INTRODUCTORY THEORY FOR SAMPLE SURVEYS

Harold F. Huddleston, Retired
Principal Research Statistician
National Agricultural Statistics Service
U.S. Department of Agriculture
Washington, DC

1990

Acknowledgment

Harold F. Huddleston has had a major impact on statistical methodology used for the official domestic agricultural statistics program of the U.S. Department of Agriculture (USDA). He worked as a statistician for the USDA's agricultural statistics agency for 33 years.

In 1965, Mr. Huddleston was selected for advanced graduate level studies in statistics at Texas A&M University. There, he studied general single frame and multiple frame sampling theory with Dr. H.O. Hartley and Dr. J.N.K. Rao.

Mr. Huddleston authored numerous research reports on topics ranging from nonsampling errors on conventional surveys of farmers to the use of earth resource satellite data in a regression estimator for crop acreage estimates and yield forecasts. In addition to his many research reports, he authored three voluminous publications designed primarily to aid in the training of younger statisticians, particularly those less advanced in sampling theory as applied to agriculture. Included were many international program participants from numerous countries.

The first of these publications was "A Training Course in Sampling Concepts for Agricultural Surveys" published in April 1976. "Sampling Techniques For Measuring and Forecasting Crop Yields" was the second one published in August 1978.

This manuscript on sampling theory is the third publication. The National Agricultural Statistics Service is indebted to Harold F. Huddleston for this manuscript and all his major contributions to the agricultural statistics program of the U.S Department of Agriculture.

CHARLES E. CAUDILL
Administrator
National Agricultural Statistics Service
U.S. Department of Agriculture

PREFACE

The extensive sampling theory paper was written after the USDA granted the author a sabbatical scholarship at Texas A&M University in 1965 to work under Professor H.O. Hartley in multiple frame theory and J.N.K. Rao in sampling theory. The purpose of this account of theory is to supplement lectures in developing the building blocks of sampling, as are likely to be needed in applied theory. Several contrasting approaches are used in showing how the theory may be developed. The logic of these alternative derivations of the basic theory and formulas is to provide the student with greater exposure to different derivations in estimating parameters is frequently helpful in complex designs since sometime there is an easier approach.

The elements of the theory covered herein might be found in either a beginning or advanced sampling theory course, but the goal is to present the topics at an introductory level assuming only some previous exposure to sampling methods for motivational purposes. In addition, some limited background in mathematical expectation and basic probability is helpful.

This particular effort is largely the result of the direct influence of H.O. Hartley and J.N.K. Rao who stimulated work and interest in sampling by their teachings. The influence of writings by Cochran, Hendricks, Des Raj, Jessen, Hansen, Hurwitz, and Madow have also been substantial. In addition, many papers have influenced both the point of view adapted as well as the material presented. I acknowledge these sources and others which I may have unintentionally omitted. This presentation is intended for the student working at the Masters Degree Level who may only take a one term course in the theory of survey sampling.

Special thanks are due to Mrs. Sue Horstkamp and Mrs. Mary Ann Lenehan for their excellent typing and help in completing this monogram.

HAROLD F. HUDDLESTON
Washington, DC

CONTENTS

Chapter I. Probability and Expectation

1.0 Introduction

Since probability forms the basis of sampling theory, we begin with a presentation of some results used in sampling. This topic is followed with some of the important results on expected values. Both topics in the context of sampling relate to all types of populations and parameters, thus the classical theory of sampling is regarded as distribution free. However, confidence interval statements about the sample statistics do assume that their derived distributions are known. In practice, reliance is placed on the central-limit theorem and estimators which approach normality.

Modern sample surveys are multicharacteristic (multiple content items) in practice and it is frequently not practical to use many of the results available from general estimation theory. Consequently, the use of specific distributions and the method of maximum-likelihood are generally not considered. Likewise, an estimator which is cheaper or operationally easier to handle, is frequently preferred to another which requires considerable computations and may have a smaller variance. However, it is not correct to assume that more powerful estimation theory can not be employed to good advantage where they are appropriate and there is a high level of expertise and resources available for their use in surveys.

1.1 Sample Space and Events

We are concerned with random samples or experiments in which the outcome depends on chance. The sample space is made up of elements which correspond to the possible outcomes of the conceptual experiment. The elements depend on the sample and frame sizes together with the probability selection procedure. The outcome of the sample selection, with the associated observed characteristics, correspond to one and only one element of the sample space(S). An event is a subset of S. The event E occurs if the outcome of the experiment corresponds to an element of the subset E.

Event E = aggregation of all points containing E

Event $\bar{E}$ = all sample points not contained in E is called the complementary event

Event E_1+E_2 = The totality of sample points contained in either E_1 or E_2 (E_1+E_2 is called the union, i.e., $E_1 \cup E_2$)

Event $E_1 \cdot E_2$ = aggregation of sample points contained in both of the events E_1 and E_2 (E_1E_2 is called the intersection, $E_1 \cap E_2$)

The events E_1 and E_2 are mutually exclusive if they have no points in common.

It is convenient, often, to speak of the union or intersection of an infinite, but countable, set of events; or a set of events as being countable. If only a finite number of positive integers is used in this counting, the set is countable and finite. If all positive integers are used, the set is infinite, but countable. An event is associated with a particular integer by setting the integer as a subscript on the symbol for the event. Thus

$A_1, A_2, \ldots\ldots, A_{N-1}, A_N$ is countable and finite set of events; while

$A_1, A_2, \ldots\ldots\ldots\ldots\ldots$ represents a countable and infinite set of events.

1.2 Probability

To each element in S is assigned a number, $P(e_i)$, called the probability of e_i which depends on the model used in setting up the experiment, subject to the restrictions:

(a) $\Sigma P(e_i) = 1$ or $P(S) = 1$

(b) $P(e_i) \geq 0$

(c) $P(e_i \cup e_j) = P(e_i) + P(e_j)$ $i \neq j$

If E is any event, then $0 \leq P(E) \leq 1$. The values 0 and 1 do not imply either an impossible event or a certain event. This is the result of some elements of S being assigned a probability of 0. In problems involving uncountably infinite sample spaces there must exist events that are not impossible but yet have probability 0. If we insisted that the probability of each element in the space be positive, i.e., $P(e_i)>0$, then only an empty event would have probability 0, and only the whole sample space would have probability 1. The assignment of probabilities to the elements in the space may vary for different real-world situations to which the theory is applied.

Some Basic Laws of Probability:

If A is a random event, we write P(A) for its probability. In terms of the elements in the probability space S, P(A) is the ratio of the number of elementary events (elements of S) favorable to A divided by the total number of elementary events.

Law of Total Probability (Pairwise Independent events). The probability of the union of a countable set of mutually exclusive events is the sum of their probabilities.

(1) $P(U_1^N A_i) = \sum_{i=1}^{N} P(A_i)$ sets are countable and finite,

(1′) $P(U_1^\infty A_i) = \sum_{i=1}^{\infty} P(A_i)$ sets countable and infinite

Certain theorems are consequences of the above law.

A. If A and B are events, and if $A \subset B$ (A subset of B), then $P(B-A) = P(B)-P(A)$

B. $P(\bar{A}) = 1-P(A)$ for every event A.

C. $P(0) = 0$

D. If A and B are events, and if $A \subset B$, then $P(A) \leq P(B)$

E. If $A_1, A_2 \ldots$ are events (not necessarily mutually exclusive)

$$P(U_1^\infty A_i) \leq \sum_{i=1}^{\infty} P(A_i)$$

F. If B_1, $B_2 \ldots$ are events, if $B_1 \subset B_2 \subset$ and if $B = U_1^\infty B_i$, then

$$P(B) = \lim_{i\to\infty} P(B_i)$$

G. If B_1, $B_2 \ldots$ are events, if $B_1 \supset B_2 \supset$ and if $B = \cap_1^\infty B_i$, then

$$P(B) = \lim_{i\to\infty} P(B_i).$$

Law of Total Probability (Arbitrary random events). Let $A_1, A_2, \ldots A_N$, where $N \geq 3$, be arbitrary random events.

(2) $P(U_1^N A_i) = \sum_{i=1}^{N} P(A_i) - \sum_{i \neq j}^{N} P(A_i A_j) + \sum_{i \neq j \neq k}^{N} P(A_i A_j A_k) + (-1)^{N+1} P(A_1 A_2 \ldots A_N)$

(There are N terms in the expression with each term smaller (or equal to) than the preceding term). The terms after the first involve compound events which will be discussed below.

Compound Events and Probabilities:

If two events A and B occur simultaneously (joint occurrence), then we have a compound event. In terms of sets, we write $A \cup B = A\bar{B} + \bar{A}B + AB = A + B - AB$.

In terms of probabilities, we write based on (2) above $P(A \cup B) = P(A) + P(B) - P(AB)$, or based on (1) $P(A \cup B) = P(A\bar{B}) + P(\bar{A}B) + P(AB)$ since $A\bar{B}$, $\bar{A}B$, and AB are non-overlapping sets. The above partitioning of sets generalizes to compound events involving N arbitrary sets.

Compound Probability. If A and B are any two events, their joint probability is $P(AB) = P(B)P(A/B) = P(A)P(B/A)$, or

$$P(AB) = P(A \cup B) - P(A\bar{B}) - P(\bar{A}B); \text{ for N events}$$

(3) $P(A_1A_2\ldots A_N) = P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\ldots P(A_N|A_1\ldots A_N-1)$

where we define $P(A|B)$, $P(A_2|A_1)$, $P(A_3|A_1A_2)$ etc. as conditional probabilities; or restated (3) becomes

(3′) $P(A_1A_2\ldots A_N) = P(A_1A_2\ldots A_{N-1})\ P(A_N|A_1A_2\ldots A_{N-1})$.

Conditional Probabilities. If we consider two events A and B, we mean by the conditional probability of A given B that we have redefined the sample space to be only those elements contained in event B, where B is a subset of the sample space S. Consequently, we define

(4) $P(A|B) = \dfrac{P(AB)}{P(B)}$ if $P(B) \neq 0$, and immediately we have

$P(AB) = P(A|B)\ P(B) = P(A)\ P(B|A)$ even if $P(A) = 0$ or $P(B) = 0$. The conditional probability $P(A|B)$ implies that the event A is of interest only if B has occurred; however, we can define $P(A|B)$ in (4) in terms of the probabilities in the total sample space S. That is, for three events from the same sample space S

(4′) $P(A_1\ A_2A_3) = \dfrac{P(A_1A_2A_3)}{P(A_2A_3)}$.

Bayes Theorem. If the events A_1, A_2...satisfy our previous assumptions and $P(B)>0$, then the posteriori probability of A_i given B has occurred is

(5) $P(A_i|B) = \dfrac{P(A_i)P(B|A_i)}{P(B)} = \dfrac{P(A_i)P(B|A_i)}{\Sigma P(A_i)P(A_i|B)}$ where $P(A_i)$ is called a priori probability.

Independent Events. Two events A and B are said to be independent events if and only if

(6) $P(AB) = P(A)P(B)$, otherwise they are said to be dependent; and for K events

(6′) $P(A_1A_2...A_K) = P(A_1)P(A_2)...P(A_K)$.

1.3 Samples and n-Tuples

A basic tool for the construction of sample description spaces of random selection is provided by the notion of an n-tuple. An n-tuple $(Z_1,Z_2,...Z_n)$ is an array of n symbols with first, second, and so on up to the n^{th} component. The order in which the components are written is of importance since sometimes we wish to speak of ordered n-tuples. Two n-tuples are identical, if and only if, they consist of the same components written in the same order. The usefulness of n-tuples derives from the fact that they are convenient devices for reporting the results of a drawing of a sample of size n.

(a) Sampling with replacement - The sample is said to be drawn with replacement (W.R.) if after each draw, the unit selected is returned to the frame so its chance of selection is the same in each successive draw as on the first draw.

(b) Sampling without replacement - The sample is said to be drawn without replacement (W.O.R.) if after each draw the unit selected is removed from the frame so its changes of selection become zero in each successive draw.

The basic principles of combinatorial analyses are useful in counting sets of n-tuples for various values of n that may arise.

The size of the set A or ordered n-tuples is given by the product of the numbers $N_1,N_2,...N_n$, or Size (A) $= N_1N_2...N_n$

where N_1 = number of objects that may be used as the first component,

N_2 = number of objects (if it exists) that may be second components,

.

.

.

.

N_n = number of objects (if it exists) that may be the n^{th} component of the n-tuple.

The number of ways in which one can draw a sample of n objects from M distinguishable objects is: $M(M-1)\ldots\ldots(M-n+1)$ or $\frac{M!}{(M-n)!}$ if sampling is done without replacement, and M^n if the sampling is done with replacement. An important application of the foregoing relations is the problem of finding the number of subsets of a set.

The number of subsets of S of size K, multiplied by the number of samples that can be drawn without replacement from a subset of size K, is equal to the number of samples of size K that can be drawn without replacement from S itself, or $X_K . K! = \frac{N!}{(N-K)!}$.

Therefore, $X_K = \frac{N!}{K!(N-K)!} = \binom{N}{K}$. These quantities are generally called the binomial coefficients where the binomial form is $(a+b)^N$.

From these coefficients, one may determine how many subsets of a set of size N that can be formed.

$$\binom{N}{0} + \binom{N}{1} + \binom{N}{2} + \ldots\ldots + \binom{N}{N} = 2^N$$

Thus, the number of events (including the impossible event) that can be formed from a sample description space of size N is 2^N (i.e., power set.

Another counting problem is that of finding the number of partitions of a set of size N into sets $S=(1,2,\ldots N)$. Let r be a positive integer, and let K_1, $K_2\ldots,K_r$ be positive integers such that $K_1+K_2+\ldots+K_r = N$, we speak of a division of S into r subsets (ordered) such that the first subset has size K_1, second size K_2, and so on.

The number of ways one can partition a set of size N into r ordered subsets is the product $\binom{N}{K_1}\binom{N-K_1}{K_2}\binom{N-K_1-K_2}{K_3}\ldots\ldots\binom{N-K_1-K_2\ldots-K_{r-1}}{K_r}$

which also may be written as $\frac{N!}{K_1!K_2!\ldots K_r!}$

This is known as the multinomial coefficient $\binom{N}{K_1K_2\ldots K_r} = \frac{N!}{K_1!K_2!\ldots K_r!}$

where the multinomial form is $(a_1+a_2+\ldots+a_r)^N$. For an event A_K, for $K=0,1,2,\ldots n$, where the sample will contain exactly K objects of a particular kind, then $N(A_K) = \binom{M_W}{K}\binom{M-M_W}{n-K}$ where M = total objects of all kinds in the frame, and M_W = the objects of the type we are interested in. Consequently, the probability of exactly K objects is:

$$P(A_K) = \frac{\binom{M_W}{K}\binom{M-M_W}{n-K}}{\binom{M}{n}} \text{ for samples drawn without replacement.}$$

This is also the probability for ordered samples drawn without replacement. However, in sampling with replacement for an unordered sample of size n,

$$P(A_K) = \frac{\binom{M_W+K-1}{K}\binom{M-M_W+n-K-1}{n-K}}{\binom{M+n-1}{n}}$$

.4 Expectation

By definition, an expected value is the population mean for a parameter. Let U be a random variable taking values μ_i $(i=1,2,\ldots,K)$ with probability $P(U=\mu_i)(i=1,\ldots,K)$, $\sum^{K} P(U=\mu_i) = 1$. Then the expected value of U is defined as

$$(7) \quad E(U) = \sum_{i=1}^{K} \mu_i \, P(U=\mu_i) = \bar{U}$$

$$(8) \quad E(U^2) = \sum_{i=1}^{K} \mu_i^2 \, P(U=\mu_i) = \sigma^2 + \bar{U}^2 \text{ where } \sigma^2 \text{ is defined in 1.5.}$$

A random variable is a characteristic of a random event.

Some useful propositions concerning operations with expected values are given now

$E(a) = a$, if a is a constant

$E(aU) = aE(U)$, if a is a constant

$E(\Sigma\mu_i) = \Sigma E(\mu_i)$

$E(\Sigma a_i\mu_i) = \sum_i a_i E(\mu_i)$, if the a_i's are constants

Using these results

$E(aU+b) = aE(U)+b$, a and b are constants

$E[\phi(U)] = \sum_{i=1}^{K} P(U=\mu_i)\phi(\mu_i)$ where $\phi(U)$ is a function of the random variable U.

If we have a second random variable W taking values $w_j (j=1,2,\ldots 1)$ with probabilities $P(W=w_j)$ and $\sum_{j=1}^{1} P(W=w_j) = 1$. The joint probability of U and W is given by $P(U=\mu_i, W=w_j)=P(UW)$ where $\sum_{ij}\sum P(U=\mu_i, W=w_j) = 1$.

Also $P(U=\mu_i) = \sum_j P(U=\mu_i, W=w_j)$, and

$P(W-w_j) = \sum_i P(U=\mu_i, W=w_j)$

Expectation of the sum of two random variables $E(U+W)=E(U)+E(W)$ which we generalize to n random variables $U_1, U_2, \ldots U_N$

(9) $E(U_1+U_2+\ldots+U_N) = \sum_{i=1}^{n} E(U_i)$

Expectation of the product of two independent random variables

(10) $E(UW) = E(U)E(W)$, and

(11) $E[f_1(U)f_2(W)] = E[f_1(U)]E[f_2(W)]$ if $f_1(U)$ and $f_2(W)$ are any two functions of the independent random variables U and W.

Expectation of one random variable divided by a second random variable

(12) $E(\frac{U}{W}) = \frac{E(U)}{E(W)} - \frac{Cov(\frac{U}{W},W)}{E(W)}$

<u>Conditional Expectation</u>. We consider the expectation of any two random variables W and U where it is known that U has occurred.

(13) $E(W/U) = \sum_j \frac{W_j \; P(W-W_j \; U=\mu_i)}{P(U=\mu_i)}$, or

$= \sum_j W_j \; P(W=W_j | U=\mu_i)$.

We may restate the expectation E(UW) for any two random variable.

(14) $E(UW) = E[UE(W|U)] = E[UE_2(W)] = E[\mu_i E(W_i|\mu_i)]$

where E_2 is the conditional expectation for a given value(s) of U which is commonly written in this way for brevity. The RHS of (13) may be written, for the same reason, in terms of conditional expectation as $E(UW) = E_1E_2$ where the subscripts 1 and 2 indicate the order in which the operations occurred.

Decomposition of total expectation given in (7)

(15) $E(U) = E[E(U|W)]$ since

$$= E[\sum_i^K \mu_i P(U=\mu_i|W_j)]$$

$$= \sum_j^1 P(W_j) \sum_i^K \mu_i P(U=\mu_i|W_j)$$

Conditional expectation is frequently easier to use in evaluating the total expectation of a random variable in complex survey designs.

1.5 Variances and Covariances

By definition, a variance of a random variable U with expectation $\bar{U}$ is

(16) $V(U) = \sum_{i=1}^{K} (\mu_i-\bar{U})^2 \; P(U=\mu_i) = E(U-\bar{U})^2 = E(U^2) - \bar{U}^2$

where σ^2 is commonly used to denote V(U).

(17) $V(aU+b) = a^2V(U)$ where a and b are constants, consequently their variance is zero.

By definition, a covariance of two random variables U and W with expectation $\bar{U}$ and $\bar{W}$ is

(18) $Cov(U,W) = \sum_{i,j} (\mu_i-\bar{U})(w_j-\bar{W})P(U=\mu_i, W=w_j)$

$$= E[(U-\bar{U})(W-\bar{W})] = E(UW)-E(U)E(W)$$

Obviously, the variance is a special case of the covariance where the same variable is involved.

The variance of a linear sum $L = a_1U_1+a_2U_2+\ldots+a_nU_n$ of random variables $U_1,U_2,\ldots,U_n$ is

(19) $V(\sum^n a_iU_i) = \sum\sum_{ij} a_ia_j \; Cov(U_i,U_j)$ which can be restated in terms of variances and covariance as

(19′) $V(\sum^{n} a_i U_i) = \sum a_i^2 V(U_i) + 2 \sum_{ij>i}\sum a_i a_j \, Cov(U_i, U_j)$

(19″) $= \sum a_i^2 V(U_i) + 2 \sum_{ij>i}\sum a_i a_j \, \rho_{ij} \, \sigma(U_i)\sigma(U_j)$

where $\rho_{ij} = \dfrac{Cov(U_i, U_j)}{\sigma(U_i)\sigma(U_j)}$

The covariance of two linear sums $U = a_1U_1 + a_2U_2 + \ldots + a_mU_m$ and $W = b_1W_1 + b_2W_2 + \ldots + b_nW_n$ is

(20) $Cov(U,W) = \sum_{ij}\sum a_i b_j \, Cov(U_i, W_j)$

The variance of a product of two independent random variables is often needed. We express each variable (X and Y) in a more useful form.

Let $X = \bar{X} + \dfrac{\bar{X}(X-\bar{X})}{\bar{X}} = \bar{X}(1+\delta x)$

$Y = \bar{Y} + \dfrac{\bar{Y}(Y-\bar{Y})}{\bar{Y}} = \bar{Y}(1+\delta y)$

(21) $V(XY) = E[XY - \bar{X}\bar{Y}]^2 = (\bar{X}\bar{Y})^2 \, E[\delta x + \delta y + \delta x \delta y]^2$

$= (\bar{X}\bar{Y})^2 \left[\dfrac{V(X)}{\bar{X}^2} + \dfrac{V(Y)}{\bar{Y}^2} + \dfrac{V(X)V(Y)}{\bar{X}^2\bar{Y}^2}\right]$

(21′) $V(XY) = [E(Y)]^2 V(X) + [E(X)]^2 V(Y) + V(X)V(Y)$

The variance of a product of any two random variables

(22) $V(XY) = \bar{Y}^2 V(X) + \bar{X}^2 V(Y) + 2\bar{X}\bar{Y}E_{11} + 2\bar{X}E_{12} + 2\bar{Y}E_{21} + E_{22} - E_{11}^2$

where $E_{ij} = E[(\Delta X)^i (\Delta Y)^j]$ and $\Delta X = X - \bar{X}$, $\Delta Y = Y - \bar{X}$.

An unbiased estimate of $\bar{X}\cdot\bar{Y}$ is given by

$\widehat{\bar{X}\bar{Y}} = (n\bar{x}\bar{y} - \dfrac{\sum x_i y_i}{n}) \div n-1$

The variance of one random variable divided by a second random variable is given under Section 1.8 below.

1.6 Conditional Variances and Covariances

Using (18) it is easy (at least possible) to obtain the covariance of two random variables in terms of conditional expectation. The covariance of two random variables U and W where they are conditioned by H may be expressed as

$$(23)\ Cov(U,W) = E_1\ Cov_2(U,W) + Cov_1(E_2U, E_2W)$$

$$\text{where } Cov_2(U,W) = E(UW|H_j) - E(U|H_j)E(W|H_j)$$

$$E_1[Cov_2(U,W)] = \sum_j P(H_j)\ Cov_2(U_1W)$$

$$E_2(U) = E(U|H_j)$$

$$E_2(W) = E(W|H_j)$$

Since the variance is a special case of the covariance, we may state the variance in terms of conditional expectations using (23).

$$(24)\ Cov(U,U) = V(U) = E_1V_2(U) + V_1E_2(U)$$

1.7 Distribution of Sample Mean

Generally, interest centers on an estimate of the population mean μ with an estimate of σ_μ being needed for probability statements on the precision of the sample mean $\bar{X}$ (or the total $N\bar{X}$). While unbiased estimates of these parameters are readily obtained for all population based on expectation operations, two fundamental principles are required for making probability statements which depend on the population from which the random variable X is selected.

Many random variables possess normal distributions, at least approximately. By using probability and distribution theory, it is possible to derive the distribution of $\bar{X}$ when X is selected from a normal distribution. The mathematical results are expressed in the form of a theorem.

Theorem: If X is selected from a normal distribution with mean μ and standard deviation σ, then the sample mean $\bar{X}$, based on a random sample of size n, will possess a normal distribution with mean μ and standard deviation $\sigma/\sqrt{n}$.

The distribution of $\bar{X}$ given by this theorem is called the sampling distribution of $\bar{X}$ because it represents the distribution of means obtained by repeated sampling from a fixed population of X's and a given sample size n.

The distribution of $\bar{X}$ when X is selected from a non-normal distribution depends on the non-normal distribution sampled. However, the

Central Limit Theorem provides a satisfactory basis for dealing with the distribution of $\bar{X}$ without being concerned about the nature of the distribution of X for most practical problems. This theorem states that under very mild assumptions (the mean and variance exist) the distribution of $\bar{X}$ approaches a normal distribution as the sample size, n, increases. The results of sampling experiments from many populations of X's and for small values of n (10 to 20) support the theorem.

