T088
26/11/84

RESTRICTED COLLECTION

CONTRIBUTIONS TO THE THEORY OF OPTIMUM UTILISATION OF AUXILIARY INFORMATION IN SURVEY SAMPLING

INDIAN STATISTICAL INSTITUTE LIBRARY
No. T 088
26 NOV 1984
CALCUTTA

By
V. N. REDDY
Indian Institute of Management, Calcutta

A thesis submitted to the Indian Statistical Institute
in partial fulfilment of the requirements
for the award of the degree of
DOCTOR OF PHILOSOPHY

CALCUTTA
1975

CONTRIBUTIONS TO
THE THEORY OF OPTIMUM UTILISATION OF AUXILIARY
INFORMATION IN SURVEY SAMPLING

By

V. N. REDDY

INDIAN INSTITUTE OF MANAGEMENT, CALCUTTA

A thesis submitted to the Indian Statistical Institute
in partial fulfilment of the requirements
for the award of the degree of

DOCTOR OF PHILOSOPY

Calcutta
1975

ACKNOWLEDGEMENTS

It is a great pleasure to express my deep sense of gratitude to Professor T.J.Rao for his valuable guidance and constant encouragement throughout the period of this work. I record my sincere thanks to him for going through this thesis and offering many valuable suggestions for its improvement.

I am grateful to Dr. C.R.Rao and Dr. M.N.Murthy who initially set me to work in Sample surveys. I am also thankful to Dr. M.N.Murthy for the many discussions I had with him.

I am deeply indebted to Professor N. Krishnaji for drawing my attention initially to the problem of mis-stratification which has been fruitfully investigated in chapters VIII and IX of this thesis and for many valuable discussions.

Dr. A.R.Rao rendered valuable help in correcting a number of errors in the language and presentation of the thesis and offered many critical comments. To him are my heartfelt thanks.

I wish to record with thanks the help I received from a number of my colleagues, especially, Sri Chandan Mukherjee and Sri Rohitaswa Das who helped me in computations by providing a number of programmes.

Thanks are due to Sri Arun Das for his very patient and elegant typing of the thesis.

Finally, I wish to thank the duplicating section of our Institute for quick and efficient cyclostyling of the thesis.

V. N. Reddy

CONTENTS

CHAPTER I

INTRODUCTION

The main problem of Survey sampling from finite populations consists of devising an appropriate procedure for selecting a representative sample from a given population and developing an appropriate procedure for the estimation of the population parameter of interest such as population mean or total or population ratio with a view to maximising the precision of the estimator within the available resources of time and cost or alternatively minimising the cost for achieving a given level of precision. Towards the solution of this problem it is only during the thirties and forties that a more systematic development has taken place owing to the outstanding and pioneering contributions of Cochran, Hansen, Hurwitz, Madow, Mahalanobis, Neyman, Sukhatme, Yates and others. Theoretical developments as well as practical techniques were considered and large-scale agricultural and socio-economic surveys played a major rôle in this context (Mahalanobis (1944, 46) Sukhatme (1945, 46) and Yates (1947)).

During the next decade followed some significant developments in the theory of sampling from finite populations mainly concentrated on the utilisation of auxiliary information at different stages (stratification, sample selection, estimation, etc.) through the notable contributions of Cochran, Dalenius, Des Raj, Hansen, Hurwitz, Lahiri, Madow, Koop, Yates and others. These developments gave rise to a number of sampling techniques (systematic sampling, probability proportional to size (PPS) sampling, stratified sampling, etc.) and estimation procedures (ratio and regression methods of estimation) appropriate to various situations in practice to estimate the population mean or total of a variate and the errors of these estimates. The most important concept of 'cost function' was introduced by Mahalanobis (1940) to judge the efficiency of various estimators per unit cost.

A number of reviews have appeared on the developments in theoretical and applied research on sampling from finite populations; the notable ones in this direction are those by Yates (1946), Cochran (1947), Stephan (1948), Seng (1951), Sukhatme (1959, 66), Dalenius (1962), Murthy (1963), Vos (1974), among others.

With the advent of various selection procedures and the corresponding estimation procedures, the need was first felt by Horvitz and Thompson (1952) to evolve a systematic theory of sampling from finite populations and besides formulating the theory neatly, they defined three classes of estimators. Later in (1955), Godambe proposed a unified theory of sampling from finite populations with a view to discussing the fundamental problems of sampling within this frame work. Further, Godambe (1955) has obtained the celebrated result that for any sample design there does not exist a uniformly minimum variance unbiased estimator of the population total in the class of all linear unbiased estimators (with some exceptions characterised by Hanurav (1965) later, termed as 'unicluster designs'). This then, has led to the choice of estimators from a sub-class of admissible estimators. Alternatively, various criteria have been put forward by many others to arrive at an optimum choice, namely, (i) Bayesness (Godambe (1955), Hajek (1958)), (ii) invariance and regular class (Roy and Chakravorti (1960)) and (iii) hyper-admissibility (Hanurav (1966)), to quote a few.

Using Bayes approach in sampling theory, it was first shown by Cochran (1946) that whenever auxiliary information on a supplementary variate closely related to the study variate is available, we can utilize this information to formulate a prior distribution for the variate under study. This is now well known as the super-population concept. As stated earlier, with the criteria of unbiasedness and minimum variance, in general, there does not exist a best estimator in the class of linear estimators. However, this desperate situation prevails only when no a priori knowledge is available on the study variate. When information on a positive valued supplementary variate is available on all units of the population in advance, Godambe (1955) has shown that in the class of all sampling strategies with a given expected number of distinct units, any strategy (sampling design together with an estimator being termed a 'strategy' (Hajek (1958))) such that

(1) every sample has the same number of distinct units,

(2) probability of including any unit in the sample is proportional to the value of the supplementary variate for that unit, and

(3) the estimator is the Horvitz and Thompson estimator

is best in a well defined Bayesian sense. These optimum designs are termed by Hanurav (1965, 67) as πPS designs and the sampling schemes to construct such designs as πPS schemes. The problem of constructing πPS schemes has attracted the attention of many and notable contributions in this direction have been made by Goodman and Kish (1950), Midjuno (1951), Horvitz and Thompson (1952), Durbin (1953), Grundy (1954), Des Raj (1956), Hajek (1959), Hanurav (1962, 65, 67), Hartley and Rao (1962), Rao, Hartley and Cochran (1962), Brewer (1963), Fellegi (1963), Stuart (1964), Vijayan (1967), Rao, T.J. (1967), Sampford (1967), Ramakrishnan (1971), Dodds and Fryer (1971), Foreman and Brewer (1971), Das and Mohanthy (1973) and many others.

Murthy (1957) proved that any estimator which depends on the order of selection of units in the sample is inadmissible while Basu (1958) proved the inadmissibility of any estimator which depends on the repetition of units in the sample. Basu also introduced the concept of 'sufficiency' in sampling theory which was later developed by Pathak (1961, 64), Ramakrishnan (1971) and others.

