be solved for a_i iteratively.

Option-2:

$$q_1 = \left[2\sum_{i=1}^{N} \sigma_i^2 a_i \left(1 - \frac{2\pi_i}{n}\right) + \frac{2}{n^2} \sum_{i=1}^{N} \sigma_i^2 \sum_{j=1}^{N} a_j \pi_j^2 + \frac{1-n}{n} \sum_{i=1}^{N} \sigma_i^2 \right] = 0.$$

Now substituting $\sigma_i^2 = \sigma^2 \pi_i^{2\gamma}$

$$2\sigma^2 \sum_{i=1}^{N} a_i \pi_i^{2\gamma} \left(1 - \frac{2\pi_i}{n}\right) + \frac{2\sigma^2}{n^2} \sum_{i=1}^{N} \pi_i^{2\gamma} \sum_{j=1}^{N} a_j \pi_j^2 + \frac{\sigma^2(1-n)}{n} \sum_{i=1}^{N} \pi_i^{2\gamma} = 0.$$

Let $\gamma = \frac{1}{2}$.

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

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

$$= 2\left[\sum_{i=1}^{N} a_i\left(\pi_i - \frac{\pi_i^2}{n}\right)\right] + (1-n) = 0$$

Therefore we can choose a_i such that

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

This linear equation has an $N-1$ dimensions solution space. Requiring that $a_i = a$ for all i leads to the following formula:

$$a = \frac{\frac{n-1}{2}}{n - \frac{1}{n}\sum_{i=1}^{N}\pi_i^2} \ . \tag{2.17}$$

2.4.2 The Second Proposed Approximate Variance Formula

Recall formula (2.7) and model (2.13):

$$Var_{SYG}(y'_{HT}) \approx \sum_{i=1}^{N}\pi_i\left(1-2a_i\right)\left(\frac{Y_i}{\pi_i} - \frac{Y}{n}\right)^2 \tag{2.7}$$

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

$$= \sum_{i=1}^{N}\pi_i(1-2a_i)\left[\varepsilon_i\left(\frac{1}{\pi_i} - \frac{1}{n}\right) - \frac{1}{n}\sum_{j\neq i=1}^{N}\varepsilon_j\right]^2 .$$

$$E_M\left(Var_{SYG}(y'_{HT})\right) \approx \sum_{i=1}^{N}\pi_i(1-2a_i)\left[\sigma_i^2\left(\frac{1}{\pi_i} - \frac{1}{n}\right)^2 + \frac{1}{n^2}\sum_{j\neq i=1}^{N}\sigma_j^2\right]$$

$$=\sum_{i=1}^{N}(1-2a_i)\left[\sigma_i^2\left(\frac{1}{\pi_i}-\frac{2}{n}\right)+\frac{\pi_i}{n^2}\sum_{j=1}^{N}\sigma_j^2\right]$$

$$=\sum_{i=1}^{N}\sigma_i^2\left[\left(\frac{1}{\pi_i}-\frac{1}{n}\right)-2a_i\left(\frac{1}{\pi_i}-\frac{2}{n}\right)-\frac{2}{n^2}\sum_{j=1}^{N}a_j\pi_j\right]$$

$$=\sum_{i=1}^{N}\sigma_i^2\left[\left(\frac{1}{\pi_i}-1\right)-2a_i\left(\frac{1}{\pi_i}-\frac{2}{n}\right)-\frac{2}{n^2}\sum_{j=1}^{N}a_j\pi_j-\frac{1-n}{n}\right]$$

$$=\sum_{i=1}^{N}\sigma_i^2\left(\frac{1}{\pi_i}-1\right)-\left[2\sum_{i=1}^{N}\sigma_i^2a_i\left(\frac{1}{\pi_i}-\frac{2}{n}\right)+\frac{2}{n^2}\sum_{i=1}^{N}\sigma_i^2\sum_{j=1}^{N}a_j\pi_j+\frac{1-n}{n}\sum_{i=1}^{N}\sigma_i^2\right]. \tag{2.18}$$

Formula (2.18) shows that the expected value of the approximate variance formula (formula 2.7) is approximately equal to Godambe and Joshi's (1965) lower bound for the expected variance i.e., the expected value of the approximate variance of the Horvitz and Thompson estimator is equal to Godambe and Joshi's (1965) lower bound for the expected variance minus some quantity, q_2. Now we will find the suitable values of a_i in the second approximate variance of the Horvitz and Thompson estimator (formula 2.7) by equating q_2 to zero in order to make the expected value of the approximate variance formula (formula 2.7) equal to Godambe and Joshi's (1965) lower bound. Two options will be considered to deal with q_2.

Option-1:

$$q_2=\left[2\sum_{i=1}^{N}\sigma_i^2a_i\left(\frac{1}{\pi_i}-\frac{2}{n}\right)+\frac{2}{n^2}\sum_{i=1}^{N}\sigma_i^2\sum_{j=1}^{N}a_j\pi_j+\frac{1-n}{n}\sum_{i=1}^{N}\sigma_i^2\right]$$

$$=\sum_{i=1}^{N}\sigma_i^2\left[-2a_i\left(\frac{1}{\pi_i}-\frac{2}{n}\right)-\frac{2}{n^2}\sum_{j=1}^{N}a_j\pi_j-\frac{1-n}{n}\right]=0$$

This leads to the following equation:

$$a_i = \frac{\frac{n-1}{n} - \frac{2}{n^2}\sum_{j=1}^{N} a_j \pi_j}{2\left(\frac{1}{\pi_i} - \frac{2}{n}\right)} . \qquad (2.19)$$

Equation (2.19) can be solved for a_i iteratively.

Option-2:

$$q_2 = \left[2\sum_{i=1}^{N} \sigma_i^2 a_i \left(\frac{1}{\pi_i} - \frac{2}{n}\right) + \frac{2}{n^2}\sum_{i=1}^{N}\sigma_i^2 \sum_{j=1}^{N} a_j \pi_j + \frac{1-n}{n}\sum_{i=1}^{N}\sigma_i^2 \right] = 0$$

Now substituting $\sigma_i^2 = \sigma^2 \pi_i^{2\gamma}$,

$$2\sigma^2 \sum_{i=1}^{N} a_i \pi_i^{2\gamma}\left(\frac{1}{\pi_i} - \frac{2}{n}\right) + \frac{2\sigma^2}{n^2}\sum_{i=1}^{N}\pi_i^{2\gamma}\sum_{j=1}^{N} a_j \pi_j + \frac{\sigma^2(1-n)}{n}\sum_{i=1}^{N}\pi_i^{2\gamma} = 0$$

Let $\gamma = \frac{1}{2}$.

$$2\sigma^2\left(\sum_{i=1}^{N} a_i - \frac{2}{n}\sum_{i=1}^{N} a_i \pi_i\right) + \frac{2\sigma^2}{n}\sum_{j=1}^{N} a_j \pi_j + \sigma^2(1-n) = 0$$

$$2\left(\sum_{i=1}^{N} a_i - \frac{2}{n}\sum_{i=1}^{N} a_i \pi_i\right) + \frac{2}{n}\sum_{j=1}^{N} a_j \pi_j + (1-n) = 0$$

Therefore we can choose a_i such that

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

This linear equation has an $N-1$ dimensions solution space. Requiring that $a_i = a$ for all i leads to the following formula:

$$a = \frac{n-1}{2(N-1)} . \qquad (2.20)$$

2.4.3 The Third Proposed Approximate Variance Formula

Recall formula (2.9) and model (2.13):

$$Var_{SYG}(y'_{HT}) \approx \sum_{i=1}^{N} \pi_i \left(1 - a\pi_i\right)\left(\frac{Y_i}{\pi_i} - \frac{Y}{n}\right)^2 \qquad (2.9)$$

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

$$= \sum_{i=1}^{N} \pi_i (1 - a\pi_i)\left[\varepsilon_i\left(\frac{1}{\pi_i} - \frac{1}{n}\right) - \frac{1}{n}\sum_{j\neq i=1}^{N} \varepsilon_j\right]^2 .$$

$$E_M\left(Var_{SYG}(y'_{HT})\right) \approx \sum_{i=1}^{N} \pi_i (1 - a\pi_i)\left[\sigma_i^2\left(\frac{1}{\pi_i} - \frac{1}{n}\right)^2 + \frac{1}{n^2}\sum_{j\neq i=1}^{N} \sigma_j^2\right]$$

$$= \sum_{i=1}^{N} (1 - a\pi_i)\left[\sigma_i^2\left(\frac{1}{\pi_i} - \frac{2}{n}\right) + \frac{\pi_i}{n^2}\sum_{j=1}^{N} \sigma_j^2\right]$$

$$= \sum_{i=1}^{N} \sigma_i^2\left[\left(\frac{1}{\pi_i} - \frac{1}{n}\right) - a\left(1 - \frac{2\pi_i}{n}\right) - \frac{a}{n^2}\sum_{j=1}^{N} \pi_j^2\right]$$

$$= \sum_{i=1}^{N} \sigma_i^2\left[\left(\frac{1}{\pi_i} - 1\right) - a\left(1 - \frac{2\pi_i}{n}\right) - \frac{a}{n^2}\sum_{j=1}^{N} \pi_j^2 - \frac{1-n}{n}\right]$$

$$= \sum_{i=1}^{N} \sigma_i^2\left(\frac{1}{\pi_i} - 1\right) - \left[a\sum_{i=1}^{N} \sigma_i^2\left(1 - \frac{2\pi_i}{n}\right) + \frac{a}{n^2}\sum_{i=1}^{N} \sigma_i^2 \sum_{j=1}^{N} \pi_j^2 + \frac{1-n}{n}\sum_{i=1}^{N} \sigma_i^2\right] . \qquad (2.21)$$

Formula (2.21) shows that the expected value of the approximate variance formula (formula 2.9) is approximately equal to Godambe and Joshi's (1965) lower bound for the expected variance, i.e., the expected value of the approximate variance of the Horvitz and Thompson estimator is equal to Godambe and Joshi's (1965) lower bound for the expected variance minus some quantity, q_3. Now we will find the suitable values of a in the third approximate variance of the Horvitz and Thompson estimator (formula 2.9) by equating q_3 to zero in order to make the expected value of the approximate variance formula (formula 2.9) equal to Godambe and Joshi's (1965) lower bound.

$$q_3 = \left[a\sum_{i=1}^{N}\sigma_i^2\left(1-\frac{2\pi_i}{n}\right)+\frac{a}{n^2}\sum_{i=1}^{N}\sigma_i^2\sum_{j=1}^{N}\pi_j^2+\frac{1-n}{n}\sum_{i=1}^{N}\sigma_i^2\right]=0 .$$

Now substituting $\sigma_i^2=\sigma^2\pi_i^{2\gamma}$,

$$a\sigma^2\left[\sum_{i=1}^{N}\pi_i^{2\gamma}-\frac{2}{n}\sum_{i=1}^{N}\pi_i^{2\gamma}\pi_i+\frac{1}{n^2}\sum_{i=1}^{N}\pi_i^{2\gamma}\sum_{j=1}^{N}\pi_j^2\right]+\frac{\sigma^2(1-n)}{n}\sum_{i=1}^{N}\pi_i^{2\gamma}=0 .$$

Let $\gamma=\frac{1}{2}$.

$$a=\frac{n-1}{n-\frac{1}{n}\sum_{i=1}^{N}\pi_i^2} . \qquad (2.22)$$

This equation is the same as equation (1.26) that was obtained by Brewer and Donadio (2003). Therefore the third approximate variance of the Horvitz and Thompson estimator (formula 2.9) is as follows:

$$Var_{SYG}(y_{HT}^{/})\approx\sum_{i=1}^{N}\pi_i\left(1-\frac{n-1}{n-\frac{1}{n}\sum_{i=1}^{N}\pi_i^2}\pi_i\right)\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2 . \qquad (2.23)$$

2.4.4 The Fourth Proposed Approximate Variance Formula

Recall formula (2.12) and model (2.13):

$$Var_{SYG}(y_{HT}^{/})\approx\left(1-a-b\right)\sum_{i=1}^{N}\pi_i\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2 \qquad (2.12)$$

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

$$= (1-a-b)\sum_{i=1}^{N}\pi_i\left[\varepsilon_i\left(\frac{1}{\pi_i}-\frac{1}{n}\right)-\frac{1}{n}\sum_{j\neq i=1}^{N}\varepsilon_j\right]^2.$$

$$E_M\left(Var(y'_{HT})\right) \approx (1-a-b)\sum_{i=1}^{N}\pi_i\left[\sigma_i^2\left(\frac{1}{\pi_i}-\frac{1}{n}\right)^2+\frac{1}{n^2}\sum_{j\neq i=1}^{N}\sigma_j^2\right]$$

$$= (1-a-b)\sum_{i=1}^{N}\left[\sigma_i^2\left(\frac{1}{\pi_i}-\frac{2}{n}\right)+\frac{\pi_i}{n^2}\sum_{j=1}^{N}\sigma_j^2\right]$$

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

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

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

$$= \sum_{i=1}^{N}\sigma_i^2\left(\frac{1}{\pi_i}-1\right)-\left[a\sum_{i=1}^{N}\sigma_i^2\left(\frac{1}{\pi_i}-\frac{1}{n}\right)+b\sum_{i=1}^{N}\sigma_i^2\left(\frac{1}{\pi_i}-\frac{1}{n}\right)+\frac{1-n}{n}\sum_{i=1}^{N}\sigma_i^2\right]. \qquad (2.24)$$

Formula (2.24) shows that the expected value of the approximate variance formula (formula 2.12) is approximately equal to Godambe and Joshi's (1965) lower bound for the expected variance, i.e., the expected value of the approximate variance of the Horvitz and Thompson estimator is equal to Godambe and Joshi's (1965) lower bound for the expected variance minus some quantity, q_4. Now we will find the suitable values of $(a+b)$ in the fourth approximate variance of the Horvitz and Thompson estimator (formula 2.12) by equating q_4 to zero in order to make the expected value of the approximate variance formula (formula 2.12) equal to Godambe and Joshi's (1965) lower bound.

$$q_4 = \left[a\sum_{i=1}^{N}\sigma_i^2\left(\frac{1}{\pi_i}-\frac{1}{n}\right)+b\sum_{i=1}^{N}\sigma_i^2\left(\frac{1}{\pi_i}-\frac{1}{n}\right)+\frac{1-n}{n}\sum_{i=1}^{N}\sigma_i^2 \right]=0.$$

Now substituting $\sigma_i^2 = \sigma^2\pi_i^{2\gamma}$,

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

$$=(a+b)\sum_{i=1}^{N}\pi_i^{2\gamma}\left(\frac{1}{\pi_i}-\frac{1}{n}\right)+\frac{1-n}{n}\sum_{i=1}^{N}\pi_i^{2\gamma}=0.$$

Let $\gamma=\frac{1}{2}$.

$$(a+b)=\frac{n-1}{N-1}. \tag{2.25}$$

Therefore the fourth approximate variance of the Horvitz and Thompson estimator (formula 2.12) is given as follow

$$Var_{SYG}(y_{HT}^{/}) \approx \frac{N-n}{N-1}\sum_{i=1}^{N}\pi_i\left(\frac{Y_i}{\pi_i}-\frac{Y}{n}\right)^2. \tag{2.26}$$

2.5 An Empirical Study to Evaluate the Proposed Variance Approximations

2.5.1 The Setup

To evaluate the four proposed variance approximations, an empirical study was carried out. To carry out this study a total of 36 populations were used to calculate the sampling variance of the Horvitz and Thompson (1952) estimator with both the exact variance and the approximated variances using Brewer's (1963a) selection procedure.

Note that formula 2.16 and formula 2.19 (option-1) were used to calculate a_i in the first two variance approximations, where formula 2.17 and formula 2.20 (option-2) were used as starting values for formula 2.16 and formula 2.19, respectively. The relative error percentage, L, was used as a criterion in the comparison of the performance of the four approximated variances.

$$L = \frac{V_a - V_e}{V_e} \text{ X } 100,$$

where V_a and V_e represent the approximated variance and the exact variance, respectively. A good approximation should have L very close to 0%. In order to test the approximated variances under different situations, 36 populations were used in this empirical study. The main features of each population are given as

1- The data generated according to model 1 as

$$Y = \beta_1 X + \varepsilon,$$

where

$X \sim GAMMA\ (\theta = 10, K)$,

$$f(X;\theta,K) = \frac{1}{\theta^k \Gamma(K)} X^K e^{-\frac{X}{\theta}},\ X>0.$$

$\varepsilon \sim$ NORMAL ($\mu = 20$, $\sigma^2 = 10$), and population size (N) is 15,

$$f(\varepsilon;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma}} e^{-[(\varepsilon-\mu)/\sigma]^2/2};\ -\infty<\varepsilon<\infty.$$

K	C.V-X	C.V-Y	ρ	β_1	Population No.
1.8	0.65	0.16	0.52	0.1	1
45	0.17	0.13	0.95	0.12	2

Where (C.V) is the population coefficients of variation of X and Y and (ρ) is the population correlation coefficient between X and Y.

2- The data generated according to model 2 as

$Y = \beta_1 X + \varepsilon,$

where

$X \sim EXPONENTIAL\ (\theta),$

$$f(X;\theta,K) = \frac{1}{\theta^k} e^{-\frac{X}{\theta}},\ X>0.$$

$\varepsilon \sim$ NORMAL ($\mu = 20$, $\sigma^2 = 10$), and population size (N) is 15,

$$f(\varepsilon;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-[(\varepsilon-\mu)/\sigma]^2/2};\ -\infty<\varepsilon<\infty.$$

θ	C.V-X	C.V-Y	ρ	β_1	Population No.
40	0.78	0.16	0.7	0.11	3

3- The data generated according to model 3 as

$Y = \beta_1 X + \varepsilon,$

where

$X \sim (CHI-SQUARE)\ \chi^2_{\upsilon},$

$$f(X;\nu) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} X^{\nu/2-1} e^{-\frac{X}{2}},\ X>0.$$

$\varepsilon \sim NORMAL$ ($\mu = 20$, $\sigma^2 = 10$), and population size (N) is 15,

$$f(\varepsilon;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-[(\varepsilon-\mu)/\sigma]^2/2};\ -\infty<\varepsilon<\infty.$$

υ	C.V-X	C.V-Y	ρ	β_1	Population No.
4	0.61	0.28	0.95	3.5	4
30	0.15	0.14	0.5	0.16	5

4- The data generated according to model 4 as

$Y = \beta_1 X + \varepsilon,$

where

$X \sim GAMMA\ (\theta = 10, K),$

$$\text{f}(\text{X};\theta,K) = \frac{1}{\theta^{k}\Gamma(K)} X^{K} e^{-\frac{X}{\theta}},\ \text{X>0}.$$

$\varepsilon \sim LOG$ NORMAL $(\mu_{\varepsilon} = e^{\mu+\frac{\sigma^2}{2}},\ \sigma_{\varepsilon}^2 = e^{2\mu+\sigma^2}\left(e^{\sigma^2} - 1\right)),\ \mu = 0$ and population size is 15,

$$\text{f}(\varepsilon;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma\varepsilon} e^{-[(\ln(\varepsilon)-\mu)/\sigma]^2/2};\ \varepsilon \geq 0$$

K	C.V-X	C.V-Y	σ^2	ρ	β_1	Population No.
1.8	0.65	0.95	1.6	0.5	0.16	6
1.8	0.65	0.84	2.5	0.65	0.56	7
1.8	0.65	0.70	5.55	0.87	7.43	8
45	0.168	0.22	4	0.71	0.3	9
45	0.168	0.17	6.5	0.95	2.88	10

5- The data generated according to model 5 as

$Y = \beta_1 X + \varepsilon,$

where

$X \sim EXPONENTIAL\ (\theta)$,

$$\mathrm{f(X;}\theta, K) = \frac{1}{\theta^{\mathrm{k}}} e^{-\frac{X}{\theta}},\ \mathrm{X>0.}$$

$\varepsilon \sim LOG$ NORMAL $(\mu_\varepsilon = e^{\mu + \frac{\sigma^2}{2}},\ \sigma_\varepsilon^2 = e^{2\mu+\sigma^2}\left(e^{\sigma^2} - 1\right))$, $\mu = 0$ and population size is 15,

$$\mathrm{f}(\varepsilon; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma\varepsilon} e^{-[(\ln(\varepsilon)-\mu)/\sigma]^2/2};\ \varepsilon \geq 0\,.$$

θ	C.V-X	C.V-Y	σ^2	ρ	β_1	Population No.
0.5	0.78	0.17	0.035	0.51	0.4	11
0.5	0.78	0.15	0.02	0.70	0.4	12
0.5	0.78	0.20	0.0091	0.95	0.76	13
10	0.78	0.57	0.39	0.95	0.4	14

6- The data generated according to model 6 as

$Y = \beta_1 X + \varepsilon,$

where

$X \sim (CHI - SQUARE)\ \chi_\upsilon^2,$

$$\mathrm{f(X;}\nu) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} X^{\nu/2-1} e^{-\frac{X}{2}},\ \mathrm{X>0.}$$

$\varepsilon \sim LOG$ NORMAL $(\mu_\varepsilon = e^{\mu + \frac{\sigma^2}{2}},\ \sigma_\varepsilon^2 = e^{2\mu+\sigma^2}\left(e^{\sigma^2} - 1\right))$, $\mu = 0$ and population size is 15,

$$\mathrm{f}(\varepsilon; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma\varepsilon} e^{-[(\ln(\varepsilon)-\mu)/\sigma]^2/2};\ \varepsilon \geq 0\,.$$

υ	C.V-X	C.V-Y	σ^2	ρ	β_1	Population No.
4	0.61	0.19	0.0073	0.95	0.1	15
4	0.61	0.95	1.9	0.50	1.041	16
4	0.61	0.79	8.9	0.70	79.5	17
4	0.61	0.41	0.082	0.95	0.38	18
30	0.15	0.47	4.15	0.50	1.7	19

7- The good mixture of natural populations used by Brewer and Donadio (2003):

C.V-X	C.V-Y	ρ	Population size	Population No
0.14	0.15	0.65	10	20
0.47	0.67	0.72	13	21
0.43	0.39	0.93	14	22
0.69	0.67	0,94	15	23
0.48	0.51	0.51	20	24
0.50	0.63	0.59	20	25
0.44	0.40	0.87	20	26
0.71	0.61	0.80	17	27
0.73	0.75	0.91	18	28

8- The data generated according to model 4 as

$Y = \beta_1 X + \varepsilon,$

where

$X \sim GAMMA\ (\theta = 10, K),$

$$f(X;\theta,K)=\frac{1}{\theta^{k}\Gamma(K)}X^{K}e^{-\frac{X}{\theta}},\ X>0.$$

$\varepsilon \sim LOG$ NORMAL ($\mu_{\varepsilon}=e^{\mu+\frac{\sigma^2}{2}}$, $\sigma_{\varepsilon}^{2}=e^{2\mu+\sigma^2}\left(e^{\sigma^2}-1\right)$), $\mu=0$ and population size is 1000,

$$f(\varepsilon;\mu,\sigma)=\frac{1}{\sqrt{2\pi}\sigma\varepsilon}e^{-[(\ln(\varepsilon)-\mu)/\sigma]^2/2};\ \varepsilon\geq 0.$$

K	C.V-X	C.V-Y	σ^2	ρ	β_1	Population No.
2.4	0.65	0.92	2	0.50	0.183	29 similar to (6)
2.4	0.65	0.84	4.4	0.70	1.85	30 similar to (7)
2.4	0.65	0.71	4.6	0.87	3.9	31 similar to (8)
45	0.15	0.22	4.6	0.67	0.5	32 similar to (9)

9- The data generated according to model 5 as

$Y=\beta_1 X+\varepsilon,$

where

$X \sim EXPONENTIAL\ (\theta)$,

$$f(X;\theta,K)=\frac{1}{\theta^{k}}e^{-\frac{X}{\theta}},\ X>0.$$

$\varepsilon \sim LOG$ NORMAL ($\mu_{\varepsilon}=e^{\mu+\frac{\sigma^2}{2}}$, $\sigma_{\varepsilon}^{2}=e^{2\mu+\sigma^2}\left(e^{\sigma^2}-1\right)$), $\mu=0$ and population size is 1000,

$$f(\varepsilon;\mu,\sigma)=\frac{1}{\sqrt{2\pi}\sigma\varepsilon}e^{-[(\ln(\varepsilon)-\mu)/\sigma]^2/2};\ \varepsilon\geq 0.$$

θ	C.V-X	C.V-Y	σ^2	ρ	β_1	Population No.
10	1.0	0.72	0.37	0.95	0.23	33 similar to (14)

10- The data generated according to model 6 as

$Y = \beta_1 X + \varepsilon,$

where

$X \sim (CHI-SQUARE)\, \chi^2_\upsilon$,

$$f(X;\nu) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} X^{\nu/2-1} e^{-\frac{X}{2}},\ X>0.$$

$\varepsilon \sim LOG$ NORMAL ($\mu_\varepsilon = e^{\mu+\frac{\sigma^2}{2}}$, $\sigma^2_\varepsilon = e^{2\mu+\sigma^2}\left(e^{\sigma^2}-1\right)$), $\mu = 0$ and population size is 1000,

$$f(\varepsilon;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma\varepsilon} e^{-[(\ln(\varepsilon)-\mu)/\sigma]^2/2};\ \varepsilon \geq 0\,.$$

υ	C.V-X	C.V-Y	σ^2	ρ	β_1	Population N0.
5	0.63	0.86	7.9	0.70	64	34 similar to (17)
5	0.63	0.52	0.4	0.95	0.8	35 similar to (18)
90	0.15	0.30	6.9	0.50	6	36 similar to (19)

Note: An idea about the processes to select the 36 populations follows. The plan was to generate populations with different characteristics. Brewer and Donadio (2003) characterize their selected populations according to the population coefficients of variation (C.V) of X and Y and correlation coefficient (ρ) between X and Y. The same criteria were followed in the data generation. Three levels of coefficients of variation were considered for X and Y in the data generation as low (0.15), medium (0.50), and high (1.0). Also, three levels of correlation coefficient between X and Y were considered in the data generation as low (0.50), medium (0.70), and high (0.95). All combinations of the levels of coefficients of variation and correlation coefficients were considered in the first place with model 1, 2, and 3. The first problem with those models was that a high level of coefficient of variation for Y could not be reached and at the same time have positive values for Y. Therefore, models 4, 5, and 6 were considered in order to solve this problem. After generating 37 populations according to those characteristics, it was found that only 27 populations out of the 37 populations had the acceptable range of slope, β_1, according to the fields of application. Models 1, 2, 3, 4, 5, and 6 represent the relationship between X and Y in different scenarios where the probability proportional to size sampling is recommended to be used. The distributions of X and ε were selected to be representative to most of the fields of application. Finally, the good mixture of natural populations used by Brewer and Donadio (2003) were used in addition to those 27 populations to provide a good mixture of populations with different characteristics.