Theorem: If X possesses a distribution with mean μ and standard deviation σ, then the sample mean $\bar{X}$, based on a random sample of size n, will possess an approximately normal distribution with mean μ and standard deviation $\sigma/\sqrt{n}$, the approximation becoming increasingly good as n increases.

These two theorems permit one to calculate the probability that μ will lie in any specified interval by transforming the observed mean to a standard normal distribution with mean zero and standard deviation of one and utilizing tables of the standardized normal distribution. However, these theorem are normally used to make interval estimates about the parameter μ. The interval estimate is constructed so the probability of the interval containing the parameter can be specified. Such intervals are normally constructed so the probability is high so that the parameter will be in the stated interval and is referred to as the confidence interval for the parameter. That is, $\bar{X} \pm \frac{Z_\alpha \hat{\sigma}}{\sqrt{n}}$ is used to define the upper and lower values of the confidence interval where ($1-\alpha$) indicates the probability that the parameter will lie in the interval in repeated sampling. If the sample size is small (variance degrees of freedom less than 30), a value t_α from the Student t-distribution is used in place of Z_α from the normal distribution.

1.8 Use of Approximation Techniques

The Taylor series is occasionally a useful devise for evaluating certain expressions approximately, such as, encountered in evaluating expectations. Since the remainder term may be evaluated in the series, the degree of approximation can be determined.

(A) If f(X) and its first n+1 derivatives are continuous in the closed interval containing x = a, then

$$f(X) = f(a) + f'(a)(x-a) + \ldots + \frac{f^n(a)}{n!}(x-a)^n + R_{n+1}$$

where the remainder is $R_{n+1} = \frac{1}{n!} \int_a^x (x-t)^n f^{n+1}(t)dt$.

The size of the remainder may be estimated if

$|f^{n+1}(t)| \leq M$ when $a \leq t \leq X$

$$|R_{n+1}| \leq \frac{M}{n!} \int_a^x (x-t)^n \, dt = \frac{M(x-a)^{n+1}}{(n+1)!}$$

However, there exist other forms of the remainder which hold under slightly less stringent assumptions. We may extend the version of Taylor's formula to several variables.

This technique is most commonly used in evaluating the expression for the ratio of two random variables or a complex function of one or several random variables and their variances. In determining a mean value for f(X), "a" is replaced by E(X) before taking expectations. The expression for f(X) is squared and written as a Taylor's series, then taking the expectations to find $Ef^2(X)$. The variance of any function is $Ef^2(X)-[Ef(X)]^2$. The use of "a" = E(X) can be justified on the basis that if E(X) is a maximum likelihood estimator, then f[E(X)] is a maximum likelihood estimator of f(X). Where X is a normally distributed variable, E(X) is the maximum likelihood estimator. The variance of the division of one random variable by a second random variable is given below.

First approximation

$$(25)\quad V\left(\frac{U}{W}\right) = \frac{1}{n}\left(1-\frac{n}{N}\right) \frac{\bar{U}^2}{\bar{W}^2} \left[\frac{V(U)}{\bar{U}^2} + \frac{V(W)}{\bar{W}^2} - \frac{2Cov(U,W)}{\bar{U}\bar{W}}\right]$$

which holds for n large enough so $\delta_W = \frac{W-\bar{W}}{\bar{W}} << 1$ holds. If W is the sample mean (i.e., $\bar{W}$), then the coefficient of variation is frequently much less than 1. The approximation is quite good if $\delta_W = .1$.

(B) An alternative way of looking at the variance of a ratio which employs the Taylor's series is:

Let $R = \frac{E(U)}{E(W)} = \frac{\bar{U}}{\bar{W}}$, and $\hat{R} = \frac{\bar{\mu}}{\bar{w}}$ for the sample

$$\hat{R} - R = \frac{\bar{\mu}}{\bar{w}} - \frac{\bar{U}}{\bar{W}} = \frac{\bar{\mu} - R\bar{w}}{\bar{w}}$$

If n is sufficiently large, we may replace $\bar{w}$ by $\bar{W}$ in the denominator so the expected bias is zero, or we may write $\bar{w} = \bar{W} + (\bar{w} - \bar{W})$.

Then $\hat{R} - R = \frac{\bar{\mu} - R\bar{w}}{\bar{W}} \quad (1 + \frac{\bar{w}-\bar{W}}{\bar{W}})^{-1}$

and expand the term in parenthesis by a Taylor's series.

$$\hat{R} - R = \frac{\bar{\mu} - R\bar{w}}{\bar{W}} [1 - \frac{\bar{w}-\bar{W}}{\bar{W}} + \frac{(\bar{w}-\bar{W})^2}{\bar{W}^2} - \ldots\ldots].$$

When we square this expression and take expectation we obtain an approximation for the variance. This expression is complicated but it simplifies if U and W have a bivariate normal distribution and provides a means of studying the nature of the approximation. The usual (or first) approximation is based on retaining only the first term in the brackets.

1.9 Three Sampling Schemes for Simple Random Sampling

Scheme A: A fixed number of n units is selected with equal probability and with replacement at each draw.

Scheme B: A fixed number of n units is selected with equal probability at each draw and without replacement. Everyone of the $\binom{N}{n}$ distinct samples s_n has an equal chance of being selected.

Scheme B′: Selection is continued with replacement and with equal probability until the desired number n of distinct units is obtained.

The scheme B and B are equivalent in the sense that the probability of selecting a sample s_n is the same for both schemes, provided the estimators are based only on distinct units. Usually scheme B is referred to as simple random sampling without replacement.

1.10 Miscellaneous Results

On Expectations:

(a) $E(X) = \mu$

(b) $E(X^2) = \sigma^2 + \mu^2$

(c) $E(\bar{X}^2) = \frac{\sigma^2}{n} + \mu^2$

(d) $E[(\sum^{n} X)^2] = n\sigma^2 + n^2\mu^2$

(e) $E[\Sigma a_i X_i] = \Sigma a_i E(x_i)$, a_i a constant

(f) $E(X_i\bar{X}) \neq E(X_i)E(\bar{X})$ due to dependency

(g) $E(X_i-\mu)(X_j-\mu) = 0 \quad i \neq j$

(h-k) $E(a) = a$, $E(a^2) = a^2$, $V(a) = 0$, and $Cov(X,a) = 0$ where a is constant

(l) $E(X_i-\bar{X})(X_j-\bar{X}) = \frac{-\sigma^2}{n}$

(m) $E(X_i\bar{X}) = \frac{\sigma^2}{n} + \mu^2$

<u>On Identities</u>:

$$\sum_i^n X_i = X_1 + X_2 + \dots + X_n$$

$$\Sigma(X_i-\bar{X}) = 0$$

$$(\Sigma X_i)^2 \div n = n\bar{X}^2$$

$$\Sigma(X_i-\bar{X})^2 = \Sigma(X_i-\bar{X})X_i$$

$$\Sigma(X_i-a)^2 = \Sigma(X_i-\bar{X})^2 + n\ (\bar{X}-a)^2$$

$\sum_{i\neq j} X_iX_j = 2\sum_{i<j} X_iX_j$ there are n(n-1) values in $\sum_{i\neq j} X_iX_j$

For $N = 2$, $\sigma^2 = (X_1-X_2)^2/2$

$$\sum_j X_{ij} = X_{i.}$$

$$\sum_i X_{ij} = X_{.j}$$

$$\prod_{i=1}^n X_i = X_1.X_2.\ \dots\ X_n$$

$$\Sigma(X_i-Y_i) = \Sigma X_i - \Sigma Y_i$$

$$\sum^n X_i{}^2 = \sum_{i=1}^n X_i^2 + \sum_{i\neq j}^{n(n-1)} X_iX_j$$

<u>Double Summation</u>:

$$\sum_{ij}\Sigma X_{ij} = \sum_{i,j} X_{ij} = X_{..}$$

$$\sum_{ij}^{ab}\sum X_{ij} = \sum_{i=1}^{a} (\sum_{j=1}^{b} X_{ij}) = \sum_{i=1}^{a} bX_{i.} = X_{..}$$

$$= \sum_{j=1}^{b} (\sum_{i=1}^{a} X_{ij}) = \sum_{j=1}^{b} aX_{.j} = X_{..}$$

$$\sum_{i} \sum_{j\neq i} X_i X_j = \sum_{i\neq j} X_i X_j$$

Chapter II. Single Stage Sampling

2.0 Introduction

Single stage selection is the basic building block in sampling theory. In practice, there are very few surveys or experiments for which the design employed could be called a single stage sample. Instead, theory of single stage sampling is applied at each of several stages. The use of stages in sampling is the result of the inability or cost of directly selecting from all N units in the frame. Even where all N units are accessible for sampling, the lack of homogeneity of the units generally dictate some other method of sampling such as stratification which will give greater precision for less cost.

Two different approaches for deriving estimators of population totals and means are presented. These are: (1) the method of weight variables, and (2) the expectation of the characteristic value. In the first case, the random variable is the weight associated with each of the N units in the universe which depends on the method of selecting units and the characteristic value of the unit is treated as a constant. In the second case, the random variable is the characteristic value which is associated with each of the N units in the universe. Method 1 is useful since it enables us to use standard results from infinite population theory in constructing estimators.

2.1 Sampling With Replacement - (Method 1)

Notation

	Universe of Distinguishable Units
Unit labels - L	1, 2, 3, ... , i, ...N
Characteristic values	$y_1, y_2, y_3, \ldots, y_i, \ldots\ldots\ldots\ldots y_N$
Weights	$\mu_1, \mu_2, \mu_3, \ldots, \mu_i, \ldots\ldots\ldots\ldots \mu_N$

The weight variables are defined as

$$\mu_i = \begin{cases} 0 \text{ if the unit is not in the sample} \\ c_i \geq 0 \text{ if the unit is in the sample} \end{cases}$$

The population total for a characteristic is: $\sum_{i=1}^{N} y_i = Y$

where for brevity Σ is defined as the summation over N units; and, mean is: $\bar{Y} = \frac{Y}{N}$

The estimator of Y is: $\hat{Y} = \Sigma\mu_i y_i = \Sigma'\mu_i y_i$

where Σ' is defined as summation over the n units in the sample since $\mu_i = 0$ for the N-n units not selected.

The estimator $\hat{Y}$ has a distribution because the μ_i have a distribution; however, the μ_i are not necessarily independent.

Theorem 1: A necessary and sufficient condition that $E\hat{Y} = Y$ is that $E(\mu_i) = 1$ since

$$\hat{Y} = E\Sigma\mu_i y_i = \Sigma E(\mu_i) y_i = \Sigma 1 \cdot y_i = Y$$

The estimator of $\bar{Y}$ is: $\hat{\bar{Y}} = \frac{E\Sigma'\mu_i y_i}{N} = \frac{\hat{Y}}{N}$

2.1.1 Equal Probability of Selection (EPS)

The probability of selection for the i^{th} unit on any draw is $\frac{1}{N}$.
For a sample of n independent draws when n is fixed, then

$$\mu_i = \begin{cases} 0 \text{ with probability } 1 - \frac{n}{N} = \frac{N-n}{N} \\ C_i \text{ with probability } \frac{n}{N} = \sum_{i=1}^{n} \frac{1}{N} \; (C_i = \text{constant } C) \end{cases}$$

$$\therefore E(\mu_i) = 0 \cdot \frac{N-n}{N} + C \cdot \frac{n}{N} = \frac{Cn}{N}$$

If the estimator is to be unbiased, Theorem 1 must be satisfied, or $E(\mu_i) = 1$.

$$\therefore \frac{Cn}{N} = 1 \text{ or } C = \frac{N}{n} \text{ (i.e., "expansion" or "jack up" factor)}$$

An unbiased estimator of the total Y is:

$$\hat{Y} = E\Sigma\mu_i y_i = \Sigma y_i \, E(\mu_i) = \frac{N}{n} \Sigma' y_i \quad (\text{or } \frac{N}{n} \overset{\nu}{\Sigma'} t_i y_i) \text{ where}$$

t_i is the number of times a unit is selected and ν is the distinct units.
The mean is estimated by:

$$\hat{\bar{Y}} = \frac{\hat{Y}}{N} = \frac{1}{N}\frac{N}{n} \Sigma' y_i = \frac{1}{n} \Sigma' y_i \quad (\text{or } \frac{1}{n} \overset{\nu}{\Sigma'} t_i y_i)$$

where it is clear that $\nu \leq n$ in the above formulas.

2.1.2 Unequal Probability of Selection (UEPS)

This implies some information (at least ordinal) is available in the frame for each unit, besides the unit labels, which is useful in assigning probabilities to all units in the frame. While there may be several different kinds of information available for each of the N units, we shall require that the information for each unit will be reduced to a single number X_i for each of the N units. The probabilities of selection P_i will depend on the X_i's and the selection procedure.

<u>Notation</u>

	Universe of Distinguishable Units	
Unit labels - L	$1,2,3,\ldots\ldots\ldots\ldots i,\ldots\ldots N$	Total
Characteristic value	$y_1,y_2,y_c,\ldots\ldots\ldots y_i,\ldots\ldots y_N$	Y
Probability of selection	$P_1,P_2,P_3,\ldots\ldots\ldots P_i,\ldots\ldots P_N$	1
Weight	$\mu_1,\mu_2,\mu_3,\ldots\ldots\ldots \mu_i,\ldots\ldots \mu_N$	
Number of times unit selected	$t_1,t_2,t_3,\ldots\ldots\ldots t_i,\ldots\ldots t_N$	n

The weight variables are defined as

$$\mu_i = \begin{cases} 0 \text{ with probability } 1-np_i \\ c_i t_i > 0 \text{ with probability } np_i \end{cases}$$

where p_i = the probability of selecting the i^{th} unit

and $E(t_i) = np_i$, and $P(t_1 \ldots t_N) = \dfrac{n!}{t_1! \ldots t_N!} p_1^{t_1} \cdot p_2^{t_2} \cdots p_N^{t_N}$

$$\therefore E(\mu_i) = E(c_i t_i) = 0\cdot(1-np_i) + c_i np_i = nc_i p_i$$

If the estimator is to be unbiased $E(\mu_i) = 1$

$$\therefore nc_i p_i = 1 \quad \text{or} \quad c_i = \frac{1}{np_i} \quad \text{hence } \mu_i = \frac{t_i}{np_i}$$

An unbiased estimator of the total is:

$$\hat{Y} = E\,\Sigma\mu_i y_i = \Sigma y_i E(\mu_i) = \frac{1}{n}\,\Sigma' \frac{t_i y_i}{p_i}$$

The estimator of $\bar{Y}$ is:

$$\hat{\bar{Y}} = \frac{\hat{Y}}{N} = \frac{\frac{1}{n}\Sigma' \frac{t_i y_i}{p_i}}{N} = \frac{1}{nN}\,\Sigma' \frac{t_i y_i}{p_i}$$

Two methods of selection which may be employed are:

(a) The Hansen-Hurwitz Method - All N values of X_i known before selection

(1) Form cumulative sums S_i of the X_i where $S_i = S_{i-1} + X_i$, $i=1,2,\ldots..N$ and $S_0 = 0$.

(2) Draw a random number R between 0 and S_N

(3) Select the i^{th} unit if $S_{i-1} < R \leq S_i$,

(4) Repeat (2) and (3) until all n units are selected.

(b) The Lahiri Method - knowledge of X_i required only for selected units

(1) Select two random numbers: one from 1 to N, called R_1, which identifies a particular unit, and the other from 0 to X*, called R_2, where X* is the maximum value possible for any of the X_i (or use a larger value).

(2) For the unit corresponding to R_1 determine if $R_2 \leq X_i$, if so select the i^{th} unit; otherwise repeat (1) until a selection is made, then

(3) Repeat steps (1) and (2) until n selections have been made in step (2).

For both schemes $P_i = \frac{X_i}{S_N}$ where $S_N = \sum_{i=1}^{N} X_i$, and N is known.

2.2 Sampling Without Replacement - Unordered Samples (Method 1)

Notation - same as introduced in 2.1

2.2.1 Equal probability of selection (EPS)

The total number of possible samples of size n is $\binom{N}{n}$. The total number of possible samples of size n which contain a particular unit, i.e., the i^{th} unit, is $\binom{N-1}{n-1}$. The total number of possible samples of size n which contain a particular pair of units, i and j, is $\binom{N-2}{n-2}$. The probability of selecting the i^{th} unit is:

$$P = \frac{\binom{N-1}{n-1}}{\binom{N}{n}} = \frac{n}{N}$$

The weight variables are defined as

$$\mu_i = \begin{cases} 0 \text{ with probability } 1 - \frac{n}{N} \\ C_i \text{ with probability } \frac{n}{N} \quad (C_i = \text{constant } C) \end{cases}$$

$$\therefore E(\mu_i) = 0 . \frac{N-n}{N} + C . \frac{n}{N} = \frac{Cn}{N}$$

For unbiasedness $\frac{Cn}{N} = 1$, $\therefore C = \frac{N}{n}$.

An unbiased estimator of the total is

$$\hat{Y} = E\sum_{i=1}^{N} \mu_i y_i = \Sigma y_i\ E(\mu_i) = \frac{N}{n}\Sigma' y_i$$

and the mean is estimated by

$$\hat{\bar{Y}} = \frac{\hat{Y}}{N} = \frac{1}{n}\Sigma' y_i$$

Both of these results are the same as we obtained in Section 2.1 for EPS samples.

Systematic sampling using EPS-WOR technique which is convenient in practice because of its simplicity in execution. The technique consists in selecting every K^{th} unit starting with the unit corresponding to a random number R from 1 to K where K is taken as the integer nearest to N/n which is referred to as the sampling interval. A sample selected by this procedure is termed a systematic sample with a random start. It may be seen by an inspection of the possible units in the sample that the selection of R determines the whole sample. This procedure amounts to selecting one of the K possible groups of units (i.e., clusters) into which the universe can be divided.

In addition to the convenience in practice, the procedure provides more efficient estimators than simple random sampling under many conditions. Namely, each group of K units may be thought of as being ordered to achieve homogeneity over the universe. That is, the universe is effectively stratified into K strata with one unit being selected from each stratum.

In many universes the units are found already arrayed in strata based on the proximity of the units in their natural ordering. A geographic ordering of the units frequently provide a natural stratification which may lead to a more efficient estimator than simple random sampling. Likewise, systematic sampling over a time interval may prove more efficient than simple random sampling over time.

The estimator for systematic sampling is the same as given above, namely

$$\hat{\bar{Y}} = \frac{1}{n}\Sigma' y_i.$$

However, the selection of only one cluster of K units based on a single random start does not permit an unbiased estimate of the variance. To overcome this difficulty, m systematic samples of size N/nm are selected using m random starts and K = N/nm. The mean is than a simple average of the m sample means for each of the systematic samples of size K. The use of m systematic sample is referred to as replicated or interpenetrating sub-samples.

2.2.2 Unequal Probability of Selection (UEPS)

This case is very difficult to handle because of the calculation of the probability p_i and the joint probability (p_{ij}) of the i^{th} and j^{th} units which vary depending on the order in which the units are selected. We shall look at several types of selection procedures which make the calculation of p_i and p_{ij} manageable.

Notation - same as introduced in 2.1.2

We consider methods that use the Horvitz-Thompson estimator (HT) which is the only unbiased estimator. Let P(s) denote the probability of selection of a fixed number of n units without replacement, and S denote the set of all $\binom{N}{n}$ possible samples of size n.

π_i = probability that i^{th} unit is in the sample

π_{ij} = probability the i^{th} and j^{th} units are both in the sample

Then $\pi_i = \sum_{s \ni i} P(s)$, and $\pi_{ij} = \sum_{s \ni i,j} P(s)$

where the sum π_i is taken over all samples of size n containing the i^{th} unit, and the sum π_{ij} is over all samples containing the i^{th} and j^{th} units.

$$\sum_{i=1}^{N} \pi_i = n , \quad \sum_{j \neq i}^{N} \pi_{ij} = (n-1)\pi_i \quad \text{and} \quad \sum\sum_{i \neq j}^{N} \pi_{ij} = n(n-1)$$

Theorem 2: A set of necessary and sufficient conditions for the estimability of any linear function $\sum^{N} \mu_i y_i$ is $\pi_i > 0$ if $\mu_i \neq 0$.

Consider the estimator $\hat{Y} = \sum_{i=1}^{N} a_i c_i y_i$ where a's are random variables defined as

$$a_i = \begin{cases} 1 \text{ if the } i^{th} \text{ unit is in the sample} \\ 0 \text{ otherwise} \end{cases}$$

$$E(a_i) = 0(1-\pi_i) + 1\cdot\pi_i = \pi_i \text{ , } E(a_i a_j) = \pi_{ij}$$

For the estimator to be unbiased

$$c_i\pi_i = 1 \text{ or } c_i = \frac{1}{\pi_i} \text{ hence } \mu_i = \frac{a_i}{\pi_i}$$

$$\therefore \hat{Y} = \sum^{N} \frac{a_i y_i}{\pi_i} = \Sigma' \frac{y_i}{\pi_i}$$

which is the Horvitz-Thompson estimator of the total.

The mean is estimated by

$$\hat{\bar{Y}} = \frac{\hat{Y}_{HT}}{N} = \frac{1}{N} \quad \Sigma' \frac{y_i}{\pi_i}$$

If the selection procedure is such that π_i is proportional to y_i, the estimator reduces to a constant, and thus has zero variance. In practice we search for measures of size X_i proportional to y_i and try to have a selection procedure based on the X_i such that π_i is proportional to X_i since y_j is unknown.

We now consider some sampling methods for which the sample estimate of $V(\hat{Y}_{HT})$ is non-negative. In addition, we would like to impose certain other minimum requirements:

(a) π_i proportional to X_i, where X_i was proportional to $y_i(\pi_i = np_i.)$. This is necessary for sampling efficiency.

(b) $V(\hat{Y}_{HT})$ is always smaller than the variance in with replacement sampling (section 2.1.2).

(c) $\pi_{ij} \leq \pi_i \pi_j$ for all $i \neq j$. This is the condition necessary for non-negativity of variance estimator.

(d) $\pi_{ij} > 0$ for all $i \neq j$. Condition for estimability.

(e) Computations relatively simple.

Methods for $n = 2$

(a) Brewer method - The procedure is based on construction of 'revised' sizes X_i' which make

$$\pi_i = 2p_i \text{ . } (\pi_i = np_i)$$

(1) Select the first unit with probability proportional to the revised sizes X_i'

(2) Select the second unit with probabilities proportional to the original probabilities p_i of the remaining units.

Durbin has obtained the same result for n = 2, but the method extends for sample sizes greater than 2.

Durbin procedure for n = 2, the first unit is drawn with probability p_i, and the second unit from the remainder of the population

$$p_i = \frac{X_i}{X}$$

$p_{j.i}$ = prob. (selection of j/i already selected)

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

and $\pi_i = 2p_i$ (n times probability of i unit)

$$\pi_{ij} = 2p_i p_{j.i} = \overbrace{2p_j p_{i.j}}^{np_{ij}} \text{ (n times compound probability of i + j)}$$

(b) Murthy's Method

(1) Draw the first unit with probability $\dfrac{X_i}{\sum^N X_i}$

(2) Draw the second unit with probability $\dfrac{X_j}{\sum^N X_i - X_j}$

However, the ordering effect on p_i is removed by considering all orderings. The estimated total for the Murthy estimator for n = 2 is:

$$\hat{Y} = \frac{1}{2-p_1-p_2}\left[\frac{Y_1}{p_1}(1-p_2) + \frac{Y_2}{p_2}(1-p_1)\right]$$ based on the ordered estimator in 2.3.1

where p_1 = probability that i^{th} unit is drawn first

p_2 = probability that j^{th} unit is drawn first.

Method for Any Size n

(a) Rao, Hartley, Cochran Method (RHC)

(1) Split the population at random into n groups of sizes $N_1, N_2, \ldots N_n$ where $N_1 + N_2 + \ldots + N_n = N$.

(2) Draw a sample of size one with probabilities proportional to size from each of these n groups independently.

(3) And $N = nR+K$, $N_i = R$ or $N_i = R+1$, $0 \leq K<n$

The unbiased estimator of the total is

$$\hat{Y} = \sum_{i=1}^{N} \mu_i y_i = \sum^{N} y_i \, E(\mu_i) = \Sigma' \frac{y_i P_t}{p_i}$$

where P_t = probability of selecting a unit from the t^{th} group (the probability of selecting the group), that is

$$P_t = \sum_{i=1}^{N_t} p_i$$

p_i = probability of selecting the i^{th} unit from the t^{th} group ($p_i = P(i/t)$)

The probability of selecting the i^{th} unit in the t^{th} group $P_i = p_i/P_t$

A disadvantage of this method is that there exists another estimator which has uniformly smaller variance but this is generally considered a theoretical rather than a practical disadvantage since the other estimator is very difficult to compute.