It has been pointed out by several authors (e.g. Des Raj (1954), Rao, J.N.K. (1966)) that for estimating the population total, PPS sampling scheme is expected to be more efficient than SRS whenever the correlation coefficient between the study variate $\underset{=}{Y}$ and the supplementary variate $\underset{=}{X}$ is high (and positive) and the size measure $\underset{=}{X}$ is approximately proportional to $\underset{=}{Y}$ Further Rao, J.N.K. (1966) has suggested the use of alternative estimators in PPS sampling if $\underset{=}{Y}$ and $\underset{=}{X}$ are poorly correlated. A modified PPS method of estimation obtained by considering a suitably transformed supplementary variate is introduced in this thesis for estimating the population total when the correlation between $\underset{=}{Y}$ and $\underset{=}{X}$ is high (<u>positive or negative</u>).

Another side of development, as pointed out earlier, was with regard to the use of information on a supplementary variate closely related to the variate under study at the stratification, sample selection and estimation stages. Cochran (1942) developed the ratio method of estimation using information on a single supplementary variate which is positively correlated with the variate under consideration and this was later extended to an estimator using information on two or more such supplementary variates by Olkin (1958). A product method of estimation, complementary to the ratio method was considered by Murthy (1964) who has suggested its use when the supplementary variate has high negative correlation with the variate under study and the extension of this product method of estimation using information on several supplementary variates was considered by Singh (1967). Further contributions in this direction are by Koop (1951, 64), Des Raj (1954, 63, 65), Robson (1957), Goodman (1960), Robson and Vithyasai (1961), Srivastava (1965, 67, 71), Rao, P.S.R.S. and Mudholkar (1967), Rao, P.S.R.S. (1969), Singh (1970), Walsh (1970), Rao, P.S.R.S. and Rao, J.N.K. (1971), Sukhatme and David (1973) and many others.

Since the ratio and product methods of estimation lead to biased estimators, attempts were made to make these estimators unbiased or 'almost unbiased'. In some cases this was achieved as in Lahiri (1951), Midjuno (1951), Sen (1952), Murthy, Nanjamma and Sethi (1959) and others and in some others by adjusting for the corresponding bias as in Hartley and Ross (1954), Quenouille (1956), Durbin (1959), Murthy and Nanjamma (1959), Mickey (1959), Pascual (1961), Rao, J.N.K. (1964), Rao, T.J. (1966), Singh(1970), Sastry (1974) and others.

Using the information on multi-auxiliary variates and certain population parameters, Srivastava (1971) has obtained a class of generalised multivariate ratio estimators whose efficiency to the second degree of approximation is the same as that of the multivariate regression estimator. In this thesis we consider this problem in a more general set up and suggest a 'transformed ratio estimator' which is almost unbiased and which generalises both ratio and product methods of estimation.

Besides the utilisation of supplementary information on a variate which is available for all the units in the population, the use of X (the population total of the supplementary variate) and β the regression coefficient of $\underline{\underline{Y}}$ on $\underline{X}$ in the estimation procedure is also well known in the literature. Searls (1965, 67) has suggested the use of C_o, the population coefficient of variation of the variate under study, which is assumed to be known, in the estimation procedure to obtain improved estimators. Des Raj (1965) has suggested the use of <u>a priori</u> knowledge on R (the population ratio) in the difference method of estimation to obtain results comparable with those obtained for the ratio method of estimation. The use of <u>a priori</u> knowledge on $k = \frac{\beta}{R}$ in the ratio method of estimation to obtain optimum estimators has been suggested by Srivastava (1967, 71) and the author (1973, 74a), while Singh, M.P. and Roy (1968) suggested the use of a priori knowledge on the population total in the estimation procedure.

In large-scale sample surveys it is quite common to use the values of certain population parameters based on a previous survey for estimating the current values of these parameters. This necessitates a study on the most important property namely the stability over time and space of these parameters. This desirable property of the stability over time and space for the population parameters R, β, k and C_0 is investigated in this thesis and a comparison of the various estimators using information on these parameters is also made.

It is well known that the technique of stratified sampling consists in classifying the population into a certain number of homogeneous groups called strata and then selecting samples independently from each group or stratum. Broadly, the important points that need careful consideration in the stratified sampling are :

(1) choice of sampling design within strata,

(2) choice of stratification variable,

(3) allocation of sample size to strata,

(4) number of strata, and

(5) demarcation of strata.

Eventhough the earlier developments in stratified sampling (Bowley (1926), Neyman (1934), Mahalanobis (1944), Hayashi and Maruyama (1948), and others) were mainly confined to questions relating to (3), and some attention was given to points (1),(2) and (4), it is only since the fifties that the problems (4) and (5) have been studied objectively by, among others, Dalenius (1950, 52), Dalenius and Gurney (1951), Hayashi, Maruyama and Isida (1951), Dalenius and Hodges (1959), Sethi (1963), Taga (1967), Isii and Taga (1969), Singh and Sukhatme (1972) and the author (1975). Mahalanobis (1952), Dalenius and Hodges(1957), Durbin (1959) and Ekman (1959) have suggested approximations to the theoretical solutions which are easier to apply in practice.

As a first step in finding out the optimum points of demarcation of strata in stratified random sampling, it is freely assumed in the literature that the $\underline{\underline{X}}$ - variate which is presumably having high correlation with the study variate $\underline{\underline{Y}}$, should be arranged in increasing (or decreasing) order of magnitude, but we are not aware of any formal proof or justification for doing the same. This problem has also been investigated in this thesis.

Besides the uses of information on a supplementary variate as mentioned earlier, it is also a common practice of the survey statisticians to make use of this at both the design and estimation stages. Examples of such uses are stratification of the population and selecting units within each stratum by PPS scheme or adopting estimation procedures like ratio, product and regression methods in each stratum. This type of problem is also considered in this thesis.

The author's contributions

We shall now give a brief summary of the results presented in this thesis. This thesis is divided into ten chapters. After the first introductory chapter, we explain in chapter II the basic concepts and definitions which will be used in this thesis.

In chapter III we study the ratio and product methods of estimation for estimating the population mean or total, the simplest methods of utilisation of auxiliary information. For any sampling design, when information on the population parameter $k = \frac{\beta}{R}$ is available, we derive two optimum ratio type estimators whose efficiency to the second degree of approximation is the same as that of the regression estimator. Some interesting relations among these estimators have been examined. Also, a better alternative to the usual product estimator is obtained.

One of the optimum estimators discussed in chapter III which is almost unbiased has been studied further in chapter IV.

It is shown that this optimum estimator can be obtained by a suitable transformation on the supplementary variate and we term this estimator as the 'optimum transformed ratio estimator'. Since the exact information on k is not usually available, the effect of the deviations of the guessed value from the true value on bias and mean square error (m.s.e.) is studied. It is found that for a fairly wide range of guessed values of k, the near optimum transformed ratio estimator has uniformly smaller absolute bias and m.s.e. than those of the usual ratio or product estimator. This optimum estimator has also been examined under a super-population model and it is shown that a ratio type estimator with a suitable transformation on the supplementary variate will have a smaller absolute bias and m.s.e. under normally valid conditions. A multivariate extension of the optimum estimator is also suggested.

Different estimationprocedures using a priori knowledge on certain population parameters are dealt with in chapter V. It is found that transformed ratio method of estimation which assumes knowledge on k fares better than all other estimators based on information on the corresponding population parameters. The desirable property of the stability (cf. p.80) of the parameters has been investigated empirically and it is found that k is remarkably more stable over time and space than other population parameters. The relative advantages and disadvantages

INDIAN STATISTICAL INSTITUTE LIBRARY
No. T 088
26 NOV 1984

of these methods are discussed and the results are illustrated by several empirical examples based on data on productivity in Indian Agriculture.