2.5.2 Results

Table A.1 shows the values of the exact variance and the four approximate variances for the 36 populations. Table 2.1 reports the value of the criterion (L) for the four variance approximations as L_1, L_2, L_3, and L_4, respectively. The results in Table 2.1 show that the four proposed variance approximations perform very well over all. The four approximate variances can be arranged according to the quality of their performance as the first approximation, the second approximation, the third approximation, and the fourth approximation, where the first approximation is the best. The first approximation over-approximates or under-approximates the variance by up to 0.1%. The second approximation over-approximates or under-approximates the variance by up to 0.25%. The third approximation over-approximates or under-approximates the variance by up to

0.47%. The fourth approximation over-approximates or under-approximates the variance by up to 6%.

Table 2.1

The Relative Error Percentage, L, for the Four Approximated Variances Using SYG Form With n = 2

Pop#	L_1	L_2	L_3	L_4	N
1	0.0023795 %	0.0884698 %	0.1333176 %	-4.135281 %	15
2	0.0039571 %	0.0451379 %	0.0821328 %	0.4457326 %	15
3	-0.000526 %	0.0344497 %	0.0459953 %	-5.841881 %	15
4	-0.000083 %	0.0682042 %	0.0997347 %	-3.959445 %	15
5	0.0095729 %	0.105128 %	0.1851903 %	0.5879113 %	15
6	-0.078507 %	-0.150817 %	-0.243723 %	1.0941482 %	15
7	-0.084322 %	-0.175206 %	-0.283544 %	1.2281084 %	15
8	-0.088713 %	-0.196617 %	-0.319234 %	1.2336463 %	15
9	-0.005276 %	-0.021323 %	-0.037732 %	-0.076353 %	15
10	-0.005401 %	-0.022234 %	-0.039389 %	-0.084881 %	15
11	-0.001328 %	0.0322348 %	0.0412319 %	-5.784571 %	15
12	-0.000753 %	0.0328576 %	0.0434404 %	-5.861717 %	15
13	-0.000431 %	0.0323466 %	0.0436175 %	-5.937172 %	15
14	-0.017477 %	-0.033413 %	-0.081326 %	-5.411188 %	15
15	0.0015519 %	0.0737778 %	0.109107 %	-4.075762 %	15
16	-0.066363 %	-0.1337 %	-0.214842 %	1.2064164 %	15
17	-0.072368 %	-0.163203 %	-0.263965 %	1.2461173 %	15
18	-0.003898 %	0.062947 %	0.0924765 %	-3.39847 %	15
19	-0.009566 %	-0.023313 %	-0.037948 %	0.2133991 %	15
20	-0.021851 %	-0.082846 %	-0.149033 %	-0.750803 %	10
21	-0.030591 %	-0.013502 %	-0.02928 %	-0.496758 %	13
22	-0.039742 %	-0.110726 %	-0.194049 %	-0.677311 %	14
23	-0.095018 %	-0.22785 %	-0.391384 %	-1.31527 %	15
24	-0.002354 %	0.0395862 %	0.0765673 %	0.5930117 %	20
25	-0.011705 %	-0.026552 %	-0.044489 %	0.1245145 %	20
26	0.0072578 %	0.1293195 %	0.2433228 %	2.1651862 %	20
27	-0.040421 %	-0.050667 %	-0.084201 %	0.3980141 %	17
28	-0.002251 %	0.2451464 %	0.4668244 %	4.137034 %	18
29	$-1.103 X10^{-7}$ %	$-9.318 X10^{-6}$ %	-0.000019 %	-0.025484 %	1000
30	$-1.211 X10^{-7}$ %	$-6.88 X10^{-6}$ %	-0.000014 %	-0.009763 %	1000
31	$-1.213 X10^{-7}$ %	$-6.647 X10^{-6}$ %	-0.000013 %	-0.009005 %	1000
32	$-7.959 X10^{-9}$ %	$-1.488 X10^{-6}$ %	$-2.97 X10^{-6}$ %	-0.000464 %	1000
33	$-4.368 X10^{-8}$ %	$8.4177 X10^{-6}$ %	0.0000167 %	-0.083379 %	1000
34	$-1.113 X10^{-7}$ %	$-2.448 X10^{-6}$ %	$-4.864 X10^{-6}$ %	-0.000362 %	1000
35	$-6.319 X10^{-8}$ %	$-1.532 X10^{-6}$ %	$-3.105 X10^{-6}$ %	-0.043782 %	1000
36	$-8.655 X10^{-9}$ %	$-1.833 X10^{-6}$ %	$-3.659 X10^{-6}$ %	-0.000678 %	1000

CHAPTER 3

PROPOSED APPROXIMATE EXPRESSIONS FOR VARIANCE OF THE HORVITZ AND THOMPSON ESTIMATOR USING THE HORVITZ AND THOMPSON VARIANCE FORM

3.1 Introduction

In this chapter some approximate expressions for the variance of the Horvitz and Thompson estimator have been derived using the variance form of Horvitz and Thompson (1952). Those approximate expressions require only marginal inclusion probabilities, π_i's. The estimator of the population total proposed by Horvitz and Thompson (1952) is given as

$$y'_{HT} = \sum_{i=1}^{n} \frac{y_i}{\pi_i} . \tag{3.1}$$

The variance of the estimator given in (3.1) has a popular form given by Horvitz and Thompson (1952) as in (1.8). This variance form is valid if the sample size is random (Hanurav (1967)).

The variance expression given in (1.8) requires the calculation of joint probabilities of inclusion, π_{ij}. These probabilities are very complicated to evaluate when the sample size increases from 2. Many efforts have been made to approximate these joint inclusion probabilities by some suitable relation with the marginal inclusion probabilities. Some of the most famous references for these approximations were given in chapter one. In this chapter, four approximations for joint inclusion probabilities, π_{ij}, as linear combination of the marginal inclusion probabilities π_i and π_j have been obtained and used to derive four approximation expressions for the variance of Horvitz and Thompson estimator using the variance form of the Horvitz and Thompson (1952).

3.2 An Alternative Expression for the Variance of the Horvitz and Thompson Estimator Using the Variance Form of Horvitz and Thompson

In this section an alternative expression for the variance of the Horvitz and Thompson estimator is derived using the variance form of Horvitz and Thompson (1952). This expression is very useful to obtain a suitable approximation for the variance using only the first order inclusion probabilities.

Recall

$$Var_{HT}(y'_{HT}) = \sum_{i=1}^{N} \frac{1-\pi_i}{\pi_i} Y_i^2 + \sum_{\substack{i=1 \\ j\neq i}}^{N} \sum_{j=1}^{N} \frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j} Y_i Y_j \ . \tag{1.8}$$

An approximate result for the variance of the Horvitz and Thompson estimator may be obtained by appropriately manipulating only the last term of formula (1.8). Let us recall the last term of formula (1.8).

$$\sum_{\substack{i=1 \\ j\neq i}}^{N} \sum_{j=1}^{N} \frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j} Y_i Y_j = \sum_{\substack{i=1 \\ j\neq i}}^{N} \sum_{j=1}^{N} \frac{\pi_{ij}}{\pi_i\pi_j} Y_i Y_j - \sum_{\substack{i=1 \\ j\neq i}}^{N} \sum_{j=1}^{N} Y_i Y_j$$

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

Therefore formula (1.8) becomes

$$Var_{HT}(y'_{HT}) = \sum_{i=1}^{N} \frac{1}{\pi_i} Y_i^2 - Y^2 + \sum_{\substack{i=1 \\ j \neq i}}^{N} \sum_{j=1}^{N} \frac{\pi_{ij}}{\pi_i \pi_j} Y_i Y_j \ . \tag{3.2}$$

In equation (3.2) only the last term contains the joint probability of inclusion, π_{ij}. An approximate result for the variance of the Horvitz and Thompson estimator may be obtained by appropriately manipulating only the last term.

3.3 Approximations for π_{ij}

To obtain the first approximation, recall the third term of formula (3.2) and substitute the first proposed approximation $\pi_{ij} = a_i \pi_i \pi_j + a_j \pi_i \pi_j$ in it:

$$\sum_{\substack{i=1 \\ j \neq i}}^{N} \sum_{j=1}^{N} \frac{\pi_{ij}}{\pi_i \pi_j} Y_i Y_j = \sum_{\substack{i=1 \\ j \neq i}}^{N} \sum_{j=1}^{N} a_i Y_i Y_j + \sum_{\substack{i=1 \\ j \neq i}}^{N} \sum_{j=1}^{N} a_j Y_i Y_j$$

$$= 2\left(Y \sum_{i=1}^{N} a_i Y_i - \sum_{i=1}^{N} a_i Y_i^2 \right) . \tag{3.3}$$

Substituting (3.3) in (3.2), the following approximate expression for the variance of the Horvitz and Thompson estimator is obtained:

$$Var_{HT}(y'_{HT}) \approx \sum_{i=1}^{N} \left(\frac{1}{\pi_i} - 2a_i \right) Y_i^2 + Y \left(2 \sum_{i=1}^{N} a_i Y_i - Y \right). \tag{3.4}$$

The second approximation is relatively simple compared to the first one. To obtain the second approximation, recall the third term of formula (3.2) and substitute the second proposed approximation $\pi_{ij} = a_i \pi_j + a_j \pi_i$ in it:

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

$$= 2\left(Y\sum_{i=1}^{N}\frac{a_i}{\pi_i}Y_i - \sum_{i=1}^{N}\frac{a_i}{\pi_i}Y_i^2 \right). \tag{3.5}$$

Substituting (3.3) in (3.2), another approximation for the variance of the Horvitz and Thompson estimator is obtained:

$$Var_{HT}(y_{HT}^{/}) \approx \sum_{i=1}^{N}\left(\frac{1}{\pi_i} - 2\frac{a_i}{\pi_i} \right)Y_i^2 + Y\left(2\sum_{i=1}^{N}\frac{a_i}{\pi_i}Y_i - Y \right). \tag{3.6}$$

For the third approximation, again recall the third term of formula (3.2) and substitute the third proposed approximation $\pi_{ij} = a\ \pi_i\ \pi_j$ in it:

$$\sum_{\substack{i=1\\ j\neq i}}^{N}\sum_{j=1}^{N}\frac{\pi_{ij}}{\pi_i\pi_j}Y_iY_j = a\sum_{\substack{i=1\\ j\neq i}}^{N}\sum_{j=1}^{N}Y_iY_j$$

$$= a\left(Y^2 - \sum_{i=1}^{N}Y_i^2 \right). \tag{3.7}$$

Substituting (3.7) in (3.2), another approximation for the variance of the Horvitz and Thompson estimator is obtained:

$$Var_{HT}(y_{HT}^{/}) \approx \sum_{i=1}^{N}\left(\frac{1}{\pi_i} - a \right)Y_i^2 + Y^2(a-1). \tag{3.8}$$

For the fourth approximation, again recall the third term of formula (3.2) and substitute the fourth proposed approximation $\pi_{ij} = a\pi_i + b\pi_j$ in it:

$$\sum_{\substack{i=1\\ j\neq i}}^{N}\sum_{j=1}^{N}\frac{\pi_{ij}}{\pi_i\pi_j}Y_iY_j = \sum_{\substack{i=1\\ j\neq i}}^{N}\sum_{j=1}^{N}\frac{a}{\pi_j}Y_iY_j + \sum_{\substack{i=1\\ j\neq i}}^{N}\sum_{j=1}^{N}\frac{b}{\pi_i}Y_iY_j$$

$$= (a+b)\left(Y\sum_{i=1}^{N}\frac{1}{\pi_i}Y_i - \sum_{i=1}^{N}\frac{1}{\pi_i}Y_i^2 \right). \tag{3.9}$$

Substituting (3.9) in (3.2), another approximation for the variance of the Horvitz and Thompson estimator is obtained:

$$Var_{HT}(y_{HT}^{/}) \approx \left(1-(a+b)\right)\sum_{i=1}^{N}\frac{1}{\pi_i}Y_i^2 + Y\left((a+b)\sum_{i=1}^{N}\frac{1}{\pi_i}Y_i - Y\right). \tag{3.10}$$

All four approximations require suitable values of a_i in (3.4) and (3.6), a in (3.8) and a and b in (3.10) for close approximations of the true variance of the Horvitz and Thompson estimator.

3.4 Assisting the Superpopulation Model

A superpopulation model frequently is used in the design of finite population to estimate the population total. The model described here is the same as that used by Smith (1938), in a survey of agricultural data. The model is described as:

$$Y_i = \beta\pi_i + \varepsilon_i\ , \tag{3.11}$$

where $E(\varepsilon_i) = E(\varepsilon_i, \varepsilon_j) = 0$ and $E(\varepsilon_i^2) = \sigma_i^2 = \sigma^2\pi_i^{2\gamma}$ and $\frac{1}{2} \le \gamma \le 1$.

Useful references for model (3.11) are Brewer (1963b), Brewer (1979), Brewer and Donadio (2003), Cassel et al. (1976), Chaudhuri and Vos (1986), Cochran (1953), Foreman and Brewer (1971), Godambe and Joshi (1965), Hanif and Brewer (1980), Hansen et al. (1983), Kalton (1983), Rao (1966), Rao (1977), Royall (1970), Royall and Herson (1973a, 1973b), Sarndal and Wright (1984), Sarndal et al. (1992), and many others.

Under the superpopulation model, it will be shown in the following section that the expected value of the proposed approximate variance formulas (formulas 3.4, 3.6, 3.8, and 3.10) derived after getting the approximations for π_{ij} will be approximately equal to Godambe and Joshi's (1965) lower bound for the expected variance, i.e.,

$$E_M[Var(y'_{HT})] = \sum_{i=1}^{N} \sigma_i^2 \left(\frac{1}{\pi_i} - 1 \right). \tag{3.12}$$

The suitable values of a_i in formulas 3.4 and 3.6 and the suitable values of a and $a+b$ in formulas 3.8 and 3.10 will be obtained using the fact that the expected value of the variance of the Horvitz and Thompson estimator is equal to the lower bound of the expected variance for any design-unbiased estimator under the superpopulation model (Godambe and Joshi, 1965).

3.4.1 The First Proposed Approximate Variance Formula

Recall formula (3.4) and model (3.11):

$$Var_{HT}(y'_{HT}) \approx \sum_{i=1}^{N} \left(\frac{1}{\pi_i} - 2a_i \right) Y_i^2 + Y \left(2\sum_{i=1}^{N} a_i Y_i - Y \right) \tag{3.4}$$

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

$$= \sum_{i=1}^{N} \left[\left(\frac{1}{\pi_i} - 1 - 2a_i + 1 \right) \left(\beta^2 \pi_i^2 + \varepsilon_i^2 + 2\beta\pi_i\varepsilon_i \right) \right]$$

$$+ \left(n\beta + \sum_{i=1}^{N} \varepsilon_i \right) \left(2\sum_{i=1}^{N} a_i (\beta\pi_i + \varepsilon_i) - n\beta - \sum_{i=1}^{N} \varepsilon_i \right)$$

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

$$-4\beta \sum_{i=1}^{N} a_i \pi_i \varepsilon_i + \sum_{i=1}^{N} \varepsilon_i^2 + 2n\beta^2 \sum_{i=1}^{N} a_i \pi_i + 2n\beta \sum_{i=1}^{N} a_i \varepsilon_i - n^2\beta^2$$

$$-n\beta \sum_{i=1}^{N} \varepsilon_i + 2\beta \left(\sum_{i=1}^{N} \varepsilon_i \right) \left(\sum_{i=1}^{N} a_i \pi_i \right) + 2 \left(\sum_{i=1}^{N} \varepsilon_i \right) \left(\sum_{i=1}^{N} a_i \varepsilon_i \right)$$

$$-n\beta\sum_{i=1}^{N}\varepsilon_i-\left(\sum_{i=1}^{N}\varepsilon_i\right)^2.$$

$$E_M\left(Var_{HT}(y_{HT}^{/})\right)\approx\sum_{i=1}^{N}\sigma_i^2\left(\frac{1}{\pi_i}-1\right)+\beta^2\left(2n\sum_{i=1}^{N}a_i\pi_i-2\sum_{i=1}^{N}a_i\pi_i^2-n(n-1)\right). \qquad (3.13)$$

Formula (3.13) shows that the expected value of the approximate variance formula (formula 3.4) is approximately equal to the Godambe and Joshi's (1965) lower bound for the expected variance, i.e., the expected value of the approximate variance of the Horvitz and Thompson estimator is equal to Godambe and Joshi's (1965) lower bound for the expected variance plus some quantity, q_1. Now we will find the suitable values of a_i in the first approximated variance of the Horvitz and Thompson estimator (formula 3.4) by equating q_1 to zero.

$$q_1=\beta^2\left(2n\sum_{i=1}^{N}a_i\pi_i-2\sum_{i=1}^{N}a_i\pi_i^2-n(n-1)\right)=0$$

$$2n\sum_{i=1}^{N}a_i\pi_i-2\sum_{i=1}^{N}a_i\pi_i^2-n(n-1)=0$$

Therefore we can choose a_i such that

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

This linear equation has an $N-1$ dimensional solution space. Requiring that $a_i=a$ for all i, leads to the following formula:

$$a=\frac{\frac{n-1}{2}}{n-\frac{1}{n}\sum_{i=1}^{N}\pi_i^2}. \qquad (3.14)$$

It is also interesting to note that this equation is the same as equation (2.17) that was obtained for option-2 of the variance form of Sen, and Yates and Grundy, but here we did not have any restriction on σ_i^2.

3.4.2 The Second Proposed Approximate Variance Formula

Recall formula (3.6) and model (3.11):

$$Var_{HT}(y'_{HT}) \approx \sum_{i=1}^{N}\left(\frac{1}{\pi_i}-2\frac{a_i}{\pi_i}\right)Y_i^2+Y\left(2\sum_{i=1}^{N}\frac{a_i}{\pi_i}Y_i-Y\right) \tag{3.6}$$

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

$$=\sum_{i=1}^{N}\left[\left(\frac{1}{\pi_i}-1-2\frac{a_i}{\pi_i}+1\right)\left(\beta^2\pi_i^2+\varepsilon_i^2+2\beta\pi_i\varepsilon_i\right)\right]$$

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

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

$$-2\sum_{i=1}^{N}\frac{a_i}{\pi_i}\varepsilon_i^2-4\beta\sum_{i=1}^{N}a_i\varepsilon_i+\sum_{i=1}^{N}\varepsilon_i^2+2n\beta^2\sum_{i=1}^{N}a_i$$

$$+2n\beta\sum_{i=1}^{N}\frac{a_i}{\pi_i}\varepsilon_i-n^2\beta^2-n\beta\sum_{i=1}^{N}\varepsilon_i+2\beta\left(\sum_{i=1}^{N}\varepsilon_i\right)\left(\sum_{I=1}^{N}a_i\right)$$

$$+2\left(\sum_{i=1}^{N}\varepsilon_i\right)\left(\sum_{i=1}^{N}\frac{a_i}{\pi_i}\varepsilon_i\right)-n\beta\sum_{i=1}^{N}\varepsilon_i-\left(\sum_{I=1}^{N}\varepsilon_i\right)^2.$$

$$E_M\left(Var_{HT}(y'_{HT})\right)\approx\sum_{i=1}^{N}\sigma_i^2\left(\frac{1}{\pi_i}-1\right)+\beta^2\left(2n\sum_{i=1}^{N}a_i-2\sum_{i=1}^{N}a_i\pi_i-n(n-1)\right). \tag{3.15}$$

Formula (3.15) shows that the expected value of the approximate variance formula (formula 3.6) is approximately equal to Godambe and Joshi's (1965) lower bound for the expected variance, i.e., the expected value of the approximate variance of the Horvitz and Thompson estimator is equal to the Godambe and Joshi's (1965) lower bound for the expected variance plus some quantity, q_2. Now we will find the suitable values of a_i in the second approximate variance of the Horvitz and Thompson estimator (formula 3.6) by equating q_2 to zero.

$$q_2 = \beta^2\left(2n\sum_{i=1}^{N} a_i - 2\sum_{i=1}^{N} a_i\pi_i - n(n-1)\right) = 0$$

$$2n\sum_{i=1}^{N} a_i - 2\sum_{i=1}^{N} a_i\pi_i - n(n-1) = 0$$

Therefore we can choose a_i such that

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

This linear equation has an $N-1$ dimensional solution space. Requiring that $a_i = a$ for all i, leads to the following formula:

$$a = \frac{n-1}{2(N-1)}\ . \tag{3.16}$$

It is also interesting to note that this equation is the same as equation (2.20) that obtained for option-2 of the variance form of Sen, and Yates and Grundy, but here we did not have any restriction on σ_i^2.