(b) Systematic Selection

A generalization can be made from systematic sampling with equal probabilities. Cumulate the measures of sizes of the units and assign them the range 1 to X_1, $X_1 + 1$ to $X_1 + X_2$, $X_1 + X_2 + 1$ to $X_1 + X_2 + X_3$, and so on. In order to select a sample of size n, a random number is taken between 1 and $K = X/n$. The units in the sample are those in whose range lie the random number i and all other numbers i+K, i+2K,......., obtained by adding K successively to i. If there is any unit whose measure of size $\geq X/n$, it is removed before the sample selection procedure begins and included in the sample with certainty. The probability of the i^{th} unit is $X_i/(X/n) = np_i$. No simple formula for π_{ij} can be written down. For a specific arrangement of units, this can be calculated by finding out which random numbers (from 1 to X/n will selected the

i^{th} and j^{th} units. If m_{ij} is the number of such random numbers, $\pi_{ij} = nm_{ij}/X$. If N is not too large, this is certainly feasible. Thus, unless π_{ij} can be computed, the variance of the estimator cannot be derived.

To insure that exactly n units, and not (n-1) or (n+1), are selected, circular systematic selection can be used. That is, a random number R between 1 and X is selected and multiples of K are added and subtracted from R to determine the n units to be included in the sample. An unbiased estimator of the population total is given by:

$$\hat{Y} = \sum_{i=1}^{n} \frac{y_i}{\pi_i}$$

(c) Murthy's Estimator

This unordered estimator is obtained by weighting all the possible "ordered estimators" (given in 2.3.1) derived by considering all possible orders of selection of the given samples with their respective probabilities. The estimator can be shown to be of the form

$$\hat{Y} = \frac{1}{P(S)} \sum_{i=1}^{n} y_i \; P(S|i)$$

where P(S) is the probability of getting the S^{th} unordered sample and P(S|i) is the conditional probability of getting the S^{th} sample given the i^{th} unit was selected on the first draw.

2.3 Sampling Without Replacement - Ordered Samples (UEPS)

For unequal probability of selection schemes, ordered estimators have been used because of the ease in calculating the conditional probabilities based on the order of selection. We consider only a few schemes.

2.3.1 Des Raj Estimator

(A) Case for $n = 2$

For the first unit drawn we have the estimate

$$E\hat{Y} = \sum^{N} \mu_i y_i = \Sigma'\left(\frac{y_1}{P_1}\right) = \frac{y_1}{P_1} = z_1$$

After the second unit is drawn, we estimate

$$z_2 = y_1 + y_2 \frac{1-P_1}{P_2} \; ; \quad E(z_2 | z_1) = y_1 + (Y-y_1) = Y$$

Considering both estimates

$$\hat{Y} = \frac{z_1+z_2}{2} = \frac{1}{2}\left[(1+P_1)\frac{y_1}{P_1} + (1-P_1)\frac{y_2}{P_2}\right]$$

(B) Case for any n

Define z_i as above and

$$z_i = y_1 + y_2 + y_3 + \ldots + y_{i-1} + y_i \frac{(1-P_1-P_2-\ldots P_{i-1})}{P_i}$$

$$\hat{Y} = \frac{1}{n} \sum^{n} z_i$$

where $P_i = P(i | i-1, i-2, \ldots 1)$

Since a sample of size n can be ordered in n! ways, an unordered sample estimate can be derived by considering all n! estimates and averaging them. Murthy's estimator corresponds to the average of all the ordered estimates for the n units selected.

2.3.2 Midzuno System of Sampling

The first unit is selected with unequal probabilities, and for all subsequent draws the units are selected with equal probabilities and without replacement. The estimator for this scheme is the Horvitz-Thompson

Estimator. The initial probability of selection P_i must satisfy the condition (i.e., a minimum value)

$$P_i > \frac{n-1}{n(N-1)} .$$

The probability of the largest unit is maximized when the other units have equal probability of inclusion. This probability barely satisfies the minimum size condition.

$$P_i = 1 - \sum_{i=1}^{N-1} P_i' \quad \text{where } P_i' \text{ for the other N-1 units must satisfy}$$

$$P_i' > \frac{n-1}{n(N-1)}$$

Hence

$$\sum_{i-1}^{N-1} P_i' > \sum_{i=1}^{N-1} \frac{n-1}{n(N-1)} = \frac{n-1}{n} \quad \text{if all probabilities are equal.}$$

Thus P_L must be smaller than $1 - \frac{n-1}{n} = \frac{1}{n}$, or

$n =$	2		5		n
$P_L =$	$\frac{1}{2}$	.....	$\frac{1}{5}$		$\frac{1}{n}$
$P_i =$	$\frac{1}{2(N-1)}$	.....	$\frac{4}{5(N-1)}$		$\frac{n-1}{n(N-1)}$

Theorem 3: For every ordered estimator there exists an unordered estimator which has smaller variance.

2.4 Sampling With Replacement - (Method 2)

We start with an estimator of the mean and obtain its expected value by treating y_i as the random variable. The estimator is unbiased if the expected value corresponds to the mean value for the parameter in the population.

Notation:

	Universe of Distinguishable Units
Unit labels - L	1, 2, 3, ..., i, ..., N
Characteristic value	$y_1, y_2, y_3, \ldots, y_i, \ldots, y_N$

2.4.1 Equal Probability of Selection (EPS)

Consider the estimator $\bar{y}_n = \frac{1}{n} \sum_{i=1}^{n} y_i$

$$E(\bar{y}_n) = E(\frac{1}{n} \sum^{n} y_i) = \frac{1}{n} \{ E(y_1') + \ldots + E(y_r') + \ldots + E(y_n') \}$$

where y_r' corresponds to the unit selected on the r^{th} draw. For with replacement sampling, the probability of selection on the r^{th} draw, P_{ir}, for a unit is $\frac{1}{N}$.

$$\therefore E(y_r') = \sum^{N} P_{ir} y_i = \frac{1}{N} \sum^{N} y_i = \frac{Y}{N} = \bar{Y}$$

Substituting this results above

$$E(\bar{y}_n) = \frac{1}{n} \{ \bar{Y} + \bar{Y} + \ldots + \bar{Y} \} = \frac{n\bar{Y}}{n} = \bar{Y} \quad n = \text{terms in sum}$$

2.4.2 Unequal Probability of Selection (UEPS)

Consider the estimator $\bar{y}_n = \frac{1}{n} \sum^{n} \frac{y_i}{NP_i}$

where $\sum^{N} P_i = 1.$

Let $z_i = \frac{y_i}{NP_i}$ for brevity

$$E(z_i) = \sum^{N} P_i z_i = \sum^{N} \frac{y_i}{N} = \frac{1}{N} \sum^{N} y_i = \frac{Y}{N} = \bar{Y}$$

$$\therefore E(\bar{z}_n) = \frac{1}{n} \sum^{n} E(z_i) = \frac{n\bar{Y}}{n} = \bar{Y}$$

2.5 Sampling Without Replacement - Unordered Samples (Method 2)

Notation:

	Universe of Distinguishable Units
Unit labels - L	1, 2, 3, ..., i, ..., N
Characteristic value	$y_1, y_2, y_3, \ldots, y_i, \ldots, y_N$

2.5.1 Equal Probability of Selection (EPS)

Consider the estimator $\bar{y}_n = \frac{1}{n} \sum^{n} y_i$

$$E(\bar{y}_n) = E(\frac{1}{n} \sum^{n} y_i) = \frac{1}{n} \{ E(y_1') + \ldots + E(y_r') + \ldots + E(y_n') \}$$

where y'_r corresponds to the unit selected on the r^{th} draw. For without replacement sampling, the probability of the selection on the r^{th} draw

$$P_{ir} = \frac{N-1}{N} \cdot \frac{N-2}{N-1} \cdots \frac{N-r+1}{N-r+2} \cdot \frac{1}{N-r+1} = \frac{1}{N} .$$

$$\therefore E(y'_r) = \sum^{N} P_{ir} y_i = \frac{1}{N} \sum^{N} y_i = \frac{Y}{N} = \bar{Y}$$

Substituting this results above

$$E(\bar{y}_n) = \frac{1}{n} \{\bar{Y} + \bar{Y} + \ldots + \bar{Y}\} = \frac{n\bar{Y}}{n} = \bar{Y}$$

2.5.2 Unequal probability of selection (UEPS)

Each of the methods of selection may indicate a different estimator (see section 2.2.2) for the mean.

2.6 Sampling for Qualitative Characteristics

We consider only sampling without replacement for equal probability of selection schemes since it will be shown later that WOR sampling is more efficient than WR sampling. We shall not consider in detail UEPS schemes since the units either assume a value of 0 or 1 as a measure of size. Consequently, it is unlikely that UEPS schemes would be considered except in those situations where quantitive data was also being collected for the same sampling units.

Notation

	Universe of Distinguishable Units
Unit labels - L	1, 2, ..., i, ..., N
Attribute value	$a_1, a_2, \ldots, a_i, \ldots, a_N$
Weights	$\mu_1, \mu_2, \ldots, \mu_i, \ldots, \mu_N$

The weights are defined as before while a_i is equal to either 1 or 0 depending on whether the unit has the attribute or not.

2.6.1 Two classes - EPS

The sampling units in the universe are divided into two mutually exclusive classes. Let p and q denote the proportion of sampling units in the population belonging to Class 1 and Class 2, respectively. In a sample of n selected out of N, n_1 units will occur in Class 1 and n_2 in Class 2. The probability $P(n_1)$ is given by

$$P(n_1) = \frac{\binom{N_p}{n_1}\binom{N_q}{n_2}}{\binom{N}{n}}, \quad \sum_{n_1} P(n_1) = 1.$$

The variate n_1 or the proportion n_1/n is said to be distributed in a hypergeometric distribution. As N tends to be large, the distribution approaches the binomial.

An unbiased estimator of the total population size is

$$\hat{N}_1 = E\sum^{N} \mu_i a_i = \sum^{N} a_i E(\mu_i) = \frac{N}{n} \Sigma' a_i = \frac{N}{n} n_1$$

based on the results in section 2.2.1, and the proportion P is estimated by considering

$$E(n_1) = \sum_{n_1} n_1 \, P(n_1) = \sum_{n_1} n_1 \frac{N_p!}{n_1!(N_p - n_1)!} \; \frac{N_q!}{n_2!(N_q - n_2)!} \; \frac{n!(N-n)!}{N!}$$

$$E(n_1) = \frac{N_{np}}{N} \sum_{n_1} \frac{(N_p - 1)!}{(n_1 - 1)!(N_p - n_1)!} \; \frac{N_q!}{n_2!(N_q - n_2)!} \; \frac{(n-1)!(N-n)!}{(N-1)!}$$

$$= np \sum_{n_1} \frac{\binom{N_p - 1}{n_1 - 1}\binom{N_q}{n_2}}{\binom{N-1}{n-1}}$$ where the summation represents the probability

that in a sample of n-1, $n_1 - 1$ will fall in Class 1 and n_2 will fall in Class 2. The sum over all values of n_1 is 1.

$$\therefore E(n_1) = np$$

Consequently, the proportion p is estimated by

$$E(\hat{P}) = \frac{1}{n} E(n_1) = \frac{n_1}{n}, \text{ and}$$

$$E(\hat{q}) = \frac{n_2}{n}.$$

2.6.2 For K classes - EPS

$$P(n_i) = \frac{\binom{N_i}{n_i}\binom{N - N_i}{n - n_i}}{\binom{N}{n}}$$

$$P(n_1, n_2, \ldots n_K) = \frac{\binom{N_1}{n_1}\binom{N_2}{n_2}\cdots\binom{N_K}{n_K}}{\binom{N}{n}}$$

$$E(n_i) = np_i \quad \text{where } P_i = \frac{N_i}{N} \quad \text{and} \quad \sum^{K} N_i = N.$$

The total population size for the i class is:

$$\hat{N}_i = \frac{N}{n} n_i \ .$$

2.6.3 For K Classes - UEPS

We briefly consider sampling without replacement and unequal probability of selection based some measure of size related to a quantitive variable which was observed for the same set of n sample units and used a the basis for selecting the units. That is, as part of a multiple characteristic or multipurpose survey. We consider only the RHC method of sample selection. For the total number of units

$$\hat{N}_i = \sum^{N} \mu_i a_i = \sum^{N} a_i E(\mu_i) = \sum^{n} a_i \frac{P_t}{P_i} = \sum^{n_i} \frac{P_t}{P_i}$$

and for the proportion of units in the i^{th} category. We use f_i to indicate this fraction to avoid confusion with the use of P_i for the probability of selecting the i^{th} unit in the t^{th} group.

$$\hat{f}_i = \frac{\hat{N}_i}{N} = \frac{1}{N} \sum^{n_i} \frac{P_t}{P_i} \ .$$

2.7 Sampling for Quantitative and Qualitative Characteristics in Subpopulations

This sampling problem is concerned with subpopulations and is commonly known under the title of "Domain Theory." We are concerned with estimating the total of a quantitative characteristic for each of K subclasses in a population where the subpopulation sizes are unknown.

Notation - Same as given in 2.6 except the quantitive variable y_i is also defined for each unit in the universe of N.

2.7.1 Sampling With Equal Probabilities (EPS)

We define the quantitative characteristic for "domain theory"

$${}_j y_i = \begin{cases} y_i & \text{if the } i^{th} \text{ unit belongs to the } j^{th} \text{ class} \\ 0 & \text{otherwise} \end{cases}$$

${}_j n$ = number of y_i in the j^{th} class

The estimated total for the j^{th} class is

$${}_j\hat{Y} = \frac{N}{n} \sum^{n} {}_j y_i = \frac{N}{n} {}_j y_i = \frac{N}{n} {}_j\bar{y} E({}_j n) = N f_j \cdot {}_j\bar{y}$$

since the sample is chosen by simple random sampling from the entire population, the subsample ${}_jn$ can also be considered a random sample from Nf_j, hence ${}_j\bar{y}$ is unbiased for a given ${}_jn$.

2.8 Inverse Sampling - (EPS)

If the proportion p of units in a given class is very small, the method of estimation given previously may be unsatisfactory. In this method the sample size n is not fixed in advance. Instead, sampling is continued until a predetermined number of units, m, possessing the rare attribute have been drawn. To estimate the proportion p, the sampling units are drawn one by one with equal probability and without replacement. Sampling is discontinued as soon as the number of units possessing the rare attribute is equal to the predetermined number m.

$$P(n) = P\left\{\begin{array}{l}\text{In a sample of n-1 units}\\ \text{drawn from N, m-1 units}\\ \text{will possess the attribute}\end{array}\right\} . P\left\{\begin{array}{l}\text{The unit drawn at the}\\ n^{th}\text{ draw will possess}\\ \text{the attribute}\end{array}\right\}$$

$$= \frac{\binom{N_P}{m-1}\binom{N_q}{n-m}}{\binom{N}{n-1}} \cdot \frac{N_P-m+1}{N-n+1}$$

Since the possible values of n are: m, m+1,..., m+Nq, we have

$$\sum_{n\geq m} P(n) = 1.$$

An unbiased estimate of P is given by

$$\hat{P} = \frac{m-1}{n-1} \text{ and } \hat{N}_P = \hat{P}N.$$

2.9 Linear Estimators and Optimality Properties

A thorough examination of linear estimators has been underway since the Horvitz-Thompson paper on sampling without replacement in 1952. Seven or eight subclasses have been proposed, three subclasses will be considered below.

$$(1) \quad T_1 = \sum_{r=1}^{n} \alpha_r y_r$$

where $\alpha_r (r=1,2,...n)$ is the coefficient to be attached to the unit appearing in the sample at the r^{th} draw (no attention is paid to the unit label μ_i.) and is defined prior to sampling.

(2) $T_2 = \sum_{i \epsilon s_n} \beta_i y_i$

where β_i is the coefficient to be attached to the unit with the label μ_i whenever it is in the sample, i.e., i = 1,2,...N, and is defined in advance of sampling.

(3) $T_3 = \gamma_{s_n} (\sum_{i=1}^{n} y_i)_{s_n}$

where γ_{s_n} is a constant to be used as a weight when the sample s_n is selected and the weights γ_{s_n} is defined in advance for all s_n.

When EPS is used, the sample mean

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$$

is the best linear unbiased estimator (BLUE) in the class T_1. The proof rest on an extension of Markoff's Theorem. The sample mean $\bar{y}$ is the only unbiased linear estimator in the subclass T_2, and corresponds to the methods of weight variables. It is known that there is no BLUE in the subclass T_3. However, a BLUE may not exist in a broad class of unbiased linear estimators for any sampling design.

Chapter III. Multistage Sampling

3.0 Introduction

In single stage sampling procedures the unit selected was completely enumerated or observed. These units could have been a cluster of people living in the same household, a section of land, a city block, a fruit tree, an individual student, or a classroom. It is frequently necessary for reasons of efficiency in sampling or cost to consider multistage sampling in which only a part of a cluster of units is enumerated. The selection of only a part of the cluster leads to the use of multistage or subsampling designs. The number of units within a cluster may be thought of as a measure of size. Since the y values are fixed under the method of weight variables, only the distribution of the μ_i will be affected. A design is characterized by its weight variables.

3.1 Two Stage Sampling

3.1.1 Equal Probability Without Replacement at Both Stages

Notation:

	Population of primary units	Total
Primary number	$1, 2, \ldots, i, \ldots N$	N
Secondary Units in Primary	$M_1, M_2, \ldots, M_i, \ldots M_N$	M_0
Primary Weight Variables	$\mu_1, \mu_2, \ldots, \mu_i, \ldots \mu_N$	n
Primary Total	$Y_1, Y_2, \ldots, Y_i, \ldots Y_N$	Y
Primary Mean	$\bar{Y}_1, \bar{Y}_2, \ldots, \bar{Y}_i, \ldots, \bar{Y}_N$	$\bar{\bar{Y}}_N$

	Secondary units for i^{th} primary
Characteristic value	$y_{i1}, y_{i2}, \ldots, y_{it}, \ldots y_{iM_i}$
Weight variables	$v_{i1}, v_{i2}, \ldots, v_{it}, \ldots v_{iM_i}$

Stage 1: Select n primaries out of N (EPS, WOR)

Stage 2: Select m_i secondaries out of M_i (EPS, WOR)

Weight coefficients

$$\mu_i = \begin{cases} 0 \text{ with probability } 1 - \frac{n}{N} \\ C_i \text{ with probability } \frac{n}{N} \end{cases}$$

$$v_{it} = \begin{cases} 0 \text{ with probability } 1 - \frac{m_i}{M_i} \\ C_{it} \text{ with probability } \frac{m_i}{M_i} \end{cases}$$

Where V_{it} is correlated with $V_{it'}$, but not with the weight in another primary, say V_{jt}.

$$E(\mu_i) = 0\cdot(1-\frac{n}{N}) + C_i\frac{n}{N} = C_i\frac{n}{N}$$

Since $C_i\frac{n}{N} = 1$ for unbiasedness $C_i = \frac{N}{n}$

$$E(V_{it}) = 0\cdot(1-\frac{m_i}{M_i}) + C_{it}\frac{m_i}{M_i} = C_{it}\frac{m_i}{M_i}$$

Since $C_{it}\frac{m_i}{M_i} = 1$ for unbiasedness $C_{it} = \frac{M_i}{m_i}$

The estimator of the population total is

$$\hat{Y} = E\sum_{i}^{N}\sum_{t}^{M_i}\mu_i V_{it} y_{it} = \frac{N}{n}\Sigma'\frac{M_i}{m_i}y_{i.} = \frac{N}{n}\sum^{n} M_i\bar{y}_{i.}$$

where $y_{i.}$ is the total of the m_i secondary units. Then

$$\bar{Y} = \frac{N}{nM_0}\sum^{n} M_i\bar{y}_i \text{ where } \sum^{N} M_i = M_0.$$

In selecting primary units with probabilities proportional to size, selection with replacement is used for simplicity. In some surveys the number of primary units in a strata or population is rather small. In such situations, it is desirable to select units without replacement to produce the reduction in variance associated with a finite population correction factor. However, these gains affect mostly the between primary unit contribution of the variance, and to a lessor extent the within primary unit variances. Consequently, the gains in multistage sampling will be smaller. This is especially true when the primary sampling fraction is large and the within primary sampling fractions are small so the major part of the total variance is due to the second or lower stage units. The calculation of π_i and π_{ij} for primary units is more manageable if N is not too large.

Rule 1: The unbiased estimator of Y in multi-stage sampling is obtained by replacing Y_i by $\hat{Y}_i$ in the corresponding unbiased estimator of Y in single-stage sampling of clusters (i.e. primaries) when the clusters are completely enumerated.

3.1.2 Primaries Sampled With Unequal Probability With Replacement

Secondaries with equal probabilities without replacement, but the same m_i secondary units are used each time a primary is selected. The weight variables are:

$$\mu_i = \begin{cases} 0 \text{ with probability } 1 - np_i \\ c_i t_i \text{ with probability } np_i \end{cases}$$

where t_i = the number of times a primary is selected (multinomial variate)
p_i = the probability of the i^{th} primary being drawn

$$E(c_i t_i) = c_i \, np_i = 1, \therefore \; c_i = \frac{1}{np_i} \quad E(\mu_i) = \frac{t_i}{np_i} = 1$$

and $$V_{it} = \begin{cases} 0 \\ \frac{M_i}{m_i} \end{cases} \quad E(V_{it}) = 1 \text{ from previous section}$$

$$\hat{Y} = E \sum\sum_{it}^{NM_i} \mu_i V_{it} y_{it} = \frac{1}{n} \Sigma' \frac{t_i}{p_i} M_i \bar{y}_i$$

The use of the same m_i units each time a primary is selected will alter the variance formula as compared to a sampling scheme selecting a different set of m_i units each time a primary is selected.

$$\hat{\bar{Y}} = \frac{\hat{Y}}{\sum^{N} M_i} = \frac{1}{nM_0} \Sigma' \frac{t_i}{p_i} M_i \bar{y}_i$$

where $M_0 = \sum^{N} M_i$ the total number of secondaries in population.

3.1.3 Primaries and Secondaries Sampled Without Replacement

Stage 1: Select n primaries out of N (PPS, WOR)

Stage 2: Select m_i secondaries out of M_i (EPS, WOR)

Weight coefficients (From section 2.2.2)

$$\mu_i = \begin{cases} 0 \text{ with probability } 1-np_i \\ c_i t_i > 0 \text{ with probability } np_i \end{cases}$$

where $E(t_i) = np_i = \pi_i \qquad E(t_i t_j) = \pi_{ij}$

$$\therefore \; E(\mu_i) = c_i np_i = 1 \text{ hence } c_i = \frac{1}{np_i} \text{ for unbiasedness}$$

$$V_{it} = \begin{cases} 0 \text{ with probability } 1 - \frac{m_i}{M_i} \\ c_{it} \text{ with probability } \frac{m_i}{M_i} \end{cases}$$

$$E(V_{it}) = c_{it}\frac{m_i}{M_i} = 1 \text{ hence } c_{it} = \frac{M_i}{m_i} \text{ for unbiasedness}$$

$$\hat{Y} = E\sum\sum^{NM_i} \mu_i V_{it} y_{it} = \frac{1}{n}\sum_i^n \frac{1}{P_i}\frac{M_i}{m_i}\sum_t^{m_i} y_{it} = \frac{1}{n}\sum^n \frac{1}{P_i} M_i \bar{y}_{i.}$$

$$\hat{\bar{Y}} = \frac{1}{nM_0}\sum^n \frac{1}{P_i} M_i \bar{y}_i$$

3.2 Three Stage Sampling

3.2.1 Equal Probability Without Replacement at All Stages

Stage 1: n primaries are selected from N

Stage 2: m_i secondaries are selected from M_i for a primary

Stage 3: k_{it} tertiaries are selected from K_{it} for a secondary

Notation: y_{ith} is the observation in the i^{th} primary, t^{th} secondary, and h^{th} tertiary.