In chapter VI, we study the efficiency of the conventional ratio estimator under a super-population model. The ratio estimator is found to be highly superior to the simple unbiased estimator under this super-population set up in almost all the situations met with in practice. The separate ratio estimator in stratified random sampling is found to have smaller absolute bias and m.s.e. than those of the combined ratio estimator under the same super-population model in most of the situations met with in practice.

A modified probability proportional to size (PPS) sampling method of estimation obtained by considering a suitable transformed supplementary variate is introduced in chapter VII. The modified PPS estimator is found to be highly superior to the conventional PPS estimator in most of the situations met with in practice. For the case when $\underline{\underline{Y}}$ and $\underline{\underline{X}}$ are negatively correlated a corresponding modified PPS method of estimation is also introduced. The empirical efficiency of the suggested modified PPS estimator with respect to Horvitz and Thompson estimator and symmetrized Des Raj estimator is also examined for both the cases. Finally, a comparison between the transformed ratio estimator and Rao, Hartley and Cochran (RHC) estimator is also made in this chapter under a linear finite population model.

Chapters VIII and IX deal with the problem of optimum demarcation of a finite population into a specified number of strata. It is shown that in the case of SRS (with or without replacement) scheme in each stratum, while it is necessary to arrange the $\underline{\underline{Y}}$-character in increasing (or decreasing) order of magnitude for optimum stratification with proportional allocation, the same need not be true with optimum allocation. However, if the coefficient of variation of the $\underline{\underline{Y}}$-variate is the same in each stratum, the necessity of arranging the $\underline{\underline{Y}}$-character in increasing (or decreasing) order of magnitude for optimum stratification with optimum allocation is established. Optimum points of demarcation of strata in the case of proportional allocation have been examined. Similar results have been obtained for optimum demarcation of strata using the supplementary variate under an appropriate super-population set up.

It is shown in several text books on sampling (for example Cochran (1963)) that stratified random sampling with proportional allocation is superior to unstratified random sampling provided the finite population correction factor (f.p.c.) in each stratum is ignored. It is shown in chapter X that the same result is true even without ignoring the f.p.c. if the stratification is such that the $\underline{\underline{Y}}$ variate is arranged in increasing (or decreasing) order of magnitude. It is also shown here that the same result need not be true in case of stratified ratio, product or regression estimators. In fact, it is noted here that combined ratio or product estimator in stratified random sampling with proportional allocation is inferior to the corresponding estimator for unstratified random sampling provided the strata ratios (R_1's) are the same. A simple condition is also obtained to show that the combined regression estimator in stratified random sampling is inferior to the corresponding estimator for unstratified random sampling.

Some of the results obtained in chapters III, IV, VI and IX of this thesis are already published and the results of chapter X have been presented in the International Symposium on recent trends of research in statistics held in ISI, Calcutta in December 1974.

CHAPTER II

CONCEPTS AND DEFINITIONS

In this chapter we explain the basic concepts and definitions which will be used in this thesis.

A 'finite population' $\underset{=}{U}$ is defined as a collection $U_1, U_2, \ldots, U_N$ where N is a known finite number and $U_1, U_2, \ldots, U_N$ are distinguishable. We denote this population by

$$\underset{=}{U} : (U_1, U_2, \ldots, U_N) \qquad (2.1)$$

A list such as (2.1) is called a 'sampling frame' and N is called the 'size of the population'.

Any finite ordered sequence of units from $\underset{=}{U}$ is called a 'sample' and is denoted by

$$s = (u_1, u_2, \ldots, u_n) \qquad (2.2)$$

where each u_i belongs to $\underset{=}{U}$. Alternatively (2.2) is written as

$$s = (U_{i_1}, U_{i_2}, \ldots, U_{i_n}) \qquad (2.3)$$

where $1 \leq i_r \leq N$ for $1 \leq r < n$.

We call n the 'sample size' when n is finite and the number of distinct units in the sample is termed as the 'effective sample size'. Let

$$\underset{=}{S} = \{ s \} \qquad (2.4)$$

denote the totality of samples from $\underline{\underline{U}}$ and $\underline{\underline{P}}$ be a probability measure on $\underline{\underline{S}}$, that is

$$p_s \geq 0, \text{ for all } s \in \underline{\underline{S}} \text{ and } \sum_{s \in \underline{\underline{S}}} p_s = 1, \qquad (2.5)$$

where p_s is the probability of selecting sample s . The pair $(\underline{\underline{S}}, \underline{\underline{P}})$ is called 'sample (sampling) design' and is denoted by

$$D = D(\underline{\underline{S}}, \underline{\underline{P}}) = (\underline{\underline{S}}, \underline{\underline{P}}). \qquad (2.6)$$

Thus the definition of the design gives us a method of selecting a sample which requires the listing down of all the possible samples and choosing one from the list with the corresponding probability. But, in practice, especially in large-scale surveys, it is very difficult to list down all possible samples and follow this procedure. Alternatively, Hanurav (1965) defined a 'sampling mechanism' of drawing units from $\underline{\underline{U}}$ one-by-one with probabilities which depend on the previous draws. A 'sampling mechanism' (or a 'drawing mechanism') is a function

$$q\ (u,\ r,\ s_{r-1}) \qquad (2.7)$$

where $u \in \underline{\underline{U}}$, r is a positive integer and s_{r-1} is a sample of size (r-1) such that

$$q(u,r,s_{r-1}) \geq 0 \text{ for all } u, r \text{ and } s_{r-1}$$

and

$$\sum_{u \in \underline{\underline{U}}} q(u,r,s_{r-1}) = 1 \text{ for all } r \text{ and } s_{r-1}. \qquad (2.8)$$

In this connection the following result is obtained by Hanurav.

Theorem 2.1 (Hanurav 1962a) : There exists one to one correspondance between sampling designs and sampling mechanisms.

Corresponding to any given design $D(\underset{=}{S}, \underset{=}{P})$ the 'inclusion probability' of a unit U_i is defined as

$$\pi_i = \sum_{s \ni i} p_s \qquad (2.9)$$

where the summation is taken over all the samples that contain U_i at least once. The 'joint inclusion probability' of a pair (U_i, U_j), $i \neq j$, is defined as

$$\pi_{ij} = \sum_{s \ni (i,j)} p_s , \qquad (2.10)$$

the summation being taken over all samples which contain both U_i and U_j.

Considering the problem of estimation in finite populations, let $\underset{=}{Y}$ be a real-valued characteristic taking value Y_i on the unit U_i of (2.1), $i = 1, 2, \ldots, N$. Let us denote $\underset{\sim}{Y} = (Y_1, Y_2, \ldots, Y_N)$. Any function of $\underset{\sim}{Y}$ is called a 'parametric function'. Any function t defined over a design $D(\underset{=}{S},\underset{=}{P})$ such that for samples $s \varepsilon D$ the function t depends only on the values of $\underset{=}{Y}$ for the units belonging to the sample is called a 'statistic'. A statistic t when used to estimate a parametric function $T(\underset{\sim}{Y})$ is called an 'estimator' of T. An estimator t is called an 'unbiased estimator' of T iff

$$E(t) = \sum_{s \in \underline{\underline{S}}} t_s \, p_s = T(\underset{\sim}{Y}) \text{ for all values of } \underset{\sim}{Y}. \qquad (2.11)$$

An estimator t which is not unbiased for T is said to be 'biased'.