3.4.3 The Third Proposed Approximate Variance Formula

Recall formula (3.8) and model (3.11):

$$Var_{HT}(y_{HT}^{/}) \approx \sum_{i=1}^{N}\left(\frac{1}{\pi_i}-a\right)Y_i^2 + Y^2(a-1) \tag{3.8}$$

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

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

$$+\left(n^2\beta^2+\left(\sum_{i=1}^{N}\varepsilon_i\right)^2+2n\beta\sum_{i=1}^{N}\varepsilon_i\right)(a-1)$$

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

$$-a\sum_{i=1}^{N}\varepsilon_i^2-2a\beta\sum_{i=1}^{N}\pi_i\varepsilon_i+an^2\beta^2+a\left(\sum_{i=1}^{N}\varepsilon_i\right)^2$$

$$+2an\beta\sum_{i=1}^{N}\varepsilon_i-n^2\beta^2-\left(\sum_{i=1}^{N}\varepsilon_i\right)^2-2n\beta\sum_{i=1}^{N}\varepsilon_i.$$

$$E_M\left(Var_{HT}(y_{HT}^{/})\right) \approx \sum_{i=1}^{N}\sigma_i^2\left(\frac{1}{\pi_i}-1\right)+\beta^2\left(an^2-a\sum_{i=1}^{N}\pi_i^2-n(n-1)\right). \tag{3.17}$$

Formula (3.17) shows that the expected value of the approximate variance formula (formula 3.8) is approximately equal to the Godambe and Joshi's (1965) lower bound for the expected variance, i.e., the expected value of the approximate variance of the Horvitz and Thompson estimator is equal to the Godambe and Joshi's (1965) lower bound for the expected variance plus some quantity, q_3. Now we will find the suitable values of a in the third approximated variance of the Horvitz and Thompson estimator (formula 3.8) by equating q_3 to zero.

$$q_3 = \beta^2\left(an^2 - a\sum_{i=1}^{N}\pi_i^2 - n(n-1)\right) = 0$$

$$an^2 - a\sum_{i=1}^{N}\pi_i^2 - n(n-1) = 0$$

$$a = \frac{(n-1)}{n - \frac{1}{n}\sum_{i=1}^{N}\pi_i^2}. \tag{3.18}$$

It is also interesting to note that this equation is the same as equation (2.22) that obtained for the variance form of Sen, and Yates and Grundy, but here we did not have any restriction on σ_i^2.

Therefore the third approximate variance of the Horvitz and Thompson estimator (formula 3.9) becomes

$$Var_{HT}(y_{HT}^{/}) \approx \sum_{i=1}^{N}\left(\frac{1}{\pi_i} - \frac{n(n-1)}{n^2 - \sum_{i=1}^{N}\pi_i^2}\right)Y_i^2 + Y^2\left(\frac{\sum_{i=1}^{N}\pi_i^2 - n}{n^2 - \sum_{i=1}^{N}\pi_i^2}\right). \tag{3.19}$$

3.4.4 The Fourth Proposed Approximate Variance Formula

Recall formula (3.10) and model (3.11):

$$Var_{HT}(y_{HT}^{/}) \approx \left(1-(a+b)\right)\sum_{i=1}^{N}\frac{1}{\pi_i}Y_i^2 + Y\left((a+b)\sum_{i=1}^{N}\frac{1}{\pi_i}Y_i - Y\right) \tag{3.10}$$

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

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

$$+\left(n\beta + \sum_{i=1}^{N}\varepsilon_i\right)\left((a+b)N\beta + (a+b)\sum_{i=1}^{N}\frac{\varepsilon_i}{\pi_i} - n\beta - \sum_{i=1}^{N}\varepsilon_i\right)$$

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

$$+\sum_{i=1}^{N}\varepsilon_i^2 + 2\beta\sum_{i=1}^{N}\varepsilon_i - 2\beta(a+b)\sum_{i=1}^{N}\varepsilon_i + Nn\beta^2(a+b)$$

$$+n\beta(a+b)\sum_{i=1}^{N}\frac{\varepsilon_i}{\pi_i} - n^2\beta^2 - n\beta\sum_{i=1}^{N}\varepsilon_i + N\beta(a+b)\sum_{i=1}^{N}\varepsilon_i$$

$$+(a+b)\left(\sum_{i=1}^{N}\varepsilon_i\right)\left(\sum_{i=1}^{N}\frac{\varepsilon_i}{\pi_i}\right) - n\beta\sum_{i=1}^{N}\varepsilon_i - \left(\sum_{i=1}^{N}\varepsilon_i\right)^2.$$

$$E_M\left(Var_{HT}(y_{HT}^{/})\right) \approx \sum_{i=1}^{N}\sigma_i^2\left(\frac{1}{\pi_i} - 1\right) + \beta^2\left(Nn(a+b) - n(a+b) - n(n-1)\right). \tag{3.20}$$

Formula (3.20) shows that the expected value of the approximate variance formula (formula 3.10) is approximately equal to the Godambe and Joshi's (1965) lower bound for the expected variance, i.e., the expected value of the approximate variance of the Horvitz and Thompson estimator is equal to Godambe and Joshi's (1965) lower bound for the expected variance plus some quantity, q_4. Now we will find the suitable values of $(a+b)$ in the fourth approximate variance of the Horvitz and Thompson estimator (formula 3.10) by equating q_4 to zero.

$$q_4 = \beta^2\left(Nn(a+b) - n(a+b) - n(n-1)\right) = 0$$

$$Nn(a+b) - n(a+b) - n(n-1) = 0$$

$$(a+b)=\frac{n-1}{N-1}. \tag{3.21}$$

It is also interesting to note that this equation is the same as equation (2.25) that obtained for the variance form of Sen, and Yates and Grundy, but here we did not have any restriction on σ_i^2.

Therefore the fourth approximate variance of the Horvitz and Thompson estimator (formula 3.10) becomes

$$Var_{HT}(y_{HT}^{\prime}) \approx \frac{N-n}{N-1}\sum_{i=1}^{N}\frac{1}{\pi_i}Y_i^2 + Y\left(\frac{n-1}{N-1}\sum_{i=1}^{N}\frac{1}{\pi_i}Y_i - Y\right). \tag{3.22}$$

3.5 An Empirical Study to Evaluate the Proposed Variance Approximations

3.5.1 The Setup

To evaluate the four proposed variance approximations, an empirical study was carried out. To carry out this study, a total of 36 populations were used to calculate the sampling variance of the Horvitz and Thompson (1952) estimator with both the exact variance and the approximated variances. The probability proportion to size with replacement (ppswr) sampling design is used as the selection procedure, but the Horvitz and Thompson (1952) estimator will depend only on the distinct units in the sample, not on the numbers of repeat selections (Thompson 1992). Note that formula 2.16 and formula 2.19 (option-1) were used to calculate a_i in the first two variance approximations, where formula 2.17 and formula 2.20 (option-2) were used as starting values for formula 2.16 and formula 2.19, respectively. The relative error percentage, L, was used as a criterion in the comparison of the performance of the four approximated variances.

$$L = \frac{V_a - V_e}{V_e} \text{ X } 100,$$

where V_a and V_e represent the approximated variance and the exact variance, respectively. A good approximation should have L very close to 0%. In order to test the approximated variances under different situations, 36 populations were used in this empirical study. The main features of each population are given in section (2.5.1).

3.5.2 Results

Table B.1, and B.2 show the values of the exact variance and the four approximate variances for the 36 populations when the sample size (n) was 2, and 5, respectively. Table B.3, and B.4 show the values of the exact variance and the four approximate variances for the last eight populations when the sample size (n) was 50, and 100, respectively. Table 3.1, and 3.2 report the value of the criterion (L) for the four approximate variances as L_1, L_2, L_3, and L_4, respectively for the 36 populations when the sample size (n) was 2, and 5, respectively. Table 3.3, and 3.4 report the value of the criterion (L) for the four approximate variances as L_1, L_2, L_3, and L_4, respectively for the last eight populations when the sample size (n) was 50, and 100, respectively. The results in Tables 3.1, 3.2, 3.3, and 3.4 show that the first three proposed variance approximations perform well over all. The fourth approximation does not perform well compared to the first three. Table 3.2 shows that the first three variance approximations are sensitive to the populations with similar coefficients of variation for X and Y, for example, populations 6-9, 14, and 17-19 which are similar to populations 29-36, respectively. The results in Tables 3.1, 3.2, 3.3, and 3.4 show that the first three proposed variance approximations perform very well over all for populations 29-36 despite of their characteristics.

In general the first three approximations perform very well over all if $\left(\frac{n}{N} X\ 100\right)\% \leq 10\%$. The first and the second variance approximations perform similarly. The second variance approximation consistently over-approximates the variance by up to 5.2%. The third variance approximation is not far from the quality of the performance of the first two variance approximations, but the third variance approximation over-approximates or under-approximates the variance by up to 8%.

Table 3.1

The Relative Error Percentage, L, for the Four Approximated Variances Using HT Form with n = 2

Pop#	L_1	L_2	L_3	L_4	N
1	-1.613023 %	0.2553455 %	1.7798721 %	114.28238 %	15
2	2.8951192 %	3.1530696 %	-8.889533 %	112.93058 %	15
3	-0.772211 %	0.1054453 %	0.900732 %	126.77672 %	15
4	-1.709629 %	0.6474327 %	1.7893544 %	166.70543 %	15
5	-0.903622 %	1.6010887 %	-1.547494 %	100.93331 %	15
6	0.1694762 %	0.3384236 %	-0.25308 %	24.754827 %	15
7	0.9614805 %	0.5671382 %	-1.159252 %	17.946891 %	15
8	4.1133296 %	1.5737656 %	-4.681738 %	21.629983 %	15
9	2.4455886 %	2.1025761 %	-6.488381 %	65.355294 %	15
10	3.857959 %	3.152171 %	-9.914522 %	96.916743 %	15
11	-0.781242 %	0.1013804 %	0.9102566 %	123.39072 %	15
12	-0.756446 %	0.1003156 %	0.8820402 %	123.8006 %	15
13	-0.765017 %	0.1303 %	0.9019479 %	140.84683 %	15
14	-0.193808 %	0.7300316 %	0.4263235 %	225.05516 %	15
15	-1.7939 %	0.439127 %	1.905538 %	140.39644 %	15
16	0.3144695 %	0.3368427 %	-0.482205 %	16.574803 %	15
17	1.5495936 %	0.6921424 %	-1.969987 %	6.4404426 %	15
18	-0.82373 %	1.2410932 %	0.7166273 %	210.48006 %	15
19	0.9421701 %	0.6819088 %	-2.086731 %	18.274092 %	15
20	4.1011674 %	4.0302345 %	-12.32678 %	94.343847 %	10
21	2.2715017 %	0.8567936 %	-3.027656 %	15.322662 %	13
22	3.9090532 %	2.6310116 %	-5.979255 %	110.87862 %	14
23	6.5106854 %	2.0955782 %	-5.35534 %	79.441285 %	15
24	-1.074225 %	0.3900601 %	0.8219997 %	90.534877 %	20
25	-0.153417 %	0.3433754 %	-0.016093 %	43.53225 %	20
26	2.3671963 %	1.3340737 %	-3.842351 %	53.270243 %	20
27	-0.66526 %	0.9137505 %	0.7664748 %	79.938056 %	17
28	3.8651236 %	1.7222575 %	-4.356733 %	34.601795 %	18
29	-0.013264 %	0.0000561 %	0.0132947 %	29.259526 %	1000
30	-0.008467 %	0.0001201 %	0.0085042 %	19.635819 %	1000
31	-0.013264 %	0.0003592 %	0.0133592 %	32.242531 %	1000
32	-0.000786 %	0.000935 %	-0.00091 %	3.0211929 %	1000
33	-0.03397 %	0.0000982 %	0.0342678 %	190.64755 %	1000
34	-0.003725 %	0.0001297 %	0.0037449 %	8.3679878 %	1000
35	-0.102752 %	0.0005924 %	0.1029845 %	202.75479 %	1000
36	-0.00037 %	0.0003756 %	-0.000312 %	1.1346344 %	1000

Table 3.2

The Relative Error Percentage, L, for the Four Approximated Variances Using HT Form with n = 5

Pop#	L_1	L_2	L_3	L_4	N
1	-1.23511 %	3.0445595 %	-1.174993 %	465.29008 %	15
2	19.891942 %	19.849265 %	-32.76581 %	520.1197 %	15
3	-0.969518 %	1.1781873 %	-0.019895 %	507.71163 %	15
4	2.6621329 %	7.0745661 %	-8.525861 %	623.97766 %	15
5	12.015497 %	15.032764 %	-21.9359 %	482.07339 %	15
6	4.9145342 %	4.8399904 %	-8.994273 %	163.90027 %	15
7	9.1245724 %	7.507287 %	-15.54004 %	168.7842 %	15
8	19.981909 %	14.743644 %	-32.89086 %	262.16145 %	15
9	18.169348 %	17.509226 %	-29.55417 %	441.2217 %	15
10	20.862112 %	19.952334 %	-33.85923 %	501.47989 %	15
11	-1.039811 %	1.1286852 %	0.1007414 %	495.13357 %	15
12	-0.98457 %	1.1192282 %	0.0491018 %	496.37139 %	15
13	-0.697811 %	1.4629095 %	-0.584085 %	559.87068 %	15
14	5.8169367 %	7.4298519 %	-12.82882 %	798.31743 %	15
15	0.4531637 %	5.121084 %	-4.577107 %	551.36901 %	15
16	5.4917292 %	5.0440027 %	-9.586666 %	136.77289 %	15
17	11.759321 %	9.0635164 %	-19.22678 %	152.00809 %	15
18	8.7083815 %	11.463615 %	-18.43797 %	704.40232 %	15
19	10.644463 %	9.7625246 %	-17.25577 %	234.43924 %	15
20	29.149392 %	28.943318 %	-46.54575 %	567.12819 %	10
21	14.333663 %	10.929035 %	-23.06367 %	214.65645 %	13
22	20.934088 %	18.55621 %	-35.59643 %	498.36826 %	14
23	21.92013 %	14.644587 %	-37.87752 %	400.43116 %	15
24	2.2950895 %	5.4421398 %	-6.284288 %	400.92196 %	20
25	3.9413448 %	4.8903972 %	-7.854354 %	243.49986 %	20
26	14.53271 %	12.055106 %	-23.60043 %	357.84516 %	20
27	7.5558424 %	9.2174543 %	-15.68783 %	366.28003 %	17
28	19.968695 %	14.459178 %	-32.59762 %	284.06786 %	18
29	-0.051482 %	0.0010245 %	0.050736 %	117.56365 %	1000
30	-0.030934 %	0.0022858 %	0.0291158 %	79.682023 %	1000
31	-0.044267 %	0.0068275 %	0.0387496 %	131.44818 %	1000
32	0.014917 %	0.0209238 %	-0.02908 %	31.847398 %	1000
33	-0.132075 %	0.0012824 %	0.1310531 %	761.12416 %	1000
34	-0.011928 %	0.0025203 %	0.009905 %	34.922086 %	1000
35	-0.386539 %	0.0108254 %	0.3784105 %	798.71172 %	1000
36	0.005975 %	0.0086745 %	-0.011846 %	12.893257 %	1000

Table 3.3

The Relative Error Percentage, L, for the Four Approximated Variances Using HT Form with n = 50

Pop#	L_1	L_2	L_3	L_4	N
29	-0.369149 %	0.1479555 %	0.215184 %	1538.0903 %	1000
30	0.0906865 %	0.3225903 %	-0.427389 %	1180.0267 %	1000
31	0.6865298 %	0.8513671 %	-1.576365 %	2007.7602 %	1000
32	1.9280644 %	1.9188552 %	-2.806707 %	2403.9552 %	1000
33	-1.034665 %	0.1581964 %	0.6819745 %	9088.2734 %	1000
34	0.3279788 %	0.3563137 %	-0.687493 %	683.41105 %	1000
35	-1.677707 %	1.1922383 %	0.4636768 %	8267.1079 %	1000
36	1.0711405 %	1.0677606 %	-1.559863 %	1334.6967 %	1000

Table 3.4

The Relative Error Percentage, L, for the Four Approximated Variances Using HT Form with n = 100

Pop#	L_1	L_2	L_3	L_4	N
29	-0.261713 %	0.558382 %	-0.315417 %	3334.1068 %	1000
30	0.992295 %	1.1802376 %	-2.208658 %	2825.5879 %	1000
31	3.0581783 %	2.8111588 %	-5.955723 %	4791.1452 %	1000
32	5.329045 %	5.2478716 %	-7.658607 %	6831.0266 %	1000
33	-1.165089 %	0.5669861 %	-0.076489 %	17962.643 %	1000
34	1.4741613 %	1.300109 %	-2.767926 %	1918.7998 %	1000
35	0.1230021 %	3.6366956 %	-3.793053 %	15057.104 %	1000
36	3.4571559 %	3.4069231 %	-4.968859 %	4429.6739 %	1000

CHAPTER 4

THE PROPOSED GENERAL SELECTION PROCEDURE

4.1 Introduction

In this chapter a general selection procedure is developed on the basis of the estimator given by Horvitz and Thompson (1952). Special cases of this selection procedure are obtained. Also, the important properties regarding the probability of inclusion and joint probability of inclusion are verified.

The general selection procedure for a sample of size 2 $(n = 2)$ is stated as:

- Select the first unit with probability proportional to $\dfrac{p_i(1-a\,p_i)}{1-2b\,p_i}$.

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

The expression of π_i for this selection procedure is derived as follows:

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

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

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

where $$d=\sum_{k=1}^{N}\frac{p_k(1-a\,p_k)}{(1-2b\,p_k)} \tag{4.1}$$

Therefore

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

$$= \frac{p_i}{d}\left[\frac{(1-a\,p_i)(1-2\,p_i)}{(1-p_i)(1-2b\,p_i)}+\sum_{j=1}^{N}\frac{p_j(1-a\,p_j)}{(1-p_j)(1-2b\,p_j)}\right]. \tag{4.2}$$

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

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

$$= \frac{\dfrac{p_i(1-a\,p_i)}{(1-2b\,p_i)}}{\sum_{k=1}^{N}\dfrac{p_k(1-a\,p_k)}{(1-2b\,p_k)}}\cdot\frac{p_j}{(1-p_i)}+\frac{\dfrac{p_j(1-a\,p_j)}{(1-2b\,p_j)}}{\sum_{k=1}^{N}\dfrac{p_k(1-a\,p_k)}{(1-2b\,p_k)}}\cdot\frac{p_i}{(1-p_j)}$$

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

$$= \frac{p_i\,p_j}{d}\left[\frac{(1-a\,p_i)}{(1-p_i)(1-2\,b\,p_i)}+\frac{(1-a\,p_j)}{(1-p_j)(1-2\,b\,p_j)}\right]. \tag{4.3}$$

4.2 Some Useful Results for the General Selection Procedure

4.2.1 **Result-1**: $\sum_{i=1}^{N}\pi_i = 2$ for this selection procedure.

Proof: To prove this result again consider the value of π_i given in (4.2) as follows:

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

Applying sum on both sides:

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

$$= \frac{1}{d}\left[\sum_{i=1}^{N}\frac{p_i(1-a\,p_i)(1-2\,p_i)}{(1-p_i)(1-2b\,p_i)}+\sum_{i=1}^{N}\frac{p_i(1-a\,p_i)}{(1-p_i)(1-2b\,p_i)}\right] = 2. \tag{4.4}$$

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

Proof: To prove this result, we consider the value of π_{ij} from equation (4.4) as follows:

$$\pi_{ij} = \frac{p_i\,p_j}{d}\left[\frac{(1-a\,p_i)}{(1-p_i)(1-2\,b\,p_i)}+\frac{(1-a\,p_j)}{(1-p_j)(1-2\,b\,p_j)}\right].$$

Applying conditional sum on both sides we have:

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

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

$$= \frac{p_i}{d}\left[\frac{(1-a\,p_i)(1-2\,p_i)}{(1-p_i)(1-2b\,p_i)} + \sum_{j=1}^{N}\frac{p_j(1-a\,p_j)}{(1-p_j)(1-2b\,p_j)}\right] = \pi_i\,. \tag{4.5}$$

4.2.3 Result-3: The quantity π_{ij} obtained under this selection procedure, satisfies the relation $\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}\pi_{ij} = 2$.

Proof: To prove this result, we consider the value of π_{ij} from equation (4.3) as follows:

$$\pi_{ij} = \frac{p_i\,p_j}{d}\left[\frac{(1-a\,p_i)}{(1-p_i)(1-2\,b\,p_i)} + \frac{(1-a\,p_j)}{(1-p_j)(1-2\,b\,p_j)}\right] \tag{4.3}$$

Applying the double summation:

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

$$= \sum_{i=1}^{N}\left[\sum_{\substack{j=1\\j\neq i}}^{N}\left\{\frac{p_i\,p_j}{d}\left(\frac{(1-a\,p_i)}{(1-p_i)(1-2\,b\,p_i)} + \frac{(1-a\,p_j)}{(1-p_j)(1-2\,b\,p_j)}\right)\right\}\right]. \tag{4.6}$$

Also from (4.5)

$$\sum_{\substack{j=1\\j\neq i}}^{N}\pi_{ij} = \sum_{\substack{j=1\\j\neq i}}^{N}\frac{p_i\,p_j}{d}\left[\frac{(1-a\,p_i)}{(1-p_i)(1-2\,b\,p_i)} + \frac{(1-a\,p_j)}{(1-p_j)(1-2\,b\,p_j)}\right] = \pi_i\,. \tag{4.7}$$

Substituting (4.7) in (4.6):

$$\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}\pi_{ij} = \sum_{i=1}^{N}\pi_i = 2. \tag{4.8}$$

4.3 Special Cases

In this section, some special cases of the general class of selection procedures are obtained by using various combinations of the constants a and b in quantities π_i and π_{ij} given in equations (4.2) and (4.3). These special cases are obtained in the following.

4.3.1 **Special Case–1**: Recall the quantities π_i and π_{ij} as:

$$\pi_i = \frac{p_i}{d}\left[\frac{(1-a\,p_i)(1-2\,p_i)}{(1-p_i)(1-2b\,p_i)} + \sum_{j=1}^{N}\frac{p_j(1-a\,p_j)}{(1-p_j)(1-2b\,p_j)}\right] \tag{4.2}$$

$$\pi_{ij} = \frac{p_i\,p_j}{d}\left[\frac{(1-a\,p_i)}{(1-p_i)(1-2\,b\,p_i)} + \frac{(1-a\,p_j)}{(1-p_j)(1-2\,b\,p_j)}\right]. \tag{4.3}$$

Now substituting $a = 0$ and $b = 0$ in (4.2) and (4.3) we have:

$$\pi_i = p_i\left[1 + \sum_{j=1}^{N}\frac{p_j}{1-p_j} - \frac{p_i}{1-p_i}\right] \tag{4.9}$$

$$\pi_{ij} = p_i p_j\left[\frac{1}{1-p_i} + \frac{1}{1-p_j}\right]. \tag{4.10}$$

The quantities given in (4.9) and (4.10) are the probability of inclusion of the *ith* unit and the joint probability of inclusion of the *ith* and the *jth* units for Yates and Grundy (1953) draw-by-draw procedure.

4.3.2 **Special Case–2**: For this special case, we again recall the quantities given in (4.2) and (4.3). Now substituting $a = 1$ and $b = 0$ in these equations, we have:

$$\pi_i = \frac{2p_i(1-p_i)}{1-\sum_{j=1}^{N} p_j^2} \qquad and \qquad \pi_{ij} = \frac{2p_i p_j}{1-\sum_{j=1}^{N} p_j^2}. \tag{4.11}$$

The quantities given in (4.11) are the probability of inclusion of the *ith* unit and joint probability of inclusion of the *ith* and the *jth* units for the Yates and Grundy (1953) rejective procedure.

4.3.3 <u>Special Case–3</u>: For this special case we use $a = 0$ and $b = 1$ in quantities given in (4.2) and (4.3). For these values of constants the quantities π_i and π_{ij} reduce to:

$$\pi_i = \frac{p_i}{k_1}\left[\frac{1}{1-p_i} + \sum_{j=1}^{N} \frac{p_j}{(1-p_j)(1-2p_j)}\right] \tag{4.12}$$

$$\pi_{ij} = \frac{p_i p_j}{k_1}\left[\frac{1}{(1-p_i)(1-2p_i)} + \frac{1}{(1-p_j)(1-2p_j)}\right] \tag{4.13}$$

where $k_1 = \sum_{i=1}^{N} \frac{p_i}{1-2p_i}$.

The quantities given in (4.12) and (4.13) are the probability of inclusion of the *ith* unit and the joint probability of inclusion of the *ith* and the *jth* units for the Shahbaz and Hanif (2003) selection procedure.