Weight cpefficients:

$$\mu_i = \begin{cases} 0 \text{ with probability } 1 - \frac{n}{N} \\ c_i \text{ with probability } \frac{n}{N} \end{cases}$$

$$E(\mu_i) = 0\cdot(1-\frac{n}{N}) + c_i\frac{n}{N} = 1 \text{ for unbiasedness}$$

$$\therefore c_i = \frac{N}{n}$$

$$V_{it} = \begin{cases} 0 \text{ with probability } 1 - \frac{m_i}{M_i} \\ c_{it} \text{ with probability } \frac{m_i}{M_i} \end{cases}$$

$$E(V_{it}) = 0\cdot(1-\frac{m_i}{M_i}) + c_{it}\frac{m_i}{M_i} = 1 \text{ for unbiasedness}$$

$$\therefore c_{it} = \frac{M_i}{m_i}$$

$$W_{ith} = \begin{cases} 0 \text{ with probability } 1-\frac{k_{it}}{K_{it}} \\ c_{ith} \text{ with probability } \frac{k_{it}}{K_{it}} \end{cases}$$

$$E(W_{ith}) = 0\cdot\ (1-\frac{k_{it}}{K_{it}}) + c_{ith}\frac{k_{it}}{K_{it}} = 1 \text{ for unbiasedness}$$

$$\therefore W_{ith} = \frac{K_{it}}{k_{it}}$$

$$\hat{Y} = E\ \Sigma\Sigma^{NM_i K_{it}}\Sigma\ \mu_i V_{it} W_{ith} y_{ith} = \frac{N}{n}\sum_i^n \frac{M_i}{m_i}\Sigma^{m_i}\frac{K_{it}}{k_{it}}\Sigma^{k_{it}} y_{ith}$$

$$= \frac{N}{n}\sum_i^n \frac{M_i}{m_i}\Sigma^{m_i}\frac{K_{it}}{k_{it}} y_{it}.$$

3.3 Use of Conditional Expectation

In the previous sections, we have relied largely on the weighted variable technique to derive unbiased estimators. The weighted variable technique is extremely useful since it provides us with a positive method of finding unbiased estimators for any design. However, a sample estimator for a parameter is frequently proposed based on certain heuristic considerations and its expectation needs to be evaluated. It is proposed to examine several alternative estimators for parameters previously derived to develop an appreciation for the usefulness of the conditional expectation technique in multistage designs. There are many situations in which the two methods are combined in deriving an estimator for a parameter.

The total expectations are always taken starting with the last stage of selection, and proceeding to the next higher level. The effect of the conditioning event (i.e., the selection of a particular unit at a given stage) is to permit us to treat the units selected as subpopulations (strata) when taking expectations over the stages below a given stage.

3.3.1. Two Stage Sampling (3.1.1 for Method 1)

We consider an alternative estimator of the mean which is based on the average of the primary means in the sample.

That is: $\hat{\bar{Y}}_1 = \frac{1}{n}\overset{n}{\Sigma}\frac{1}{m_i}\overset{m_i}{\Sigma}\; y_{it} = \frac{1}{n}\overset{n}{\Sigma}\;\bar{y}_i$

$$\hat{\bar{Y}}_1 = E(\frac{1}{n}\overset{n}{\Sigma}\bar{y}_i) = E\{\frac{1}{n}\overset{n}{\Sigma}E(\bar{y}_i|i)\}$$

where we now treat the i^{th} primary as a "subpopulation" by taking the conditional expectation of the random variable y_{it} conditioned by the random event i. Using EPS within the primary, $E(\bar{y}_i|i)=E(\frac{1}{m_i}\overset{m_i}{\Sigma}y_{it}|i)=$ $\frac{1}{M_i}\overset{M_i}{\Sigma}y_{it}$, the expectation over all y_{it} in the i^{th} primary.

$$\therefore\ \hat{\bar{Y}}_1 = E(\frac{1}{n}\overset{n}{\Sigma}\bar{Y}_i)$$

Now taking the expectation over all N primaries

$\hat{\bar{Y}}_1 = \frac{1}{N}\overset{N}{\Sigma}\bar{Y}_i = \bar{\bar{Y}}_N$ which is the average primary mean in the population of N primaries, but is a biased estimator for the population mean of the $\overset{N}{\Sigma}M_i = M_0$ units.

Another estimator that could be considered is:

$$\hat{\bar{Y}}_2 = \frac{\overset{n}{\Sigma}M_i\bar{y}_i}{\overset{n}{\Sigma}M_i}$$

which is based on the estimated totals for the primaries selected divided by the number of secondary units in the n selected primary units. Both the numerator and demonimator will vary from sample to sample due to different secondary units being chosen. We start by taking expectations at the lowest level or stage in the design.

$$\hat{\bar{Y}}_2 = E\left\{\frac{\overset{n}{\Sigma}M_iE(\bar{y}_i|i)}{\overset{n}{\Sigma}M_i}\right\} = E\left\{\frac{\overset{n}{\Sigma}M_i\bar{Y}_i}{\overset{n}{\Sigma}M_i}\right\} = E\left\{\frac{\overset{n}{\Sigma}Y_i}{\overset{n}{\Sigma}M_i}\right\}$$

which is now in the form of a ratio estimator. The numerator and demonimator can both be divided by n so each will resemble the last estimator ($\hat{\bar{Y}}_1$). The average primary total divided by the average primary size:

$\hat{\bar{Y}}_2 = E\{\frac{y_n}{\bar{M}_n}\}$ which based on Chapter I gives

$$\hat{\bar{Y}}_2 = \frac{Ey_n}{E\bar{M}_n} - \frac{Cov(\frac{y_n}{\bar{M}_n},\bar{M}_n)}{E\bar{M}_n}$$

$$= \frac{Y_N}{\bar{M}_N} - \frac{Cov(\frac{Y_N}{\bar{M}_N},\bar{M}_N)}{\bar{M}_N}$$ which is biased unless all M_i are equal or the covariance is zero.

3.3.2. Two Stage Sampling (3.1.2 for Method 1)

We generalize section 3.1.2 with regard to the second stage of sampling by specifying only that the sub-sampling scheme within a primary provides an unbiased estimate of the primary total.

The estimator proposed for the population total was:

$$\hat{Y} = \frac{1}{n} \Sigma' \frac{t_i}{p_i} M_i \bar{y}_i = \frac{1}{n} \Sigma' \frac{t_i}{p_i} \hat{Y}_i$$

Now

$$\hat{Y} = E\{\frac{1}{n} \Sigma' \frac{t_i}{p_i} E(\hat{Y}_i | i)\}$$

$$= E\{\frac{1}{n} \Sigma' \frac{t_i}{p_i} Y_i\}$$

$$= \frac{1}{n} \Sigma' t_i \; E(\frac{Y_i}{p_i}) = \frac{1}{n} \cdot nY = Y$$

which is an unbiased estimator of the population total.

3.3.3. Three Stage Sampling (3.2.1 for Method 1)

The basic technique is illustrated for three stages using the estimator derived in 3.2.1.

The estimator for the population total was:

$$\hat{Y} = \frac{N}{n} \Sigma^{n} \frac{M_i}{m_i} \Sigma^{m_i} \frac{K_{it}}{k_{it}} \Sigma^{k_{it}} y_{ith}$$

$$\hat{Y} = E(\frac{N}{n} \Sigma^{n} \frac{M_i}{m_i} \Sigma^{m_i} \frac{K_{it}}{k_{it}} \Sigma^{k_{it}} y_{ith})$$

$$= E[\frac{N}{n} \sum^{n} \frac{M_i}{m_i} \sum^{m_i} \frac{K_{it}}{k_{it}} \sum^{k_{it}} E_3(y_{ith} | it)]$$

where

$$E_3(y_{ith} | t) = \bar{Y}_{it.} \qquad \sum^{k_{it}} \bar{Y}_{it.} = k_{it} \bar{Y}_{it.}$$

or

$$\hat{Y} = E[\frac{N}{n} \sum^{n} \frac{M_i}{m_i} \sum^{m_i} E(K_{it}\bar{Y}_{it} | i)]$$

$E(K_{it}\bar{Y}_{it.} | i) = E_2(\hat{Y}_{it.} | i) = \bar{Y}_{i..}$. Where $\hat{Y}_{it}$ is the sample total for the t^{th} secondary in the i^{th} primary.

$$\hat{Y} = E[\frac{N}{n} \sum^{n} M_i\bar{Y}_{i..}]$$

Since $M_i\bar{Y}_{i..}$ is the total for the i^{th} primary

$$\hat{Y} = \frac{N}{n} \sum^{n} E_1(M_i\bar{Y}_{i..}) = \frac{N}{n} \sum^{n} E_1(\hat{Y}_{i..}) = \frac{N}{n} (n\bar{Y}...)$$

where $\bar{Y}...$ is the average primary total

$$\hat{Y} = N\bar{Y}... = Y.$$

Chapter IV. Estimation of Variances

4.0 Introduction

The estimation of variances is developed in terms of deriving the expression for the population variance and then seeking an estimator of the population variance based on the sample data. An estimator of the population total or mean require knowledge of the probability of selection for each unit in the population or the units in the sample depending on class of estimators being considered. The estimation of variances requires the knowledge of joint probabilities of each pair of units as well as the probabilities of the individual units for estimability. The requirement is that both p_i and p_{ij} be greater than zero for a finite population. Variances will be derived for the class of estimators T_1 and T_2 introduced in Chapters II and III.

Of particular importance in these derivations will be the use of conditional expectations and probabilities, especially for multistage sampling designs. The notation established in Chapters II and III will be followed. It should be noted that we are concerned with the variance of estimators and not the variance of the population characteristic which was defined in Chapter I, Section 1.5 (16).

However, a definite relationship between the variance of the estimator of the population parameter and the variance of the characteristic measured (or observed) in the sample does exist.

4.1 Single Stage Designs - Population Variances

4.1.1 Equal Probability of Selection With Replacement

Notation from Section 2.1.1

a. For the subclass T_2 of linear estimators,

$$\hat{Y} = \Sigma \mu_i y_i = \frac{N}{n} \Sigma' y_i = \frac{N}{n} \overset{\nu}{\Sigma} t_i y_i$$

Taking expected values in terms of distinct units and fixed sample size n; t_i is distributed as a binomial variable.

$$E(t_i) = n.(\frac{1}{N}) = \frac{n}{N}$$

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

$$E(t_i t_j) = n(n-1)(\frac{1}{N})(\frac{1}{N}) = \frac{n(n-1)}{N^2}$$

$$V(t_i) = \frac{n(N-1)}{N^2} + (\frac{n}{N})^2 - (\frac{n}{N})^2 = \frac{n(N-1)}{N^2}$$

$$Cov(t_i t_j) = \frac{n(n-1)}{N^2} - (\frac{n}{N})^2 = -\frac{n}{N^2}$$

Using (19) of Chapter I, Section 1.5

$$V(\hat{Y}) = \frac{N^2}{n^2} V(\sum^N t_i y_i) = \frac{N^2}{n^2} [\sum^N y_i^2 V(t_i) + \sum_{i \neq j}^N \sum^N y_i y_j Cov(t_i t_j)]$$

$$= \frac{N^2}{n^2} [\sum^N y_i^2 \frac{n(N-1)}{N^2} + \sum_{i \neq j}^N \sum^N y_i y_j (\frac{-n}{N^2})]$$

$$= \frac{N^2 n}{n^2 N^2} [N\sum^N y_i^2 - \sum^N y_i^2 - (\sum^N y_i)(\sum^N y_j) + \sum^N y_i^2]$$

(1) $$= \frac{N^2 n}{n^2 N^2} [N\sum^N y_i^2 - (\sum^N y_i)^2] = \frac{N^2}{n} [\frac{\sum^N y_i^2 - \frac{(\sum^N y_i)^2}{N}}{N}]$$

$$= \frac{N^2}{n} [E(y_i^2) - \{E(y_i)\}^2] = \frac{N^2}{n} \sigma_y^2$$

b. For the subclass T_1 of linear estimators

$$\hat{Y} = \frac{N}{n} \sum^n y_r'$$

$$V(\hat{Y}) = \frac{N^2}{n^2} \sum^n V(y_r') = \frac{N^2}{n^2} . n\sigma_y^2 = \frac{N^2}{n} \sigma_y^2$$

The variance of the estimator of the population total is related to the variance of an individual unit of the population by the factor $\frac{N^2}{n}$ and it follows that the variance of the mean is related by the factor $\frac{1}{n}$.

4.1.2 Equal Probability of Selection Without Replacement

Notation from Section 2.2.1

a. From the subclass T_2 we have

$$\hat{Y} = \Sigma\mu_i y_i = \frac{N}{n}\Sigma' y_i$$

since all units are distinct $t_i = 1$, i.e. a constant.

$$E(\mu_i) = 1$$

$$E(\mu_i^2) = \frac{N}{n}$$

$$E(\mu_i \mu_j) = \binom{N-2}{n-2} \cdot \left(\frac{N}{n}\right)^2 = \frac{N(n-1)}{n(N-1)}$$

$$V(\mu_i) = \frac{N}{n} - 1 = \frac{N-n}{n}$$

$$Cov(\mu_i \mu_j) = \frac{N(n-1)}{n(N-1)} - 1 = -\frac{N-n}{n(N-1)}$$

where the total number of possible samples of size n

(1) Are $\binom{N}{n}$, and

(2) Which contain a particular unit are $\binom{N-1}{n-1}$,

(3) Which contain a particular pair of units are $\binom{N-2}{n-2}$.

Therefore

$$V(\hat{Y}) = \sum^{N} y_i^2 \, V(\mu_i) + \sum_{i \neq j}^{N} \sum^{N} y_i y_j \, Cov(\mu_i \mu_j)$$

$$= \sum^{N} y_i^2 \left(\frac{N-n}{n}\right) + \sum_{i \neq j}^{N} \sum^{N} y_i y_j \left(-\frac{N-n}{n(N-1)}\right)$$

$$= \frac{N-n}{n}\left[\sum^{N} y_i^2 - \frac{1}{N-1}\left\{\left(\sum^{N} y_i\right)\left(\sum^{N} y_j\right) - \sum^{N} y_i^2\right\}\right]$$

$$= \frac{N-n}{n}\left[\frac{N}{N-1}\sum^{N} y_i^2 - \frac{1}{N-1}\left(\sum^{N} y_i\right)^2\right]$$

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

$$(2) \quad = \frac{N^2}{n}\left(\frac{N-n}{N}\right)\left[\frac{\sum^{N} y_i^2 - \frac{(\sum^{N} y_i)^2}{N}}{N-1}\right]$$

$$= \frac{N^2}{n}\left(\frac{N-n}{N}\right) S_y^2$$

We conclude this section with a theorem on the estimability of any quadratic function. This result is a companion to the theorem on the estimability of a linear function - Theorem 2.

Since $V(\hat{Y})$ is a quadratic function of y_s', we state the necessary and sufficient conditions for the estimability of any quadratic function, $\sum_{i}^{N}\sum_{j}^{N} \mu_{ij} y_i y_j$; i.e., estimability implies unbiasedness.

Theorem 3: A set of necessary and sufficient conditions for the estimability of $\Sigma\Sigma\mu_{ij}y_iy_j$ is:

$$\left.\begin{array}{l} \pi_{ij}>0 \text{ if } \mu_{ij} \neq 0 \\ \pi_i>0 \text{ if } \mu_{ii} \neq 0 \end{array}\right\} \text{ where } i\neq j \text{ and ranges all N units.}$$

Corollary: The variance of an unbiased estimator of Y is not estimable unless $\pi_{ij}>0$ for all i and j in the population.

These conditions have been satisfied for equal probability sampling and for unequal probability sampling with replacement, but are critical assumptions in unequal probability sampling without replacement. We shall show the consequencies of this in the next section.

4.1.3 Unequal Probability of Selection With Replacement

$\hat{Y} = \Sigma\mu_i y_i = \frac{1}{n}\Sigma' \frac{t_i}{p_i} y_i$ where the t_i's indicate the number of times a unit is selected and follows the multinomial distribution.

$$E(\mu_i) = 1$$

$$E(\mu_i^2) = \frac{(1-p_i)}{np_i} + 1$$

$$V(\mu_i) = \frac{1-p_i}{np_i}$$

$$Cov(\mu_i\mu_j) = -\frac{np_ip_j}{n^2p_ip_j} = -\frac{1}{n}$$

$$V(\hat{Y}) = \Sigma y_i^2 \cdot \frac{1}{n}(\frac{1}{p_i} - 1) + \Sigma y_i y_j (-\frac{1}{n})$$

Using (19) of Chapter I, Section 1.5

$$V(\hat{Y}) = \frac{1}{n}\Sigma y_i^2(\frac{1}{p_i} - 1) - \frac{1}{n}\{(\Sigma y_i)(\Sigma y_j) + \Sigma y_i^2\}$$

(3) $$= \frac{1}{n}\Sigma p_i(\frac{y_i}{p_i})^2 - \frac{1}{n}(\Sigma y_i)^2$$

$$= \frac{1}{n}\Sigma p_i(\frac{y_i}{p_i} - Y)^2 = \frac{\sigma_Y^2}{n}$$

Note that the subscript on σ^2 is a capital Y to indicate the variance of the total of the characteristic while a small y was used in the previous section, i.e., $Y = N\bar{Y}$.

4.1.4 Unequal Probability of Selection Without Replacement

Notation from Section 2.2.2

$\hat{Y} = \Sigma E(a_i)c_iy_i = \Sigma' \frac{y_i}{\pi_i}$ is the HT estimator where $a_i = 0$ or 1 depending on whether the i^{th} unit is included in the sample or not. The variance of $\hat{Y}$ is given by

$$V(\hat{Y}) = \Sigma \frac{y_i^2}{\pi_i^2} V(a_i) + \sum_i \sum_{j\neq i} \frac{y_i}{\pi_i}\frac{y_j}{\pi_j} Cov(a_ia_j)$$

$$E(a_i) = \pi_i \ , \ E(a_i^2) = \pi_i$$

$$E(a_ia_j) = \pi_{ij}$$

$$V(a_i) = \pi_i(1-\pi_i)$$

$$Cov(a_i a_j) = \pi_{ij} - \pi_i \pi_j$$

$$(4)\quad V(\hat{Y}) = \Sigma \frac{\pi_i(1-\pi_i)y_i^2}{\pi_i^2} + \sum_i \sum_{j\neq i} \frac{(\pi_{ij}-\pi_i\pi_j)}{\pi_i\pi_j} y_i y_j$$

which is the HT expression for the variance.

An alternate expression is derived from writing $V(a_i)$ and $Cov(a_i a_j)$ as two terms

$$(4')\quad V(\hat{Y}) = \Sigma \frac{y_2^2}{\pi_i^2} E(a_i^2) - \Sigma \frac{y_i^2}{\pi_i^2} [E(a_i)]^2 + \sum_i \sum_{j\neq i} \frac{y_i}{\pi_i}\frac{y_j}{\pi_j} E(a_i a_j)$$

$$- \sum_i \sum_{j\neq i} \frac{y_i}{\pi_i}\frac{y_j}{\pi_j} [E(a_i)E(a_j)]$$

where the second and fourth terms combine and equal Y the population total. That is

$$(4'')\quad V(\hat{Y}) = \Sigma \frac{y_i^2}{\pi_i^2} + \sum_i \sum_{j\neq i} \frac{\pi_{ij}}{\pi_i\pi_j} y_i y_j - Y^2$$

Next, we examine several sampling schemes considered in 2.2 and 2.3. The Rao, Hartley, Cochran estimator in 2.2.2

$$\hat{Y}_{RHC} = \Sigma' \frac{y_i P_t}{p_i}$$

$V(\hat{Y}_{RHC}) = E_1V_2(\hat{Y}) + V_1E_2(\hat{Y})$ where the variance is conditioned by the random split.

Since $\hat{Y}_{RHC}$ is conditional unbiased, i.e., $E_2(\hat{Y})$ = constant

hence $V(\hat{Y}_{RHC}) = E_1V_2(\hat{Y})$ since $V_1E_2(\hat{Y}) = 0$.

Now $V(\hat{Y}_{RHC}) = E_1V_2(\hat{Y}) = E_1\Sigma' V_2(\frac{y_i P_i}{p_i}) = \Sigma' E_1V_2(\frac{y_i P_i}{p_i})$

where $V_2(\frac{y_i P_i}{p_t}) = \Sigma^{N_i} \frac{p_t}{P_i} (\frac{y_i}{p_t/P_i} - Y_i)^2 = \sum_{t<t'}^{N_i}\Sigma\, p_t p_{t'} (\frac{y_t}{p_t} - \frac{y_{t'}}{p_{t'}})^2$

$$= \sum_{t<t'}^{N}\sum a_{ti}a_{t'i}p_t p_{t'} \left(\frac{y_t}{p_t} - \frac{y_{t'}}{p_{t'}}\right)^2$$

where $a_{ti} = \begin{cases} 0 \text{ if the } t^{th} \text{ unit is not in group } i \\ 1 \text{ if the } t^{th} \text{ unit is in group } i \end{cases}$

t and t´ are two units from the same group.

Now $E_1(a_{ti}a_{t'i}) = \frac{N_i(N_i-1)}{N(N-1)}$

Therefore $E_1V_2\left(\frac{y_i P_i}{p_t}\right) = \frac{N_i(N_i-1)}{N(N-1)} \sum_{t<t'}^{N}\sum p_t p_{t'} \left(\frac{y_t}{p_t} - \frac{y_{t'}}{p_{t'}}\right)^2$

(5) $$= \frac{N_i(N_i-1)}{N(N-1)} \left[\sum^{N} p_t \left(\frac{y_t}{p_t} - Y\right)^2\right]$$

(6) or $$V(\hat{Y}_{RHC}) = \frac{\sum^{n} N_i^2 - N}{N(N-1)}\, n\, V(\hat{Y}) = \frac{\sum^{n} N_i^2 - N}{N(N-1)} \sum_{i=1}^{N} P_i \left(\frac{Y_i}{P_i} - Y\right)^2 .$$

where $V(\hat{Y})$ is the with replacement variance in 4.1.3.

We conclude this section with an intuitive proof of the Corollary to Theorem 3.

Consider any unbiased estimator of the population total for n>1 (so we have a sum), the population variance may be written in the general form

$$V(T_s) = \sum_{s \in S} T_s^2\, P(s) - Y^2$$

If $\pi_{ij} = 0$ for some i and j, there is no sample containing μ_i and μ_j so that $\Sigma T^2 P(s)$ cannot contain $y_i y_j$. Therefore the coefficient of $y_i y_j$ in the Horvitz-Thompson estimator will be equal to -2.

$$\sum_{i\neq j}\sum \frac{(\pi_{ij}-\pi_i\pi_j)}{\pi_i\pi_j} y_i y_j = \frac{0-\pi_i\pi_j}{\pi_i\pi_j} y_i y_j + \frac{0-\pi_j\pi_i}{\pi_j\pi_i} y_j y_i = 2y_i y_j$$

Hence, from Theorem 3, $V(T_g)$ is not estimable; thus the need for the 5 conditions stated on page 6 of Chapter II.

4.2 Multistage Designs - Population Variances

4.2.1 Two Stage Design - EPS and WOR at both stages

Stage 1: Select n primaries out of N with EPS and WOR

Stage 2: Select m_i secondaries out of M_i with EPS and WOR

$$\hat{Y} = \sum_i \sum_t \mu_i V_{it} y_{it} \qquad \text{rewriting}$$

$$= \sum_i \mu_i \sum_t V_{it} (y_{it}-\bar{Y}_i) + \sum_i \mu_i Y_i$$

$$= \Sigma \mu_i \hat{U}_i + \Sigma \mu_i Y_i$$

$$= \quad W \quad + \quad B$$

Now

$$V(B) = \frac{N^2}{n} (1 - \frac{n}{N}) \sum_{i=1}^{N} \frac{(Y_i-\bar{\bar{Y}})^2}{N-1} \quad \text{where } \bar{\bar{Y}} = \frac{1}{N} \sum^{N} Y_i$$

And

$$\hat{U}_i = \sum_t V_{it} y_{it} - Y_i \qquad E(\hat{U}_i) = 0$$

$$V(W) = \sum^{N} V(\mu_i \hat{U}_i) + 2 \sum_{i<j}\sum \text{Cov}(\mu_i\hat{U}_i, \mu_j\hat{U}_j)$$

$$= \Sigma E(\mu_i^2)\, V(\hat{U}_i)$$

$$= \frac{N}{n} \Sigma \frac{M_i^2}{m_i} (1 - \frac{m_i}{M_i})\, S_i^2$$

using $E(\mu_i^2) = (\frac{N}{n})^2 \cdot \frac{n}{N} = \frac{N}{n}$ and $E(\mu_i \hat{U}_i) = 0$

$$2\sum_{i<i} \text{Cov}(\mu_i \hat{U}\mu_i, \hat{U}_i) = 2\sum_{i<i} E(\mu_i \hat{U}_i \mu_i, \hat{U}_i)$$

Now $Cov(WB) = E(WB)$ since $E(W) = 0$

$$= \sum_{i=i} E(\mu_i \hat{U}_i \mu_i, Y_i)$$

$$= \sum_{i=i} Y_i E(\mu_i \mu_i) E(\hat{U}_i) = 0$$

$$(7) \quad \therefore \; V(\hat{Y}) = \frac{N}{n} \Sigma \frac{M_i^2}{m_i} (1 - \frac{m_i}{M_i}) \; S_i^2 + \frac{N^2}{n} \; (1 - \frac{n}{N}) \; \Sigma \frac{(Y_i - \bar{\bar{Y}})^2}{N-1}$$

If all $M_i = \bar{M}$

$$\hat{Y} = \frac{N\bar{M}}{n} \Sigma' \bar{y}_i$$

and the "large" between primary component reduces to

$$V(B) = \frac{N^2}{n} \; (1 - \frac{n}{N}) \bar{M}^2 \; \Sigma \frac{(\bar{Y}_i - \bar{Y})^2}{N-1}$$

Since it is rarely possible to achieve primaries of equal size, it is desirable to reduce this between primary component when the M_i vary considerably by either (1) changing the design, or (2) changing the estimator to one using auxiliary information.