In the estimation of T, the deviation $(t_s^{*} - T)$ is taken as the 'error' on the basis of sample s. Any convex function $f(t_s - T)$ is taken to be the 'loss function' and $E(f)$ is called the 'expected loss'. A loss function which is used quite often is the mean square error' (m.s.e.) given by

$$M(t) = E(t - T)^2 = \sum_{s \in \underline{\underline{S}}} (t_s - T)^2 \, p_s \qquad (2.12)$$

where t_s is used as an estimator of T based on the sample s and p_s is the probability of selecting the sample s. If t is an unbiased estimator of T then $E(t - T)^2$ is called the 'variance' of t which may also be written as

$$V(t) = \sum_{s \in \underline{\underline{S}}} t_s^2 \, p_s - T^2. \qquad (2.13)$$

A design $D(\underline{\underline{S}}, \underline{\underline{P}})$ together with an estimator t of T defined over D is called a 'sampling strategy' for the estimation of T and is denoted by

$$H = H(D, t) = H(\underline{\underline{S}}, \underline{\underline{P}}, t). \qquad (2.14)$$

This definition is due to Hajek (1958). A strategy H when used for the estimation of T is called an 'unbiased strategy' f, t is an unbiased estimator of T. Otherwise, it is called 'biased strategy'. The expectation and the mean square error variance) of a strategy H are defined as the expectation and he mean square error (variance) respectively of the estimator over D.

Of the two estimators t_1 and t_2 of a parametric function $T(\underset{\sim}{Y})$, both defined over a design $D(\underset{=}{S}, \underset{=}{P})$, t_1 is said o be 'uniformly better' than t_2 if and only if

$$M(t_1) \leq M(t_2) \text{ for all } \underset{\sim}{Y} \tag{2.15}$$

ith a strict inequality at least for one $\underset{\sim}{Y}$. Also in a class of estimators of T defined over a given design $D(\underset{=}{S}, \underset{=}{P})$, an stimator $t_1 \in \underset{=}{C}$ is the 'best estimator' if and only if t_1 is niformly better than t_2 for all t_2 different from t_1 and elonging to $\underset{=}{C}$.

It is well known that in the case of infinite populations the sample mean $\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$ is the minimum variance unbiased estimator of μ in the class of all linear unbiased estimators $\sum_{i=1} w_i y_i$, w_i's being constants independent of the observations y_i and $\sum_{i=1}^{n} w_i = 1$) where $E(y_i) = \mu$ for all i. The 'distinguishability of units' which plays an important role in the theory of finite populations is the main difference between the theory f finite and infinite populations. Thus, it is important to know whether two equal values of $\underline{\underline{Y}}$ in the sample belong to the same unit repeated, or to two different units. This fact was first observed by Des Raj and Khamis (1958) and Basu (1958) who found that in simple random sampling with replacement (SRSWR) of sample size n, the mean of the 'effective sample', i.e., the average of the $\underline{\underline{Y}}$ -values corresponding to the distinct units in the sample is better than the usual sample mean $\bar{y}$.

This necessitated a general definition of linear estimator and Horvitz and Thompson (1952) defined the following three classes of estimators :

$$t_1 = \sum_{i=1}^{n} \beta_i \, y_{(i)},$$

where $y_{(i)}$ is the value of the $\underline{\underline{Y}}$ -variate for the unit selected at the i-th draw and β_i is the corresponding weight ;

$$t_2 = \sum_{i=1}^{n} w_i \, y_i,$$

where w_i is the weight attached to the i-th unit whenever it is selected in the sample and

$$t_3 = \surd_s \sum_{i=1}^{n} y_i,$$

where the estimator has a single coefficient $\surd_s$ for all the sample values of $\underline{\underline{Y}}$.

The ratio estimator which we study in some detail in this thesis falls in the t_3 class.

Godame (1955) generalised the above classes of estimators by considering

$$t_s = \sum_{i \epsilon s} \beta_{st} \, y_i \qquad (2.16)$$

and proved the following theorem.

Theorem 2.2 (Godambe (1955)) : There does not exist a uniformly minimum variance unbiased estimator for any sampling design D.

Later, Hanurav (1966) pointed out some exceptions to this result by constructing some non-trivial designs (called 'uni-cluster designs') where a best estimator exists. This leads to the search for optimum estimators in a sub-class of designs. Even though the criterion 'admissibility' helps in eliminating bad (inadmissible) estimators, it does not help much in obtaining optimum estimators. We say that an estimator t_1 of T is 'admissible' if and only if there does not exist another estimator $t_2 \neq t_1$ of T which is better than t_1.

An estimator is said to be 'inadmissible' if it is not admissible. The results of Murthy (1957), Des Raj and Khamis (1958) and Basu(1958) showed that estimators which depend on the order in which the units appear or which have a unit in the sample repeated are inadmissible. Godambe and Joshi (1965) and Joshi (1965a, 1965b) considered admissibility removing the restriction of linearity and later relaxing the criterion of unbiasedness. Three other criteria, namely 'linear invariance' and 'regular estimators' by Roy and Chakravorti (1960) and 'hyper-admissibility' due to Hanurav (1966) are defined as follows :

An estimator t of T is called 'regular estimator' if

$$V(t) = K \sigma^2, \tag{2.17}$$

where K is a constant and $\sigma^2 = \frac{1}{N} (\sum_{i=1}^{N} Y_i^2 - N \bar{Y}^2)$.

Roy and Chakravorti have shown that a best estimator exists in the class of regular estimators.

An estimator t of T is said to be possessing the property of 'linear invariance' if it is invariant under linear transformations of $\underset{\sim}{Y}$.

In a class $\underline{\underline{C}}$ of unbiased estimators of $T(\underset{\sim}{Y})$, an estimator $t_1 \epsilon \underline{\underline{C}}$ is said to be 'hyperadmissible' if given any other estimator $t_2 \epsilon \underline{\underline{C}}$ in every hyper plane of R^N, there exists at least one point $\underset{\sim}{Y}$ at which

$$V(t_1) < V(t_2). \tag{2.18}$$

It must now be stressed that a search for optimum should be considered amongst strategies which are equally costly. We consider a simple linear 'cost function'

$$c(s) = a_o + b_o \; n(s) \tag{2.19}$$

where a_o is the overhead cost, b_o is the cost for collecting data on one sample unit and $n(s)$ is the effective size of sample s. The expected cost of a strategy $H(\underset{=}{S}, \underset{=}{P}, t)$ or equivalently of a design $D(\underset{=}{S}, \underset{=}{P})$ is defined as

$$C(H) = C(D) = \sum_{s\epsilon \underset{=}{S}} C(s)\, p_s = a_0 + b_0\, \mu(d) \qquad (2.20)$$

where $\mu(d) = \sum_{s\epsilon \underset{=}{S}} n(s)\, p_s$, the expected effective sample size of D. Hence, under this cost set up, two designs (or strategies) are equally costly iff they have the same expected effective sample size.