4.3.4 <u>Special Case–4</u>: This special case is obtained by using the combination $a = 1$ and $b = 1$ in (4.2) and (4.3). For this combination of a and b, we obtain:

$$\pi_i = 2p_i \qquad and \qquad \pi_{ij} = \frac{2p_i p_j}{k}\left[\frac{1}{1-2p_i} + \frac{1}{1-2p_j}\right] \tag{4.14}$$

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

The quantities given in (4.14) are the probability of inclusion of the *ith* unit and the joint probability of inclusion of the *ith* and the *jth* units for the Brewer (1963a) selection procedure.

4.4 An Empirical Study to Specify the General Selection Procedure

4.4.1 The Setup

For this section, an empirical study was carried out to obtain the best combination of constants a and b involved in the proposed general selection procedure. To carry out this study, a total of 36 populations were selected and the sampling variance of the Horvitz and Thompson (1952) estimator for these populations was calculated with selected combinations of values for a and b. After calculating these variances, the rankings were performed for each population where the smallest rank was given to the combination with least variance and so on; then the average rank over all populations for each combination of a and b was calculated to obtain the best combination for the constants involved, such that the smallest average rank provides the best combination of a and b. The selected combinations of a and b were chosen so that the special cases of the general selection procedure are included in the comparison. The ranges of a and b were also selected so that the following condition was satisfied.

$$1 > \frac{p_i(1-ap_i)}{1-2bp_i} > 0 \Rightarrow p_i(1-ap_i) > 0 \text{ and } 1-2bp_i > 0 \Rightarrow$$

$$\frac{p_i(1-ap_i)}{1-2bp_i} > 0 \Rightarrow p_i(1-ap_i) > 0 \text{ and } 1-2bp_i > 0 \Rightarrow a < \frac{1}{p_i} \text{ and } b < \frac{1}{2p_i}$$

and

$$1 > \frac{p_i(1-ap_i)}{1-2bp_i} \Rightarrow 1-2bp_i > p_i(1-ap_i) \Rightarrow b < \frac{1-p_i+ap_i^2}{2p_i}.$$

To satisfy those conditions with the selected combinations of a and b, one restriction on p_i was needed to be satisfied which is $p_i < 0.21$. This restriction on p_i is satisfied in most of the fields of application including the 36 populations. The main features of each of the 36 populations are given in section (2.5.1).

4.4.2 Results

The sampling variance of the Horvitz and Thompson (1952) estimator for the 36 populations is given in Tables C.1-C.36. The selected set of combinations for a and b are given in Table 4.1. The average ranks of the various combinations of a and b are given in Table 4.1. The results in Table 4.1 show that the combination of $a = 1.5$ and $b = -1.5$ has the best performance, as this combination has the smallest average rank among the selected set of combinations for a and b. Table 4.1 shows that the best value of b is $b = -1.5$ which is located at the beginning of the selected range of b for the selected set of combinations for a and b. The location of the value of b may interpreted as there is a better value of b if the selected range of b was extended from its beginning. Therefore, the range of b was extended to provide an extended selected set of combinations for a and b which are given in Table 4.2. The results in Table 4.2 show that the combination $a = 1$ and $b = -3$ has the best performance, as this combination has the smallest average rank. Table 4.2 shows that the best value of b is $b = -3$ which is still located at the beginning of the selected range of b for the extended selected set of combinations for a and b. Therefore, the conclusion has a limitation that the combination $a = 1$ and $b = -3$ has the best performance among the extended selected set of combinations for a and b. This combination has no extra restrictions on p_i other than the regular restriction on p_i which is $o < p_i < 1$. Therefore, the specified selection procedure with this combination has an advantage in that it can be used with any populations including those have units with large probability of selection, p_i.

Table 4.1

Average Rank of the Selected Set of Combinations

a	*b*						
	-1.50	-1.00	-0.50	0.00	0.50	1.00	1.50
-3.0	41.4722	49.9167	60.4722	70.2361	79.0972	85.9583	90.7361
-2.5	36.5556	44.9861	55.7500	66.2778	76.2083	84.1667	89.7639
-2.0	32.0556	41.4722	51.6528	62.8889	72.7083	82.2639	88.6528
-1.5	28.6528	35.6111	46.1528	57.8611	69.6389	80.0417	87.4444
-1.0	25.9583	30.7778	41.4722	53.3472	65.4444	76.3472	85.9583
-0.5	23.3056	27.0417	34.4444	47.2083	60.3056	72.9167	83.5417
0.0	22.0000	24.1806	29.4722	41.4722	55.3472	68.8333	80.3194
0.5	20.8194	22.3889	25.6806	33.2500	48.7500	63.5278	77.1667
1.0	20.6250	21.1944	22.8194	28.0278	41.4722	57.7083	73.1944
1.5	**19.6944**	20.6528	21.4444	23.9306	31.9167	50.2083	67.1944
2.0	20.0000	20.3750	20.5556	21.7083	26.4167	41.4722	61.3750
2.5	21.0694	20.1667	20.5139	20.9444	22.7083	30.3333	52.9583
3.0	22.2361	21.1667	21.0417	20.6389	21.5139	24.7778	41.4722

Table 4.2

Average Rank of the Extended Selected Set of Combinations

a	*b*									
	-3.0	-2.5	-2.0	-1.50	-1.00	-0.50	0.00	0.50	1.00	1.50
-3.0	51.083	56.736	63.250	72.750	82.167	93.944	105.014	115.431	123.569	129.153
-2.5	47.736	52.736	58.958	67.097	76.819	89.278	100.639	112.097	121.528	128.042
-2.0	44.278	48.917	55.181	61.417	72.750	84.236	96.806	108.125	119.264	126.819
-1.5	40.597	45.819	50.875	57.014	66.069	78.097	91.778	104.611	116.736	125.417
-1.0	38.931	43.000	47.194	52.764	59.861	72.750	86.236	99.861	112.514	123.681
-0.5	37.458	39.792	43.875	48.431	55.069	64.653	79.292	94.639	108.583	121.014
0.0	36.278	38.264	41.569	45.056	50.194	58.500	72.750	88.569	103.944	117.403
0.5	35.292	36.653	38.847	42.278	46.181	52.764	63.236	81.194	98.361	113.694
1.0	**34.653**	35.833	37.264	40.250	43.167	47.778	56.458	72.750	91.347	108.972
1.5	35.083	35.264	36.208	37.514	40.958	44.347	50.375	61.764	82.958	102.778
2.0	35.7500	35.6806	36.0556	36.7083	39.0278	41.2778	45.1944	54.3333	72.7500	95.6250
2.5	36.5556	36.0278	36.0139	37.1250	37.0694	39.7361	42.6806	47.5694	59.6528	86.1528
3.0	38.0000	37.5000	36.7917	36.8889	37.4306	39.1111	40.4722	43.8889	51.3056	72.7500

4.4.3 Illustration for Applying the Specified Selection Procedure

The following is an illustration of selecting a sample by using the specified selection procedure:

- Select first unit with probability proportional to $q_i = \dfrac{p_i(1-p_i)}{1+6p_i}$.
- Select second unit with probability proportional to size of the remaining units.

The example that was shown in Chapter 1 was used to select a sample of size two units with the specified selection procedure. The Y variable (Variable of Interest) is the number of households and the X variable (Measure of Size) is the eye-estimated number of households. Table 4.3 shows the values of Y, X, p_i, q_i, q_i^*, Cumulative of q_i^* and Range. To select the first sample unit, a random number is selected between zero and one, then the last column in Table 4.3 will be used to determine the interval, which includes the selected random number. The first sample unit is the corresponding unit of that interval. For example, let the selected random number be 0.4290. Therefore the first sample unit is unit number 9 which has Y=20 and X=17. To select the second sample unit, the first selected sample unit (unit number 9) will be discarded from the population, then Table 4.4 will be constructed. To select the second sample unit, a random number is selected between zero and one, then the last column in Table 4.4 will be used to determine the interval, which includes the selected random number. The second sample unit is the corresponding unit of that interval. For example, let the selected random number be 0.1540. Therefore the second sample unit is unit number 5 which has Y=21 and X=24. In this way, a sample of size two was selected by using the specified selection procedure.

Table 4.3

Table Used to Select the First Sample Unit

Y_i	X_i	p_i	q_i	$q_i^* = \frac{q_i}{\sum_{i=1}^{N} q_i}$	Cumulative of q_i^*	Range
19	18	0.04569	0.034218	0.048358	0.04836	0.0000 - 0.04836
9	9	0.02284	0.019630	0.027742	0.07610	0.04837 - 0.07610
17	14	0.03553	0.028248	0.039920	0.11602	0.07611 – 0.11602
14	12	0.03046	0.024967	0.035283	0.15130	0.11603 – 0.15130
21	24	0.06091	0.041892	0.059202	0.21051	0.15131 – 0.21051
22	25	0.06345	0.043040	0.060824	0.27133	0.21052 – 0.27133
27	23	0.05838	0.040709	0.057531	0.32886	0.27134 – 0.32886
35	24	0.06091	0.041892	0.059202	0.38806	0.32887 – 0.38806
20	**17**	**0.04315**	**0.032795**	**0.046347**	**0.43441**	**0.38807 – 0.43441**
15	14	0.03553	0.028248	0.039920	0.47433	0.43442 – 0.47433
18	18	0 .04569	0.034218	0.048358	0.52269	0.47434 – 0.52269
37	40	0.10152	0.056686	0.080109	0.60280	0.52270 – 0.60280
12	12	0.03046	0.024967	0.035283	0.63808	0.60281 – 0.63808
47	30	0.07614	0.048285	0.068237	0.70632	0.63809 – 0.70632
27	27	0.06853	0.045233	0.063924	0.77024	0.70633 – 0.77024
25	26	0.06599	0.044153	0.062398	0.83264	0.77025 – 0.83264
25	21	0.05330	0.038232	0.054030	0.88667	0.83265 – 0.88667
13	9	0.02284	0.019630	0.027742	0.91441	0.88668 – 0.91441
19	19	0.04822	0.035598	0.050307	0.96472	0.91442 – 0.96472
12	12	0.03046	0.024967	0.035283	1.00000	0.96473 – 1.00000
Sums:	394	1.00000	0.70761	1.00000		

Table 4.4

Table Used to Select the Second Sample Unit

Y_i	X_i	p_i	Cumulative of p_i	Range
19	18	0.04775	0.04775	0.00000 – 0.04775
9	9	0.02387	0.07162	0.04776 – 0.07162
17	14	0.03714	0.10875	0.07163 – 0.10875
14	12	0.03183	0.14058	0.10876 – 0.14058
21	**24**	**0.06366**	**0.20424**	**0.14059 – 0.20424**
22	25	0.06631	0.27056	0.20425 – 0.27056
27	23	0.06101	0.33156	0.27057 – 0.33156
35	24	0.06366	0.39523	0.33157 – 0.39523
15	14	0.03714	0.43236	0.39524 – 0.43236
18	18	0.04775	0.48011	0.43237 – 0.48011
37	40	0.10610	0.58621	0.48012 – 0.58621
12	12	0.03183	0.61804	0.58622 – 0.61804
47	30	0.07958	0.69761	0.61805 – 0.69761
27	27	0.07162	0.76923	0.69762 – 0.76923
25	26	0.06897	0.83820	0.76924 – 0.83820
25	21	0.05570	0.89390	0.83821 – 0.89390
13	9	0.02387	0.91777	0.89391 – 0.91777
19	19	0.05040	0.96817	0.91778 – 0.96817
12	12	0.03183	1.00000	0.96818 – 1.00000
Sums:	377	1.00000		

CHAPTER 5

CONCLUSIONS AND FUTURE WORK

The first purpose of this thesis was to find approximate expressions for the variance of the Horvitz and Thompson estimator free of the joint inclusion probabilities when the sample size is fixed and random. In the case of fixed sample size, the four proposed variance approximations perform very well. The four approximate variances can be arranged according to the quality of their performance as the first approximation, the second approximation, the third approximation, and the fourth approximation where the first approximation is the best. In the case of random sample size, the first three variance approximations perform very well if $\left(\frac{n}{N} X\ 100\right)\% \leq 10\%$. The first and the second variance approximations perform similarly. The second variance approximation consistently over-approximates the variance. The third variance approximation is not far from the quality of the performance of the first two variance approximations. Research is needed to find estimators for the proposed variance approximations for both fixed and random sample sizes. Also, Monte Carlo simulation studies are needed to examine the statistical properties of those variance estimators.

The last purpose of this thesis was to develop a general selection procedure on the basis of the estimator given by Horvitz and Thompson (1952). Special cases of this selection procedure were obtained. Also, the important properties regarding the probability of inclusion and joint probability of inclusion were verified. Finally, the empirical study specified the general selection procedure by determining the best values of a and b as $a = 1$ and $b = -3$. Research is needed to generalize the selection procedure for more than sample size two.

REFERENCES

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, 1645-667.

Beg, M.A. and Hanif, M. (1991), "Comparison of sampling with and without replacement using unequal probability sampling," *Bull. Inter. Stat. Inst.,* 48(4), 118-119.

Berger, Y.G. (2003), "A simple variance estimator for unequal probability sampling without replacement." To appear in the *Journal of Applied Statistics*.

Brewer, K.R.W. (1963a), "A model of systematic sampling with unequal probabilities," *Aust. J. Stat.,* 5, 5-13.

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.

Brewer, K.R.W. (1979), "A class of robust sampling design for large scale surveys," *J. Amer. Stat. Assoc.,* 74(4), 911-915.

Brewer, K.R.W. and Hanif, M. (1983), *Sampling with Unequal Probabilities*, Lecture notes to Statistics, No. 15, Springer-Verlag.

Brewer, K.R.W. (2002), *Combined Survey Sampling Inference,* Weighing Basu's Elephants, Arnold, London.

Brewer, K.R.W. and Donadio, M.E. (2003), "The High Entropy Variance of the Horvitz-Thompson Estimator," *Survey Methodology,* 29 (2), 189-196.

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.

Chaudhuri, A. and Vos, J.W.E. (1986), *Unified Theory and Strategies of Survey Sampling*, North Holland series in Statistics and Probability. Elsevier Science, New York.

Cochran, W.G. (1953, 1963, 1977), *Sampling Techniques*, John Wiley and Sons, Inc., New York.

Das, A.C. (1951), "On two phase sampling and sampling with varying probabilities," *Bull. Inter. Stat. Inst.*, 33(2), 105-112.

Deshpande, M.N., Prabhu-Ajgonkar, S.G. (1982), "An IPPS sampling scheme," *Statistica*, 36(4), 209-212.

Deville, J.C. and Tille, Y. (2002), "Variance approximation under balanced sampling," Submitted to *Journal of Statistical Planning and Inference*.

Durbin, J. (1953), "Some results in sampling when the units are selected with unequal probabilities," *J. Roy. Stat. Soc.* B, 15, 262-269.

Durbin, J. (1967), "Design of multistage surveys for estimation of sampling errors," *Appl. Stat.*, 16, 152-164.

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.

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.

Hajek, J. (1981), *Sampling from a Finite Population*, Marcel Dekker, New York.

Hajek, J. (1964), "Asymptotic theory of rejective sampling with varying probabilities from a finite population," *Ann. Math. Stat.*, 35, 1491-1523.

Hanif, M. and Brewer, K.R.W. (1980), "Sampling with unequal probabilities without replacement; a review," *Inter. Stat. Rev.*, 48(3), 317-335.

Hansen, M.H. and Hurwitz, W.N. (1943), "On the theory of sampling from a finite population," *Ann. Math. Stat.*, 14, 333-362.

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.

Hanurav, T.V. (1967), "Optimum utilization of auxiliary information: πps sampling of two units from a stratum," *J. Roy. Stat. Soc. Ser. B, 29*, 374-391.

Hartley, H.O. and Rao, J.N.K. (1962), "Sampling with unequal probabilities and without replacement," *Ann. Math. Stat.,* 33, 350-374.

Herzel, A. (1986), "Sampling without replacement with unequal probabilities with pre-assigned joint inclusion probabilities of any order," *Metron*, 44, 1-4, 49-68.

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.

Kalton, G. (1983), "Model in the practice of survey sampling," *Inter. Stat. Review*, 51, 175-193.

Lanke, J. (1974), "On non-negative variance estimators in survey sampling," *Sankhya*, C, 36(1), 33-42.

Madow, W.G. (1949), "On the theory of systematic sampling II," *Ann. Math. Stat.*, 20, 333-354.

Matei, A. and Tille, Y. (2003), "Evaluation of variance estimators in unequal probability sampling," Submitted to *Journal of Official Statistics.*.

Mukhopadhyay, P. (2000), *Theory and Methods of Survey Sampling,* Prentice-Hall of India, New Delhi.

Murthy, M.N. (1957), "Ordered and unordered estimators in sampling without replacement," *Sankhya*, 18, 379-390.

Narain, R.D. (1951), "On sampling without replacement with varying probabilities," *J. Ind. Soc, Agri. Stat.*, 3, 169-175.

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.

Pathak, P.K. (1967a), "Asymptotic efficiency of Des Raj's strategy-I," *Sankhya* A, 29, 283-298.

Pathak, P.K. (1967b), "Asymptotic efficiency of Des Raj's strategy-II," *Sankhya* A, 29, 299-304.

Raj, D. (1956), "Some estimators in sampling with varying probabilities without replacement," *J. Amer. Stat. Assoc.*, 60, 278-284.

Rao, J.N.K. (1963), "On two systems of unequal probability sampling without replacement," *Ann. Inst. Stat. Math.*, 15, 67-72.

Rao, J.N.K. (1965), "On two simple schemes of unequal probability sampling without replacement," *J. Ind. Stat. Assoc.*, 3, 173-180.

Rao, J.N.K. (1966), "On the relative efficiency of some estimators in pps sampling for multiple characteristics," *Sankhya*, A, 28, 61-70.

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.

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.

Rao, J.N.K. and Singh, M.P. (1973), "On the choice of estimators in survey sampling," *Aust. Jour. Stat.*, 15, 95-104.

Rao, T.J. (1977), "Estimating the variance of the ratio estimator for the Midzuno-Sen sampling scheme," *Metrika*, 24, 203-208.

Royall, R.M. (1970), "On finite population sampling theory under certain regression models," *Biometrika*, 57, 377-387.

Royall, R.M. and Herson, J. (1973a), "Robust estimation in finite populations-I," *J. Amer. Stat. Assoc.,* 68, 880-889.

Royall, R.M. and Herson, J. (1973b), "Robust estimation in finite populations-II," *J. Amer. Stat. Assoc.* 68, 890-893.

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.

Sampford, M.R. (1967), "On Sampling without replacement with unequal probabilities of selection," *Biometrika*, 54, 499-513.

Sarndal, C.E. and Wright, R.L. (1984), "Cosmetic form of estimators in survey sampling," *Scan. J. Stat.*, 11, 146-156.

Sarndal, C.W., Swensson, B., and Wretman, J.H. (1992), *Model Assisted Survey Sampling*, Springer-Verlag.

Sarndal, C.E. (1996), "Efficient estimators with simple variance in unequal probability sampling," *J. Amer. Stat. Assoc.,* 91, 1289-1300.

Sen, A.R. (1953), "On the estimate of the variance in sampling with varying probabilities," *J. Ind. Soc. Agri. Stat.*, 5, 119-127.

Shahbaz, M.Q. and Hanif, M. (2003), "A simple procedure for unequal probability sampling without replacement sample size 2," *Pak. J. Stat.,* 19(1), 151-160.

Smith, H.F. (1938), "An experimental law describing heterogeneity in the yields of agricultural crops," *J. Agri. Science*, 28, 1-23.

Thompson, S.K. (1992), *Sampling*, John Wiley and Sons, New York.

Yates, F. and Grundy, P.M. (1953), "Selection without replacement from within strata with probability proportional to size," *J. Roy. Stat. Soc.* B, 15, 153-161.

APPENDIXES

APPENDIX A

Table A.1: The Variance and the Approximated Variances of SYG Form with n = 2

Pop#	Var_SYG	Approx_1	Approx_2	Approx_3	Approx_4	N
1	39790.65	39791.597	39825.853	39843.698	38145.195	15
2	1790.1371	1790.2079	1790.9451	1791.6074	1798.1163	15
3	159502.19	159501.35	159557.14	159575.56	150184.26	15
4	32691.852	32691.825	32714.15	32724.457	31397.436	15
5	2811.2506	2811.5198	2814.2061	2816.4568	2827.7783	15
6	1590.5669	1589.3182	1588.1681	1586.6903	1607.9701	15
7	6424.4686	6419.0514	6413.2126	6406.2525	6503.3681	15
8	207285.8	207101.92	206878.24	206624.08	209842.98	15
9	49954.714	49952.079	49944.062	49935.865	49916.572	15
10	588156.82	588125.06	588026.05	587925.15	587657.59	15
11	412.08973	412.08426	412.22257	412.25965	388.25211	15
12	409.35174	409.34866	409.48625	409.52957	385.3567	15
13	411.23073	411.22895	411.36374	411.41009	386.81525	15
14	925.19331	925.03161	924.88418	924.44089	875.12936	15
15	86.697115	86.698461	86.761078	86.791708	83.163547	15
16	2581.1673	2579.4544	2577.7163	2575.6218	2612.3069	15
17	3452703.4	3450204.7	3447068.5	3443589.5	3495728.1	15
18	80.961723	80.958567	81.012686	81.036594	78.210263	15
19	56499.954	5649.549	56486.782	56478.513	56620.524	15
20	6373.3188	6371.9262	6368.0388	6363.8204	6325.4678	10
21	11657.685	11654.119	11656.111	11654.272	11599.775	13
22	37211.204	37196.415	37170.001	37138.996	36959.168	14
23	83779.678	83700.072	83588.786	83451.778	82677.749	15
24	805749.94	805730.97	806068.9	806366.88	810528.13	20
25	48513960	48508281	48501078	48492376	48574367	20
26	3010.591	3010.8095	3014.4843	3017.9165	3075.7759	20
27	25536.944	25526.622	25524.005	25515.442	25638.585	17
28	18324.015	18323.602	18368.936	18409.556	19082.086	18
29	21164766	21164766	21164764	21164762	21159372	1000
30	554581504	554581503	554581466	554581428	554527361	1000
31	704689908	704689907	704689861	704689814	704626447	1000
32	763835099	763835099	763835088	763835076	763831556	1000
33	6408189	6408189	6408189.5	6408190.1	6402845.9	1000
34	$2.2841 X10^{10}$	$2.2841 X10^{10}$	$2.2841 X10^{10}$	$2.2841 X10^{10}$	$2.2841 X10^{10}$	1000
35	1062859.4	1062859.4	1062859.3	1062859.3	1062394	1000
36	$1.1193 X10^{10}$	$1.1193 X10^{10}$	$1.1193 X10^{10}$	$1.1193 X10^{10}$	$1.1193 X10^{10}$	1000