<u>4.2.2 Two Stage Designs - Vary Selection by Stages</u>

Stage 1: Primaries sampled with UEPS and WR

Stage 2: Secondaries sampled with EPS and WOR

Scheme A: From 3.1.2 of Chapter III - The same m_i secondary units are used each time a primary is selected.

$$\hat{Y} = \sum_{it}\sum \mu_i \; V_{it} \; y_{it} = \frac{1}{n} \sum_i{}' \frac{t_i}{P_i} \sum_t{}' V_{it} y_{it}$$

$$= \frac{1}{n} \Sigma' \frac{t_i}{P_i} M_i \bar{y}_i$$

As in 4.2.1,

$$\hat{Y} = \sum_i \mu_i \hat{U}_i + \sum_i \mu_i y_i = W + B$$

$$V(B) = \frac{1}{n} \sum_i P_i (\frac{y_i}{P_i} - Y)^2 \text{ by section 4.1.3}$$

$$V(W) = \sum_i E(\mu_i^2)\ V(\hat{U}_i) \text{ as in 4.2.1}$$

$$= \sum_i (1 + \frac{q_i}{np_i}) \frac{M_i^2}{m_i} (1 - \frac{m_i}{M_i}) S_i^2$$

since

$$E(\mu_i^2) = \frac{np_i q_i}{(np_i)^2} + \frac{n^2 p_i^2}{n^2 p_i^2} = 1 + \frac{q_i}{np_i}$$

and as 4.2.1 Cov(WB) = 0

$$(8)\ \therefore\ V(\hat{Y}) = \sum_i (1 + \frac{q_i}{np_i}) \frac{M_i^2}{m_i} (1 - \frac{m_i}{M_i}) S_i^2 + \frac{1}{n} \sum_i P_i (\frac{y_i}{P_i} - Y)^2$$

In this scheme, there is variance due to

1. y variation
2. variation in primary sizes
3. allocation of different probabilities

These last two sources of variation will cancel by choosing appropriate probabilities.

Choose $p_i = \frac{M_i}{M}$, then $\frac{y_i}{P_i} = M\bar{Y}_i$ and the primary component becomes

$\frac{M^2}{n} \sum P_i (\bar{y}_i - \bar{Y})^2$ which does not vary with varying primary size.

Scheme B: The subsampling is done independently each time a primary is selected in the sample, and the subsampling permits unbiased estimates of the primary total. The estimator of the population total is based on all primaries whether distinct or not.

Let E_1 and V_1 denote the expectation and variance over samples at the first stage.

$$\hat{Y} = \frac{1}{n} \sum' \frac{M_i \bar{y}_i}{P_i} \qquad E\hat{Y} = E_1 E_2 \sum^n \frac{M_i \bar{y}_i}{np_i}$$

$$= E_1 \sum^n \frac{Y_i}{np_i} = Y$$

$$V(\hat{Y}) = E_1V_2(\hat{Y}) + V_1E_2(\hat{Y})$$

Now

$$V_1E_2(\hat{Y}) = V_1 \sum^n \frac{y_i}{np_i} = \frac{1}{n} \sum^N P_i(\frac{y_i}{P_i} - Y)^2$$

from single stage theory. Also

$$V_2(\hat{Y}) = \sum^n \frac{V_2(\hat{Y}_i)}{n^2P_i^2}$$

Therefore

$$E_1V_2(\hat{Y}) = E_1 \sum^n \frac{V(Y_i)}{n^2p_i^2} = E_1 \sum^N \frac{t_iV(Y_i)}{n^2p_i^2} = \sum^N \frac{V(Y_i)}{np_i}$$

where t_i is the number of times the i^{th} primary is selected in the sample and $V(Y_i)$ is the variance of the estimator of $\hat{Y}_i$ for the i^{th} primary total.

Now if an independent sample of m_i second-stage units is selected without replacement each time the i^{th} primary is selected in the sample after replacing the whole subsample (i.e., a secondary may be selected more than once),

Then

$$\hat{Y}_i = M_i\bar{y}_i \text{ and } V_2(\hat{Y}_i) = \frac{M_i^2}{m_i}(1 - \frac{m_i}{M_i})S_i^2$$

Hence, combining this with the between primary component

$$(9)\quad V(\hat{Y}) = \frac{1}{n}\Sigma P_i(\frac{y_i}{P_i} - Y)^2 + \frac{M_i^2}{m_i}(1 - \frac{m_i}{M_i})S_i^2$$

which differs from (8) by the factor

$$\frac{q_i}{np_i}\frac{M_i^2}{m_i}(1 - \frac{m_i}{M_i})S_i^2 .$$

When the probabilities are chosen as indicated on the top of page 15 and Scheme B used, a self-weighting sample is obtained.

If $p_i = \frac{M_i}{M_0}$ and $\frac{1}{f_0} = \frac{M_0}{mn}$, then

$$V(\hat{Y}) = \frac{M_0^2}{n(n-1)m^2} \sum^{n} (Y_i - \hat{\bar{Y}})^2$$

where $\hat{\bar{Y}} = (\hat{Y}_1 + \ldots + \hat{Y}_n) \div n$

and Y_i = the sample total in the selected primary on the r^{th} draw.

In comparing schemes A and B, it should be noted that in scheme B the number of secondaries actually selected is a random variable which in scheme A the number of secondaries is fixed for a particular primary. Therefore, in scheme A, the optimum allocations can be found by equating the actual fixed subsample sizes to the optimum values while in scheme B only expected subsample sizes can be equated to the optimum values.

4.2.3 Three Stage Sampling

Notation from section 3.2.1

Stage 1: n primaries selected with EPS-WOR from N

Stage 2: m_i secondaries selected with EPS-WOR from M_i

Stage 3: k_{it} tertiaries selected with EPS-WOR from K_{it}

$$\hat{Y} = \frac{N}{n} \sum^{n} \frac{M_i}{m_i} \sum^{m_i} \frac{K_{it}}{k_{it}} \sum^{k_{it}} y_{ith}, \text{ or}$$

$$\hat{Y} = \sum_{ith} \mu_i V_{it} W_{ith} y_{ith}$$

To obtain the variance of $\hat{Y}$, write

$$\hat{Y} = \sum_i \mu_i \underbrace{\sum_{th} V_{it} W_{ith} y_{ith} - Y_i} + \sum_i \mu_i Y_i$$

$$= \sum_i \mu_i \hat{U}_i + \sum_i \mu_i Y_i$$

$= W + B$ where $B\hat{U}_i = 0$

and μ_i and $\hat{U}_i$ are independent.

$$V(B) = \frac{N^2}{n} \left(1 - \frac{n}{N}\right) \frac{\sum_i (Y_i - \bar{\bar{Y}})^2}{N-1} \quad \text{where } \bar{\bar{Y}} = \frac{1}{N} \sum Y_i$$

Cov(WB) = 0 as before.

$$V(W) = \sum_i E(\mu_i^2)\ V(\hat{U}_i)$$ and using the results in 4.2.1

Note: The pattern of the term for each stage is "expanded" by all the stages above.

$$V(\hat{Y}) = \frac{N}{n} \sum_i \frac{M_i}{m_i} \sum_t \frac{K_{it}^2}{k_{it}} \left(1 - \frac{k_{it}}{K_{it}}\right) s_{it}^2 +$$

$$\frac{N}{n} \sum_i \frac{M_i^2}{m_i} \left(1 - \frac{m_i}{M_i}\right) \frac{\sum_t (Y_{it\cdot} - \bar{Y}_{i\cdot})^2}{M_i - 1} +$$

$$\frac{N^2}{n} \left(1 - \frac{n}{N}\right) \frac{\sum_i (Y_{i\cdot} - \bar{\bar{Y}})^2}{N-1} .$$

where $\bar{Y}_{i\cdot} = \frac{1}{M_t} \sum_t Y_{it\cdot}$ (Average secondary total)

$\bar{\bar{Y}} = \frac{1}{N} \sum_i Y_{i\cdot}$ (Average primary total)

$s_{it}^2 = \frac{1}{K_{it}-1} \sum_h (y_{ith} - \bar{y}_{it\cdot})^2$ (within secondary)

4.3 Conditional Expectation in Multistage Designs

The general formula

$$V(\hat{Y}) = E_1V_2(\hat{Y}) + V_1E_2(\hat{Y})$$

can be extended to additional stages by using the same identity to express $V_2(\hat{Y})$ as

$$V_2(\hat{Y}) = E_2V_3(\hat{Y}) + V_2E_3(\hat{Y})$$

and using formula 15 of section 1.4 to express

$$E_2(\hat{Y}) = E_2E_3(\hat{Y})$$

Combining these two results, an expression for the variance in a three stage design is obtained.

$$V(\hat{Y}) = E_1E_2V_3(\hat{Y}) + E_1V_2E_3(\hat{Y}) + V_1E_2E_3(\hat{Y})$$

These same two identities can be used repeatedly to get the variance for any number of stages by expressing the expectation and variance operation of the last stage units in the same manner as going from a two stage to a three stage design.

For a three stage design with equal size units at all stages, we have

Stage 1: n primaries selected with EPS-WOR from N

Stage 2: m secondaries selected with EPS-WOR from M

Stage 3: k tertiaries selected with EPS-WOR from K

The estimator of the mean is:

$$\hat{\bar{Y}} = \frac{\sum\sum\sum^{nmk} y_{ijh}}{nmk}$$

$$E(\hat{\bar{Y}}) = E_1E_2E_3 \frac{1}{nm} \sum\sum^{nm} \bar{y}_{ij.}$$

$$= E_1E_2 \frac{1}{nm} \sum\sum^{nm} E_3(\bar{y}_{ij.})$$

taken over all third stage units in the n.m "strata" over the selected primaries and secondaries

$$= E_1E_2 \frac{1}{nm} \sum\sum^{nm} \bar{Y}_{ij}$$

$$= E_1 \frac{1}{n} \sum^{n} E_2 \frac{1}{m} \sum^{m} \bar{Y}_{ij.}$$

taken over all second stage units in the n selected primaries

$$= E_1 \frac{1}{n} \sum^{n} \frac{1}{M} \sum_{j}^{M} \bar{Y}_{ij.}$$

taken over all primary units in the population

$$= \frac{1}{N} \sum^{N} \bar{Y}_{i..} = \bar{Y}$$

The variances is derived as follows:

$$V_1[E_2E_3\bar{Y}] = V_1[E_2 \frac{1}{nm} \sum\sum^{nm} E_3(\bar{y}_{ij.})]$$

$$= V_1[E_2\frac{1}{nm}\sum\sum^{nm}\bar{Y}_{ij}]$$

$$= V_1[\frac{1}{n}\sum^{n} E_2\bar{Y}_{ij}]$$

$$= V_1[\frac{1}{n}\sum^{n}\bar{Y}_i]$$

$$= (\frac{1}{n} - \frac{1}{N})\frac{1}{N-1}\sum^{N}(\bar{Y}_i - \bar{Y})^2$$

Now work on $E_1[E_2V_3(\hat{Y}) + V_2E_3(\hat{Y})]$ or first two terms at bottom of page 13. Replace the term in the bracket by two stage results:

For two stages

$$E_1V_2 + V_1E_2 = \frac{1}{n}(\frac{1}{m} - \frac{1}{M})\ S_w^2 + (\frac{1}{n} - \frac{1}{N})\ \frac{S_b^2}{M}$$

which for three stages becomes

$$E_1[E_2V_3 + V_2E_3] = \frac{1}{n}(\frac{1}{mk} - \frac{1}{MK})\ S_{ww}^2 + \frac{1}{n}(\frac{1}{m} - \frac{1}{M})S_w^2$$

Adding this results to that for $V_1E_2E_3$ on page 21

$$V(\hat{\bar{Y}}) = (\frac{1}{n} - \frac{1}{N})\ \frac{S_b^2}{MK} + \frac{1}{n}(\frac{1}{m} - \frac{1}{M})S_w^2 + \frac{1}{n}(\frac{1}{mK} - \frac{1}{MK})S_{ww}^2$$

Or, in terms of variance components

$$V(\hat{\bar{Y}}) = (\frac{1}{n} - \frac{1}{N})S_1^2 + (\frac{1}{nm} - \frac{1}{NM})S_2^2 + (\frac{1}{nmk} - \frac{1}{NMK})S_3^2$$

where

$$S_1^2 = \frac{\sum^{N}(\bar{Y}_i-\bar{Y})^2}{N-1} \quad \text{from } V_1E_2E_3 \text{ term}$$

$$S_2^2 = \frac{\sum\sum^{NM}(\bar{Y}_{ij}-\bar{Y}_i)^2}{N(M-1)} \quad \text{from } E_1V_2E_3 \text{ term}$$

$$S_3^2 = \frac{\sum\sum\sum^{NMK}(y_{ijh}-\bar{Y}_{ij})^2}{NM(K-1)} \quad \text{from } E_1E_2V_3 \text{ term}$$

If $n = N$ we have stratified sampling and $m = M$ we have cluster sample of size n.

4.4 General Formulas For Designs Where Unbiased Estimators of Primary Totals Are Available

4.4.1 Selecting Primaries With Replacement

Stage 1: UCPS-WR

Stage 2 and lower - Sampling is done independently each time the primary is selected and permits unbiased estimators of primary totals. All secondaries are replaced after selection of units for the r^{th} primary selection. Note: It is possible for a secondary (or tertiary) to get selected more than once if a primary is selected more than once.

The estimator of the total is based on all primaries whether distinct or not, i.e., $r = 1, 2, \ldots n$.

Let $\hat{Y}_r$ be an unbiased estimator of the population total based on the primary drawn at the r^{th} selection. $\hat{Y}_r$ is based on subsampling at the second and subsequent stages and is such that

$$E_2(\hat{Y}_r) = Y_r \text{ and } V_2(\hat{Y}_r) = V_r$$

where E_2 and V_2 denote conditional expectation and variance over the second and subsequent stages. Let E_1 and V_1 denote expectation and variance over samples at the first stage.

An unbiased estimator of the population total is

$\hat{Y} = \frac{1}{n} \Sigma' \frac{\hat{Y}_r}{P_r}$ where P_r is the probability of the primary selected at the r^{th} draw and Y_i is the i^{th} primary total.

$$E\,\hat{Y} = E_1 E_2 \sum^{n} \frac{\hat{Y}_r}{np_r} = E_1 \frac{1}{n} \sum^{n} \frac{Y_r}{P_r} = Y$$

using single stage results since Y_r is a constant. The variance

$$V(\hat{Y}) = E_1 V_2(\hat{Y}) + V_1 E_2(\hat{Y})$$

Now

$$V_1 E_2(\hat{Y}) = V_1 \sum^{n} \frac{Y_r}{np_r} = \frac{1}{n} \sum^{N} P_i \left(\frac{Y_i}{P_i} - Y\right)^2$$

from single stage theory, and

$$V_2(\hat{Y}) = \frac{\sum^{n} V_2(\hat{Y}_r)}{n^2 p_r^2} = \sum^{n} \frac{V_r}{n_2 P_r^2}$$

If we replace the whole subsample each time after the r^{th} draw

$$E_1 V_2(\hat{Y}) = E_1 \sum^{n} \frac{V_r}{n^2 p_i^2} = E_1 \sum^{N} \frac{t_i V_i}{n^2 p_i^2} = \sum^{N} \frac{V_i}{np_i}$$

where $E(t_i) = np_i$ and t_i is the number of times the i^{th} primary is selected in the sample and V_i is the variance of the estimator $\hat{Y}_i$ for the i^{th} primary total. Hence

$$V(\hat{Y}) = \frac{1}{n} \sum^{N} P_i (\frac{Y_i}{P_i} - Y)^2 + \sum^{N} \frac{V_i}{np_i}$$

For a two stage design with unequal size primaries

$$V_i = \frac{M_i^2}{m_i} (1 - \frac{m_i}{M_i}) S_i^2 .$$

4.4.2 Selecting Primaries Without Replacement

The estimator and its unbiased variance estimator for a design can be provided where the corresponding parameter for the single stage design are known. Rules can be written down by considering a very general estimator of Y given by Des Raj (1966). Durbin (1953) proposed similar rules based on using the Horvitz-Thompson estimator. The necessary requirements for writing down rules are:

(1) An unbiased estimator $\hat{Y}_i$ of the i^{th} primary total Y_i based on sampling at the second and subsequent stages is available, i.e., $E_2(\hat{Y}_i) = Y_i$.

(2) The primaries are subsampled independently in a known manner so that $E_2(\hat{Y}_i \hat{Y}_j) = Y_i Y_j$ and $V_2(\hat{Y}_i) = V_i$ where V_i is a constant. An unbiased variance estimator or $v_i = v(\hat{Y}_i)$ of $V_i = V_2(\hat{Y}_i)$ based on sampling at the second and subsequent stages is available, i.e., $E_2 v_i = V_i$. Since V_i has to be a constant, the designs in which subsample sizes are random variables do not fit into this set up.

The estimator of the population total, Y, is:

$$\hat{Y}_G = \sum^{\nu} a_{is} \hat{Y}_i$$

where a_{is} are predetermined numbers for every sample s with the restriction that $a_{is} = 0$ whenever s does not contain the i^{th} primary, and ν is the number of distinct primaries in the sample.

Rule 1: The unbiased estimator of Y in multistage sampling is obtained by replacing Y_i by $\hat{Y}_i$ in the corresponding unbiased estimator of Y in a single stage sampling of clusters when the clusters are completely enumerated. Note $\hat{Y}_G$ is unbiased if and only if $E(a_{is}) = 1$ for every $i = 1, 2, \ldots N$.

The variance of $\hat{Y}_G$

$$V(\hat{Y}_G) = E_1 V_2(\hat{Y}_G) + V_1 E_2(\hat{Y}_G)$$

$$= E_1[\overset{\nu}{\Sigma} a_{is}^2 \, V_i] + V_1[\overset{\nu}{\Sigma} a_{is} Y_i]$$

$$= \overset{N}{\Sigma} E(a_{is}^2) \, V_i + V_1 \, [\overset{\nu}{\Sigma} a_{is} Y_i]$$

Therefore, knowing the variance in single stage sampling, we need only $E(a_{is}^2)$ or $E_1[\overset{\nu}{\Sigma} a_{is}^2 \, V_i]$. Sometimes it is convenient to evaluate the latter quantity.

$$V(\hat{Y}_G) = \sum_{i} \sum_{j \neq i} (\pi_i \pi_j - \pi_{ij})\left(\frac{Y_i}{\pi_i} - \frac{Y_j}{\pi_j}\right)^2 + \sum_{i} \frac{\sigma_i^2}{\pi_i}$$ where the first term is the Yates-Grundy variance of the HT estimator.

4.5 Estimation of Variances in Single Stage Sample Surveys

The property of unbiased estimators of variances will be used in all derivations. In the preceding sections of this chapter, we have derived the population variances for certain schemes of sampling. For these schemes of sampling, it is clear the variance exists and have the desirable property of being always non-negative. Our task is to find unbiased sample estimators of the various parameters or expressions that appear in the variance formulas.

That is of

σ_y^2 in 4.1.1, or

S_y^2 in 4.1.2, or

π_i, π_{ij}, $V(\hat{Y})$, and Y in 4.1.4.

We start by looking at one of the expressions for the population variance and choose the parameter(s) to estimate since you may have a choice if you look at different formulation of the same variance formula. For example, the population variance has the general form

$$V(\hat{Y}) = \Sigma \frac{y_i^2}{\pi_i^2} V(a_i) + \sum_i \sum_{j \neq i} \frac{y_i}{\pi_i} \frac{y_j}{\pi_j} \mathrm{Cov}(a_i a_j)$$

which was given in 4.1.4, but this may be rewritten as:

$$V(\hat{Y}) = \Sigma \frac{y_i^2}{\pi_i^2} \{E(a_i^2) - [E(a_i)]^2\} + \sum_i \sum_{j \neq i} \frac{y_i}{\pi_i} \frac{y_j}{\pi_j} \{E(a_i a_j) - E(a_i)E(a_j)\}$$

which reduces to

$$(10)\quad V(\hat{Y}) = \Sigma \frac{y_i^2}{\pi_i^2} E(a_i^2) + \sum_i \sum_{j \neq i} \frac{y_i}{\pi_i} \frac{y_j}{\pi_j} E(a_i a_j) - Y^2$$

since

$$\Sigma \frac{y_i^2}{\pi_i^2} [E(a_i)]^2 + \sum_i \sum_{j \neq i} \frac{y_i}{\pi_i} \frac{y_j}{\pi_j} E(a_i)E(a_j) = (\Sigma y_i)^2 = Y^2$$

where

$E(a_i) = \pi_i$ and $E(a_j) = \pi_j$.

Hence in (10), we need unbiased estimates for each of the three terms.

Before proceeding to develop sample estimators of the variance, we present two general quadratic functions of the y's which will be useful. These forms are generalizations of the results given in Theorems 2 and 3 earlier.

(11) $L_1(S) = \Sigma' c_i f(y_i) = \Sigma t_i c_i f(y_i)$ where t_i is a random variable defined as 1 if the unit U_i is in the sample and 0(zero) otherwise. The c_i's are constants for each of the N units in the population.

$E\, L_1(S) = \Sigma E(t_i) C_i f(y_i) = \Sigma \pi_i c_i f(y_i)$ which is unbiased if $c_i = \frac{1}{\pi_i}$.

For the sample estimate of $\Sigma \pi_i C_i f(y_i)$, we need to have $\Sigma' C_i f(y_i)$, or we need to determine the coefficient of $f(y_i)$, i.e., $\pi_i C_i$, and then divide by π_i.

Hence, we divide this product by π_i to obtain C_i for our sample estimator $\Sigma' C_i f(y_i)$. Commonly, $f(y_i)$ is some power of y_i and we are interested in moments of the variable y_i; i.e., $f(y_i) = y_i^2$.

(12) $L_2(S) = \sum_i' \sum_{j\neq i}' C_{ij}\ f(y_i)f(y_j) = \sum_i \sum_{j\neq i} t_{ij}\ C_{ij}\ f(y_i)f(y_j)$

where t_{ij} is a random variable defined as 1 if both the i and j units are in the sample and 0 otherwise. The C_{ij} are constants assigned to each pair of units in the population.

$$E\ L_2(S) = \sum_i \sum_{j\neq i} E(t_{ij})\ C_{ij}\ f(y_i)\ f(y_j) = \sum_i \sum_{j\neq i} \pi_{ij}\ C_{ij}\ f(y_i)f(y_j)$$

For the sample estimate of $\sum_i \sum_{j\neq i} \pi_{ij}\ C_{ij}\ f(y_i)f(y_j)$, we need to have

$\sum_i' \sum_{j\neq i}' C_{ij}\ f(y_i)f(y_j)$, or we need to determine the coefficient of

$f(y_i)f(y_j)$ which must equal $\pi_{ij}\ C_{ij}$. Hence, we divide this product by π_{ij} to obtain C_{ij}, and the sample estimator.

The forms (11) and (12) are frequently combined in obtaining the variance. That is, the terms

$$\Sigma t_i C_i y_i^2 + \sum_i \sum_{j\neq i} t_{ij}\ C_{ij}\ y_i y_j,\ (\text{i.e., } = \sum_{ij}\sum t_{ij}\ C_{ij}\ y_i y_j)$$

and are estimated from the sample data by letting $f(y_i) = y_i^2$ in (11) and $f(y_i) = y_i$ and $f(y_j) = y_j$ in (12). Hence, we must have $\pi_{ij}>0$ if $C_{ij}\neq 0$ and $\pi_i>0$ if $C_{ij} = C_i \neq 0$.