If optimality is judged from uniform minimization of the variance of a strategy, then we can see that there does not exist such a one. But, whenever some auxiliary information on a characteristic $\underset{=}{X}$ which takes value X_i on unit U_i, $i = 1, 2, \ldots, N$ is available closely related to the study variate $\underset{=}{Y}$, taking value Y_i on U_i, $i = 1, \ldots, N$, it is possible to use this information in setting up a criterion of optimality. We have already mentioned this in the introductory chapter and here we shall explain it further.

The information on $\underset{=}{X}$, known before hand can be used to assume a reasonable a priori distribution over $\underset{\sim}{Y}$. According to this 'super-population concept' as termed by Cochran (1946), $\underset{\sim}{Y} = (Y_1, Y_2, \ldots, Y_N)$ is assumed to be a realisation of a random N-vector with certain distribution depending upon

$\underset{\sim}{X} = (X_1, X_2, \ldots, X_N)$ and some unknown parameters. This distribution is denoted by ξ and we can talk of expectations, variances and covariances taken over ξ. We now minimise the expected variance over ξ namely

$$\int V(H)\, d\xi \tag{2.21}$$

for H varying over $\underline{\underline{C}}(H)$, the class of all equi-cost strategies. A H_0 which minimises (2.21) uniformly with respect to all the parameters of the distribution ξ is called 'ξ-optimum strategy' in $\underline{\underline{C}}(H)$.

Let Δ_g be the class of all prior distributions ξ satisfying

$$Y_i = b\, X_i + \varepsilon_i, \quad i = 1, 2, \ldots, N$$

with $$\mathcal{E}(\varepsilon_i | X_i) = 0$$

$$\mathcal{V}(\varepsilon_i | X_i) = \sigma^2 X_i^g \tag{2.22}$$

and $$\mathcal{C}(\varepsilon_i, \varepsilon_j \mid X_i, X_j) = 0, \quad \text{for } i \neq j$$

where $\mathcal{E}$, $\mathcal{V}$ and $\mathcal{C}$ denote respectively the expectation, variance and covariance ; b, σ^2 and g are unknown constants. This Δ_g class of prior distributions has been widely used in survey sampling. In Δ_g-class we have taken $\mathcal{E}(Y_i | X_i) = b\, X_i$, but there are many situations where this is not so (eg. Dodds and Fryer (1971)); in fact, it may be more reasonable to assume that

$$Y_i = a + b X_i + \varepsilon_i, \quad i = 1,2,\ldots,N$$

with $$\mathcal{E}(\varepsilon_i|X_i) = 0$$

$$\mathcal{V}(\varepsilon_i|X_i) = \sigma^2 X_i^g \qquad (2.23)$$

and $$\mathcal{C}(\varepsilon_i,\varepsilon_j|X_i,X_j) = 0 \quad \text{for } i \neq j,$$

and denote the corresponding class of prior distributions by Λ_g'. Since this seems to be a more reasonable set up as far as the expectation of ε_i is concerned, we study some of the problems in this thesis under Λ_g'.

It is a common practice for survey statisticians to make use of auxiliary information at the design stage or estimation stage or both. Examples of such uses are stratification of the population, selection of the sample with probability proportional to size (PPS) scheme and estimation procedures like ratio, product and regression methods. Let $\hat{Y}$ and $\hat{X}$ be unbiased estimators of the population totals Y and X of the study variate $\underline{Y}$ and auxiliary variate $\underline{X}$ respectively, based on any sample design. Then, the usual ratio, product, optimum difference and regression estimators of the population total Y are respectively given by,

$$\hat{Y}_R = \frac{\hat{Y}X}{\hat{X}},\ \hat{Y}_p = \frac{\hat{Y}\hat{X}}{X},\ \hat{Y}_r' = \hat{Y} + \beta(X - \hat{X}) \text{ and } \hat{Y}_r = \hat{Y} + \hat{\beta}(X-\hat{X})$$

where $\beta = \frac{Cov(\hat{Y}, \hat{X})}{V(\hat{X})}$ and $\hat{\beta}$ is an estimator of β. It is easy to see that $B(\hat{Y}_R) = - Cov(\frac{\hat{Y}}{\hat{X}}, \hat{X})$, $B(\hat{Y}_P) = \frac{Cov(\hat{Y}, \hat{X})}{X}$,

$$B(\hat{Y}_r') = 0 \text{ and } B(\hat{Y}_r) = - Cov(\hat{\beta}, \hat{X}),$$

and

$$V(\hat{Y}_r') = V(\hat{Y})\ (1 - \rho^2),$$

where ρ is the correlation coefficient between $\hat{Y}$ and $\hat{X}$. Usually the approximate bias and m.s.e. of $\hat{Y}_R$ are evaluated by assuming $|\frac{\hat{X} - X}{X}| < 1$; that is $\hat{X}$ lies between 0 and 2X which might be valid either if the sample size is large or the variation in $\underline{\underline{X}}$ character is small. Writing $\hat{Y} = Y(1 + e)$, $\hat{X} = X(1 + e_1)$, where $E(e) = E(e_1) = 0$ and assuming $|e_1| < 1$ and neglecting the terms with powers of e, e_1 greater than 2, (i.e., up to the second degree of approximation) we have

$$B(\hat{Y}_R) = Y(1 - k)C_x^2$$

and

$$M(\hat{Y}_R) = Y^2[C_y^2\ (1 - \rho^2) + (1 - k)^2\ C_x^2]$$

where C_y and C_x are the coefficients of variation of Y and $\hat{X}$ respectively and $k = \frac{\beta}{R}$. Similarly, up to the second degree of approximation (without assuming $|e_1| < 1$)

$$M(\hat{Y}_P) = Y^2\ [C_y^2\ (1 - \rho^2) + (1 + k)^2\ C_x^2]$$

and

$$M(\hat{Y}_r) = Y^2\ C_y^2\ (1 - \rho^2).$$

Since, usually $\hat{Y}_R$ and $\hat{Y}_P$ are biased, several attempts have been made by many to make these estimators either unbiased or 'almost unbiased' (the bias is zero to the second degree of approximation) by suitably selecting the sample or by correcting their bias. It is only recently that work has begun to reduce simultaneously the bias and m.s.e. of $\hat{Y}_R$ and $\hat{Y}_P$ by assuming knowledge on certain population parameters (Srivastava (1967, 1971), Singh (1970), the author (1973, 1974a) and others). We study this problem in some detail in this thesis.

The probability proportional to size (PPS) sampling scheme which is one of the simplest methods of utilization of auxiliary information consists of selecting units in the sample with probability proportional to a given measure of size, where the size measure is usually the value of an auxiliary variate $\underset{=}{X}$ which is highly related with the variate $\underset{=}{Y}$ under study. Suppose a sample $(u_1, u_2, \ldots, u_n)$ of size n is selected using PPS with replacement sampling scheme. Then an unbiased estimator of the population total $Y = \sum_{i=1}^{N} Y_i$ of the study variate $\underset{=}{Y}$ and its variance are given by

$$\hat{Y}_{PPS} = X \sum_{i=1}^{n} \frac{y_i}{x_i} \quad \text{and} \quad V(\hat{Y}_{PPS}) = \frac{1}{n} \left[\sum_{i=1}^{N} \frac{Y_i^2}{X_i} X - Y^2 \right].$$

It is clear from the above that $V(\hat{Y}_{PPS})$ will reduce to zero when Y_i and X_i are exactly proportional. On the other hand $\hat{Y}_{PPS}$ can be inferior to $\hat{Y}$ either if $\underline{\underline{Y}}$ and $\underline{\underline{X}}$ are linearly related (Des Raj (1954))or are poorly correlated (Rao, J.N.K. (1966)). We examine this problem in this thesis.