APPENDIX B

Table B.1: The Variance and the Approximated Variances of HT Form with n = 2

Pop#	Var_HT	Approx_1	Approx_2	Approx_3	Approx_4	N
1	43028.767	42334.703	43138.639	43794.624	92203.065	15
2	23720.064	24406.788	24467.974	21611.461	50507.271	15
3	164413.41	163143.8	164586.78	165894.34	372851.34	15
4	39044.434	38376.919	39297.22	39743.077	104133.62	15
5	5377.2335	5328.6436	5463.3278	5294.0211	10804.653	15
6	1796.2246	1799.2688	1802.3034	1791.6787	2240.8769	15
7	7662.5311	7736.2049	7705.9883	7573.7031	9037.7173	15
8	326957.23	340406.05	332102.77	311649.95	397678.02	15
9	134579.87	137871.14	137409.51	125847.81	222534.94	15
10	7695031.6	7991902.7	7937592.1	6932106	15152805	15
11	424.59972	421.28257	425.03018	428.46467	948.51637	15
12	421.46949	418.2813	421.89229	425.18702	943.25124	15
13	425.66391	422.40751	426.21855	429.50317	1025.198	15
14	1086.922	1084.8154	1094.8568	1091.5558	3533.0959	15
15	98.193343	96.431852	98.624536	100.06445	236.0533	15
16	2922.2744	2931.464	2932.1178	2908.183	3406.6356	15
17	4271406.5	4337596	4300970.7	4187260.4	4546504	15
18	114.51803	113.57471	115.9393	115.33869	355.55565	15
19	72854.684	73541.099	73351.486	71334.403	86168.216	15
20	32628.003	33966.132	33942.988	28606.023	63410.517	10
21	14893.292	15231.593	15020.897	14442.374	17175.341	13
22	102406.64	106409.77	105100.97	96283.489	215953.72	14
23	150891.18	160715.23	154053.23	142810.45	270761.08	15
24	954384.41	944132.17	958107.08	962229.44	1521101.9	20
25	56104352	56018278	56297000	56095323	80527838	20
26	5828.0222	5965.9829	5905.7723	5604.0891	8932.6238	20
27	33609.11	33385.522	33916.213	33866.715	60475.579	17
28	32879.142	34149.962	33445.406	31446.686	44255.916	18
29	21189527	21186717	21189539	21192344	27389482	1000
30	555772984	555725927	555773652	555820249	664903560	1000
31	708595346	708501359	708597891	708690008	937064421	1000
32	778435252	778429135	778442530	778428166	801953283	1000
33	6415016.3	6412837.1	6415022.6	6417214.5	18645088	1000
34	$2.2895 X10^{10}$	$2.2894 X10^{10}$	$2.2895 X10^{10}$	$2.2896 X10^{10}$	$2.4811 X10^{10}$	1000
35	1072264.2	1071162.4	1072270.5	1073368.4	3246331.1	1000
36	$1.1281 X10^{10}$	$1.1281 X10^{10}$	$1.1281 X10^{10}$	$1.1281 X10^{10}$	$1.1409 X10^{10}$	1000

Table B.2: The Variance and the Approximated Variances of HT Form with n = 5

Pop#	Var_HT	Approx_1	Approx_2	Approx_3	Approx_4	N
1	19020.18	18785.26	19599.261	18796.694	107519.19	15
2	30685.529	36789.477	36776.381	20631.166	190287.01	15
3	68202.375	67541.141	69005.927	68188.806	414473.77	15
4	20475.666	21020.755	21924.231	18729.939	148239.25	15
5	4409.3344	4939.1378	5072.1792	3442.1074	25665.562	15
6	761.60019	799.02929	798.46157	693.09979	2009.865	15
7	3585.7406	3912.9241	3854.9325	3028.5151	9637.9042	15
8	217627.07	261113.11	249713.22	146047.65	788161.35	15
9	132081.16	156079.45	155207.55	93045.674	714851.92	15
10	9936613.4	12009601	11919200	6572152.9	59766731	15
11	175.7269	173.89967	177.7103	175.90393	1045.8097	15
12	174.47834	172.76048	176.43115	174.56401	1040.5389	15
13	178.54523	177.29932	181.15719	177.50237	1178.1676	15
14	561.19942	593.84404	602.89571	489.20419	5041.3523	15
15	47.338362	47.552883	49.7626	45.171635	308.34742	15
16	1242.9441	1311.2032	1305.6383	1123.7872	2942.9547	15
17	2125965.1	2375964.2	2318652.3	1717210.5	5357604.2	15
18	73.59256	80.001281	82.028928	60.023589	591.98026	15
19	40195.96	44474.604	44120.1	33259.836	134431.06	15
20	34875.055	45040.922	44969.053	18642.199	232661.32	10
21	7816.5288	8936.9236	8670.8	6013.7507	24595.212	13
22	98200.138	118757.44	116422.36	63244.393	587598.46	14
23	113781.81	138722.93	130444.68	70684.077	569399.62	15
24	479230.31	490229.07	505310.69	449114.09	2400569.9	20
25	27098139	28166170	28423345	24969755	93082067	20
26	4914.4356	5628.6363	5506.876	3754.6075	22500.505	20
27	19138.887	20584.992	20903.006	16136.411	89240.81	17
28	24544.186	29445.339	28093.073	16543.364	94266.329	18
29	8486490.6	8482121.5	8486577.5	8490796.3	18463519	1000
30	223133037	223064013	223138138	223198004	400929957	1000
31	287336836	287209640	287356454	287448178	665035881	1000
32	327933564	327982482	328002180	327838201	432371871	1000
33	2572876	2569477.9	2572909	2576247.8	22155657	1000
34	$9.19495 X10^9$	$9.19385 X10^9$	$9.19518 X10^9$	$9.19586 X10^9$	$1.2406 X10^9$	1000
35	439432.86	437734.28	439480.43	441095.72	3949234.7	1000
36	$4.60487 X10^9$	$4.60514 X10^9$	$4.60527 X10^9$	$4.60432 X10^9$	$5.19858 X10^9$	1000

Table B.3: The Variance and the Approximated Variances of HT Form with n = 50

Pop#	Var_HT	Approx_1	Approx_2	Approx_3	Approx_4	N
29	863533.45	860345.72	864811.09	865391.63	14145457	1000
30	23476187	23497477	23551919	23375853	300501467	1000
31	34286793	34522182	34578700	33746308	722683391	1000
32	56805632	57900881	57895650	55211264	$1.42239 X10^9$	1000
33	266959.76	264197.62	267382.08	268780.35	24528992	1000
34	971705696	974892685	975168017	965025285	$7.61245 X10^9$	1000
35	59035.882	58045.433	59739.73	59309.617	4939595.9	1000
36	594290448	600656134	600636048	585020330	$8.52627 X10^9$	1000

Table B.4: The Variance and the Approximated Variances of HT Form with n= 100

Pop#	Var_HT	Approx_1	Approx_2	Approx_3	Approx_4	N
29	438840.75	437692.24	441291.16	437456.57	15070260	1000
30	12307478	12429604	12452735	12035648	360066077	1000
31	19917835	20526958	20477757	18731584	974210230	1000
32	40857194	43034493	43001328	37728102	$2.83182 X10^9$	1000
33	138199.11	136588.97	138982.68	138093.41	24962413	1000
34	511395386	518934179	518044084	497240339	$1.0324 X10^{10}$	1000
35	37164.59	37210.303	38516.153	35754.917	5633075.7	1000
36	366377844	379044097	378860055	348173047	$1.6596 X10^{10}$	1000

APPENDIX C

Table C.1: Sampling Variance of Horvitz and Thompson Estimator for Population 1

a	*b*									
	-3.0	-2.5	-2.0	-1.50	-1.00	-0.50	0.00	0.50	1.00	1.50
-3.0	33211.895	34388.281	35676.08	37095.075	38671.092	40438.965	42447.717	44770.334	47524.271	50920.583
-2.5	32469.24	33638.341	34919.14	36331.611	37901.873	39665.212	41671.401	43994.768	46755.441	50170.799
-2.0	31694.735	32855.31	34127.751	35532.212	37095.075	38852.055	40853.635	43175.436	45940.213	49371.672
-1.5	30886.37	32037.023	33299.579	34694.342	36247.935	37996.455	39991.049	42308.559	45074.279	48518.174
-1.0	30041.987	31181.152	32432.097	33815.25	35357.43	37095.075	39079.918	41389.926	44152.792	47604.581
-0.5	29159.277	30285.192	31522.576	32891.942	34420.263	36144.244	38116.116	40414.834	43170.285	46624.349
0.0	28235.783	29346.461	30568.075	31921.178	33432.834	35139.93	37095.075	39378.028	42120.58	45569.967
0.5	27268.906	28362.106	29565.445	30899.46	32391.227	34077.713	36011.741	38273.63	40996.68	44432.792
1.0	26255.927	27329.116	28511.334	29823.033	31291.207	32952.761	34860.531	37095.075	39790.65	43202.839
1.5	25194.053	26244.363	27402.224	28687.915	30128.227	31759.83	33635.311	35835.041	38493.502	41868.548
2.0	24080.494	25104.678	26234.498	27489.953	28897.479	30493.293	32329.395	34485.417	37095.075	40416.525
2.5	22912.603	23906.987	25004.573	26224.948	27594.007	29147.237	30935.618	33037.32	35583.971	38831.288
3.0	21688.118	22648.551	23709.141	24888.897	26212.942	27715.682	29446.527	31481.242	33947.599	37095.075

Table C.2: Sampling Variance of Horvitz and Thompson Estimator for Population 2

a	*b*									
	-3.0	-2.5	-2.0	-1.50	-1.00	-0.50	0.00	0.50	1.00	1.50
-3.0	1217.8648	1297.1315	1394.9489	1516.3464	1668.0845	1859.4131	2103.2462	2418.0428	2830.9274	3383.095
-2.5	1169.1456	1242.7312	1334.2463	1448.6107	1592.4476	1774.8323	2008.4534	2311.4731	2710.6174	3246.5317
-2.0	1121.5835	1189.1481	1273.9823	1380.8894	1516.3464	1689.2426	1912.026	2202.5426	2587.0922	3105.7334
-1.5	1075.6788	1136.8519	1214.5928	1313.5801	1440.135	1602.9482	1814.211	2091.4307	2460.4517	2960.7038
-1.0	1032.0484	1086.4256	1156.6224	1247.1845	1364.2664	1516.3464	1715.3409	1978.3946	2330.8634	2811.5021
-0.5	991.45399	1038.5926	1100.7518	1182.3352	1289.3185	1429.9524	1615.8583	1863.7918	2198.5838	2658.2627
0.0	954.83833	994.2527	1047.8328	1119.8297	1216.0271	1344.4312	1516.3464	1748.1095	2063.9856	2501.2198
0.5	923.37169	954.52778	998.93303	1060.6743	1145.3289	1260.6397	1417.5694	1632.0029	1927.5944	2340.742
1.0	898.51198	920.82107	955.39435	1006.1417	1078.4179	1179.6813	1320.5255	1516.3464	1790.1371	2177.3768
1.5	882.08318	894.89462	918.90934	957.84598	1016.8185	1102.9774	1226.5163	1402.301	1652.6061	2011.9119
2.0	876.37837	878.97125	891.62191	917.84171	962.48344	1032.3635	1137.2407	1291.4043	1516.3464	1845.4569
2.5	884.29667	875.87021	876.26148	888.75659	917.92424	970.21706	1054.92	1185.6931	1383.1732	1679.5568
3.0	909.52682	889.1893	876.3235	873.97	886.38749	919.63131	982.46837	1087.8695	1255.5338	1516.3464

Table C.3: Sampling Variance of Horvitz and Thompson Estimator for Population 3

a	*b*									
	-3.0	-2.5	-2.0	-1.50	-1.00	-0.50	0.00	0.50	1.00	1.50
-3.0	133945.79	138584.33	143628.74	149147.24	155228.5	161991.82	169605.53	178323.35	188566.3	201144.99
-2.5	131048.58	135663.29	140687.04	146189.38	152260.85	159023.59	166650.39	175402.63	185716.09	198436.08
-2.0	128030.23	132615.64	137612.81	143092.56	149147.24	155901.76	163533.28	172310.78	182684.84	195536.03
-1.5	124883.08	129433.07	134396.99	139846.84	145876.72	152614.14	160240.55	169032.31	179454.7	192423.83
-1.0	121598.9	126106.58	131029.71	136441.35	142437.2	149147.24	156756.95	165549.82	176005.44	189075.27
-0.5	118168.8	122626.43	127500.23	132864.18	138815.41	145486.14	153065.52	161843.72	172313.97	185462.33
0.0	114583.21	118982.07	123796.86	129102.31	134996.73	141614.24	149147.24	157891.87	168353.91	181552.36
0.5	110831.87	115162.13	119906.91	125141.49	130965.04	137513.15	144980.83	153669.2	164094.96	177307.18
1.0	106903.84	111154.37	115816.65	120966.2	126702.64	133162.46	140542.42	149147.24	159502.19	172681.82
1.5	102787.54	111511.31	116559.56	116559.56	122190.09	128539.53	135805.23	144293.66	154535.22	167623.03
2.0	98470.919	102522.54	106975.13	111903.41	117406.23	123619.39	130739.32	139071.68	149147.24	162067.42
2.5	93941.738	97870.582	102191.66	106978.47	112328.26	118374.73	125311.4	133439.65	143283.98	155939.13
3.0	89188.165	92975.873	97144.262	101764.83	106932.17	112776.23	119484.93	127350.75	136882.76	149147.24

Table C.4: Sampling Variance of Horvitz and Thompson Estimator for Population 4

a	*b*									
	-3.0	-2.5	-2.0	-1.50	-1.00	-0.50	0.00	0.50	1.00	1.50
-3.0	26010.88	27181.551	28474.08	29911.752	31525.339	33356.862	35466.155	37943.196	40933.571	44697.824
-2.5	25272.523	26430.933	27710.883	29135.725	30736.355	32555.017	34651.971	37118.044	40100.697	43865.263
-2.0	24505.736	25650.351	26916.032	28326.173	29911.752	31715.191	33797.105	36249.083	39220.333	42980.837
-1.5	23709.291	24838.401	26087.925	27481.259	2904.423	30834.953	32898.736	35332.998	38288.528	42039.726
-1.0	22882.005	23993.703	25224.954	26599.114	28147.189	29911.752	31953.857	34366.21	37300.945	41036.547
-0.5	22022.793	23114.952	24325.556	25677.877	27202.834	28942.946	30959.296	33344.871	36252.841	39965.292
0.0	21130.743	22200.984	23388.282	24715.758	26214.164	27925.847	29911.752	32264.89	35139.056	38819.264
0.5	20205.224	21250.891	22411.898	23711.14	25179.101	26857.812	28807.868	31121.978	33954.028	37591.032
1.0	19246.05	20264.178	21395.546	22662.732	24095.83	25736.376	27644.353	29911.752	32691.852	36272.412
1.5	18253.719	19241.004	20338.983	21569.81	22963.036	24559.487	26418.201	28629.926	31346.428	34854.514
2.0	17229.77	18182.545	19242.95	20432.587	21780.278	23325.871	25127.052	27272.652	29911.752	33327.931
2.5	16177.334	17091.554	18109.744	19252.797	20548.586	22035.649	23769.808	25837.133	28382.494	31683.188
3.0	15101.975	15973.235	16944.126	18034.642	19271.44	20691.349	22347.692	24322.708	26755.083	29911.752

Table C.5: Sampling Variance of Horvitz and Thompson Estimator for Population 5

a	*b*									
	-3.0	-2.5	-2.0	-1.50	-1.00	-0.50	0.00	0.50	1.00	1.50
-3.0	2324.6905	2404.0343	2494.3373	2597.9789	2718.0815	2858.8297	3025.9727	3227.6489	3475.8111	3788.8436
-2.5	2272.5175	2350.4642	2439.2236	2541.1493	2659.3318	2797.9143	2962.5923	3161.4332	3406.2928	3715.4207
-2.0	2218.1861	2294.6427	2381.7547	2481.8483	2597.9789	2734.2453	2896.2847	3092.0869	3333.4021	3638.3318
-1.5	2161.5899	2236.4544	2321.8053	2419.939	2533.873	2667.6578	2826.8673	3019.4067	3256.9101	3557.3165
-1.0	2102.6217	2175.782	2259.2468	2355.2799	2466.8581	2597.9789	2754.1472	2943.1756	3176.5712	3472.0933
-0.5	2041.1755	2112.5084	2193.9492	2287.7267	2396.773	2525.0282	2677.9214	2863.1639	3092.1228	3382.3585
0.0	1977.1501	2046.5192	2125.784	2217.1345	2323.4542	2448.6201	2597.9789	2779.1298	3003.2855	3287.7859
0.5	1910.453	1977.7079	2054.628	2143.3615	2246.7385	2368.567	2514.1025	2690.8211	2909.7642	3188.0262
1.0	1841.0073	1905.981	1980.3695	2066.275	2166.4689	2284.6835	2426.0736	2597.9789	2811.2506	3082.7089
1.5	1768.7608	1831.2678	1902.9167	1985.7592	2082.5022	2196.794	2333.6782	2500.3432	2707.428	2971.4452
2.0	1693.699	1753.5328	1822.2104	1901.7276	1994.7201	2104.7434	2236.7175	2397.6621	2597.9789	2853.8342
2.5	1615.8637	1672.7941	1738.2414	1814.1401	1903.0468	2008.4137	2135.0233	2289.7064	2482.5981	2729.4737
3.0	1535.3803	1589.15	1651.0778	1723.0294	1807.4739	1907.7485	2028.4814	2176.2918	2361.0137	2597.9789

Table C.6: Sampling Variance of Horvitz and Thompson Estimator for Population 6

a	*b*									
	-3.0	-2.5	-2.0	-1.50	-1.00	-0.50	0.00	0.50	1.00	1.50
-3.0	1520.9185	1528.4487	1537.9395	1550.0092	1565.5521	1585.9132	1613.2155	1651.0305	1705.9119	1791.4893
-2.5	1515.473	1522.5495	1531.5385	1543.0482	1557.9602	1577.602	1604.0727	1640.9111	1694.6279	1778.813
-2.0	1509.8516	1516.4375	1524.8826	1535.7843	1550.0092	1568.865	1594.4234	1630.1845	1682.6064	1765.2228
-1.5	1504.0571	1510.1114	1517.9662	1528.2061	1541.6811	1559.6762	1584.231	1618.8004	1669.7779	1750.6208
-1.0	1498.0963	1503.5738	1510.7863	1520.3044	1532.9592	1550.0092	1573.4572	1606.7042	1656.0655	1734.8959
-0.5	1491.9818	1496.8319	1503.3444	1512.0736	1523.829	1539.8389	1562.0629	1593.8381	1641.3844	1717.9221
0.0	1485.7339	1489.9005	1495.6485	1503.5138	1514.2808	1529.1429	1550.0092	1580.1412	1625.6419	1699.5565
0.5	1479.3842	1482.8049	1487.7163	1494.6333	1504.3117	1517.9041	1537.2603	1565.5515	1608.7378	1679.6379
1.0	1472.9798	1475.5853	1479.5798	1485.4542	1493.9307	1506.1151	1523.7869	1550.0092	1590.5669	1657.9862
1.5	1466.5908	1468.3041	1471.2922	1476.0182	1483.1658	1493.7851	1509.5731	1533.4629	1571.0232	1634.4034
2.0	1460.3194	1461.0559	1462.9386	1466.3985	1472.0742	1480.9511	1494.6276	1515.8797	1550.0092	1608.679
2.5	1454.3161	1453.9834	1454.652	1456.7157	1460.7609	1467.696	1479.0022	1497.2652	1527.4549	1580.605
3.0	1448.8025	1447.302	1446.6397	1447.1663	1449.4069	1454.1797	1462.8253	1477.6979	1503.3534	1550.0092

Table C.7: Sampling Variance of Horvitz and Thompson Estimator for Population 7

a	*b*									
	-3.0	-2.5	-2.0	-1.50	-1.00	-0.50	0.00	0.50	1.00	1.50
-3.0	6193.4482	6209.5836	6233.9484	6269.4872	6320.5389	6393.7438	6499.767	6656.8521	6899.0019	7297.8605
-2.5	6180.9107	6194.408	6215.8325	6248.064	6295.3553	6364.2303	6465.1926	6616.2616	6851.1384	7241.1644
-2.0	6168.6942	6179.3539	6197.6095	6226.2685	6269.4872	6333.6573	6429.0985	6573.5691	6800.4085	7180.5514
-1.5	6156.9776	6164.5805	6179.4154	6204.2087	6243.0078	6302.0547	6391.4579	6528.6719	6746.6015	7115.6461
-1.0	6145.993	6150.2982	6161.4345	6182.0374	6216.0311	6269.4872	6352.2712	6481.4823	6689.5029	7046.0367
-0.5	6136.0419	6136.7847	6143.9161	6159.9689	6188.728	6236.0699	6311.5796	6431.9396	6628.9027	6971.2766
0.0	6127.5182	6124.4085	6127.1977	6138.3026	6161.3494	6201.9911	6269.4872	6380.0299	6564.6114	6890.8909
0.5	6120.9408	6113.6613	6111.7378	6117.4554	6134.2592	6167.5448	6226.1931	6325.8164	6496.4849	6804.3906
1.0	6116.9975	6105.2037	6098.1627	6098.0102	6107.9834	6133.1809	6182.0408	6269.4872	6424.4686	6711.3034
1.5	6116.6105	6099.9318	6087.3354	6080.7862	6083.283	6099.5795	6137.5948	6211.4323	6348.6705	6611.2327
2.0	6121.03	6099.0752	6080.4572	6066.9457	6061.265	6067.7679	6093.7624	6152.3697	6269.4872	6503.9705
2.5	6131.9754	6104.3447	6079.2231	6058.1579	6043.5566	6039.3055	6051.9914	6093.5555	6187.825	6389.7138
3.0	6151.8517	6118.1608	6086.0655	6056.8591	6032.5855	6016.5879	6014.6016	6037.149	6105.4995	6269.4872

Table C.8: Sampling Variance of Horvitz and Thompson Estimator for Population 8

a	*b*									
	-3.0	-2.5	-2.0	-1.50	-1.00	-0.50	0.00	0.50	1.00	1.50
-3.0	203237.18	202413.9	201997.26	202169.72	203208.42	205550.2	209917.35	217581.09	230975.68	255348.37
-2.5	203836.63	202771.82	202084.4	201950.37	202638.24	204573.16	208461.23	215551.27	228247.98	251771.51
-2.0	204625.32	203301.97	2023224.05	201860.35	202169.72	203664.06	207031.02	213493.32	225420	247993.71
-1.5	205642.72	204042.4	202752.52	201933.82	201834.27	202850.76	205649.77	211423.58	222498.23	244005.95
-1.0	206937.82	205040.62	203415.5	202214.15	201672.28	202169.72	204348.57	209365.52	219494.72	239801.79
-0.5	208571.88	206356.43	204370.89	202756.82	201736.04	201668.94	203169.48	207352.58	216429.62	235379.4
0.0	210622.23	208065.71	205692.75	203633.44	202093.88	201412.18	202169.72	205432.25	213335.13	230744.79
0.5	213187.38	210265.81	207476.73	204937.35	202835.88	201484.82	201427.75	203672.26	210261.43	225916.9
1.0	216394.32	213082.94	209847.86	206791.7	204082.25	202002.59	201052.19	202169.72	207285.8	220936.05
1.5	220408.72	216682.91	212971.65	209361.02	205995.52	203124.32	201195.32	201065.21	204527.2	215877.94
2.0	225449.9	221286.73	217070.48	212868.55	208798.9	205071.48	202073.91	200565.05	202169.72	210877.39
2.5	231812.95	227193.94	222448.39	217622.65	212804.46	208158.63	204002.22	200977.33	200501.95	206170.23
3.0	239902.52	234818.6	229529.49	224058.1	218458.05	212842.52	207446.35	202772.61	199985.35	202169.72

Table C.9: Sampling Variance of Horvitz and Thompson Estimator for Population 9

a	*b*									
	-3.0	-2.5	-2.0	-1.50	-1.00	-0.50	0.00	0.50	1.00	1.50
-3.0	49656.971	49612.796	49606.343	49651.875	49769.246	49986.526	50344.135	50901.557	51748.646	53025.475
-2.5	49697.551	49631.345	49600.422	49618.609	49705.212	49887.604	50205.3	50716.591	51509.717	52722.522
-2.0	49755.04	49665.504	49608.662	49597.877	49651.875	49797.284	50072.658	50535.014	51270.872	52415.696
-1.5	49832.417	49718.135	49633.796	49592.269	49611.659	49717.803	49948.226	50358.586	51033.568	52106.084
-1.0	49933.226	49792.656	49679.098	49604.896	49587.491	49651.875	49834.474	50189.491	50799.645	51795.117
-0.5	50061.713	49893.167	49748.507	49639.515	49582.923	49602.815	49734.441	50030.446	50571.436	51484.658
0.0	50222.979	50024.609	49846.785	49700.684	49602.278	49574.68	49651.875	49884.832	50351.887	51177.128
0.5	50423.191	50192.968	49979.712	49793.954	49650.846	49572.458	49591.413	49756.876	50144.734	50875.662
1.0	50669.84	50405.529	50154.346	49926.121	49735.128	49602.309	49558.819	49651.875	49954.714	50584.319
1.5	50972.081	50671.218	50379.351	50105.557	49863.16	49671.882	49561.293	49576.496	49787.858	50308.352
2.0	51341.177	51001.034	50665.433	50342.635	50044.934	49790.732	49607.875	49539.174	49651.875	50054.581
2.5	51791.077	51408.633	51025.914	50650.298	50292.96	49970.867	49709.989	49550.642	49556.663	49831.885
3.0	52339.195	51911.091	51477.492	51044.814	50623.013	50227.49	49882.172	49624.642	49515.01	49651.875