4.5.1 Single Stage - EPS-WOR (Section 4.1.2, also 4.1.1)

In formula (2) of section 4.1.2, the only parameter which must be estimated from the sample is S_y^2. Consequently, we look at the sample statistic derived as a "copy" of the population parameter.

Theorem 4. $E(s_y^2) = S_y^2$ where

$$S_y^2 = \frac{1}{N-1}\Sigma(y_i-\bar{Y})^2 \text{ and } s_y^2 = \frac{1}{n-1}\Sigma'(y_i-\bar{y})^2$$

By definition

$(n-1)s^2 = \Sigma'(y_i-\bar{y})^2$ which we rewrite so as to replace $\bar{y}$ by $\bar{Y}$ inside parenthesis.

$= \Sigma'(y_i-\bar{Y})^2 - n(\bar{y}-\bar{Y})^2$ replacing the first term by the weighted variable representation we get

$= \Sigma\mu_i(y_i-\bar{Y})^2 - n(\bar{y}-\bar{Y})^2$

Taking expectations

$$E[(n-1)s^2] = \frac{n}{N}\Sigma(y_i-\bar{Y})^2 - nV(\bar{y})$$

$$= \frac{n}{N}(N-1)S^2 - (1-\frac{n}{N})\ S^2 = (n-1)S^2$$ or the sample s_y^2 is an unbiased estimate of S_y^2.

Therefore $V(\hat{Y}) = \frac{N^2}{n}(1-\frac{n}{N})\ S_y^2$ is estimated by replacing the population parameter S_y^2 by the sample statistic s_y^2, or

$$v(\hat{Y}) = \frac{N^2}{n}(1-\frac{n}{N})s_y^2 \ .$$

For this method of sampling, the sample estimator of the variance is a "copy" of the parameter obtained by replacing $\bar{Y}$ by $\bar{y}$. The results above is also useful for the variance in 4.1.1 since σ_y^2 differs from S_y^2 only by the constant $\frac{N}{N-1}$ which is unimportant for moderate size populations.

4.5.2 Single Stage - UEPS - WR (Section 4.1.3)

The estimator

$$\hat{Y} = \Sigma' \frac{t_i y_i}{np_i}$$ and variance

$$V(\hat{Y}) = \frac{1}{n}\Sigma P_i(\frac{y_i}{p_i} - Y)^2 = \frac{N^2}{n}\sigma_y^2 = \frac{\sigma_T^2}{n}$$

Consider the sample quantity

$$N^2 s_y^2 = \frac{n}{n-1}\Sigma'\frac{t_i}{n}(\frac{y_i}{p_i} - \hat{Y})^2$$ which we rewrite as

$$= \frac{n}{n-1}\{\Sigma'\frac{t_i}{n}(\frac{y_i}{p_i} - Y)^2 - (\hat{Y}-Y)^2\}$$

$$\frac{n-1}{n}N^2 s_y^2 = \Sigma'\frac{t_i}{n}(\frac{y_i}{p_i} - Y)^2 - (\hat{Y}-Y)^2$$

$$= \Sigma\frac{np_i\mu_i}{n}(\frac{y_i}{p_i} - Y)^2 - (\hat{Y}-Y)^2$$ upon replacing the first term by the weighted variable representation.

Taking expectations

$$E[\frac{n-1}{n} N^2 s_y^2] = N^2 \sigma_y^2 - V(\hat{Y}) = N^2 \sigma_y^2 (1- \frac{1}{n}) = N^2 \sigma_y^2 (\frac{n-1}{n})$$

The sample estimator is

$$v(\hat{Y}) = \frac{1}{n} N^2 s_y^2 = \frac{1}{n} \Sigma' \frac{t_i}{n-1} (\frac{y_i}{p_i} - \hat{Y})^2$$

In this case the sample estimator is not an exact "copy" of the population variance, but differs from the form given at the beginning of this section in that we have $\frac{t_i}{n-1}$ instead of p_i.

There is an alternate form for the sample estimator of the variance where the summation is over all different pairs in the sample:

$$v(\hat{Y}) = \frac{1}{n} \sum_i{}' \sum_{j \neq i}{}' \frac{(\frac{y_i}{p_i} - \frac{y_j}{p_j})^2}{n(n-1)} .$$

4.5.3 Single Stage - UEPS - WOR (Section 4.1.4)

If we use the HT expression for the variance (4), we must have a means of calculating the two terms. The two general estimators, $L_1(S)$ and $L_2(S)$, can be used to immediately obtain the sample estimators. The first term

$$\Sigma \frac{\pi_i (1-\pi_i) Y_i^2}{\pi_i^2} \text{ is estimated by } \Sigma' \frac{(1-\pi_i) y_i^2}{\pi_i^2}$$

and the second term

$$\sum_i \sum_{j \neq i} \{ \frac{(\pi_{ij} - \pi_i \pi_j)}{\pi_i \pi_j} Y_i Y_j \frac{\pi_{ij}}{\pi_{ij}} \} \text{ by } \sum_i{}' \sum_{j \neq i}{}' \frac{(\pi_{ij} - \pi_i \pi_j)}{\pi_i \pi_j} \frac{y_i y_j}{\pi_{ij}}$$

In addition, the sampling scheme must be such that $\pi_{ij} > 0$ for all pairs in the population if the estimator is to be unbiased.

π_i and π_{ij} are subject to the following relations

$$\Sigma^N E(a_i) = \Sigma^N \pi_i = n, \text{ and } \sum_{j \neq i}^N E(a_i a_j) = (n-1)E(a_i) = (n-1)\pi_i$$

and $$\sum_i^N \sum_{j \neq i}^N \pi_{ij} = n(n-1)$$

Since the HT variance estimator is not always non-negative, it is useful to consider the Yates and Grundy estimator of the variance for the HT estimator of the total. This form may be obtained by using the fact that

$$\sum_{j\neq i}(\pi_{ij}-\pi_i\pi_j) = (n-1)\pi_i-\pi_i(n-\pi_i) = -\pi_i(1-\pi_i)$$

The Yates and Grundy estimator is

$$V(\hat{Y}) = \sum_i \sum_{j\neq i} (\pi_i\pi_j-\pi_{ij})\left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j}\right)^2$$

and the sample estimator using $L_2(S)$ is

$$v(\hat{Y}) = \sum_i \sum_{j\neq i} \frac{(\pi_i\pi_j-\pi_{ij})}{\pi_{ij}} \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j}\right)^2$$

This expression for the variance of the HT estimator is not always non-negative for all sampling schemes, but is the simplest of the unbiased variance estimators which has been identified for several sampling schemes. In addition, when all y_i are equal, the variance is zero which is not the case for the HT expression for the variance.

The estimator of the variances of $\hat{Y}_{RHC}$

$$E\left(\sum^n \frac{y_i^2 P_i}{p_t^2}\right) = \sum^n [E_1(\sum^{N_i} \frac{y_t^2}{P_t})] = \sum^n \frac{N_i}{N} \left(\sum^N \frac{y_t^2}{P_t}\right) = \sum^N \frac{y_t^2}{P_t}$$

An estimate of $Y^2 = \hat{Y}^2_{RHC} - v(\hat{Y}_{RHC})$ (by definition)

where $v(\hat{Y}_{RHC})$ is the unbiased estimator of $V(\hat{Y}_{RHC})$.

Hence

$$v(\hat{Y}_{RHC}) = \frac{(\sum^n N_i^2-N)}{N(N-1)} \left\{\sum^n \frac{y_i^2 P_i}{p_i^2} - \hat{Y}^2_{RHC}+v(\hat{Y}_{RHC})\right\} = \frac{\sum^n N_i^2-N}{N(N-1)}\left\{\sum^n P_i\left(\frac{y_i}{p_i} - \hat{Y}_{RHC}\right)^2\right\}.$$

An unbiased estimate of the sample variance for the Murthy Estimator given in 2.2.2 (c) is:

$$v(\hat{Y}) = \frac{1}{[P(S)]^2} \sum_{i=1}^n \sum_{j>i}^n \{P(S)P(S|ij)-P(S|i)P(S|j)\}.\ P_iP_j\left(\frac{y_i}{P_i} - \frac{y_j}{P_j}\right)^2 .$$

where $P(S|ij)$ is the conditional probability of getting the S^{th} sample given that the units L_i and L_j have been selected in the first two draws.

A result which will be useful in multi-stage sampling is given here to indicate a general method for estimating the variance.

Rule 2: Find an unbiased estimator of the variance for the single-stage design. Obtain a "copy" of it by substituting $\hat{Y}_i$ for Y_i (primary total). Also, find a copy of the estimator of Y in single-stage sampling by substituting v_i for V_i. The sum of the two copies is an unbiased estimator of the variance.

4.6 Estimation of Variances in Multistage Sample Surveys

4.6.1 Two Stage Designs - EPS and WOR at both stages (4.2.1)

In formula (7) there are two terms to be estimated. We rely on the results from single-stage theory and use conditional expectation. From 4.2.1

$$(13) \quad V(\hat{Y}) = \frac{N}{n} \Sigma \frac{M_i^2}{m_i} (1 - \frac{m_i}{M_i}) S_i^2 + \frac{N^2}{n} (1 - \frac{n}{N}) \Sigma \frac{(Y_i - \bar{\bar{Y}})^2}{N-1}$$

consider

$$\frac{1}{n-1} \Sigma' (\hat{Y}_i - \hat{Y})^2 = \frac{1}{n-1} \Sigma' (M_i \bar{y}_i - \hat{Y})^2$$

since we observed in single-stage using EPS-WOR sampling that the sample "copy" provided an unbiased estimator.

We consider the conditional expectation for a given set of V_{it} which implies $M_i \bar{y}_i$ are fixed and seek to derive the between primary component. Rewriting in weighted variable notation, we have

$$\frac{1}{n-1} \sum_i {}' \mu_i \frac{n}{N} (M_i \bar{y}_i - \hat{Y})^2 \text{ where } \hat{Y} = \frac{1}{n} \Sigma' M_i \bar{y}_i$$

$$\therefore E/_{V_{it}} \frac{1}{n-1} \Sigma' (M_i \bar{y}_i - \hat{Y}')^2 = \frac{1}{N-1} \sum_i (M_i \bar{y}_i - \hat{Y}')^2$$

$$= \frac{1}{N-1} \sum_i M_i^2 \bar{y}_i^2 - N \bar{\bar{Y}}'^2$$

Now $E\bar{\bar{Y}}'^2 = V(\bar{\bar{Y}}') + (E\bar{\bar{Y}}')^2$ (by definition)

and $V(\bar{\bar{Y}}') = \frac{1}{N^2} \sum_i M_i^2 V(\bar{y}_i)$

$$\therefore E \frac{1}{n-1} \Sigma'(M_i\bar{y}_i - \hat{Y}')^2 = \frac{1}{N-1} \Sigma M_i^2 (1 - \frac{1}{N})(1 - \frac{m_i}{M_i}) \frac{S_i^2}{m_i} + \frac{1}{N-1} \Sigma (Y_i - \frac{Y}{N})^2$$

$$(14) \qquad = \frac{1}{N-1} \Sigma M_i^2 V(\bar{y}_i) + \Sigma M_i^2 \frac{Y_i^2}{M_i^2} - NV(\bar{\bar{Y}}') - N \frac{Y^2}{N^2}$$

since $V(\bar{y}_i) = (1 - \frac{m_i}{M_i}) \frac{S_i^2}{m_i}$. This accounts for the between primary component except for the factor $\frac{N^2}{n}(1 - \frac{n}{N})$.

We now estimate the within primary component of $V(\hat{Y})$ minus the second term in (14) which has arisen.

To estimate the first term of (13) use $\Sigma_i \mu_i \frac{M_i^2}{m_i} (1 - \frac{m_i}{M_i}) s_i^2$ which has expectation $\Sigma_i \frac{M_i^2}{m_i} (1 - \frac{m_i}{M_i}) S_i^2$ based on single stage results.

So

$$v(\hat{Y}) = \frac{N}{n} \Sigma' \frac{M_i^2}{m_i} (1 - \frac{m_i}{M_i}) s_i^2 + \frac{N^2}{n} (1 - \frac{n}{N}) \Sigma' \frac{(M_i\bar{y}_i - \hat{Y}')^2}{n-1}$$

is an unbiased estimate of $V(\hat{Y})$, and a "copy" of (13). However, the magnitude of the between term is the same as in $V(\hat{Y})$ since it is summed over n and divided by n-1 while the within term is summed only over n and not N.

4.6.2 Two Stage Designs - Selecting Primaries With Replacement (4.4.1)

We examine the general estimator described in (4.4.1) and state the following theorem.

Theorem 5: An unbiased variance estimator of $\hat{Y}$ is

$$v(Y) = \frac{1}{n(n-1)} \overset{n}{\Sigma} (\frac{\hat{y}_i}{p_i} - \hat{Y})^2 .$$

where $\hat{y}_i$ is the primary total and p_i the probability of selecting the primary in WR sampling.

Since each of the n estimates, i.e., $\frac{\hat{y}_i}{p_i}$, of the population total are independent.

$$v(\hat{Y}) = \frac{1}{n} v(\frac{\hat{y}_i}{p_i})$$

but

$$v(\frac{\hat{y}_i}{p_i}) = \frac{1}{n-1} \Sigma'(\frac{\hat{y}_i}{p_i} - \frac{1}{n} \Sigma' \frac{\hat{y}_i}{p_i})^2 = \frac{1}{n-1} \Sigma'(\frac{\hat{y}_i}{p_i} - \hat{Y})^2$$

$$\therefore\ v(\hat{Y}) = \frac{1}{n(n-1)} \Sigma'\left(\frac{\hat{y}_i}{P_i} - \hat{Y}\right)^2$$

4.6.3 Two Stage Designs - Selecting Primaries Without Replacement

(A) We now examine the general estimator given in 4.4.2. Using $L_2(S)$ for the first term and Theorem 4 for the second term with EPS at the second stage.

$$v(\hat{Y}_G) = \sum_i{}' \sum_{j\neq i}{}'\left(\frac{\pi_i\pi_j-\pi_{ij}}{\pi_{ij}}\right)\left(\frac{\hat{y}_i}{\pi_i} - \frac{\hat{y}_j}{\pi_j}\right)^2 + \Sigma'\frac{s_i^2}{\pi_i}$$

When sampling is with equal probabilities at the first stage

$$\pi_{ij} = \frac{n(n-1)}{N(N-1)}$$

and

$$\pi_i = \frac{n}{N}.$$

In this case $\pi_i\pi_j-\pi_{ij} = \frac{n}{N}\left[\frac{N-n}{N} \cdot \frac{1}{N-1}\right]>0$,

hence the variance is always positive.

(B) We now examine the estimator $\hat{\bar{Y}}_1$ discussed in section 3.3.1 which was shown to be biased unless all primaries were of the same size.

$\hat{\bar{Y}}_1 = \frac{1}{n}\Sigma'\bar{y}_{i.}$ or is the average of the estimated primary means.

An unbiased estimate of the bias is provided by

$$\widehat{Bias} = -\frac{N-1}{N\bar{M}(n-1)} \Sigma'(M_i-\hat{\bar{M}}_n)(\bar{y}_{i.}-\hat{\bar{Y}}_1)$$

It follows that an unbiased estimate of the population mean is obtained from the primary means by

$$\hat{\bar{Y}} = \hat{\bar{Y}}_1 + \frac{N-1}{N\bar{M}} \cdot \frac{1}{n-1} \Sigma'(M_i-\hat{\bar{M}}_n)(\bar{y}_{i.}-\hat{\bar{Y}}_1)$$

These results on the bias will be required for the mean square error by adding the bias squared to the variance.

To find the variance

$$V(\hat{\bar{Y}}_1) = V_1[E_2(\hat{\bar{Y}}_1|n)] + E_1[V_2(\hat{\bar{Y}}_1|n)]$$

Using (7) of section 4.2.1, modified for the mean (i.e., division by N^2)

$$V(\hat{\bar{Y}}_1) = \frac{1}{n}(1 - \frac{n}{N})\ S_b^2 + \frac{1}{nN}\ \Sigma\frac{1}{m_i}(1 - \frac{m_i}{M_i})\ S_i^2$$

since $\hat{\bar{Y}}_1$ and $\bar{y}_i$ are unbiased estimates of $\bar{\bar{Y}}_N$ and $\bar{Y}_i$.

where

$$S_b^2 = \frac{1}{N-1}\sum^{N}(\bar{Y}_i-\bar{\bar{Y}}_N)^2$$

and

$$S_i^2 = \frac{1}{M_i-1}\sum^{M_i}(y_{ij}-\bar{Y}_i)^2$$

The sample estimator of the variance based on 4.2.1 page 39 is a "copy" of the above population variance

$$v(\hat{\bar{Y}}_1) = \frac{1}{n}(1 - \frac{n}{N})\hat{S}_b^2 + \frac{1}{nN}\ \Sigma'\ \frac{1}{m_i}(1 - \frac{m_i}{M_i})\hat{S}_i^2$$

and the estimated mean square error is

$$M.S.E.(\hat{\bar{Y}}_1) = v(\hat{\bar{Y}}_1) + [\frac{1}{N\bar{M}}\cdot\frac{1}{n-1}\ \Sigma'\ (M_i-\bar{M}_n)(\bar{y}_i-\hat{\bar{Y}}_1)]^2$$

4.6.4 Three Stage Sampling - Unequal 1st and 2nd Stage Units with EPS-WOR sampling at all stages (3.2.1)

Consider the unbiased estimator of the mean

$$\hat{\bar{Y}} = \frac{1}{n}\sum_{i}^{n}\frac{M_i}{m_i}\sum_{t}^{m_i}\frac{K_{it}}{k_{it}}\sum_{h}^{k_{it}}y_{ith}$$

As pointed out in 4.3.1

$V_2(\hat{\bar{Y}}) = E_2V_3(\hat{\bar{Y}}) + V_2E(\hat{\bar{Y}})$, hence we can extend the results any two stage design to get the variance for a three stage design where the first stage contribution to the variance will be unchanged. Since simple random sampling is employed at each stage, we may repeatedly apply Theorem 4 and write the variance of the mean from 4.3.1 by dividing by N^2 or in terms of means at each stage as below:

$$v(\hat{\bar{Y}}) = \frac{1}{n}(1 - \frac{n}{N})\hat{S}_1^2 + \frac{1}{nN}\ \Sigma'\ \frac{M_i^2}{m_i}(1 - \frac{m_i}{M_i})S_{2i}^2 + \frac{1}{nN}\sum_t{}'\frac{M_i}{m_i}\sum_t{}'\frac{K_{it}^2}{k_{it}}(1 - \frac{k_{it}}{K_{it}})\hat{S}_{3it}^2$$

where

$$\hat{S}_1^2 = \frac{1}{n-1}\sum^{n}(w_i\bar{y}_{i..}-\hat{\bar{Y}})^2$$

$$\hat{S}_{2i}^2 = \frac{1}{m_i-1}\sum^{m_i}(v_{ij}\bar{y}_{ij.}-\bar{y}_{i..})^2$$

$$\hat{S}^2_{3it} = \frac{1}{k_{it}-1} \sum^{k_{it}} (y_{ith}-\bar{y}_{it.})^2$$

and the subscripts 1, 2, 3 indicate the stage which gives rise to the contribution to the variance, and W_i and v_{ij} weights based on the number of third stage units in the primary and secondary.

$$w_i = \frac{\text{Total No. 3rd stage units in the } i^{th} \text{ primary}}{\text{Average No. 3rd stage units per primary}}$$

$$v_{ij} = \frac{\text{No. 3rd stage units in the } j^{th} \text{ secondary and } i^{th} \text{ primary}}{\text{Average No. 3rd stage units per secondary in the } i^{th} \text{ primary}}$$

4.7 Effect of Change in Size of Primary Units

We consider a special case in which all primary units have the same number of secondaries; that is M_i = constant = M. We also suppose that the primary units can be combined to give N ÷ C new primary units of size C·M. The variance of the mean of the original population with N primaries and M secondaries can be expressed as

$$\text{(A)} \quad V(\bar{y}) = \frac{NM-1}{NM} \frac{S^2}{nm} [1 - \frac{m(m-1)}{M(N-1)} + P_1\{\frac{(N-n)m}{(N-1)M} (M-1) - \frac{M-m}{M}\}]$$

and for the variance of the altered primary size

$$\text{(B)} \quad V'(\bar{y}) = \frac{NM-1}{NM} \frac{S^2}{nm} [1 - \frac{m(n-1)}{M(N-C)} + P_2\{\frac{N-nC}{N-C} \frac{m}{MC} (MC-1) - \frac{MC-m}{MC}\}]$$

Subtracting (B) from (A) we conclude that

$$V(\bar{y}) - V'(\bar{y}) \geq 0$$

whenever $P_1 > P_2$ provided both P_1 and P_2 are positive, and where P_1 and P_2 are the intra-class correlation within the primary units. That is, a gain in precision is brought about by enlarging first stage units whenever the intra-class correlation (1) is positive and (2) decreases as the size of the first stage unit increases. Also the smaller P_2 the larger is the gain.

Chapter V. Stratified Sampling

5.0 Introduction

We have studied schemes of selecting sampling units from the entire universe in order to estimate the mean or total. If the population characteristic understudy is heterogeneous or cost limit the size of the sample, it may be found impossible to get a sufficiently precise estimate by taking a random sample from the entire universe. In practice, the main reasons for stratification are: (1) variance considerations, (2) cost constraints, and (3) the need for information by subdivisions of the universe (i.e., States, counties, size groups, etc.).

We suppose it is possible to divide the universe into parts or strata on the basis of some characteristic(s) or information which will make the parts more homogeneous than the whole; that is, information must be available for classifying each sampling unit in the universe into more homogeneous groups or strata. As a result, it should be possible, by properly allocating the sample to the strata, to obtain a better estimate of the population total. Therefore, we propose to answer as best we can the following questions in this chapter:

(1) How should the sample data be analyzed?

(2) How should the strata be constructed?

(3) How many strata should there be?

and in Chapter VIII, we answer

(4) How should the total sample be allocated to strata?

This treatment of stratified sampling is somewhat brief and a departure from the more detailed development generally given. However, the theory is largely a straight forward application of the theory previously developed for the entire universe, but applied to individual stratum. These results are confined to a single survey variable but it should be realized that in practice surveys are multivariate in nature. We have attempted to setforth only the principles to be considered since there is either appreciable "art" involved in applying the techniques in practice or considerable prior data is required to apply the theory directly for multivariate surveys.

5.1 Estimation in Stratified Sampling

A universe of N units is divided into L strata so each unit is in one and only one strata where the h^{th} stratum contains N_h units with a total Y_h for the survey characteristic y, and $\sum^{L} N_h = N$. In each strata a probability sample is selected, the sampling in one stratum being independent of the sample selected in the other strata. Let $\hat{Y}_h$ be an unbiased estimate of Y_h, based on a sample of size n_h; also let $\hat{V}(\hat{Y}_h)$ be an unbiased sample estimate of the stratum variance $V(\hat{Y}_h)$. Applying the theory derived in Chapter II, III and IV to strata, we have

$$(1) \quad \hat{Y}_S = \sum^{L} \hat{Y}_h \text{ , and } \hat{\bar{Y}}_S = \frac{1}{N} \sum^{L} \hat{Y}_h = \frac{1}{N} \sum N_h \hat{\bar{Y}}_h \text{ ,}$$

$$(2) \quad V(\hat{Y}_S) = \sum^{L} V(\hat{Y}_h) = \sum^{L} N_h^2 \left(\frac{N_h - n_h}{N_h}\right) \frac{S_h^2}{n_h} \text{ ,}$$

$$(3) \quad \hat{V}(\hat{Y}_S) = \sum^{L} \hat{V}(\hat{Y}_h) = \sum^{L} N_h^2 \left(\frac{N_h - n_h}{N_h}\right) \frac{s^2}{n_h} \text{ ,}$$

That is, the estimates of the strata totals and variances add up to the population total and variance for each characteristic. Thus, no new principles are involved in analyzing the data if estimates can be made within each stratum. Of course, the usual relationships hold between the variances of totals and means based on the division by N^2 and N_h^2 in (2) and (3) for the universe and individual strata respectively.