The technique of stratified random sampling consists in dividing the population into groups called strata and then selecting the samples independently from each stratum. Suppose a population of N units is divided into h strata. Let N_i be the number of units in i-th stratum, $i = 1,2, \ldots, h$ and let Y_{ij} be the value of the study variate $\underline{\underline{Y}}$ for j-th unit in the i-th stratum, $j = 1, 2, \ldots, N_i$. The population total

$$Y = \sum_{i=1}^{h} \sum_{j=1}^{N_i} Y_{ij} = \sum_{i=1}^{h} Y_i$$

where $Y_i = \sum_{j=1}^{N_i} Y_{ij}$. Denote $\frac{Y_i}{N_i}$ by $\bar{Y}_i$. An unbiased estimator of Y can be obtained by estimating unbiasedly the stratum totals $\{Y_i\}$ on the basis of random samples drawn independently from each stratum. Suppose $\hat{Y}_i$ is an unbiased estimator of Y_i. Then an unbiased estimator of Y and its variance are given by

$$\hat{Y}_{st} = \sum_{i=1}^{h} \hat{Y}_i \quad \text{and} \quad V(\hat{Y}_{st}) = \sum_{i=1}^{h} V(\hat{Y}_i) \text{ respectively.}$$

Suppose a sample of size n_i is drawn using SRSWR scheme in each stratum so that $\sum_{i=1}^{h} n_i = n$. In this case

$$\hat{Y}_{st} = \sum_{i=1}^{h} N_i \bar{y}_i \quad \text{and} \quad V(\hat{Y}_{st}) = \sum_{i=1}^{h} \frac{N_i^2 \sigma_i^2}{n_i} ,$$

where $\bar{y}_i$ is the average of sample $\underline{Y}$-values of i-th stratum and $\sigma_i^2 = \frac{1}{N_i} \sum_{j=1}^{N_i} (Y_{ij} - \bar{Y}_i)^2$. Usually, the cost function in stratified random sampling is taken as

$$C = a_o + \sum_{i=1}^{h} b_i n_i,$$

where a_o is the overhead cost and b_i is the average cost of surveying one unit in the i-th stratum. Minimising $V(\hat{Y}_{st})$ for a fixed cost $C = C'$ we obtain

$$n_{i,opt.} = \frac{(C'-a_o) N_i(\sigma_i / \sqrt{b_i})}{\sum_{i=1}^{h} N_i(\sigma_i / \sqrt{b_i})} .$$

In case b_i are same for each stratum the above optimum n_i reduces to

$$n_i,opt. = \frac{n N_i \sigma_i}{\sum_{i=1}^{h} N_i \sigma_i}$$

and this is the usual optimum allocation that is considered in the literature. When σ_i's are equal the above optimum allocation reduces to

$$n_i = \frac{n N_i}{N}$$

and this allocation is called proportional allocation. In the case of proportional and optimum allocations the corresponding variances of $\hat{Y}_{st}$ are respectively

$$V_{prop.}(\hat{Y}_{st}) = \frac{N}{n} \sum_{i=1}^{h} N_i \sigma_i^2$$

and $$V_{opt.}(\hat{Y}_{st}) = \frac{1}{n} (\sum_{i=1}^{h} N_i \sigma_i)^2.$$

Thus when simple random sampling is used in each stratum, if the number of strata are pre-specified say 'h', then the problem of optimum demarcation of strata in the case of proportional and optimum allocations reduces to finding h-1 points of demarcation which minimise

$$\sum_{i=1}^{h} N_i \sigma_i^2 \quad \text{and} \quad \sum_{i=1}^{h} N_i \sigma_i \quad \text{respectively.}$$

We investigate this problem towards the end of this thesis.

CHAPTER III

OPTIMUM RATIO AND PRODUCT ESTIMATORS

3.0 Summary

After a brief introduction to the usual ratio and product methods of estimation, we study the two estimators

$$\hat{Y}^{(a)} = \hat{Y}\left(\frac{X}{\hat{X}}\right)^{a}$$

and $$\hat{Y}_{\theta} = \frac{\hat{Y}X}{X + \theta(\hat{X} - X)}$$

to improve upon the conventional ratio and product estimators Y_R and $\hat{Y}_P$ respectively, where $\hat{Y}$ and $\hat{X}$ are unbiased estimators of Y and X, the population totals of the $\underline{\underline{Y}}$ and $\underline{\underline{X}}$ characteristics ; and a and θ are given scalars. For any sampling design, the optimum estimators are obtained by minimising the m.s.e. of $\hat{Y}^{(a)}$ and $\hat{Y}_{\theta}$ with respect to a and θ respectively. Some interesting relations among these estimators have been obtained. Also it is found that

$$\frac{\hat{Y}X}{2X - \hat{X}}$$

is a better alternative to the usual product estimator $\hat{Y}_p$. Comparison of the bias and m.s.e., up to the fourth degree of approximation, of different estimators has been made assuming the population to be large and $(\underline{\underline{Y}}, \underline{\underline{X}})$ follows a bivariate

Normal distribution. The results are also illustrated by empirical examples.

3.1 Introduction

Consider a finite population $\underset{=}{U} = (U_1, U_2, \ldots, U_N)$ of size N. Let $\underset{=}{Y}$ be the study variate taking the value Y_i on the unit U_i, $i = 1, 2, \ldots, N$; we are interested in estimating parametric functions of $\underset{\sim}{Y} = (Y_1, Y_2, \ldots, Y_N)$.

Whenever information is available on a suitable supplementary variate $\underset{=}{X}$ highly correlated with the study variate $\underset{=}{Y}$ and taking the value X_i on the unit U_i, $i = 1, 2, \ldots, N$, it is possible to improve upon the usual unbiased estimator of the population total $Y = \sum_{i=1}^{N} Y_i$ by the simple ratio method of estimation (or product method of estimation) provided the correlation coefficient ρ, between the study variate under consideration and the auxiliary variate has high positive value (or high negative value, respectively). In fact, for the better utilization of a given supplementary variate, Murthy (1964) has suggested the use of

$$\text{ratio estimator} \quad \hat{Y}_R = \frac{\hat{Y} X}{\hat{X}} \;, \quad \text{if} \quad k > \frac{1}{2} \;,$$

$$\text{product estimator} \quad \hat{Y}_P = \frac{\hat{Y} \hat{X}}{X} \quad \text{if} \quad k < -\frac{1}{2} \quad \text{and}$$

$$\text{usual unbiased estimator} \quad \hat{Y} \quad \text{if} \quad -\frac{1}{2} \leq k \leq \frac{1}{2} \;,$$

whenever the ratio $R = \frac{Y}{X}$ is positive, where $X = \sum_{i=1}^{N} X_i$, $\hat{Y}$ and $\hat{X}$ are unbiased estimators of Y and X respectively, $k = \rho \frac{C_y}{C_x}$, C_y and C_x are coefficients of variation of the estimators $\hat{Y}$ and $\hat{X}$ respectively, and ρ is the coefficient of correlation between $\hat{Y}$ and $\hat{X}$.