Table C.10: Sampling Variance of Horvitz and Thompson Estimator for Population 10

a	*b*									
	-3.0	-2.5	-2.0	-1.50	-1.00	-0.50	0.00	0.50	1.00	1.50
-3.0	612468.25	601219.94	592076.06	586007.44	584387.89	589188.91	603288.71	630977.62	678815.2	757146.37
-2.5	621379.48	608285.5	597091.05	588730.27	584530.88	586405.55	597156.18	620972.3	664277.29	737227.41
-2.0	632143.45	617095.24	603729.15	592940.34	586007.44	584780.52	591980.26	611688.71	650184.02	717421.05
-1.5	645060.22	627939.69	612270.23	598905.43	589071.66	584552.23	587981.11	603325.64	636708.76	697869.87
-1.0	660485.47	641163.89	623047.38	606945.21	594027.9	586007.44	585425.12	596125.54	624065.43	678753.23
-0.5	678842.73	657179.47	636458.91	617442.89	601242.22	589492.51	584635.76	590384.99	612518.52	660296.68
0.0	700638.76	676479.85	652983.26	630860.03	611156.9	595427.57	586007.44	586468.08	602395.94	642784.13
0.5	726483.07	699659.58	673198.18	647755.34	624308.99	604324.72	590023.19	584823.62	594105.58	626573.61
1.0	757112.92	727439.06	697805.14	668808.88	641354.08	616811.35	597277.56	586007.44	588156.82	612117.88
1.5	793425.49	760696.51	727661	694853.3	663097.17	633660.55	608506.44	590711.4	585188.74	599991.48
2.0	836519.86	800509.56	763819.26	726914.74	690533.06	655830.96	624626.18	599801.81	586007.44	590926.75
2.5	887751.99	848209.98	807584.42	766266.68	724899.74	684519.44	646785.52	614370.34	591635.82	585862.09
3.0	948807.97	905456.5	860584.31	814501.85	767749.62	721231.69	676435.36	635802.72	603380.85	586007.44

Table C.11: Sampling Variance of Horvitz and Thompson Estimator for Population 11

a	*b*									
	-3.0	-2.5	-2.0	-1.50	-1.00	-0.50	0.00	0.50	1.00	1.50
-3.0	346.42015	358.329	371.28269	385.45699	401.08021	418.45947	438.02797	460.43825	486.77261	519.11427
-2.5	338.97454	350.82304	363.7245	377.85806	393.45686	410.83528	430.43796	452.93707	479.45281	512.15724
-2.0	331.21557	342.98982	355.82398	369.90044	385.45699	402.81509	422.43063	444.99525	471.66711	504.70869
-1.5	323.12331	334.80769	347.55754	361.55827	377.05213	394.36731	413.97064	436.57277	463.36953	496.71449
-1.0	314.67621	326.25317	338.89948	352.80328	368.21094	385.45699	405.01858	427.62466	454.50781	488.11227
-0.5	305.85103	317.30081	329.82178	343.60448	358.89892	376.04536	395.53045	418.10027	445.0224	478.82981
0.0	296.62268	307.92302	320.29393	333.92792	349.078	366.08937	385.45699	407.94229	434.84513	468.78298
0.5	286.9642	298.08994	310.28274	323.73641	338.70623	355.54118	374.74294	397.0858	423.8977	457.87329
1.0	276.84672	287.76944	299.75216	312.98929	327.73734	344.34759	363.32627	385.45699	412.08973	445.98467
1.5	266.23961	276.92707	288.66329	301.64231	316.12049	332.44953	351.13726	372.97186	399.31664	432.97961
2.0	255.11072	265.52638	276.97452	289.64757	303.80007	319.78162	338.09777	359.53473	385.45699	418.69418
2.5	243.42715	253.5295	264.642	276.95396	290.71589	306.27206	324.12057	345.03686	370.36968	402.93198
3.0	231.15654	240.89842	251.6209	263.50821	276.80406	291.84315	309.10939	329.35556	353.89116	385.45699

Table C.12: Sampling Variance of Horvitz and Thompson Estimator for Population 12

a	*b*									
	-3.0	-2.5	-2.0	-1.50	-1.00	-0.50	0.00	0.50	1.00	1.50
-3.0	344.04725	355.90832	368.80495	382.91092	398.45192	415.73163	435.17843	457.43794	483.58182	515.67551
-2.5	336.64077	348.44144	361.28584	375.35112	390.86781	408.14681	427.62789	449.97617	476.30091	508.75564
-2.0	328.92466	340.65101	353.42812	367.43635	382.91092	400.16973	419.66379	442.07753	468.55789	501.34799
-1.5	320.87936	332.51579	345.2086	359.14119	374.55318	391.76925	411.25124	433.70248	460.30727	493.39897
-1.0	312.48375	324.01273	336.60202	350.43779	365.76377	382.91092	402.35136	424.80662	451.49739	484.84686
-0.5	303.71504	315.11684	327.58089	341.29572	356.5087	373.55653	392.92077	415.33993	442.06941	475.62019
0.0	294.54867	305.8011	318.11524	331.6816	346.75057	363.66372	382.91092	405.24588	431.95599	465.63577
0.5	284.95828	296.03627	308.17254	321.55894	336.44811	353.18541	372.26737	394.46045	421.07981	454.79624
1.0	274.91565	285.79089	297.71749	310.88787	325.55592	342.0693	360.92906	382.91092	409.35174	442.98694
1.5	264.39089	275.03134	286.71202	299.62503	314.02411	330.25739	348.82746	370.51455	396.66865	430.0721
2.0	253.35273	263.72205	275.11549	287.72357	301.79823	317.68553	335.88579	357.17726	382.91092	415.88998
2.5	241.76919	251.82617	262.88517	275.13359	288.81942	304.2834	322.01851	342.79222	367.93971	400.24702
3.0	229.60902	239.30691	249.97753	261.80327	275.02538	289.97513	307.13139	327.23913	351.5944	382.91092

Table C.13: Sampling Variance of Horvitz and Thompson Estimator for Population 13

a	*b*									
	-3.0	-2.5	-2.0	-1.50	-1.00	-0.50	0.00	0.50	1.00	1.50
-3.0	343.96917	356.16727	369.43642	383.95786	399.96743	417.7832	437.85493	460.86214	487.93471	521.24882
-2.5	336.35815	348.48953	361.70006	376.17439	392.15329	409.96221	430.06266	453.15433	480.4063	514.08751
-2.0	328.43292	340.48285	353.61877	368.0286	383.95786	401.73917	421.84555	444.99665	472.4008	506.42124
-1.5	320.17425	332.1262	345.16949	359.49508	375.35296	393.08262	413.16816	436.34871	463.87147	498.19461
-1.0	311.56154	323.39696	336.32726	350.54614	366.30772	383.95786	403.99111	427.16527	454.76531	489.34386
-0.5	302.57277	314.27081	327.06509	341.15169	356.78834	374.32661	394.2706	417.39547	445.02207	479.79524
0.0	293.18446	304.72166	317.35386	331.27902	346.75781	364.14662	383.95786	406.98205	434.57298	469.46302
0.5	283.37181	294.72172	307.16228	320.89268	336.17569	353.37134	372.99851	395.86047	423.33931	458.24706
1.0	273.10886	284.24162	296.45698	309.95452	324.998	341.9495	361.33206	383.95786	411.23073	446.02974
1.5	262.36887	273.25078	285.20276	298.42379	313.17721	329.82504	348.89135	371.19205	398.14336	432.6723
2.0	251.12514	261.71807	273.36325	286.25773	300.66258	316.93715	335.60232	357.47072	383.95786	418.0104
2.5	239.35235	249.61322	260.90216	273.41269	287.40128	303.22111	321.3844	342.69104	368.53756	401.84911
3.0	227.02905	236.90925	247.7858	259.84662	273.34065	288.61043	306.15227	326.74063	351.72786	383.95786

Table C.14: Sampling Variance of Horvitz and Thompson Estimator for Population 14

a	*b*									
	-3.0	-2.5	-2.0	-1.50	-1.00	-0.50	0.00	0.50	1.00	1.50
-3.0	740.40722	772.23221	807.42426	846.65171	890.81344	941.16749	999.56856	1068.9496	1154.4219	1266.2108
-2.5	720.49029	751.89929	786.66786	825.46844	869.20702	919.15426	977.18854	1046.2911	1131.6858	1243.9192
-2.0	699.89808	730.83346	765.11461	803.41685	846.65171	896.10035	953.66199	1022.3623	1107.5328	1220.0406
-1.5	678.62436	709.02166	742.74349	780.46656	823.1059	871.95049	928.91651	997.0686	1081.8386	1194.4075
-1.0	656.67206	686.45918	719.54084	756.59327	798.53248	846.65171	902.87907	970.3107	1054.4675	1166.8311
-0.5	634.05743	663.15376	695.50435	731.78261	772.90236	820.15661	875.47881	941.98644	1025.2734	1137.0984
0.0	610.81592	639.13144	670.64885	706.03584	746.20007	792.42863	846.65171	911.99456	994.10148	1104.971
0.5	587.0107	614.44513	645.01488	679.37839	718.43218	763.45015	816.34817	880.24126	960.79282	1070.1843
1.0	562.74514	589.18732	618.68162	651.87288	689.6403	733.23547	784.54561	846.65171	925.19331	1032.4509
1.5	538.18142	563.50914	591.78642	623.6392	659.92167	701.85159	751.26921	811.19012	887.17138	991.47052
2.0	513.56896	537.64979	564.55534	594.88613	629.46198	669.45233	716.62703	773.8952	846.65171	946.95529
2.5	489.28861	511.98287	537.35175	565.96266	598.58952	636.33564	680.87078	734.94404	803.68033	898.68781
3.0	465.92394	487.09172	510.7565	537.44355	567.86737	603.04346	644.5046	694.77114	758.5538	846.65171

Table C.15: Sampling Variance of Horvitz and Thompson Estimator for Population 15

a	*b*									
	-3.0	-2.5	-2.0	-1.50	-1.00	-0.50	0.00	0.50	1.00	1.50
-3.0	70.665302	73.518125	76.652337	80.119105	83.985334	88.341303	93.313841	99.090797	105.97099	114.47938
-2.5	68.865649	71.695808	74.807366	78.251803	82.096507	86.432541	91.388066	97.153356	104.03213	112.56089
-2.0	66.992298	69.79658	72.88197	76.300209	80.119105	84.430422	89.363541	95.111016	101.98129	110.52223
-1.5	65.041178	67.815993	70.871273	74.258954	78.047186	82.328326	87.23283	92.955333	99.808724	108.35193
-1.0	63.008059	65.74939	68.770127	72.122329	75.874384	80.119105	84.987837	90.677029	97.50363	106.03712
-0.5	60.888594	63.591939	66.573149	69.884306	73.593913	77.79507	82.619767	88.265915	95.053994	103.56323
0.0	58.678394	61.338703	64.274769	67.538582	71.198601	75.347999	80.119105	85.710833	92.446473	100.91384
0.5	56.373147	58.984751	61.869338	65.078664	68.680954	72.769182	77.47563	82.999617	89.666283	98.070409
1.0	53.968803	56.525333	59.351292	62.498028	66.033294	70.049535	74.678486	80.119105	86.697115	95.012003
1.5	51.461869	53.956169	56.715429	59.790379	63.248007	67.179809	71.716352	77.055249	83.521128	91.715097
2.0	48.849858	51.273897	53.957356	56.950089	60.317959	64.150993	68.577806	73.793396	80.119105	88.153457
2.5	46.132004	48.476794	51.074215	53.972923	57.237221	60.955021	65.251993	70.318903	76.470924	84.298308
3.0	43.310394	45.565937	48.065861	50.857251	54.002303	57.586021	61.729882	66.618349	72.556662	80.119105

Table C.16: Sampling Variance of Horvitz and Thompson Estimator for Population 16

a	*b*									
	-3.0	-2.5	-2.0	-1.50	-1.00	-0.50	0.00	0.50	1.00	1.50
-3.0	2499.6598	2507.1103	2517.1645	2530.6835	2548.914	2573.7285	2608.0675	2656.8296	2728.8449	2841.925
-2.5	2494.0312	2500.7858	2510.0692	2522.7266	2539.9844	2563.687	2596.7364	2643.9757	2714.156	2824.983
-2.0	2488.2943	2494.2999	2502.752	2514.4787	2530.6835	2553.1792	2584.8242	2630.3986	2698.5614	2806.8898
-1.5	2482.4679	2487.6657	2495.2198	2505.9391	2521.0013	2542.1842	2572.2963	2616.0454	2681.9833	2787.5315
-1.0	2476.5786	2480.9036	2487.4858	2497.1124	2510.9323	2530.6835	2559.1179	2600.8604	2664.3369	2766.7803
-0.5	2470.6634	2474.0437	2479.5717	2488.0106	2500.4767	2518.6628	2545.2564	2584.7862	2645.5308	2744.4937
0.0	2464.7732	2467.129	2471.5116	2478.6568	2489.6443	2506.1154	2530.6835	2567.7661	2625.4677	2720.5132
0.5	2458.978	2460.221	2463.3565	2469.0899	2478.4587	2493.0458	2515.3791	2549.7466	2604.0468	2694.6642
1.0	2453.3737	2453.406	2455.1816	2459.3711	2466.9643	2479.4769	2499.3378	2530.6835	2581.1673	2666.757
1.5	2448.0926	2446.8059	2447.0964	2449.5948	2455.2366	2465.4597	2482.5782	2510.5504	2556.7365	2636.5913
2.0	2443.3171	2440.5927	2439.2596	2439.9035	2443.3974	2451.0892	2465.1592	2489.3547	2530.6835	2603.9663
2.5	2439.3023	2435.0104	2431.9023	2430.5118	2431.6397	2436.5298	2447.2051	2467.1627	2502.9842	2568.7019
3.0	2436.4078	2430.4096	2425.3631	2421.7427	2420.266	2422.0554	2428.949	2444.1454	2473.7082	2530.6835

Table C.17: Sampling Variance of Horvitz and Thompson Estimator for Population 17

a	*b*									
	-3.0	-2.5	-2.0	-1.50	-1.00	-0.50	0.00	0.50	1.00	1.50
-3.0	3390018.7	3387628.4	3388900	3395325.9	3409124.9	3433713.5	3474588.7	3541120.3	3650568.2	3838413.4
-2.5	3390727.5	3386737.8	3386223	3390634.7	3402137.4	3424073.4	3461836.2	3524647.1	3629552.2	3811746.2
-2.0	3392178	3386476.7	3384046.9	3386295.1	3395325.9	3414398.7	3448790.8	3507555.3	3607491.8	3783451.2
-1.5	3394535.6	3387000.1	3382514.3	3382435.2	3388800.2	3404776.5	3435510.4	3489863.5	3584350.1	3753408.5
-1.0	3398005.6	3388502	3381806.1	3379219.4	3382704.7	3395325.9	3422080.7	3471612.4	3560103.7	3721496.6
-0.5	3402845	3391227.7	3382152.8	3376860.5	3377230.1	3386208.7	3408626.2	3452875	3534751.2	3687598.4
0.0	3409378.4	3395488.5	3383850.4	3375635.2	3372628.7	3377646.1	3395325.9	3433771.5	3508326.4	3651610
0.5	3418018.5	3401684.5	3387282.5	3375907	3369237.8	3369940.8	3382435.5	3414491.3	3480918.3	3613457.7
1.0	3429296.6	3410333.8	3392951.4	3378157.2	3367511.6	3363510.3	3370321.4	3395325.9	3452703.4	3573124.4
1.5	3443903.4	3422116.3	3401522	3383030.9	3368068.1	3358935	3359509.7	3376719.7	3423995.3	3530695.6
2.0	3462749.6	3437934.7	3413886	3391403.7	3371759.2	3357031.5	3350763	3359349.5	3395325.9	3486437.2
2.5	3487052.8	3459007	3431258.2	3404482.8	3379777.2	3358965.2	3345200	3344251.4	3367579.6	3440932.3
3.0	3518470.3	3487004.5	3455323.1	3423962.5	3393820.8	3366429	3344488.7	3333031.3	3342222.5	3395325.9

Table C.18: Sampling Variance of Horvitz and Thompson Estimator for Population 18

a	*b*									
	-3.0	-2.5	-2.0	-1.50	-1.00	-0.50	0.00	0.50	1.00	1.50
-3.0	61.173203	64.492704	68.212537	72.419025	77.229169	82.807319	89.394985	97.368318	107.36011	120.55274
-2.5	59.091945	62.348312	66.000996	70.135902	74.869591	80.366029	86.866593	94.748137	104.64684	117.7578
-2.0	56.950463	60.137284	63.715669	67.770945	72.419025	77.823204	84.224315	91.999325	101.78692	114.79368
-1.5	54.749939	57.860124	61.356272	65.322945	69.875162	75.175195	81.462844	89.114447	98.770073	111.64608
-1.0	52.492831	55.518535	58.923626	62.791687	67.236542	72.419025	78.577294	86.086146	95.585597	108.29951
-0.5	50.183346	53.115889	56.42013	60.178409	64.503002	69.552803	75.563586	82.907478	92.222611	104.73733
0.0	47.828097	50.657888	53.850406	57.486446	61.676307	66.576337	72.419025	79.57242	88.670456	100.94194
0.5	45.437028	48.153487	51.222238	54.722162	58.761074	63.492044	69.143176	76.076691	84.919387	96.89526
1.0	43.024707	45.616214	48.547897	51.896292	55.766126	60.306303	65.739196	72.419025	80.961723	92.579527
1.5	40.612195	43.066067	45.846085	49.025928	52.706507	57.031497	62.215888	68.603206	76.793768	87.97894
2.0	38.229761	40.532312	43.144808	46.137469	49.606532	53.689152	58.590923	64.641325	72.419025	83.082536
2.5	35.920919	38.057659	40.485714	43.271146	46.504493	50.314855	54.895977	60.559134	67.853676	77.889506
3.0	33.748558	35.704661	37.930799	40.488061	43.460106	46.966151	51.185143	56.405025	63.136119	72.419025

Table C.19: Sampling Variance of Horvitz and Thompson Estimator for Population 19

a	*b*									
	-3.0	-2.5	-2.0	-1.50	-1.00	-0.50	0.00	0.50	1.00	1.50
-3.0	56346.802	56313.149	56297.789	56308.164	56354.86	56453.229	56626.083	56908.308	57355.249	58058.847
-2.5	56367.24	56324.123	56298.199	56296.695	56329.923	56412.875	56567.879	56829.158	57251.112	57924.291
-2.0	56393.604	56340.457	56303.332	56289.229	56308.164	56374.747	56510.784	56749.786	57145.139	57785.896
-1.5	56426.93	56363.139	56314.122	56286.635	56290.38	56339.551	56455.401	56670.673	57037.658	57643.795
-1.0	56468.459	56393.356	56331.694	56289.967	56277.541	56308.164	56402.492	56592.439	56929.118	57498.217
-0.5	56519.684	56432.539	56357.409	56300.508	56270.838	56281.67	56353.012	56515.885	56820.124	57349.518
0.0	56582.409	56482.423	56392.927	56319.83	56271.742	56261.417	56308.164	56442.039	56711.486	57198.229
0.5	56658.824	56545.128	56440.281	56349.867	56282.071	56249.089	56269.474	56372.228	56604.285	57045.111
1.0	56751.611	56623.25	56501.975	56393.014	56304.095	56246.806	56238.877	56308.164	56499.954	56891.238
1.5	56864.075	56720.004	56581.116	56452.257	56340.657	56257.242	56218.846	56252.072	56400.4	56738.105
2.0	57000.315	56839.387	56681.588	56531.347	56395.349	56283.802	56212.557	56206.847	56308.164	56587.787
2.5	57165.461	56986.42	56808.285	56635.029	56472.745	56330.851	56224.124	56176.287	56226.643	56443.15
3.0	57365.991	57167.464	56967.431	56769.369	56578.72	56404.036	56258.914	56165.41	56160.403	56308.164

Table C.20: Sampling Variance of Horvitz and Thompson Estimator for Population 20

a	*b*									
	-3.0	-2.5	-2.0	-1.50	-1.00	-0.50	0.00	0.50	1.00	1.50
-3.0	5931.6462	5983.7397	6049.0544	6132.1487	6239.684	6381.6978	6573.876	6841.8397	7229.7898	7819.4832
-2.5	5897.3899	5945.7065	6006.693	6084.7774	6186.446	6321.4956	6505.2725	6762.9015	7137.8298	7710.6137
-2.0	5863.6799	5907.9019	5964.195	6036.8453	6132.1487	6259.64	6434.2951	6680.6973	7041.4708	7595.8587
-1.5	5831.063	5870.8314	5922.0184	5988.7553	6077.13	6196.3907	6361.1082	6595.273	6940.6069	7474.9107
-1.0	5800.2784	5835.1858	5880.7982	5941.0774	6021.8828	6132.1487	6286.0004	6506.7769	6835.2066	7347.4998
-0.5	5772.3277	5801.9094	5841.4127	5894.6127	5967.1174	6067.5147	6209.4391	6415.5097	6725.3564	7213.4271
0.0	5748.5739	5772.2969	5805.0778	5850.4849	5913.8486	6003.3731	6132.1487	6321.9961	6611.3238	7072.6173
0.5	5730.8825	5748.132	5773.4823	5810.2727	5863.5241	5941.0142	6055.2271	6227.0925	6493.6545	6925.2002
1.0	5721.8275	5731.8896	5748.987	5776.2026	5818.2121	5882.3154	5980.3177	6132.1487	6373.3188	6771.6378
1.5	5724.9957	5727.036	5734.9184	5751.4382	5780.8824	5830.0133	5909.8702	6039.2541	6251.9395	6612.9252
2.0	5745.4455	5738.4815	5736.0158	5740.5178	5755.8348	5788.1198	5847.5426	5951.6203	6132.1487	6450.9125
2.5	5790.4132	5773.2799	5759.1207	5750.0343	5749.3663	5762.5738	5798.8322	5874.1868	6018.1592	6288.8279
3.0	5870.4295	5841.7342	5814.2719	5789.7158	5770.8346	5762.2842	5772.0928	5814.6037	5916.701	6132.1487

Table C.21: Sampling Variance of Horvitz and Thompson Estimator for Population 21

a	*b*									
	-3.0	-2.5	-2.0	-1.50	-1.00	-0.50	0.00	0.50	1.00	1.50
-3.0	11495.608	11507.078	11527.611	11560.676	11611.368	11687.41	11800.95	11972.008	12235.571	12657.638
-2.5	11494.023	11500.344	11515.082	11541.571	11584.734	11652.065	11755.409	11914.38	12163.425	12567.882
-2.0	11496.909	11497.709	11506.222	11525.634	11560.676	11618.587	11710.87	11856.677	12089.832	12474.881
-1.5	11505.329	11500.202	11502.02	11513.805	11540.073	11587.777	11668.036	11799.472	12015.189	12378.784
-1.0	11520.613	11509.116	11503.725	11507.278	11524.05	11560.676	11627.836	11743.548	11940.075	12279.881
-0.5	11544.434	11526.085	11512.922	11507.579	11514.062	11538.641	11591.502	11689.968	11865.321	12178.662
0.0	11578.909	11553.186	11531.639	11516.673	11511.991	11523.453	11560.676	11640.183	11792.105	12075.906
0.5	11626.747	11593.088	11562.493	11537.111	11520.303	11517.462	11537.555	11596.176	11722.101	11972.816
1.0	11691.438	11649.244	11608.889	11572.235	11542.253	11523.806	11525.109	11560.676	11657.685	11871.218
1.5	11777.537	11726.18	11675.313	11626.471	11582.185	11546.706	11527.386	11537.475	11602.249	11773.863
2.0	11891.05	11829.893	11767.738	11705.752	11645.963	11591.922	11549.977	11531.894	11560.676	11684.9
2.5	12040.013	11968.44	11894.234	11818.143	11741.622	11667.415	11600.702	11551.498	11540.081	11610.62
3.0	12235.34	12152.82	12065.877	11974.786	11880.348	11784.376	11690.688	11607.226	11550.983	11560.676