5.2 Formation of Strata

If we look at the difference of the variances for $\hat{Y}$ and $\hat{Y}_S$ or $\hat{\bar{Y}}$ and $\hat{\bar{Y}}_S$ using EPS-WOR sampling, we have

$$V(\hat{\bar{Y}}) - V(\hat{\bar{Y}}_S) = \frac{N-n}{nN} S^2 - \sum^{L} \frac{N_h - n_h}{n_h N_h} S_h^2$$

where

$$S^2 = \sum^{L} \frac{N_h S_h^2}{N} + \sum^{L} \frac{N_h}{N} (\bar{Y}_h - \bar{Y})^2$$

From this last equation, we can see that the smaller the values of S_h^2 , the smaller the variance will be. Also, the larger the differences in the strata means, the larger S^2 will be, and the gain in sampling from a stratified population over sampling from the entire population will be increased. Thus, if all prior knowledge, statistical judgment, and available data can be brought into play to achieve similarity within strata and increase differences between strata, a reduction in variance can be obtained. The best information for stratification is usually data on the characteristic y being estimated for some previous time. However, a search is usually necessary just to find some variable which is highly correlated to y, possibly from a previous census. Commonly, a geographic or political subdivision information may provide the only basis for forming strata.

5.2.1 "Exact" Solution

If the distribution of y is known from previous data and a given method of allocating the sample to the strata is specified, the variance to be minimized is a function of the strata boundaries or division points. Consequently, the boundary points will have to be found by iterative procedures until a minimum variance for the population characteristic is obtained. With a high speed computer, this type of solution is feasible though seldom known to be applied.

5.2.2 Approximate Solution

A solution due to Dalenius and Hodges is based on the argument that the distribution of y within strata can be assumed to be rectangular if the number of strata are large. This means that the points y_h (strata boundaries) are to be obtained by taking equal intervals of the cummulatives of $\sqrt{f(y)}$ (i.e., square root of cum. frequencies). Ekman proposes that the points y_h satisfying

$$\frac{N_h}{N}(y_h - y_{h-1}) = \text{constant}$$

will provide approximately the optimum points of stratification. The above is applicable when the unbiased estimator for $\hat{\bar{Y}}_S$ is used. While these approximate procedures are iterative in nature, they readily yield solutions with a computer. The necessity of having the frequencies of y by size available for the population of sampling units is seldom realized. At best, a variable x which is highly correlated with y is all that is usually available. In geographical stratification, the selection of strata boundaries is less amenable to mathematics and are usually based on available data.

5.3 Number of Strata

In 5.2.2 the procedures require a large number of strata which for a given sample size implies L cannot exceed n, if all strata are to be sampled. Many survey statisticians favor the use of this large a number of strata. However, when the stratification for y is made on the basis of another characteristic x a large number of strata may not bring about a proportionate reduction in variance. At best, as L increases the variance decreases inversely as the square of the number of strata, i.e.,

$$V(\hat{\bar{Y}}_S) = \frac{V(\hat{\bar{Y}})}{L^2}$$

when the L strata are of equal size based on y, with $\frac{N_h}{N} = \frac{1}{L}$ and $n_h = \frac{n}{L}$. Letting $L = \frac{n}{2}$ would appear to be a useful upper limit on the number of strata (See 5.5). With a related variable x and a linear relationship between x and y, Cochran gives some evidence that little reduction in variance is to be expected beyond L = 6. When geographic areas or political subdivision are used as strata, the fact that information is needed by strata may determine the minimum number of strata. Likewise if an increase in number of strata leads to reduced survey costs, an increase of L beyond 6 may be advantageous.

5.4 Latin Square Stratification

This topic is also referred to as "deep" stratification or two-way stratification. It is designed for small samples where it is desired for the sample to give proportional representation of each criterion of stratification. This requires that each of the N universe units be classified into a two-way table so the frequencies of the N units in each of the R·C cells can be determined. To achieve the proportional representation, the sample size n leads to the construction of a two-way table with n rows and columns derived from the R-rows and C-columns corresponding to the two criterion for stratification. To estimate the mean, a value for $n \geq R \cdot C$ is required and $n \geq 2RC$ to estimate the variance. The work by Bryant, Hartley and Jessen is described in detail by both the original authors and texts on sampling.

Earlier techniques for this problem were called <u>controlled selection</u>, but while these ideas were doubtlessly attractive as a means of securing preferred types of samples they did not always lead to designs with known precision. In situations where the selection between strata is not made independently but in a <u>dependent manner</u> has been used by Goodman and Kish. In this procedure the joint probability of the preferred combinations of units from two different strata (p_{ij}) is different than zero while the non-preferred pairs have $p_{ij} = 0$.

5.5 Method of Collapsed Strata

As indicated in 5.3, survey statisticians favor using a large number of strata for highly heterogeneous populations. When the number of strata used is equal to n, only one sampling unit can be drawn from each stratum. In this case, it is not possible to estimate the variability within each stratum. In such a case an approximate estimate of the variance of the estimated mean is obtained. The method consists of grouping pairs of strata whose means do not differ very much from each other. Assume that L is an even number so we have $k=L\div 2$ pairs. Suppose the selection within strata is EPS and N_j and $N_{j'}$ are equal. That is, the two paired strata j and j′ are of equal size so the probability of selection is the same. Then consider for the variance

$$V(\bar{Y}) = \frac{1}{N^2} \sum_j^k N_j^2 (y_j - y_{j'})^2$$

which has expectation as follows

$$(4) \quad E[V(\bar{Y})] = \frac{1}{N^2} \sum_j^k N_j^2 (\bar{y}_{N_j} - \bar{y}_{N_{j'}})^2 + \frac{1}{N^2} \sum_j^L N_j (N_j - 1) S_j^2$$

which shows that our variance is over-estimated. The extent of the over-statement is such that it is debatable whether the smaller strata are preferable. If the selection within a stratum is with probability proportional to some variable x, then

$$V(\hat{Y}) = \sum_j^k (\hat{Y}_j - \hat{Y}_{j'})^2$$ is likewise an over-estimate of the variance

where the capital letter $\hat{Y}$ represents a estimated total. If the above variance (4) is compared with the variance for a sample with $L\div 2$ strata and two units per strata, the variance (4) overstates not only the true variance with one unit per stratum but also the variance if the strata were twice as large. It is probably preferable to have only n÷2 strata which are larger with 2 units per strata.

5.6 Post Stratification

In some surveys, it is not possible (or very costly) to know the stratum to which individual sampling units belong until after the survey data has been collected. This technique may also be useful for a sample that has been stratified by one factor, such as geographic regions, and post stratified on a second factor within each of the first factor strata. The stratum size N_h may be known from official statistics, but the stratification characteristic for the units may not be available. In this case, a probability sample from the entire population is selected and the units are classified into strata based on survey data collected for this purpose.

The population total is estimated by

$$\hat{Y} = \sum^{L} N_h \hat{\bar{Y}}_h$$

This total (or mean) is almost as precise as proportional stratified sampling (i.e., $n_h \div N_h$ = constant) if the sample units classified into each stratum is reasonably large, say greater than 20. Let m_h be the number of units falling in the h^{th} strata for a particular selection where m_h will vary from sample to sample even though n is constant. Since m_h is a variable while n_h is fixed, the variance will be increased over stratification which is imposed prior to selecting the sample. If n is moderately large so the probability of m_h being zero is very small, an approximation of order n^{-2} is available. Since

$$E\left(\frac{1}{m_h}\right) = \frac{1}{\frac{nN_h}{N}} - \frac{1 - \frac{N_h}{N}}{n^2 \frac{N_h^2}{N^2}} = \frac{N}{nN_h}\left(1 - \frac{N-N_h}{nN_h}\right)$$

hence,

$$\hat{V}(\hat{Y}) = \frac{N}{n}\left(\frac{N-n}{N}\right) \sum^{L} N_h S_h^2 + \left(\frac{N}{n}\right)^2 \sum^{L} \left(\frac{N-N_h}{N}\right) S_h^2$$

where the first term is the same as for proportional stratification and the second term arise because the m_h's are not distributed proportionally.

5.7 "Domains" or Subpopulations in Stratified Surveys

We discuss domain estimation in this chapter because it is an extension of the post-stratification principles and the theory for simple random sampling is a special case of the theory for stratified random sampling. But, stratified domain estimators are not generalization of simple random sampling. This subject is treated more fully by Hartley (1959) in Analytic Studies of Survey Data. Where subpopulations or "domains" are represented in all strata, we may wish to estimate the domain total or mean. The circumstances are similar to post-stratification in that we cannot identify which domain a sampling unit belongs to until after the survey has been completed. However, it differs in that N_h is known in post-stratification and can be used in the estimators of the total and variance, but the corresponding subpopulation size is unknown in domain estimation theory. Of course, the stratification estimator should be used if the units can be classified before sample selection or post-stratification theory if N_h is known but sampling units cannot be classified until after the survey.

The domain notation is indicated on variables and population parameters by preceding the letter by a subscript $j(j = 1, 2, ...K)$ for each of the subpopulations. Other notation is the same as in previous sections. The strata might be geographic regions and the domains irrigated and non-irrigated farms.

The survey characteristic is defined from the standard stratification theory as ${}_j y_{hi}$ where

$$ {}_j y_{hi} = \begin{cases} y_{hi} \text{ if the } i^{th} \text{ unit in the } h^{th} \text{ strata belongs in the } \\ \quad j^{th} \text{ sub-population} \\ 0 \text{ otherwise} \end{cases} $$

${}_j n_h$ = the number of y_{hi} in the h^{th} strata belonging in the j^{th} subpopulation

where both ${}_j y_{hi}$ and ${}_j n_{hi}$ are treated as random variables.

The domain total is estimated by

$$ {}_j\hat{Y} = \sum_h^L N_h \left\{\sum_i^{n_h} {}_jy_{hi}/n_h\right\} $$

The variance of the total is estimated by

$$ \hat{V}({}_j\hat{Y}) = \sum_h^L N_h^2 \left(\frac{N_h-n_h}{N_h}\right) \left\{{}_j\sum_i^{n_h} \frac{{}_jy_{hi}^2 - {}_j\bar{y}_h^2/n_h}{n_h(n_h-1)}\right\} $$

The sample mean is estimated by first estimating the domain size ${}_jN$ and deriving the mean from

$$ {}_j\hat{\bar{Y}} = \frac{{}_j\hat{Y}}{{}_jN} $$ which is the ratio of two random variables.

The domain size is estimated by first defining a "count" variable ${}_j\mu_{hi}$ as

$$ {}_j\mu_{hi} = \begin{cases} 1 \text{ if the } i^{th} \text{ unit in the } h^{th} \text{ strata belongs in the } j^{th} \text{ subpopulation} \\ 0 \text{ otherwise} \end{cases} $$

The estimator for ${}_jN$ is

$$ {}_j\hat{N} = \sum^L \frac{N_h}{n_h}\, {}_jn_h \quad \text{provided} \quad {}_j\hat{N} \neq 0. $$

Hence, the mean is estimated by a ratio combined over strata

$$ {}_j\hat{\bar{Y}}_c = \frac{{}_j\hat{Y}}{{}_j\hat{N}} $$ (i.e., a combined ratio estimate)

The variance of the mean is estimated by

$$ \hat{V}({}_j\hat{\bar{Y}}) = \sum^L N_h^2 \left(\frac{N_h-n_h}{N_h}\right) \left\{\frac{({}_jn_h-1)\,{}_jS_h^2}{(n_h-1)n_h} + \frac{{}_jP_h\,{}_jq_h}{(n_h-1)}\ {}_j\bar{y}_h^2\right\} $$

where

$$ {}_jP_h = \frac{{}_jn_h}{n_h} \text{ and } {}_jq_h = 1 - {}_jP_h\,. $$

The above theory is the direct result of substituting ${}_jy_{hi}$ for y_{hi} in the standard theory. The results for simple random sampling is obtained as a special case when $L = 1$, $N_h = N$, $n_h = n$, ${}_jy_{hi} = {}_jy_i$ and ${}_j\bar{y}_h = {}_j\bar{y}$.

Chapter VI. Use of Auxiliary Data in Estimators

6.0 Introduction

In Chapters II and III the use of auxiliary data was employed or could have been employed in selecting sampling units with probabilities proportional to size. In Chapter V, we considered employing auxiliary variable(s) or information in the construction of strata. In this chapter, we consider a third way in which we can use auxiliary data-- in the estimator of the population total or mean. The use of auxiliary information brings about consideration of the use of biased estimators of totals, means, and ratios.

The use of ratio estimators will be explored first because ratios are of interest in two respects: (1) the ratio itself is of interest since we may wish to know the pounds of rice per acre, or (2) the ratio of pounds of rice produced, y, to acres of rice, x, may be less variable than the y's themselves and hence the ratio may be utilized in estimating production where there is a known total acreage of rice; that is, we shall use auxiliary information to achieve higher precision based on a ratio estimate.

As an alternative to using auxiliary information in a ratio, we can also consider difference or regression type estimators where a linear relationship between y and x exist. Both of these estimators may be preferable to a ratio when the linear relationship does not pass through the origin. In recent years, the discussion of biased, unbiased and approximately unbiased ratio estimators has received considerable attention in the literature. We will discuss each of these briefly before turning to the consideration of regression type estimators.

6.1 Ratio Estimators

Notation and Definitions:

	Population of Distinguishable Units				Sample	
Unit labels	$\mu_1, \mu_2, \mu_3,$	μ_N	Total	Mean	Total	Mean
Characteristic	$y_1, y_2, y_3,$	y_N	Y	$\bar{Y}$	y	$\bar{y}$
Auxiliary variable	$x_1, x_2, x_3,$	x_N	X	$\bar{X}$	x	$\bar{x}$
Ratios	$r_1, r_2, r_3,$	r_N	R	$\bar{R}$	r	$\bar{r}$

where $r_i = \dfrac{y_i}{x_i}$, $R_M = \dfrac{\bar{Y}}{\bar{X}}$ and $r_M = \dfrac{\bar{y}}{\bar{x}}$

Y or $\bar{Y}$ is to be estimated from a sample of size n, and X and $\bar{X}$ are known exactly. To use the ratio estimator, the hypothesis is needed that the relationship between y and x passes through the origin.

6.1.1 Ratio of Means

The ratio of means $R_M = \frac{\bar{Y}}{\bar{X}}$ is the "classical" ratio and is biased except under special conditions. We should emphasize that R_M is an overall rate while $\bar{R}$ is an average rate per unit in the population. This distinction is of importance if we are interested in the use cited in (1) rather than (2) above.

We consider the bias of $\hat{Y}$ and $\hat{R}_M$ which can be found using the results of Goodman and Hartley from the expression for the covariance of $\hat{R}_M$ and $\bar{x}$. For EPS, we have

$$Cov(\frac{\bar{y}}{\bar{x}} , \bar{x}) = E(\frac{\bar{y}}{\bar{x}} . \bar{x}) - E(\frac{\bar{y}}{\bar{x}})E(\bar{x})$$

Rearranging terms

$$E(\frac{\bar{y}}{\bar{x}})E(\bar{x}) = E(\bar{y}) - Cov(\frac{\bar{y}}{\bar{x}} , \bar{x})$$

Or

$$\bar{X}\ E(\hat{R}_M) = \bar{Y} - Cov(\frac{\bar{y}}{\bar{x}} , \bar{x})$$

$$E(\hat{R}_M) = R_M - \frac{1}{\bar{X}} Cov(\frac{\bar{y}}{\bar{x}} , \bar{x})$$

The bias is:

$$B(\hat{R}_M) = E(\hat{R}_M) - R_M = - \frac{1}{\bar{X}} Cov(\hat{R}_M , \bar{x})$$

which is zero if $Cov(R_M , \bar{x}) = 0$, i.e., the correlation is zero.

If we express the $Cov(\hat{R}_M , \bar{x})$ as $\rho\sigma(\hat{R}_M)\sigma(\bar{x})$ then an upper bound on the bias can be readily found in terms of the coefficient of variation of $\bar{x}$. That is:

$$B(\hat{R}_M) = - \rho \frac{\sigma(\hat{R}_M)\sigma(\bar{x})}{\bar{X}}$$

or

$$\frac{B(\hat{R}_M)}{\sigma(\hat{R}_M)} = - \rho \frac{\sigma(\bar{x})}{\bar{X}} = - \rho\ C.V.(\bar{x})$$

Since $|\rho| \leq 1$, the relative and absolute bias are:

$$\frac{|B(\hat{R}_M)|}{\sigma(\hat{R}_M)} \leq \text{C.V.}(\bar{x}), \text{ and } |B(\hat{R}_M)| \leq \frac{\sigma(\hat{R}_M)\sigma(\bar{x})}{\bar{X}}$$

Hence, if the C.V.($\bar{x}$) is small, the bias will be negligibly small.
We make

$$\text{C.V.}(\bar{x}) = \sqrt{\frac{N-n}{nN}} \quad \frac{S_x}{\bar{X}}$$ small by our choice of the sample size n.

The bias of $\hat{Y}$ is $XB(\hat{R}_M)$. In the foregoing discussion $\bar{y}$ and $\bar{x}$ may be replaced by y and x.
The $|B(\hat{R}_M)|$ is of order $(\frac{1}{n})$ since both $\sigma(\hat{R}_M)$ and $\sigma(\bar{x})$ contain the factor $\frac{1}{\sqrt{n}}$.

An approximate expression for the bias of $\hat{R}_M$ based on a sample of size n is obtained by retaining the first two terms of the Taylor expansion of

$$f(\theta) = \frac{E(\bar{y}-R_M\bar{x})}{\bar{X}+\theta(\bar{x}-\bar{X})} \quad \text{around } \theta = 0.$$

or

$$b(\hat{R}_M) \doteq -\frac{\rho\sigma(\bar{y})\sigma(\bar{x})-R_M\sigma^2(\bar{x})}{\bar{X}^2}$$

<u>6.1.2 The Variance of $\hat{Y} = V(X\hat{R}_M)$</u>

$$V(\hat{Y}) = X^2V(\hat{R}_M) = \frac{N^2}{n}(\frac{N-n}{N})\ [S_y^2 + R_M^2S_x^2 - 2R_M\rho S_yS_x]$$

where the $V(\hat{R}_M)$ was given in Chapter 1. The $V(\hat{R}_M) = 0$ if y is proportional to x.
In general, the variance of <u>any ratio</u> (in this case R_M) can be obtained by "plugging" in $y_i - R_Mx_i = z_i$ for y_i in V(Y) <u>regardless of the type of sampling</u> that has been used, i.e.,

$$V(R_M) = \frac{N-n}{\bar{X}^2Nn}\ S_{z_i}^2$$

This can be seen by examining

$$E(\hat{R}_M - R_M) = E(\frac{\bar{y}-R_M\bar{x}}{\bar{x}})^2 = E(\frac{\bar{y}-R_M\bar{x}}{\bar{X}+\delta\bar{x}})^2$$

where $\delta\bar{x} = \bar{x} - \bar{X}$

Then the mean square error is the value at $\Theta = 1$ of the function

$$f(\Theta) = E\left(\frac{\bar{y}-R_M\bar{x}}{\bar{X}+\Theta\delta\bar{x}}\right)^2$$

Developing Taylor's expansion of $f(\Theta)$ we get

$$E(R_M-R_M)^2 = \frac{E(\bar{y}-R_M\bar{x})^2}{\bar{X}^2} - \frac{2E[(\bar{y}-R_M\bar{x})^2\delta\bar{x}]}{\bar{X}^3} + \ldots$$

A first approximation to the mean square error is obtained by retaining only the first term which is of order $\frac{1}{n}$. A second approximation is obtained by retaining both terms.

<u>6.1.3 Mean of Ratios</u>

The total is estimated by

$$\hat{Y} = \bar{r}\,\bar{X} \quad \text{where } \bar{r} = \frac{1}{n}\Sigma' r_i$$

However, the behavior of the r_i's may be erratic and $\hat{Y}$ is very badly biased.

It should be noted that the use of the covariance in 6.1.1 to obtain the bias of the ratio can be generalized to any type ratio. Hence,

$$B(\bar{R}) = -\frac{1}{\bar{X}}\,\mathrm{Cov}\left(\frac{y_i}{x_i}, x_i\right) = -\frac{1}{\bar{X}}\,\rho\sigma(r_i)\sigma(x_i)$$

and

$$\frac{|B(\bar{R})|}{\sigma(r_i)} \leq \mathrm{C.V.}(x_i)$$

However, the bias of the total $\hat{Y}$ is obtained as

$$\left|\frac{B(\hat{Y})}{\sigma(r_i)\bar{X}}\right| \leq \mathrm{C.V.}(x_i)$$

but $\sigma(\hat{Y}) = \bar{X}\,\sigma(\bar{r}) = \bar{X}\,\frac{\sigma(r_i)}{\sqrt{n}}$, replacing $\sigma(r_i)$ above

$$\therefore \left|\frac{B(\hat{Y})}{\sqrt{n}\,\sigma(\hat{Y})}\right| \leq \mathrm{C.V.}(x_i)$$

and $\left|\frac{B(\hat{Y})}{\sigma(\hat{Y})}\right| \leq \sqrt{n}\ \mathrm{C.V.}(x_i).$

That is, the bias as a proportion increases with $\sqrt{n}$. Consequently, R_M is much preferred to $\bar{R}$ in estimating the population total Y even though the variance is of about the same magnitude. The approximate variance can be obtained by "plugging" in $y_i - \bar{R}x_i = z_i$ for y_i to find the variance as was done for $V(R_M)$ earlier. However, we shall return to the consideration of r_i when we seek approximately unbiased estimators of ratios.

6.2 Unbiased Ratio Estimation

In view of the fact that, under simple random sampling, the ratio estimator $\hat{Y} = X\frac{\bar{y}}{\bar{x}}$ is biased, we wish to consider modifying the sampling procedure so the same estimator becomes unbiased. This can be accomplished by selecting the sample with probability proportionate to its aggregate size. This can be best done by selecting the first unit in the sample with pp to x and the other (n-1) units with equal probabilities without replacement. Under this procedure of selection, the probability of selecting a particular sample(s) of size n is given by

$$P(S_n) = \frac{\Sigma' x_i}{X \binom{N-1}{n-1}} = \frac{(n-1)!(N-n)!}{(N-1)!} \quad \frac{\Sigma' x_i}{\Sigma x_i}$$

6.2.1 The Estimator of the Population Total

$$\hat{Y} = X \frac{\Sigma' y_i}{\Sigma' x_i} = X \frac{\bar{y}}{\bar{x}}$$

which can be shown to be unbiased since the expectation over all possible samples

$$E(\hat{Y}) = \overset{S}{\Sigma^n} X \cdot \frac{\Sigma' y_i}{\Sigma' x_i} \cdot \frac{\Sigma' x}{X\binom{N-1}{n-1}} = \overset{S}{\Sigma^n} \frac{\Sigma' y_i}{\binom{N-1}{n-1}} = Y$$

6.2.2 The Population Variance

Using the most general form given on page 11, Section 4.1.4 for unbiased estimator

$$V(\hat{Y}) = \sum_{i=1}^{N'} T_i^2 \, P(S_i) - Y^2$$

$$= \frac{X}{\binom{N-1}{n-1}} \sum^{N'} \frac{(\Sigma' y_i)^2}{\Sigma' x_i} - Y^2$$

which is zero if y_i is proportionate to x_i. In the variance $P(S_i)$ is the same as $P(S_n)$ for a particular sample of size n. The total number of samples of size n is N'.

6.2.3 Sample Estimate of the Variance

An unbiased estimate of the second term Y^2 is given by

$$\widehat{Y^2} = \left[\frac{\Sigma' y_i^2}{\binom{N-1}{n-1}} + \frac{2\Sigma'_{i<j} y_i y_j}{\binom{N-2}{n-2}} \right] \div P(S_n)$$

Hence

$$v(\hat{Y}) = \hat{Y}^2 - \widehat{Y^2}$$

since

$$E(\hat{Y}^2 - \widehat{Y^2}) = E(\hat{Y}^2) - Y^2 = V(\hat{Y})$$

The estimator of the variance may assume negative values for some of the S_n samples.

6.3 Approximately Unbiased Ratio Estimators

We now return to the ratio estimators considered in 6.1 and try to remove the bias. This work follows that of Hartley and many others. We consider $\bar{r}$

$$\bar{r} = \frac{\Sigma' r_i}{n} = \frac{1}{n} \Sigma' \frac{y_i}{x_i}$$

and remove the bias.