3.2 Optimum methods of estimation

Let $\hat{Y}$ and $\hat{X}$ be unbiased estimators of Y and X respectively based on any sample design. We then have the usual ratio estimator of the population total given by

$$\hat{Y}_R = \frac{\hat{Y} X}{\hat{X}} \tag{3.2.1}$$

and the usual product estimator given by

$$\hat{Y}_P = \frac{\hat{Y} \hat{X}}{X} \tag{3.2.2}$$

where X is assumed to be known. Here, we also assume that R is different from zero. We now consider the following alternative estimators of Y namely

$$\hat{Y}^{(a)} = \hat{Y} \left(\frac{X}{\hat{X}} \right)^a \tag{3.2.3}$$

and

$$\hat{Y}_\theta = \frac{\hat{Y} X}{X + \theta(\hat{X} - X)} \tag{3.2.4}$$

where a and θ are any scalars. Following Murthy (1967, pp. 363-365) we write $\hat{Y} = Y(1 + e)$ and $\hat{X} = X(1 + e_1)$ where $E(e) = E(e_1) = 0$. Assuming $|e_1| < 1$ and also $|\theta\ e_1| < 1$, we have

$$\hat{Y}^{(a)} = Y(1 + e)\ (1 + e_1)^{-a}$$
$$= Y(1+e)[1-ae_1+\frac{a(a+1)e_1^2}{2!} - \frac{a(a+1)(a+2)e_1^3}{3\ !} + \frac{a(a+1)(a+2)(a+3)e_1^4+..}{4\ !} \tag{3.2.5}$$

and
$$\hat{Y}_\theta = Y(1+e)\ (1 + \theta\ e_1)^{-1}$$
$$= Y(1+e)[1-\theta\ e_1+\theta^2e_1^2-\theta^3e_1^3+\theta^4e_1^4 + \ldots.]. \tag{3.2.6}$$

Hence, to the second degree of approximation, we obtain

$$B(\hat{Y}^{(a)}) = \frac{a}{2X}\ [(a+1)\ R\ V(\hat{X}) - 2\ Cov(\hat{Y},\ \hat{X})], \tag{3.2.7}$$

$$B(\hat{Y}_\theta) = \frac{\theta}{X}\ [\theta\ R\ V(\hat{X}) - Cov(\hat{Y},\ \hat{X})], \tag{3.2.8}$$

$$M(\hat{Y}^{(a)}) = V(\ \hat{Y}\ ) - 2a\ R\ Cov(\hat{Y},\ \hat{X}) + a^2R^2\ V(\ \hat{X}\ ) \tag{3.2.9}$$

and
$$M(\hat{Y}_\theta) = V(\hat{Y}) - 2\theta R\ Cov(\hat{Y},\ \hat{X}) + \theta^2\ R^2\ V(\ \hat{X}\ ). \tag{3.2.10}$$

Now minimising (3.2.9) with respect to a we get the optimum value of a as

$$a_{opt.} = \frac{Cov\ (\hat{Y},\ \hat{X})}{R\ V(\ \hat{X}\ )} = \rho\ \frac{C_y}{C_x}\ . \tag{3.2.11}$$

We will use the symbol k to denote $\rho\ \frac{C_y}{C_x}$.

imilarly minimising (3.2.10) with respect to θ we obtain the ptimum value of θ as

$$\theta_{opt.} = \frac{Cov(\hat{Y}, \hat{X})}{R\ V(\hat{X})} = k . \qquad (3.2.12)$$

substituting k for a in (3.2.7) and (3.2.9) and for θ in (3.2.8) and (3.2.10) we obtain after some simplification the bias and m.s.e. of the optimum estimators to be

$$B(\hat{Y}^{(k)}) = \frac{1}{2X} k(1-k)\ R\ V(\hat{X}),$$

$$B(\hat{Y}_k) = 0,$$

and $$M(\hat{Y}^{(k)}) = V(\hat{Y})\ (1 - \rho^2) = M(\hat{Y}_k).$$

Thus we have proved the following

<u>Theorem 3.2.1</u> : Let $\hat{Y}$ and $\hat{X}$ be unbiased estimators of Y and X respectively based on any sample design and let

$$\hat{Y}^{(k)} = \hat{Y}\left(\frac{X}{\hat{X}}\right)^k \quad \text{and} \quad \hat{Y}_k = \frac{Y\ X}{X + k(X - X)} .$$

Then to the second degree of approximation, assuming $|k\ e_1| < 1$ where $e_1 = \frac{\hat{X} - X}{X}$,

$$B(\hat{Y}^{(k)}) = \frac{1}{2X} k(1-k)\ R\ V(\hat{X}),$$

$$B(\hat{Y}_k) = 0 \quad \text{and}$$

$$M(\hat{Y}^{(k)}) = M(\hat{Y}_k) = V(\hat{Y})\ (1 - \rho^2).$$

Definition 3.2.1 : We say that an estimator $\hat{Y}^*$ of Y which uses information on a supplementary variate X is $\underset{=}{X}$-optimum if $\hat{Y}^*$ is almost unbiased and $M(\hat{Y}^*) = V(\hat{Y})(1 - \rho^2)$.

From the above theorem 3.2.1 follows immediately

Corollary 3.2.1a : The usual ratio estimator $\frac{\hat{Y}X}{\hat{X}}$ is $\underset{=}{X}$-optimum if and only if $k = 1$.

Corollary 3.2.1b : The usual product estimator $\frac{\hat{Y}\hat{X}}{X}$ has bias $\frac{-RV(\hat{X})}{X}$ and the m.s.e. $V(\hat{Y})(1 - \rho^2)$ iff $k = -1$.

If k is an integer, then the computation of $\hat{Y}^{(k)}$ is simple and immediate. Consider next the case $s < k < s + 1$, where s is an integer. Then as an estimator of Y consider

$$\hat{Y}^{(s,s+1)} = d\,\hat{Y}^{(s)} + (1-d)\,\hat{Y}^{(s+1)} \qquad (3.2.13)$$

where $0 \leq d \leq 1$. We then have

$$B(\hat{Y}^{(s,s+1)}) = E[d\,\hat{Y}^{(s)} + (1-d)\,\hat{Y}^{(s+1)} - Y]$$

$$= E[d(\hat{Y}^{(s)} - Y) + (1 - d)\,(\hat{Y}^{(s+1} - Y)]$$

$$= \frac{RV(\hat{X})}{2X}[(s+1)(s+2)-2(s+1)\,k-2d(s+1\,-k)] \qquad (3.2.14)$$

and

$$M(\hat{Y}^{(s,s+1)})=E[d(\hat{Y}^{(s)}-Y)+(1-d)(\hat{Y}^{(s+1)}-Y)]^2$$

$$= d^2[V(\hat{Y}) - 2sR\ Cov(\hat{Y},\hat{X}) + s^2R^2\ V(\hat{X})]$$

$$+(1-d)^2\ [V(\hat{Y})-2(s+1)R\ Cov(\hat{Y},\hat{X})+(s+1)^2\ R^2\ V(\hat{X})]$$

$$+2d(1-d)[V(\hat{Y})-R(2s+1)\ Cov(\hat{Y},\hat{X})+s(s+1)R^2\ V(\hat{X})]$$

$$= V(\hat{Y}) + R^2\ V(\hat{X})[(s+1)^2-2(s+1)k+d^2-2d(s+1-k)], \qquad (3.2.15)$$

n minimising (3.2.15) with respect to d, we obtain

$$d = d^* = (s + 1 - k)$$

is the value that minimises (3.2.15). Now substituting d^* for d in $\hat{Y}^{(s,s+1)}$ we obtain

$$\hat{Y}^{*(s,s+1)} = (s+1-k)\ \hat{Y}^{(s)} + (k - s)\ \hat{Y}^{(s+1)}$$

which is a convex combination of the two estimators $\hat{Y}^{(s)}$ and $^{(s+1)}$.