Table C.22: Sampling Variance of Horvitz and Thompson Estimator for Population 22

a	*b*									
	-3.0	-2.5	-2.0	-1.50	-1.00	-0.50	0.00	0.50	1.00	1.50
-3.0	30994.111	31775.79	32783.925	34090.16	35795.915	38048.925	41072.08	45216.791	51069.524	59686.801
-2.5	30568.776	31259.984	32166.637	33358.216	34933.377	37036.275	39885.108	43825.087	49434.548	57759.991
-2.0	30194.179	30788.694	31586.715	32655.339	34090.16	36031.349	38691.833	42409.786	47754.154	55759.436
-1.5	29884.961	30376.019	31057.612	31994.205	33277.994	35044.696	37501.31	40978.001	46032.845	53685.996
-1.0	29659.439	30039.692	30596.365	31391.006	32512.028	34090.16	36325.712	39539.711	44277.627	51542.466
-0.5	29540.646	29802.131	30224.646	30866.499	31811.883	33185.928	35181.365	38108.772	42498.961	49334.438
0.0	29557.718	29691.831	29970.166	30447.425	31203.077	32355.944	34090.16	36704.307	40712.11	47071.547
0.5	29747.749	29745.243	29868.577	30168.426	30718.96	31631.878	33081.52	35352.658	38939.049	44769.263
1.0	30158.325	30009.351	29966.085	30074.703	30403.412	31055.845	32195.152	34090.16	37211.204	42451.539
1.5	30851.021	30545.223	30323.078	30225.716	30314.622	30684.266	31484.985	32967.126	35573.47	40154.747
2.0	31906.287	31433.02	31019.247	30700.436	30530.477	30593.414	31024.86	32053.691	34090.16	37933.643
2.5	33430.415	32779.142	32160.942	31604.927	31156.399	30887.516	30916.943	31448.534	32854.029	35870.593
3.0	35565.637	34726.641	33891.94	33083.511	32336.955	31710.872	31304.399	31292.179	32000.237	34090.16

TableC.23: Sampling Variance of Horvitz and Thompson Estimator for Population 23

a	*b*									
	-3.0	-2.5	-2.0	-1.50	-1.00	-0.50	0.00	0.50	1.00	1.50
-3.0	76252.878	76880.204	77859.141	79321.036	81459.567	84571.024	89130.609	95949.6	106536.43	124055.32
-2.5	76051.927	76492.224	77260.437	78483.463	80349.518	83148.289	87347.44	93751.23	103868.01	120897.63
-2.0	76004.15	76244.062	76785.67	77750.731	79321.036	81778.318	85580.736	91522.598	101107.78	117565.88
-1.5	76147.389	76172.638	76470.538	77156.893	78405.953	80489.889	83854.98	89281.974	98264.673	114054.24
-1.0	76529.028	76324.475	76360.361	76745.608	77645.6	79321.036	82203.451	87055.577	95353.994	110360.22
-0.5	77208.801	76758.589	76513.073	76573.204	77093.953	78322.268	80671.471	84880.77	92400.393	106486.94
0.0	78262.589	77550.419	77003.287	76712.906	76822.014	77561.09	79321.036	82810.705	89442.307	102446.85
0.5	79787.61	78797.242	77927.948	77260.741	76923.984	77128.45	78237.513	80921.194	86538.725	98267.592
1.0	81909.627	80625.73	79414.286	78343.925	77526.128	77148.115	77539.534	79321.036	83779.678	94001.668
1.5	84793.126	83202.702	81631.215	80132.989	78799.73	77790.551	77393.884	78167.928	81302.9	89742.667
2.0	88655.925	86750.685	84805.98	82859.697	80980.466	79293.985	78038.506	77693.613	79321.036	85653.189
2.5	93790.627	91570.959	89249.067	86844.133	84398.093	81997.217	79819.051	78244.87	78167.588	82014.749
3.0	100596.82	98078.509	95392.415	92536.839	89523.356	86392.399	83249.013	80352.915	78377.782	79321.036

Table C.24: Sampling Variance of Horvitz and Thompson Estimator for Population 24

a	*b*									
	-3.0	-2.5	-2.0	-1.50	-1.00	-0.50	0.00	0.50	1.00	1.50
-3.0	744166.3	755365.66	767822.96	781760.69	797458.98	815276.37	835680.9	859298.27	886989.62	919984.4
-2.5	737357.84	748346.24	760579.13	774277.6	789720.11	807263.28	827372.99	850672.42	878019.85	910641.64
-2.0	730391.75	741155.44	753149	766591.88	781760.69	799009.9	818802.83	841759.72	868735.43	900952.47
-1.5	723269.36	733793.5	745531.55	758701.18	773576.86	790510.6	809962.82	832550.3	859123.84	890901.2
-1.0	715993.61	726262.15	737727.2	750604.39	765165.83	781760.69	800846.07	823034.75	849172.71	880471.92
-0.5	708569.5	718565.08	729738.16	742302.1	756526.31	772756.79	791446.79	813204.46	838870.1	869648.76
0.0	701004.66	710708.46	721569	733797.07	747659.04	763497.3	781760.69	803052.01	828204.93	858416.18
0.5	693310.02	702701.67	713227.33	725094.91	738567.39	753983.01	771785.61	792571.79	817167.44	846759.44
1.0	685500.77	694558.17	704724.69	716205.02	7292258.27	744218.01	761522.34	781760.69	805749.94	834665.26
1.5	677597.47	686296.67	696077.7	707141.63	719743.22	734210.7	750975.65	770619.23	793947.78	822122.7
2.0	669627.64	677942.66	687309.58	697925.39	710039.9	723975.34	740155.76	759152.86	781760.69	809124.43
2.5	661627.66	669530.37	678452.11	688585.26	700174.08	713533.93	729080.29	747373.92	769194.62	795668.47
3.0	653645.4	661105.39	669548.25	679161.21	690182.23	702918.88	717776.81	735304.19	756264.26	781760.69

Table C.25: Sampling Variance of Horvitz and Thompson Estimator for Population 25

a	*b*									
	-3.0	-2.5	-2.0	-1.50	-1.00	-0.50	0.00	0.50	1.00	1.50
-3.0	46108708	46531138	47009364	47554203	48179612	48904012	49752390	50759812	51977605	53485084
-2.5	45855673	46265555	46730248	47260417	47869825	48576650	49405560	50391192	51584289	53063350
-2.0	45600279	45996872	46447216	46961808	47554203	48242308	49050440	50012761	51179367	52627841
-1.5	45343216	45725727	46160846	46658887	47233177	47901326	48687262	49624624	50762783	52178294
-1.0	45085360	45452939	45871893	46352337	46907346	47554203	48316407	49227017	50334597	51714545
-0.5	44827815	45179553	45581334	46043056	46577515	47201638	47938447	48820358	49895030	51236564
0.0	44571969	44906892	45290420	45732212	46244758	46844589	47554203	48405295	49444520	50744511
0.5	44319565	44636635	45000755	45421320	45910486	46484347	47164815	47982789	48983792	50238808
1.0	44072791	44370903	44714382	45112335	45576547	46122629	46771845	47554203	48513960	49720232
1.5	43834405	44112387	44433915	44807777	45245352	45761712	46377404	47121440	48036653	49190048
2.0	43607882	43864498	44162690	44510894	44920037	45404599	45984326	46687116	47554203	48650195
2.5	43397625	43631585	43904986	44225880	44604695	45055253	45596407	46254807	47069886	48103532
3.0	43209236	43419206	43666310	43958174	44304676	44718906	45218723	45829380	46588275	47554203

Table C.26: Sampling Variance of Horvitz and Thompson Estimator for Population 26

a	*b*									
	-3.0	-2.5	-2.0	-1.50	-1.00	-0.50	0.00	0.50	1.00	1.50
-3.0	3044.6963	3024.9078	3009.7337	3000.8319	3000.4673	3011.7891	3039.2704	3089.4322	3172.0914	3302.6331
-2.5	3059.7711	3036.5147	3017.552	3004.4909	2999.534	3005.7525	3027.5196	3071.2264	3146.5162	3268.5367
-2.0	3077.8893	3050.9938	3028.0527	3010.62	3000.8319	3001.6759	3017.4181	3054.3094	3121.8058	3234.7974
-1.5	3099.4668	3068.7503	3041.6282	3019.5973	3004.7227	2999.9019	3009.2857	3038.9739	3098.2198	3201.6337
-1.0	3124.984	3090.2525	3058.7338	3031.8629	3011.6286	3000.8319	3003.4993	3025.5674	3076.0698	3169.3123
-0.5	3154.9971	3116.0445	3079.8991	3047.9302	3022.0443	3004.9384	3000.505	3014.5041	3055.7305	3138.1584
0.0	3190.1531	3146.7602	3105.7433	3068.4009	3036.5511	3012.779	3000.8319	3006.2784	3037.6538	3108.57
0.5	3231.2067	3183.1407	3136.9918	3093.9823	3055.8351	3025.0142	3005.11	3001.4832	3022.3863	3081.0343
1.0	3279.0427	3226.0567	3174.4989	3125.5097	3080.709	3042.43	3014.0928	3000.8319	3010.591	3056.1504
1.5	3334.7021	3276.5352	3219.275	3163.9737	3112.1402	3065.9655	3028.6851	3005.1867	3003.0764	3034.6565
2.0	3399.4164	3335.7933	3272.5207	3210.5552	3151.2854	3096.7485	3049.9787	3015.5945	3000.8319	3017.4667
2.5	3474.6491	3405.2813	3335.6701	3266.6687	3199.5356	3136.1405	3079.2984	3033.3335	3005.0749	3005.7179
3.0	3562.1499	3486.7369	3410.4461	3334.019	3258.5727	3185.7949	3118.2611	3059.973	3017.3124	3000.8319

Table C.27: Sampling Variance of Horvitz and Thompson Estimator for Population 27

a	*b*									
	-3.0	-2.5	-2.0	-1.50	-1.00	-0.50	0.00	0.50	1.00	1.50
-3.0	21687.271	22245.23	22905.969	23694.397	24643.737	25799.323	27224.788	29012.688	31304.098	34328.47
-2.5	21322.732	21857.912	22494.153	23256.235	24177.272	25302.553	26695.799	28449.941	30707.068	33699.137
-2.0	20950.663	21461.219	22070.894	22804.3	23694.397	24786.368	26143.951	27860.374	30078.64	33033.135
-1.5	20571.966	21055.87	21636.701	22338.853	23195.075	24250.375	25568.422	27242.628	29416.791	32327.595
-1.0	20187.912	20642.939	21192.411	21860.452	22679.531	23694.397	24968.54	26595.419	28719.458	31579.446
-0.5	19800.262	20223.969	20739.309	21370.072	22148.367	23118.577	24343.88	25917.611	27984.592	30785.447
0.0	19411.419	19801.125	20279.276	20869.251	21602.7	22523.516	23694.397	25208.335	27210.256	29942.233
0.5	19024.636	19377.407	19815.004	20360.305	21044.382	21910.482	23020.622	24467.168	26394.77	29046.417
1.0	18644.305	18956.939	19350.29	19846.621	20476.286	21281.701	22323.94	23694.397	25536.944	28094.767
1.5	18276.341	18545.368	18890.438	19333.071	19902.733	20640.772	21607.011	22891.421	24636.449	27084.501
2.0	17928.731	18150.424	18442.843	18826.61	19330.093	19993.295	20874.39	22061.364	23694.397	26013.8
2.5	17612.298	17782.716	18017.812	18337.126	18767.678	19347.78	20133.469	21210.029	22714.278	24882.664
3.0	17341.8	17456.881	17629.774	17878.714	18229.071	18717.044	19395.928	20347.396	21703.483	23694.397

Table C.28: Sampling Variance of Horvitz and Thompson Estimator for Population 28

a	*b*									
	-3.0	-2.5	-2.0	-1.50	-1.00	-0.50	0.00	0.50	1.00	1.50
-3.0	18010.649	17883.529	17821.648	17848.03	17994.693	18307.222	18852.458	19732.182	21109.373	23263.85
-2.5	18115.455	17947.341	17840.782	17818.332	17911.478	18165.213	18645.745	19454.252	20753.334	22823.209
-2.0	18253.399	18041.572	17887.248	17812.445	17848.03	18038.294	18448.669	19179.544	20392.891	22369.011
-1.5	18430.41	18171.937	17966.508	17835.522	17809.126	17930.778	18264.965	18911.062	20030.116	21902.101
-1.0	18653.604	18345.332	18085.185	17893.853	17800.644	17848.03	18099.353	18652.705	19667.837	21423.891
-0.5	18931.571	18570.116	18251.354	17995.156	17829.858	17796.762	17957.828	18409.538	19309.898	20936.576
0.0	19274.733	18856.48	18474.912	18148.958	17905.815	17785.41	17848.03	18188.159	18961.501	20443.441
0.5	19695.813	19216.928	18768.071	18367.087	18039.846	17824.645	17779.762	17997.203	18629.689	19949.301
1.0	20210.453	19666.903	19146.002	18664.339	18246.236	17928.063	17765.676	17848.03	18324.015	19461.12
1.5	20838.019	20225.628	19627.698	19059.365	18543.141	18113.116	17822.235	17755.693	18057.504	18988.929
2.0	21602.678	20917.22	20237.134	19575.877	18953.844	18402.422	17971.051	17740.309	17848.03	18547.167
2.5	22534.86	21772.216	21004.868	20244.324	19508.501	18825.583	18240.793	17829.039	17720.35	18156.717
3.0	23673.244	22829.672	21970.247	21104.221	20246.613	19421.808	18669.956	18059.003	17709.15	17848.03

Table C.29: Sampling Variance of Horvitz and Thompson Estimator for Population 29

a	b									
	-3.0	-2.5	-2.0	-1.50	-1.00	-0.50	0.00	0.50	1.00	1.50
-3.0	21125180	21134997	21144850	21154739	21164666	21174629	21184629	21194667	21204742	21214856
-2.5	21120243	21130054	21139901	21149784	21159704	21169661	21179655	21189687	21199756	21209864
-2.0	21115302	21125107	21134948	21144825	21154739	21164690	21174678	21184704	21194767	21204869
-1.5	21110358	21120157	21129992	21139864	21149772	21159716	21169698	21179718	21189775	21199870
-1.0	21105412	21115204	21125033	21134899	21144801	21154739	21164715	21174729	21184780	21194868
-0.5	21100462	21110249	21120071	21129931	21139826	21149759	21159729	21169736	21179781	21189863
0.0	21095509	21105289	21115106	21124959	21134849	21144776	21154739	21164740	21174779	21184855
0.5	21090553	21100327	21110138	21119985	21129869	21139789	21149747	21159741	21169774	21179844
1.0	21085593	21095362	21105167	21115008	21124885	21134799	21144751	21154739	21164766	21174830
1.5	21080631	21090394	21100192	21110027	21119898	21129807	21139752	21149734	21159754	21169812
2.0	21075665	21085422	21095215	21105043	21114908	21124810	21134749	21144726	21154739	21164791
2.5	21070697	21080447	21090234	21100056	21109915	21119811	21129744	21139714	21149721	21159767
3.0	21065725	21075469	21085250	21095066	21104919	21114809	21124735	21134699	21144700	21154739

Table C.30: Sampling Variance of Horvitz and Thompson Estimator for Population 30

a	*b*									
	-3.0	-2.5	-2.0	-1.50	-1.00	-0.50	0.00	0.50	1.00	1.50
-3.0	553818926	554007532	554197145	554387772	554579421	554772102	554965823	555160592	555356419	555553311
-2.5	553724119	553912473	554101831	554292203	554483597	554676020	554869483	555063993	555259559	555456189
-2.0	553629314	553817414	554006518	554196634	554387772	554579938	554773142	554967392	555162697	555359066
-1.5	553534510	553722356	553911206	554101066	554291946	554483855	554676800	554870790	555065833	555261940
-1.0	553439708	553627300	553815894	554005499	554196122	554387772	554580457	554774187	554968969	555164813
-0.5	553344908	553532246	553720584	553909932	554100297	554291689	554484114	554677583	554872103	555067684
0.0	553250110	553437193	553625276	553814366	554004474	554195606	554387772	554580979	554775237	554970555
0.5	553155315	553342143	553529969	553718802	553908651	554099524	554291429	554484375	554678371	554873424
1.0	553060523	553247095	55343665	553623240	553812831	554003443	554195088	554387772	554581504	554776293
1.5	552965734	553152050	553339363	553527680	553717011	553907364	554098747	554291169	554484637	554679162
2.0	552870949	553057008	553244064	553432123	553621194	553811287	554002408	554194567	554387772	554582032
2.5	552776168	552961970	553148768	553336568	553525380	553715211	553906070	554097966	554290906	554484901
3.0	552681390	552866936	553053475	553241016	553429568	553619138	553809734	554001366	554194043	554387772

Table C.31: Sampling Variance of Horvitz and Thompson Estimator for Population 31

a	*b*									
	-3.0	-2.5	-2.0	-1.50	-1.00	-0.50	0.00	0.50	1.00	1.50
-3.0	703578015	703851770	704127806	704406144	704686806	704969814	705255190	705542956	705833135	706125749
-2.5	703440701	703713682	703988940	704266497	704546375	704828595	705113179	705400150	705689531	705981344
-2.0	703303593	703575798	703850277	704127052	704406144	704687575	704971368	705257543	705546125	705837135
-1.5	703166692	703438120	703711820	703987811	704266117	704546758	704829757	705115136	705402917	705693124
-1.0	703030000	703300650	703573569	703848776	704126294	704406144	704688349	704972929	705259909	705549311
-0.5	702893517	703163389	703435526	703709948	7033986677	704265735	704547144	704830926	705117103	705405699
0.0	702757246	703026338	703297692	703571328	703847267	704125532	704406144	704689126	704974500	705262289
0.5	702621188	702889500	703160069	703432917	703708066	703985537	704265351	704547532	704832101	705119082
1.0	702485344	702752874	703022658	703294718	703569075	703845750	704124766	704406144	704689908	704976079
1.5	702349716	702616463	702885461	703156731	703430295	703706174	703984390	704264965	704547922	704833283
2.0	702214306	702480268	702748479	703018959	703291728	703566810	703844225	704123996	704406144	704690694
2.5	702079113	702344291	702611714	702881402	703153376	703427659	703704272	703983237	704264577	704548314
3.0	701944141	702208533	702475166	702744061	703015240	703288723	703564533	703842692	704123221	704406144

Table C.32: Sampling Variance of Horvitz and Thompson Estimator for Population 32

a	*b*									
	-3.0	-2.5	-2.0	-1.50	-1.00	-0.50	0.00	0.50	1.00	1.50
-3.0	763734982	763759342	763784105	763809272	763834846	763860830	763887226	763914035	763941262	763968908
-2.5	763722848	763747037	763771628	763796622	763822024	763847834	763874056	763900692	763927744	763955216
-2.0	763710786	763734803	763759222	763784044	763809272	763834909	763860957	763887419	763914297	763941593
-1.5	763698796	763722641	763746887	763771536	763796592	763822055	763847929	763874217	763900920	763928041
-1.0	763686878	763710551	763734624	763759101	763783982	763809272	763834972	763861085	763887613	763914559
-0.5	763675033	763698533	763722433	763746737	763771445	763796561	763822086	763848025	763874378	763901148
0.0	763663260	763686587	763710315	763734445	763758979	763783921	763809272	763835035	763861213	763887808
0.5	763651559	763674714	763698269	763722225	763746585	763771353	763796530	763822118	763848120	763874539
1.0	763639932	763662914	763686295	763710078	763734264	763758857	763783859	763809272	763835099	763861342
1.5	763628378	763651187	763674395	763698003	763722016	763746434	763771261	763796498	763822149	763848216
2.0	763616898	763639533	763662567	763686002	763709840	763734083	763758735	763783797	763809272	763835163
2.5	763605491	763627953	763650813	763674074	763697737	763721805	763746282	763771168	763796467	763822181
3.0	763594159	763616447	763639133	763662219	763685708	763709601	763733902	763758612	763783735	763809272

Table C.33: Sampling Variance of Horvitz and Thompson Estimator for Population 33

a	*b*									
	-3.0	-2.5	-2.0	-1.50	-1.00	-0.50	0.00	0.50	1.00	1.50
-3.0	6375459.9	6383551.5	6391684	6399857.6	6408073	6416330.7	6424631.1	6432974.9	6441362.5	6449794.4
-2.5	6371380.8	6379465.9	6387591.6	6395758.6	6403967.2	6412218	6420511.6	6428848.4	6437229.1	6445654
-2.0	6367298.1	6375376.5	6383495.6	6391655.8	6399857.6	6408101.6	6416388.3	6424718.2	6433091.9	6441509.8
-1.5	6363211.8	6371283.5	6379395.8	6387549.3	6395744.3	6403981.5	6412261.3	6420584.3	6428950.9	6437361.8
-1.0	6359121.8	6367186.8	6375292.4	6383439.1	6391627.3	6399857.6	6408130.5	6416446.5	6424806.1	6433210
-0.5	6355028.1	6363086.5	6371185.3	6379325.2	6387506.6	6395730	6403995.9	6412304.9	6420657.5	6429054.3
0.0	6350930.8	6358982.5	6367074.6	6375207.6	6383382.2	6391598.7	6399857.6	6408159.6	6416505.2	6424894.8
0.5	6346829.9	6354874.8	6362960.1	6371086.3	6379254	6387463.5	6395715.6	6404010.5	6412349	6420731.5
1.0	6342725.3	6350763.4	6358841.9	6366961.3	6375122.1	6383324.7	6391569.7	6399857.6	6408189	6416564.4
1.5	6338617.1	6346648.4	6354720.1	6362832.6	6370986.5	6379182.1	6387420.1	6395701	6404025.2	6412393.4
2.0	6334505.2	6342529.7	6350594.6	6358700.2	6366847.1	6375035.8	6383266.7	6391540.5	6399857.6	6408218.6
2.5	6330389.7	6338407.4	6346465.4	6354564.1	6362704.1	6370885.7	6379109.6	6387376.3	6395686.2	6404040
3.0	6326270.5	6334281.4	6342332.5	6350424.3	6358557.3	6366731.9	6374948.7	6383208.3	6391511.1	6399857.6