6.3.1 An Unbiased Estimator of the Population Mean ($\bar{X}$ known)

$$E(r_i) = \frac{E(y_i)}{E(x_i)} - \frac{\mathrm{Cov}(\frac{y_i}{x_i}, x_i)}{E(x_i)}$$

$$E(r_i) = \frac{\bar{Y}}{\bar{X}} - \frac{\mathrm{Cov}(\frac{y_i}{x_i}, x_i)}{\bar{X}}$$

$$E(\bar{X}\bar{r}) = \bar{X}\, E(r_i) = \bar{Y} - \mathrm{Cov}(\frac{y_i}{x_i}, x_i)$$

From earlier results we note that

S_y^2 and S_x^2 are estimated unbiasedly. Also,

$$S_{y+x}^2 = \frac{1}{N-1} \sum_i (y_i+x_i-\bar{Y}-\bar{X})^2 \text{ is estimated unbiasedly by}$$

$$s_{y+x}^2 = \frac{1}{n-1} \sum_i{}' (y_i+x_i-\bar{y}-\bar{x})^2, \text{ i.e.,}$$

$$S_y^2 + S_x^2 + \frac{2}{N-1} \sum_i (Y_i-\bar{Y})(X_i-\bar{X}) \text{ is estimated unbiasedly by}$$

$$s_y^2 + s_x^2 + \frac{2}{n-1} \sum_i{}' (y_i-\bar{y})(x_i-\bar{x})$$

$$\therefore \operatorname{Cov}\left(\frac{y_i}{x_i}, x_i\right) = \frac{N\left(\frac{y_i}{x_i} - \bar{R}\right)(x_i - \bar{X})}{N-1} \cdot \frac{N-1}{N}$$

is estimated unbiasedly by

$$\sum_i^n{}' \frac{\left(\frac{y_i}{x_i} - \bar{r}\right)(x_i-\bar{x})}{n-1} \cdot \frac{N-1}{N}$$

Hence the mean is estimated unbiasedly by

$$\hat{\bar{Y}}' = \bar{X}\bar{r} + \frac{N-1}{N} \cdot \frac{1}{n-1} \left(\sum{}' \frac{y_i x_i}{x_i} - n\bar{r}\bar{x}\right)$$

Or

$$\hat{\bar{Y}}' = \bar{X}\bar{r} + \frac{N-1}{N} \frac{n}{n-1} (\bar{y} - \bar{r}\,\bar{x}) \qquad \text{[Hartley-Ross]}$$

Consequently an unbiased estimator of R_M is available upon dividing $\hat{\bar{Y}}'$ by $\bar{X}$.

$$\hat{R}_M' = \frac{\hat{\bar{Y}}'}{\bar{X}}$$

The variance for large N is given by

$$V(\hat{\bar{Y}}) = \frac{1}{n} [S_y^2 + \bar{R}^2 S_x^2 - 2\bar{R} \operatorname{Cov}(y,x)] + \frac{1}{n(n-1)} [S_r^2 S_x^2 + \operatorname{Cov}^2(r,x)]$$

However, the second term is usually negligible and the large sample approximation to the variance of the ratio takes the usual form.

$$V(\hat{R}_M') = \frac{S_y^2 + \bar{R}^2 S_x^2 - 2\bar{R}\rho S_y S_x}{n\bar{X}^2}$$

<u>6.3.2 An Unbiased Estimator for R_M for Large n ($\bar{X}$ unknown)</u>

$$E\left(\frac{\bar{y}}{\bar{x}}\right) = \frac{\bar{Y}}{\bar{X}} - \frac{\mathrm{Cov}\left(\frac{\bar{y}}{\bar{x}}, \bar{x}\right)}{\bar{X}}$$

Assuming $N \to \infty$ and $\delta x_i = \frac{x_i - \bar{X}}{\bar{X}}$ is small then it can be shown that

$$\mathrm{Cov}\left(\frac{\bar{y}}{\bar{x}}, \bar{x}\right) \doteq \frac{1}{n}\,\mathrm{Cov}\left(\frac{y_i}{x_i}, x_i\right)$$

If $C_1 + C_2 = 1$, then

$$E\left(C_1\bar{r} + C_2\frac{\bar{y}}{\bar{x}}\right) = \frac{\bar{Y}}{\bar{X}} - \frac{\mathrm{Cov}\left(\frac{y_i}{x_i}, x\right)}{\bar{X}}$$

and determine C's such that $C_1 + \frac{1}{n}C_2 = 0$. Solving these two equations for C_1 and C_2 we obtain

$$C_1 = -\frac{1}{n-1} \quad \text{and} \quad C_2 = \frac{n}{n-1}$$

Therefore an unbiased estimator of $\frac{\bar{Y}}{\bar{X}} = R_M$ is

$$\hat{R}_M'' = \frac{n}{n-1}\left(\frac{\bar{y}}{\bar{x}}\right) - \frac{1}{n-1}\bar{r}$$

This is unbiased only as far as the approximation of the covariances is correct. An alternative derivation is also available for an approximate unbiased estimator of $\frac{\bar{Y}}{\bar{X}}$. This development follows from 6.3.1 where $\bar{X}$ is replaced by $\bar{x}$. If N is large in the expression on page 7 for the Hartley-Ross estimator, then

$$\frac{\hat{\bar{Y}}'}{\bar{x}} = \bar{r} + \frac{N-1}{N} \cdot \frac{n}{n-1}\left(\frac{\bar{y}}{\bar{x}} - \bar{r}\right)$$

which reduces to $\frac{n}{n-1}\left(\frac{\bar{y}}{\bar{x}}\right) - \frac{1}{n-1}\bar{r}$.

However, this is an approximately unbiased estimator which is an improvement over $\frac{\bar{y}}{\bar{x}}$ if the coefficient of variation of $\bar{x}$ is small because the bias is now of a smaller order.

The difference in the variance of $\hat{R}_M$ and $\hat{R}_M'$ is

$$V(\hat{R}_M) - V(\hat{R}_M') = \frac{(R_M^2 - \bar{R}^2)S_X^2 - 2pS_XS_Y(R_M-\bar{R})}{n\bar{X}^2}$$

$$= \frac{[(R_M-\beta)^2-(\bar{R}-\beta)^2]S_X^2}{n\bar{X}^2}$$

where β is the regression coefficient of y on x. Consequently, for large n we see $\hat{R}_M'$ will be more efficient than $\hat{R}_M$ if and only if β is nearer to $\bar{R}$ than to R_M. If the two ratio are equal, the two variances are equal. In practice, it will be unlikely that this will be known.

6.3.3 Quenouille Method of Bias Reduction

A random sample of size 2n is split at random into two subsamples each of size n. Based on the two subsamples and the entire sample, we construct an estimator of the ratio

$$R_Q = W_1 \frac{\bar{y}_1}{\bar{x}_1} + W_2 \frac{\bar{y}_2}{\bar{x}_2} + (1 - W_1-W_2) \frac{\bar{y}}{\bar{x}}$$

which simplifies because of equal sample sizes so $W_1 = W_2 = W$ hence

$$R_Q = W\frac{\bar{y}_1}{\bar{x}_1} + W\frac{\bar{y}_2}{\bar{x}_2} + (1 - 2W) \frac{\bar{y}}{\bar{x}}$$

The bias in the estimate to the first degree approximation will be zero if

$$W = - \frac{(N-2n)}{2N}$$

Hence, an approximately unbiased ratio estimator is

$$\hat{R}_Q = \frac{(2N-2n)}{N} \frac{\bar{y}}{\bar{x}} - \frac{(N-2n)}{2N} \frac{\bar{y}_1}{\bar{x}_1} - \frac{(N-2n)}{2N} \frac{\bar{y}_2}{\bar{x}_2}$$

$$\doteq 2\,\frac{\bar{y}}{\bar{x}} - \frac{1}{2}\,\frac{\bar{y}_1}{\bar{x}_1} - \frac{1}{2}\,\frac{\bar{y}_2}{\bar{x}_2}$$

With some effort, it can be shown that the mean square error of $\hat{R}_Q$ is approximated by

$$\text{M.S.E.}(\hat{R}_Q) = \left(\frac{1}{2n} - \frac{1}{N}\right)\left(\frac{\bar{Y}}{\bar{X}}\right)^2 [C_y^2 + C_x^2 - 2\rho C_y C_x]$$

where C_y^2 and C_x^2 are the square coefficients of variation of y and x. Since this approximation is of the same order as the M.S.E.$(\hat{R})$, this latter estimator may be preferred. To estimate the mean $\bar{Y}$, we still require knowledge of $\bar{X}$.

The Quenouille method is probably best when we use groups of size one, that is, the estimator becomes

$$\hat{R}_Q = n\frac{\bar{y}}{\bar{x}} - \frac{n-1}{n} \sum_i{}' \frac{n\bar{y} - y_i}{n\bar{x} - x_i}$$

However, the variance must be obtained by Taylor's Expansion.

6.3.4 Mickey's Estimator - A generalized estimator of the mean

$$W_\alpha = a(z_\alpha)\,\bar{X} + \frac{N-\alpha}{N(n-\alpha)}\,[y_n - a(z_\alpha) + X_n] - \frac{N-n}{N(n-\alpha)}\,[y_\alpha - a(z_\alpha)\,x_\alpha]$$

where the choice of $a(z_\alpha)$ leads to a specific estimator.

(A) Let $a(z_\alpha) = \frac{1}{\alpha}\sum^{\alpha} \frac{y_i}{x_i}$

if $\alpha = 1$ $a(z_\alpha) = \frac{y_1}{x_1}$

Then $W_1 = \frac{y_1}{x_1}\,\bar{X} + \frac{N-1}{N(n-1)}\,[n\bar{y}_n - \frac{y_1}{x_1}\,n\,\bar{x}_n]$

Averaging over all possible selection of units

$$W_1^* = \bar{X}\,\frac{1}{n}\sum^{n} \frac{y_i}{x_i} + \frac{(N-1)n}{N(n-1)}\,(\bar{y}_n - \bar{r}\,\bar{x}_n)$$

$$= \bar{X}\,\bar{r} + \frac{(N-1)n}{N(n-1)}\,(\bar{y} - \bar{r}\,\bar{x}) \quad \text{[Hartley-Ross]}$$

(B) Let $\alpha \neq 1$, then $a(\bar{z}_\alpha)$ in the estimator becomes r_α and

$$W_\alpha = r_\alpha \bar{X} + \frac{(N-\alpha)n}{N(n-\alpha)} (\bar{y}_n - r_\alpha \bar{x}_n)$$

Averaging over all samples of $\binom{n}{\alpha}$

$$W^*_\alpha = \bar{r}_\alpha \bar{X} + \frac{(N-\alpha)n}{N(n-\alpha)} (\bar{y}_n - \bar{r}_\alpha \bar{x}_n)$$

For $\alpha = n-1$, we probably have the best estimator

$$W^*_{n-1} = \bar{r}_{n-1} \bar{X} + \frac{(N-n+1)n}{N} (\bar{y}_n - \bar{r}_{n-1} \bar{x}_n)$$

If X is normal and regression passes through origin this estimator is more efficient than Hartley-Ross.

6.3.5 The Product Estimator

Although similar to the ratio estimator, it is much less frequently used. Generally, the mean per establishment is desired and we wish to estimate estimator of the mean in EPS-WOR is:

$$\hat{\bar{Y}} = \bar{y} \, . \, \bar{x} \left(\frac{1}{\bar{X}}\right)$$

which is biased, but useful for many purposes.
The bias is given by

$$\text{Bias}(\bar{y}.\bar{x}) = \left(\frac{N-n}{Nn}\right) \sigma_{xy} = \left(\frac{N-n}{Nn}\right) \rho \sigma_x \sigma_y \, .$$

The approximate variance is given by:

$$V(\hat{\bar{Y}}) = \bar{X}^2 V(\bar{y}) + \bar{Y}^2 \, V(\bar{x}) + 2\bar{Y}\bar{X}\rho \, [V(\bar{y})V(\bar{x})]^{\frac{1}{2}}$$

In the product estimator the position of $\bar{X}$ is in the denominator while in the ratio estimator $\bar{X}$ is in the numerator. The product estimator depends on a negative correlation between y and x to be more efficient than the simple estimator of $\hat{\bar{Y}}$. The sample variance can be obtained by using analogous sample estimators for the parameters. The approximate variance is:

$$\hat{V}(\hat{\bar{Y}}) = (\frac{N-n}{Nn}) [S_y^2 + \frac{\bar{Y}^2}{\bar{X}^2} S_x^2 + 2 \frac{\bar{Y}}{\bar{X}} S_{xy}]$$

6.4 Ratio Estimator in Multistage Sampling

We consider the ratio estimator

$\hat{R}_M = \frac{\hat{Y}}{\hat{X}}$ where $\hat{Y}$ and $\hat{X}$ are unbiased estimators of the totals for the sampling scheme employed.

The variance of the ratio estimator is obtained by "plugging" in $y_i - \hat{R}x_i = z_i$ for y_i in the formula for the variance of a total, V(Y), resulting in $V(\hat{R}_M) = \frac{1}{\bar{X}^2} V(\hat{Y}-\hat{R}\hat{X})$.

6.4.1 Two Stage Sampling - EPS-WOR at both stages

The totals Y and X are estimated by the method of 3.1.1., that is

$$\hat{Y} = \frac{N}{n} \Sigma' M_i \bar{y}_i$$

Hence

$$V(R) = \frac{1}{\bar{X}^2} \frac{N}{n} \Sigma \frac{M_i^2}{m_i} (1 - \frac{m_i}{M_i}) S_{R_i}^2 + \frac{N^2}{n} (1 - \frac{n}{N}) \frac{1}{N-1} \Sigma (Y_i-\hat{R}_M X_i-\bar{Y} + \hat{R}_M\bar{X})^2$$

where

$$S_{R_i}^2 = \Sigma(y_{it} - R_M x_{it} - \bar{Y}_i + R_M \bar{X}_i)^2$$

and

$$\bar{Y} - \hat{R}_M \bar{X} = \frac{1}{N} \Sigma(Y_i-\hat{R}_M X_i)$$

6.5 Ratio Estimator in Double Sampling

It happens frequently that the population mean $\bar{X}$ is not known, hence the usual ratio estimate $\hat{R}_M\bar{X}$ cannot be made. It is common in such a situation to use the technique known as double sampling. The technique consist in taking a large sample of size n′ to estimate the population mean $\hat{\bar{X}}$ (assuming X is cheaper to observe than Y) while a subsample of size n is drawn from n′ to observe the characteristic y under study. The simplest estimate of the mean is the usual biased ratio estimator.

6.5.1 The Classical Ratio: $\hat{\bar{Y}}' = \hat{R}_M \bar{x}_{n'}$

where $\hat{R}_M = \frac{\bar{y}_n}{\bar{x}_n}$ is based on the small sample of size n. The relative bias of this estimator is:

$$B(\hat{\bar{Y}}') = (\frac{1}{n} - \frac{1}{n'})(C_x^2 - \rho\ C_x C_y)$$

where C_x and C_y are the coefficients of variation. The bias is negligible if the sample size n is sufficiently large so $C_{\bar{x}}$ is small. It will be zero if the regression of y on x is linear and passes through the origin.

The variance of $\hat{\bar{Y}}'$ may be more efficient than the estimate of $\hat{\bar{Y}}$ based on the small sample n. The variance is given by:

$$V(\hat{\bar{Y}}') = (\frac{1}{n} - \frac{1}{n'})(S_y^2 + R_M^2 S_X^2 - 2R_M\ \rho\ S_y S_x) + (\frac{1}{n} - \frac{1}{n'})S_y^2$$

which will be smaller than the variance of

$$V(\hat{\bar{Y}}),\ \text{if}\ \rho \frac{C_y}{C_x} > \frac{1}{2}\ .$$

6.5.2 An Unbiased Ratio-Type Estimate

$$\hat{\bar{Y}}'' = \bar{r}_n\ \bar{x}_{n'} + \frac{n(n'-1)}{n'(n-1)}\ (\bar{y}_n - \bar{r}_n\ \bar{x}_n)$$

and the variance is approximated by

$$V(\hat{\bar{Y}}'') = (\frac{1}{n} - \frac{1}{n'})(S_y^2 + \bar{R}^2\ S_x^2 - w\bar{R}\ \rho\ S_y S_X) + \frac{S_y^2}{n'}$$

6.6 Regression Type Estimators

The ratio estimator is best when the relationship between y and x is a straight line through the origin, so y - kx = 0. If the relationship is of the type y - kx = a, it is more appropriate to try an estimator based on differences of the form $y_i - kx_i = \mu_i$. Such estimators are called difference estimators or the "working" regression slope theory. The value k is determined or guessed a priori and we expect the $V(\bar{\mu})$ to be less than the $V(\bar{y})$.

6.6.1 Difference Estimation - EPS- WOR

This unbiased estimator is called a regression estimator only because it can be put in the form

$$\bar{y}_k = \bar{y} + k(\bar{X} - \bar{x})$$

which resembles a regression estimator evaluated at $\bar{X}$. Now the variance of $\bar{y}_k$ equal the variance of $\bar{\mu}$ regardless of the sampling scheme, i.e.

$V(\bar{y}_k) = V(\bar{\mu})$. The standard formulas which apply to $\bar{\mu}$, all apply to $\bar{y}_k$. Further

$$V(\mu_i) = V(y_i) + k^2 V(x_i) - 2k \, Cov(y_i, x_i)$$

and

$$V(\bar{\mu}) = \left(\frac{N-n}{N}\right) \frac{S_y^2 + k^2 S_X^2 - 2k\rho S_x S_y}{n}$$

The estimate of k is determined a priori and must not be revised after sampling has begun. The value of k which will minimize $V(\bar{\mu})$ is

$$k = \beta = \frac{\rho S_y}{S_x} \quad \text{the population regression coefficient.}$$

The difference estimator is superior to the simple average $\bar{y}$ if

$k S_x^2 (k - 2\rho \frac{S_y}{S_x}) < 0$ or $k(k - 2\beta)<0$. That is, if k lies between 0 and 2β.

If we consider the ratio of the difference between $V(\bar{\mu})$ and $V(\bar{\mu}_{min})$ divided by $V(\bar{\mu}_{min})$ we obtain

$(\frac{k}{\beta} - 1)^2 \frac{\rho^2}{1-\rho^2} \leq \Theta$. If we can specify $\Theta = .1$ and $\hat{\rho} = .7$, then

$|\frac{k}{\beta} - 1| \leq \sqrt{\frac{\Theta(1-\rho)^2}{\rho^2}}$ and k should be between .68 and 1.32.

6.6.2 Difference Estimation in Stratified Sampling

A different value of k can be used in each stratum, i.e., we denote the value as k_h for the h^{th} stratum.

$$\mu_{hi} = y_{hi} - k_h x_{hi}$$

The population total is:

$$U = Y - \sum_h^L k_h X_h \quad \text{for L strata}$$

and the mean:

$$\bar{U} = \bar{Y} - \sum_h^L \frac{k_h N_h \bar{X}_h}{N}$$

$\therefore \bar{y}_k = \bar{\mu} + \sum_h^L \frac{N_h}{N} k_h \bar{X}_h$ where $\bar{X}_h$ is known.

6.6.3 Regression Estimation

Instead of determining k a priori, the population regression coefficient β is estimated from the sample. The sample estimate of β is

$$b = \frac{\Sigma'(x_i-\bar{x})(y_i-\bar{y})}{\Sigma'(x_i-\bar{x})^2}$$

and the estimator is

$$\bar{y}_\beta = \bar{y} - b(\bar{x}-\bar{X})$$

Since b is a random variable, exact expressions for the expected value and the variance of the regression estimator are hard to find. The large sample approximation to the variance is

$$V(\bar{y}_\beta) = \frac{S_y^2(1-\rho^2)}{n} = \frac{S_y^2 + \beta^2 S_X^2 - 2\beta\rho S_x S_y}{n}$$

6.7 Multivariate Auxiliary Data

Instead of a single x variable, we now consider two or more x's.

6.7.1 Difference Estimators

Consider forming a difference estimator for the mean of y based on each of the x-variables, and then combining them using appropriate weights. We form each estimator of y as

$$t_i = \bar{y} - k_i\ (\bar{x}_i - \bar{X}_i) \qquad i = 1, 2..P$$

Let W_i be weights adding to one. Then

$\bar{y} = \sum^{P} W_i t_i$ is an unbiased estimator of $\bar{Y}$. Its variance is given by

$$V(\bar{y}) = \sum_{ij}\sum W_i W_j \ \mathrm{Cov}(t_i,t_j) \qquad \text{Chapter 1.}$$

Defining $S_{\mu v}$ as the covariance between μ and v and letting 0,1,..P stand for the variates $y_1, x_1, x_2, \ldots x_P$ respectively, we have

$$\mathrm{Cov}(t_i,t_j) = \frac{N-n}{Nn}(S_{00} - k_i S_{0i} - k_j\ S_{0j} + k_i k_j\ S_{ij})$$

An unbiased estimator of the variance is given by

$$V(\bar{y}) = \frac{N-n}{Nn} \cdot \frac{1}{n-1} \sum_{j=1}^{n}{}' [y_j - \sum_{i=1}^{P} W_i k_i (x_{ij} - \bar{x}_i)]^2$$

6.7.2 Ratio Estimation

We form ratio estimators instead of difference estimators as in the preceding section and weight these together so $\sum^{P} W_i = 1$.

$$\bar{y}_{R_M} = \sum^{P} W_i \bar{X}_i \hat{R}_{M_i}$$

and

$$V(\bar{y}_{R_M}) = \sum_{ij}\sum W_i W_j \bar{X}_i \bar{X}_j \operatorname{Cov}(\hat{R}_{M_i}, \hat{R}_{M_j})$$

The estimator is biased and the expression for the variance is only approximate in the same way that $\hat{R}_M$ and $V(\hat{R}_M)$ were.

6.7.3 Mickeys Estimator

Let

$$a(\bar{z}_\alpha) = \frac{\sum_{2}^{\alpha} (x_i - \bar{x}_\alpha)(y_i - \bar{y}_\alpha)}{\sum_{2}^{\alpha} (x_i - \bar{x}_\alpha)^2}$$

6.7.4 Remark: The estimators $\bar{y}$, $\frac{\bar{y}}{\bar{x}} \bar{X}$, $\bar{y} - k(\bar{x} - \bar{X})$ and $\bar{y} - b(\bar{x} - \bar{X})$ all belong to the class of estimators $\bar{y} - h(\bar{x} - \bar{X})$ where h is a random variable converging to some finite value. Thus, we have:

$h = 0$ for the estimator $\bar{y}$,

$h = \frac{\bar{y}}{\bar{x}}$ for the estimator $\hat{R}_M$,

$h = k$ for the difference estimator,

$h = b$ for the regression estimator.

6.8 Alternative Uses of Auxiliary Information

To this point, we have employed auxiliary information specifically in three different ways: (1) Constructing the estimator, (2) Construction of strata, and (3) Assigning selection probabilities to sampling units. A brief summary of these methods of improving the efficiency is now presented.

6.8.1 Choice of Estimator

To facilitate comparisons, the estimators of the population total are stated in a slightly modified form.

1. Simple: $N[(1)(\bar{y}) + 0(\bar{X}-0)]$

2. Classical Ratio: $N[(0)(\bar{y}) + \frac{\bar{y}}{\bar{x}}(\bar{X}-0)]$

3. Regression: $N[(1)(\bar{y}) + b(\bar{X}-\bar{x})]$

4. Difference: $N[(1)(\bar{y}) + b_0(\bar{X}-\bar{x})]$

5. Product: $N[(0)(\bar{y}) + \bar{y}\bar{x}(\frac{1}{\bar{X}} - 0)]$

In considering the problem of estimating the total from a sample, what is required besides the sample means $\bar{y}$?

(1) For the simple estimator, we need only N,

(2) For the ratio estimator, we need only X and not N since $X = N\bar{X}$,

(3) For the regression estimator, we need to know both N and X (or $\bar{X}$),

(4) For the difference estimator, we need to know N, X and b_0,

(5) For the product estimator, we need to know N and X.

It is of some importance to realize that the total cannot be estimated unless number of units in the frame, N, is known with the exception of the ratio estimator which is, in general, biased.

The efficiency of these estimators for moderate size samples (biases negligible) depends on the magnitude of the variances. Consequently, variance efficiency of an estimator A compared to B is defined as follows:

$$VE(A/B) = \frac{Var(B)}{Var(A)} .$$

Where estimator A will be relatively more efficient than B if the ratio is greater than 1. The efficiency of estimators 1, 2, and 3 above are compared for the special case where VAR(X) $\doteq$ Var(Y). The variance efficiencies under this condition are:

$$VE(2/1) = \frac{1}{2(1-\rho)}$$

and

$$VE(3/1) \doteq \frac{1}{1-\rho^2} .$$

Therefore, the ratio estimator is always more efficient than the simple estimator whenever $\rho > (\frac{1}{2}) \frac{V(x)}{V(y)} = \frac{1}{2}$, and the regression estimator is always more efficient than the simple estimator when $\rho > 0$. In general, the