Table C.34: Sampling Variance of Horvitz and Thompson Estimator for Population 34

a	b									
	-3.0	-2.5	-2.0	-1.50	-1.00	-0.50	0.00	0.50	1.00	1.50
-3.0	2.2819 $X10^{10}$	2.2825 $X10^{10}$	2.283 $X10^{10}$	2.2835 $X10^{10}$	2.2841 $X10^{10}$	2.2847 $X10^{10}$	2.2852 $X10^{10}$	2.2858 $X10^{10}$	2.2863 $X10^{10}$	2.2869 $X10^{10}$
-2.5	2.2816 $X10^{10}$	2.2822 $X10^{10}$	2.2827 $X10^{10}$	2.2833 $X10^{10}$	2.2838 $X10^{10}$	2.2844 $X10^{10}$	2.2849 $X10^{10}$	2.2855 $X10^{10}$	2.2861 $X10^{10}$	2.2866 $X10^{10}$
-2.0	2.2814 $X10^{10}$	2.2819 $X10^{10}$	2.2825 $X10^{10}$	2.283 $X10^{10}$	2.2835 $X10^{10}$	2.2841 $X10^{10}$	2.2847 $X10^{10}$	2.2852 $X10^{10}$	2.2858 $X10^{10}$	2.2864 $X10^{10}$
-1.5	2.2811 $X10^{10}$	2.2816 $X10^{10}$	2.2822 $X10^{10}$	2.2827 $X10^{10}$	2.2833 $X10^{10}$	2.2838 $X10^{10}$	2.2844 $X10^{10}$	2.2849 $X10^{10}$	2.2855 $X10^{10}$	2.2861 $X10^{10}$
-1.0	2.2808 $X10^{10}$	2.2814 $X10^{10}$	2.2819 $X10^{10}$	2.2824 $X10^{10}$	2.283 $X10^{10}$	2.2835 $X10^{10}$	2.2841 $X10^{10}$	2.2847 $X10^{10}$	2.2852 $X10^{10}$	2.2858 $X10^{10}$
-0.5	2.2806 $X10^{10}$	2.2811 $X10^{10}$	2.2816 $X10^{10}$	2.2822 $X10^{10}$	2.2827 $X10^{10}$	2.2833 $X10^{10}$	2.2838 $X10^{10}$	2.2844 $X10^{10}$	2.2849 $X10^{10}$	2.2855 $X10^{10}$
0.0	2.2803 $X10^{10}$	2.2808 $X10^{10}$	2.2814 $X10^{10}$	2.2819 $X10^{10}$	2.2824 $X10^{10}$	2.283 $X10^{10}$	2.2835 $X10^{10}$	2.2841 $X10^{10}$	2.2847 $X10^{10}$	2.2852 $X10^{10}$
0.5	2.28 $X10^{10}$	2.2805 $X10^{10}$	2.2811 $X10^{10}$	2.2816 $X10^{10}$	2.2822 $X10^{10}$	2.2827 $X10^{10}$	2.2833 $X10^{10}$	2.2838 $X10^{10}$	2.2844 $X10^{10}$	2.285 $X10^{10}$
1.0	2.2797 $X10^{10}$	2.2803 $X10^{10}$	2.2808 $X10^{10}$	2.2813 $X10^{10}$	2.2819 $X10^{10}$	2.2824 $X10^{10}$	2.283 $X10^{10}$	2.2835 $X10^{10}$	2.2841 $X10^{10}$	2.2847 $X10^{10}$
1.5	2.2795 $X10^{10}$	2.28 $X10^{10}$	2.2805 $X10^{10}$	2.2811 $X10^{10}$	2.2816 $X10^{10}$	2.2822 $X10^{10}$	2.2827 $X10^{10}$	2.2833 $X10^{10}$	2.2838 $X10^{10}$	2.2844 $X10^{10}$
2.0	2.2792 $X10^{10}$	2.2797 $X10^{10}$	2.2803 $X10^{10}$	2.2808 $X10^{10}$	2.2813 $X10^{10}$	2.2819 $X10^{10}$	2.2824 $X10^{10}$	2.283 $X10^{10}$	2.2835 $X10^{10}$	2.2841 $X10^{10}$
2.5	2.2789 $X10^{10}$	2.2795 $X10^{10}$	2.28 $X10^{10}$	2.2805 $X10^{10}$	2.2811 $X10^{10}$	2.2816 $X10^{10}$	2.2822 $X10^{10}$	2.2827 $X10^{10}$	2.283 $X10^{10}3$	2.2838 $X10^{10}$
3.0	2.2787 $X10^{10}$	2.2792 $X10^{10}$	2.2797 $X10^{10}$	2.2803 $X10^{10}$	2.2808 $X10^{10}$	2.2813 $X10^{10}$	2.2819 $X10^{10}$	2.2824 $X10^{10}$	2.283 $X10^{10}$	2.2835 $X10^{10}$

Table C.35: Sampling Variance of Horvitz and Thompson Estimator for Population 35

a	*b*									
	-3.0	-2.5	-2.0	-1.50	-1.00	-0.50	0.00	0.50	1.00	1.50
-3.0	1056657.6	1058193.1	1059735.9	1061286	1062843.6	1064408.6	1065981.1	1067561.2	1069148.9	1070744.4
-2.5	1055885.9	1057419.7	1058960.7	1060509.1	1062064.8	1063628.1	1065198.8	1066777.1	1068363.1	1069956.7
-2.0	1055114.1	1056646.1	1058185.4	1059732	1061286	1062847.5	1064416.4	1065993	1067577.2	1069169
-1.5	1054342.3	1055872.5	1057410	1058954.9	1060507.1	1062066.8	1063634	1065208.7	1066791.1	1068381.1
-1.0	1053570.3	1055098.8	1056634.6	1058177.6	1059728.1	1061286	1062851.4	1064424.4	1066004.9	1067593.2
-0.5	1052798.3	1054325	1055859	1057400.3	1058949	1060505.1	1062068.8	1063639.9	1065218.7	1066805.1
0.0	1052026.2	1053551.1	1055083.4	1056622.9	1058169.9	1059724.2	1061286	1062855.4	1064432.3	1066017
0.5	1051254	1052777.2	1054307.7	1055845.5	1057390.6	1058943.2	1060503.2	1062070.7	1063645.9	1065228.7
1.0	1050481.7	1052003.2	1053531.9	1055067.9	1056611.3	1058162	1059720.3	1061286	1062859.4	1064440.3
1.5	1049709.4	1051229.1	1052756.1	1054290.3	1055831.8	1057380.8	1058937.2	1060501.2	1062072.7	1063651.9
2.0	1048937.1	1050455	1051980.1	1053512.6	1055052.3	1056599.5	1058154.2	1059716.3	1061286	1062863.4
2.5	1048164.6	1049680.8	1051204.1	1052734.8	1054272.8	1055818.2	1057371	1058931.3	1060499.2	1062074.7
3.0	1047392.1	1048906.5	1050428.1	1051956.9	1053493.1	1055036.7	1056587.7	1058146.3	1059712.3	1061286

Table C.36: Sampling Variance of Horvitz and Thompson Estimator for Population 36

a	*b*									
	-3.0	-2.5	-2.0	-1.50	-1.00	-0.50	0.00	0.50	1.00	1.50
-3.0	1.1192 $X10^{10}$	1.1192 $X10^{10}$	1.1192 $X10^{10}$	1.1193 $X10^{10}$	1.1193 $X10^{10}$	1.1193 $X10^{10}$	1.1193 $X10^{10}$	1.1194 $X10^{10}$	1.1194 $X10^{10}$	1.1194 $X10^{10}$
-2.5	1.1192 $X10^{10}$	1.1192 $X10^{10}$	1.1192 $X10^{10}$	1.1192 $X10^{10}$	1.1193 $X10^{10}$	1.1193 $X10^{10}$	1.1193 $X10^{10}$	1.1194 $X10^{10}$	1.1194 $X10^{10}$	1.1194 $X10^{10}$
-2.0	1.1192 $X10^{10}$	1.1192 $X10^{10}$	1.1192 $X10^{10}$	1.1192 $X10^{10}$	1.1193 $X10^{10}$	1.1193 $X10^{10}$	1.1193 $X10^{10}$	1.1193 $X10^{10}$	1.1194 $X10^{10}$	1.1194 $X10^{10}$
-1.5	1.1191 $X10^{10}$	1.1192 $X10^{10}$	1.1192 $X10^{10}$	1.1192 $X10^{10}$	1.1192 $X10^{10}$	1.1193 $X10^{10}$	1.1193 $X10^{10}$	1.1193 $X10^{10}$	1.1194 $X10^{10}$	1.1194 $X10^{10}$
-1.0	1.1191 $X10^{10}$	1.1192 $X10^{10}$	1.1192 $X10^{10}$	1.1192 $X10^{10}$	1.1192 $X10^{10}$	1.1193 $X10^{10}$	1.1193 $X10^{10}$	1.1193 $X10^{10}$	1.1193 $X10^{10}$	1.1194 $X10^{10}$
-0.5	1.1191 $X10^{10}$	1.1191 $X10^{10}$	1.1192 $X10^{10}$	1.1192 $X10^{10}$	1.1192 $X10^{10}$	1.1192 $X10^{10}$	1.1193 $X10^{10}$	1.1193 $X10^{10}$	1.1193 $X10^{10}$	1.1194 $X10^{10}$
0.0	1.1191 $X10^{10}$	1.1191 $X10^{10}$	1.1192 $X10^{10}$	1.119 $X10^{10}$	1.1192 $X10^{10}$	1.1192 $X10^{10}$	1.1193 $X10^{10}$	1.1193 $X10^{10}$	1.1193 $X10^{10}$	1.1193 $X10^{10}$
0.5	1.1191 $X10^{10}$	1.1191 $X10^{10}$	1.1191 $X10^{10}$	1.1192 $X10^{10}$	1.1192 $X10^{10}$	1.1192 $X10^{10}$	1.1192 $X10^{10}$	1.1193 $X10^{10}$	1.1193 $X10^{10}$	1.1193 $X10^{10}$
1.0	1.1191 $X10^{10}$	1.1191 $X10^{10}$	1.1191 $X10^{10}$	1.1192 $X10^{10}$	1.1192 $X10^{10}$	1.1192 $X10^{10}$	1.1192 $X10^{10}$	1.1193 $X10^{10}$	1.1193 $X10^{10}$	1.1193 $X10^{10}$
1.5	1.1191 $X10^{10}$	1.1191 $X10^{10}$	1.1191 $X10^{10}$	1.1191 $X10^{10}$	1.1192 $X10^{10}$	1.1192 $X10^{10}$	1.1192 $X10^{10}$	1.1192 $X10^{10}$	1.1193 $X10^{10}$	1.1193 $X10^{10}$
2.0	1.1191 $X10^{10}$	1.1191 $X10^{10}$	1.1191 $X10^{10}$	1.1191 $X10^{10}$	1.1192 $X10^{10}$	1.1192 $X10^{10}$	1.1192 $X10^{10}$	1.1192 $X10^{10}$	1.1193 $X10^{10}$	1.1193 $X10^{10}$
2.5	1.119 $X10^{10}$	1.1191 $X10^{10}$	1.1191 $X10^{10}$	1.1191 $X10^{10}$	1.1191 $X10^{10}$	1.1192 $X10^{10}$	1.1192 $X10^{10}$	1.1192 $X10^{10}$	1.1192 $X10^{10}$	1.1193 $X10^{10}$
3.0	1.119 $X10^{10}$	1.1191 $X10^{10}$	1.1191 $X10^{10}$	1.1191 $X10^{10}$	1.1191 $X10^{10}$	1.1192 $X10^{10}$	1.1192 $X10^{10}$	1.1192 $X10^{10}$	1.1192 $X10^{10}$	1.1193 $X10^{10}$

APPENDIX D

SAS Program to Evaluate the Proposed Variance Approximations in Chapter 2

```
dm'log;clear;output;clear;';
options pageno=1;

                        /* comparing the approximated variances of SYG */


data one;       /*generating the data with GAMMA Dist. */
seed=1;
N=15;
B0=0;
B1=1.9;
MU=20;
sigmasq=10;
seta=10;
K=45;
do i=1 to N;
X1=seta*rangam(seed,K);
E=MU+sqrt(sigmasq)*rannor(seed);
output;
end;
data one1;set one (drop=seed i N);
data one2;set one1;
Y=B0+B1*X1+E;
proc print data=one2;
proc corr data=one2;
var X1 Y;

proc iml;
use one2;
read all var{x1} into x;
read all var{y} into y;

N = nrow(x);
sum_x = sum(x);
sum_Y= sum (y);
P = x/sum_x;

a = 1;
b = 1;
d = 0;
t = 0;
ai0=0.5#J(1,N,1);
ai=ai0//J(100,N,.);
a2i0=0.5#J(1,N,1);
a2i=a2i0//J(100,N,.);
Paif=J(N,1,0);
Paif1=J(N,1,0);
Vi4=J(N,1,0);
Vi3=J(N,1,0);
Vi1=J(N,1,0);
```

```
Vi2=J(N,1,0);
Pai2=J(N,1,0);
Pai = J(N,1,0);
Paij = J(N,N,.);
Var_Y = J(N,N,.);


do i = 1 to N;
      d = ((P[i] *(1- a*P[i]))/(1-2*b*P[i]))+ d;
      t = ((P[i] *(1- a*P[i]))/((1-P[i])*(1-2*b*P[i])))+ t;
end;


do i = 1 to N;
      Pai[i,1] = (P[i]/d) *((((1-a*P[i])*(1-2*P[i]))/((1-P[i])*(1-
2*b*P[i]))) + t);
end;
Sum_Pai=sum(Pai);

do i = 1 to N;
      Vi4[i,1] = Pai[i]*((Y[i]/Pai[i])-(sum_Y/2))**2;
end;
sum_Vi4=sum(Vi4);
Varapp4=((N-2)/(N-1))*Sum_Vi4; /*Calculating the fourth approximation*/

do i = 1 to N;
      Pai2[i,1] = pai[i]**2;
end;

sum_Pai2=sum(pai2);

a1=1/(2-((1/2)*sum_pai2));

do i = 1 to N;
      Vi3[i,1] = Pai[i]*(1-a1*pai[i])*((Y[i]/Pai[i])-(sum_Y/2))**2;
end;

Varapp3=sum(Vi3);               /* Calculating the third approximation*/

star1=0.5/(2-(0.5*sum_pai2));
ai0=star1#J(1,N,1);
ai=ai0//J(100,N,.);

do i = 1 to N;
      Paif[i,1] = pai[i]-(Pai2[i]/2);
end;
sum_Paif=sum(Paif);

do i=1 to 100;
do j = 1 to N;

     ai[i+1,j] =(0.5-(ai[i,j]*0.5*(sum_Pai2)))/(2*(1-pai[j]));
    end;
end;


 do i = 1 to N;
```

```
      Vi1[i,1] = Pai[i]*(1-2*ai[100,i]*pai[i])*((Y[i]/Pai[i])-
(sum_Y/2))**2;
end;

Varapp1=sum(Vi1);            /* Calculating the first approximation */

star2=1/(2*(N-1));
a2i0=star2#J(1,N,1);
a2i=a2i0//J(100,N,.);

do i = 1 to N;
      Paif1[i,1] = 1-(Pai[i]/2);
end;

sum_Paif1=sum(Paif1);

do i=1 to 100;
do j = 1 to N;

     a2i[i+1,j] =(0.5-(a2i[i,j]*0.5*(sum_Pai)))/(2*((1/Pai[j])-1));
    end;
end;

 do i = 1 to N;
      Vi2[i,1] = Pai[i]*(1-2*a2i[100,i])*((Y[i]/Pai[i])-(sum_Y/2))**2;
end;

Varapp2=sum(Vi2);        /* Calculating the second approximation */


do i = 1 to N;
      do j = i+1 to N;
            Paij[i,j] = ((P[i]*P[j])/d) *(((1-a*P[i])/((1-P[i])*(1-
2*b*P[i]))) + ((1-a*P[j])/((1-P[j])*(1-2*b*P[j]))));
      end;
end;

do i = 1 to 15;
      do j = i+1 to 15;
            Var_Y[i,j] = ((Pai[i]*Pai[j]) - Paij[i,j])
*(((Y[i]/Pai[i])- (Y[j]/Pai[j]))**2);
    end;
end;
VarY=sum (Var_Y);              /* Calculating the exact variance */

L1=((Varapp1-VarY)/VarY)*100; /*Calculating the comparison criterion*/

L2=((Varapp2-VarY)/VarY)*100;

L3=((Varapp3-VarY)/VarY)*100;

L4=((Varapp4-VarY)/VarY)*100;

print VarY  Varapp1 Varapp2 Varapp3 Varapp4 L1 L2 L3 L4;

run;
quit;
```

APPENDIX E

SAS Program to Evaluate the Proposed Variance Approximations in Chapter 3

```
dm'log;clear;output;clear;';
options pageno=1;

                  /* comparing the approximated variances of HT */


data one;          /*generating the data with GAMMA Dist. */
seed=1;
N=15;
B0=0;
B1=.1;
MU=20;
sigmasq=10;
seta=10;
K=1.8;
do i=1 to N;
X1=seta*rangam(seed,K);
E=MU+sqrt(sigmasq)*rannor(seed);
output;
end;
data one1;set one (drop=seed i N);
data one2;set one1;
Y=B0+B1*X1+E;

proc iml;
use one2;
read all var{x1} into x;
read all var{y} into y;

N = nrow(x);
sum_x = sum(x);
sum_Y=sum(Y);
P = x/sum_x;


Parti=J(N,1,0);
Parti111=J(N,1,0);
Parti121=J(N,1,0);
Parti211=J(N,1,0);
Parti221=J(N,1,0);
Parti311=J(N,1,0);
Parti411=J(N,1,0);
Parti421=J(N,1,0);
Parti41=J(N,1,0);
Parti42=J(N,1,0);
Paif1=J(N,1.0);
Paif2=J(N,1.0);
Partij = J(N,N,.);
Pai = J(N,1,0);
Pai2 = J(N,1,0);
Paij = J(N,N,.);
```

```
Var_Y = J(N,N,.);
n1=5;



do i = 1 to N;
      Pai[i,1] =1-((1-p[i])**n1);
end;
sum_pai=sum(pai);

do i = 1 to N;
      Pai2[i,1] = (pai[i])**2;
end;
sum_Pai2=sum(pai2);

do i = 1 to N;
      Parti[i,1] = ((1-Pai[i])/pai[i])* (Y[i])**2;
end;
Sum_Parti=sum(Parti);

do i = 1 to N;
      do j = i+1 to N;
            Paij[i,j] =Pai[i]+Pai[j]-(1-((1-P[i]-P[j])**n1));
      end;
end;


do i = 1 to N;
      do j = i+1 to N;
          Partij[i,j]=((paij[i,j]-
(pai[i]*pai[j]))/(pai[i]*pai[j]))*Y[i]*Y[j];
    end;
end;
Sum_Partij=2*sum(Partij);


do i = 1 to N;
      Paif1[i,1] = pai[i]-(Pai2[i]/2);
end;
sum_Paif1=sum(Paif1);


star11=0.5*(n1-1)/(n1-((1/n1)*sum_pai2));
ai01=star11#J(1,N,1);
ai1=ai01//J(100,N,.);

do i=1 to 100;
do j = 1 to N;
        ai1[i+1,j] =(((n1-1)/n1)-
(ai1[i,j]*(2/(n1)**2)*(sum_Pai2)))/(2*(1-(2*pai[j])/n1));
   end;
end;

do i = 1 to N;
      Parti111[i,1] = ((1/Pai[i])-(2*ai1[100,i]))* (Y[i])**2;
end;
Sum_Parti111=sum(Parti111);
```

```
do i = 1 to N;
      Parti121[i,1] = (ai1[100,i])* (Y[i]);
end;
Sum_Parti121=sum(Parti121);

Varppro11=sum_parti111+ (sum_Y*2*sum_Parti121) - sum_Y**2;
/*Calculating the first approximation*/


do i = 1 to N;
      Paif2[i,1] = 1-(Pai[i]/sum_pai);
end;
sum_Paif2=sum(Paif2);

star21=(n1-1)/(2*(N-1));
a2i01=star21#J(1,N,1);
a2i1=a2i01//J(100,N,.);

do i=1 to 100;
do j = 1 to N;
    a2i1[i+1,j] =(((n1-1)/n1)-
(a2i1[i,j]*(2/(n1**2))*(sum_Pai)))/(2*((1/Pai[j])-(2/n1)));

    end;
end;

do i = 1 to N;
      Parti211[i,1] = ((1-(2*a2i1[100,i]))/Pai[i])* (Y[i])**2;
end;
Sum_Parti211=sum(Parti211);

do i = 1 to N;
      Parti221[i,1] = ((a2i1[100,i])/Pai[i])* (Y[i]);
end;
Sum_Parti221=sum(Parti221);

Varppro21=sum_parti211+ (sum_Y*2*sum_Parti221) - sum_Y**2;
/*Calculating the second approximation*/


do i = 1 to N;
      Pai2[i,1] = (pai[i])**2;
end;
sum_Pai2=sum(pai2);

a31=(n1-1)/(n1-((1/n1)*sum_pai2));

do i = 1 to N;
      Parti311[i,1] = ((1/Pai[i])-a31)* ((Y[i])**2);
end;
Sum_Parti311=sum(Parti311);

Varppro31=sum_parti311+ ((sum_Y**2)*(a31-1));
/*Calculating the third approximation*/
do i = 1 to N;
```

```
      Parti41[i,1] = (1/Pai[i])* (Y[i])**2;
end;
Sum_Parti41=sum(Parti41);

do i = 1 to N;
      Parti42[i,1] = (1/Pai[i])* (Y[i]);
end;
Sum_Parti42=sum(Parti42);

Varppro41=(((N-n1)/(N-1))*sum_parti41)+ (sum_Y*((n1-1)/(N-
1))*sum_Parti42) - sum_Y**2;
/*Calculating the fourth approximation*/


VarY=Sum_Parti+Sum_Partij;
/*Calculating the exact variance HT*/


L11=((Varppro11-VarY)/VarY)*100;

L21=((Varppro21-VarY)/VarY)*100;

L31=((Varppro31-VarY)/VarY)*100;

L41=((Varppro41-VarY)/VarY)*100;
/*Calculating the comparison criterion*/

print VarY Varppro11 Varppro21 Varppro31 Varppro41 L11 L21 L31 L41;

run;
quit;
```

APPENDIX F

SAS Program to Specify the General Selection Procedure

```
dm'log;clear;output;clear;';
options pageno=1;

    /*calculating the variance using the general selection procedure*/

data one;
seed=1;
N=15;
B0=0;
B1=1.9;
MU=20;
sigmasq=10;
seta=10;
K=45;
do i=1 to N;
X1=seta*rangam(seed,K);
E=MU+sqrt(sigmasq)*rannor(seed);
output;
end;

data one1;set one (drop=seed i N);

data one2;set one1;
Y=B0+B1*X1+E;  /* generating the data with GAMMA DIS. */

proc iml;

use one2;
read all var{x1} into x;
read all var{y} into y;

N = nrow(x);
sum_x = sum(x);
P = x/sum_x;

do a = -3.0 to 3.0 by 0.5;
      do b =-1.5 to 1.5 by 0.5;
            d = 0;
            t = 0;

            Pai = J(N,1,0);
            Paij = J(N,N,.);
            Var_Y = J(N,N,.);

            do i = 1 to N;
                  d = ((P[i] *(1- a*P[i]))/(1-2*b*P[i]))+ d;
                  t = ((P[i] *(1- a*P[i]))/((1-P[i])*(1-2*b*P[i])))+ t;
            end;

            do i = 1 to N;
                  Pai[i,1] = (P[i]/d) *((((1-a*P[i])*(1-2*P[i]))/((1-
P[i])*(1-2*b*P[i]))) + t);
```

```
            end;



            do i = 1 to N;
                  do j = i+1 to N;
                        Paij[i,j] = ((P[i]*P[j])/d) *(((1-a*P[i])/((1-
P[i])*(1-2*b*P[i]))) + ((1-a*P[j])/((1-P[j])*(1-2*b*P[j]))));
                  end;
            end;


            do i = 1 to N;
                  do j = i+1 to N;
                        Var_Y[i,j] = ((Pai[i]*Pai[j]) - Paij[i,j])
*(((Y[i]/Pai[i])- (Y[j]/Pai[j]))**2);
                  end;
            end;

            VarY=sum (Var_Y);
            print VarY a b;
      end;
end;

run;
quit;
```

VITA

2

Ali Hussein Bagour

Candidate for the Degree of

Doctor of Philosophy

Dissertation: PROBABILITY PROPORTIONAL TO SIZE SAMPLING

Major Field: Statistics

Biographical:

Personal: Born in Jeddah, Saudi Arabia, August 24, 1966.

Education: Graduated from Al-Sadiq High School, Jeddah, Saudi Arabia, July 1983; received a Bachelor of Science degree in Statistics from King Abdul Aziz University, Jeddah, Saudi Arabia, August, 1990; received a Master of Science degree in Experimental Statistics from New Mexico State University, Las Cruces, New Mexico, July 1997. Completed the requirements for the Doctor of Philosophy degree with a major in Statistics at Oklahoma State University in July 2004.

Experience: Teaching assistant, Department of Statistics, King Abdul Aziz University, Saudi Arabia, 1990-1994.

Professional Memberships: American Statistical Association.