Also, we obtain

$$(\hat{Y}^{*(s,s+1)}) = \frac{R\ V(X)}{2X}[(k+d^*)(d^*-k+1)-2d^{*2}]$$

$$= \frac{R\ V(\hat{X})}{2X}[k(1-k)+d^*(1-d^*)],$$

and

$$(\hat{Y}^{*(s,s+1)})=V(\hat{Y})+R^2V(\hat{X})[(s+1)^2-2(s+1)\ k - (s+1-k)^2]$$

$$=V(\hat{Y}) - R^2\ k^2\ V(\hat{X})$$

$$=V(\hat{Y}) - \rho^2\ V(\hat{Y}) \qquad \text{(by def. of k)}$$

$$=V(\hat{Y})\ (1 - \rho^2).$$

Thus we have proved the following :

<u>Theorem 3.2.2</u> : For any $s \le k \le s+1$, the estimator

$$\hat{Y}^{*(s,s+1)} = (s+1-k)\ \hat{Y}^{(s)} + (k-s)\ \hat{Y}^{(s+1)}$$

of Y has bias and m.s.e. given by

$$B(\hat{Y}^{*(s,s+1)}) = \frac{1}{2X}[k(1-k)+(s+1-k)\ (k-s)]\ R\ V(\hat{X})$$

and $$M(\hat{Y}^{*(s,s+1)}) = V(\hat{Y})\ (1-\rho^2).$$

From theorems 3.2.1 and 3.2.2 follows immediately

<u>Corollary 3.2.2</u> : $B(\hat{Y}^{*(s,s+1)}) = 2B(\hat{Y}^{(k)})$, when $s = 0$

and $|B(\hat{Y}^{*(s,s+1)})| \le |B(\hat{Y}^{(k)})|$ when $s \neq 0$.

The assumption in theorem 3.2.1 that $|k\ e_1| < 1$ is not a serious restriction since even if $|k| > 1$, in practice $|e_1|$ is small for moderately large n and also recently Sukhatme and David (1973) established the fact that the condition $|e_1| < 1$ is necessary to obtain the approximate bias and m.s.e. for the usual ratio estimator.

Also it is interesting to note here that

$$\hat{Y}^{(a)} = \begin{cases} \hat{Y}_R & \text{when} \quad a = 1 \\ \hat{Y}_P & \text{when} \quad a = -1 \end{cases}$$

and

$$\hat{Y}_\theta = \begin{cases} \hat{Y}_R & \text{when } \theta = 1 \\ \dfrac{\hat{Y}\,X}{2X - \hat{X}} = \hat{Y}_{R'} & \text{when } \theta = -1. \end{cases}$$

Now from (3.2.7) to (3.2.11) it can be easily seen that

$$M(\hat{Y}_{R'}) = M(\hat{Y}_P),$$

and $$|B(\hat{Y}_{R'})| \leq |B(\hat{Y}_P)| \quad \text{whenever} \quad k \leq -\frac{1}{2}.$$

Also $$\hat{Y}_{R'} - \hat{Y}_P = \hat{Y}\Big[\frac{X}{2X-\hat{X}} - \frac{\hat{X}}{X}\Big]$$

$$= \frac{\hat{Y}\,(X - \hat{X})^2}{(2X - \hat{X})X}$$

$$\geq 0, \qquad \text{if} \quad |e_1| < 1.$$

Thus we note here that $\hat{Y}_{R'}$ is a better alternative to $\hat{Y}_P$ whenever $k \leq -\frac{1}{2}$ and in fact the use of product estimator is suggested only when $k \leq -\frac{1}{2}$. But the advantage of $\hat{Y}_P$ over $\hat{Y}_{R'}$ is that even though it under estimates Y, the exact expressions for its bias and m.s.e. can be obtained easily, while it is required that $|e_1| < 1$ to obtain the approximate bias and m.s.e. of $\hat{Y}_{R'}$.

3.3. Remarks

From theorems 3.2.1 and 3.2.2 and corollary 3.2.2 we see that

Case 1 : $k = 0 \Rightarrow \rho = 0$ and the optimum estimator turns out to be $\hat{Y}$ as it should be.

Case 2 : When k belongs to $[-1, 1]$, the optimum estimator to be used is given by $\hat{Y}_k = \frac{\hat{Y}\, X}{X + k(\hat{X} - X)}$.

Case 3 : When k does not belong to $[-1, 1]$, without loss of generality assume that k belongs to $(s, s+1)$ where s is an integer different from 0 as well as -1. In this case the optimum estimator to be used is given by

$$\hat{Y}_k = \frac{\hat{Y}\, X}{X+k(\hat{X}-X)} \qquad \text{if } |k\, e_1| < 1$$

and $$\hat{Y}^{*(s,s+1)} = (s+1-k)\, \hat{Y}^{(s)} + (k-s)\, \hat{Y}^{(s+1)} \qquad \text{if } |k\, e_1| \geq 1.$$

In this connection it may be noted that these results are of considerable importance as they show that for any given supplementary variate, one can decide whether to use $\hat{Y}$ or $\hat{Y}_k$ or $\hat{Y}^{*(s,s+1)}$ on the basis of k. Also, it is easy to see the computational simplicity, of $\hat{Y}_k$ or $\hat{Y}^{*(s,s+1)}$ over $\hat{Y}^{(k)}$.

By considering the estimator $\hat{Y}(\frac{\hat{X}}{X})^{\alpha}$ for estimating Y, and using simple random sampling scheme, Srivastava (1967) has obtained the optimum estimator when $\alpha = -k$. Clearly our theorems 3.2.1 and 3.2.2 are much more general and are improvements over the results mentioned by Srivastava (1967).

For estimating the total Y of a finite population, Gupta (1971) has recently considered the linear combination of ratio type estimators given by

$$\hat{Y}_{R*} = d\ \hat{Y}\left(\frac{X}{\hat{X}}\right) + (1-d)\ \hat{Y}\ \left(\frac{X}{\hat{X}}\right)^2 , \qquad (3.3.1)$$

and by minimising the m.s.e. of $\hat{Y}_{R*}$ with respect to d, he has obtained the optimum value of d as $(2-k)$ and the corresponding $M(\hat{Y}_{R*}) = V(\hat{Y})\ (1 - \rho^2)$. Clearly, when k does not belong to $[1, 2]$ either d or $1 - d$ becomes negative and consequently, the optimum estimator $\hat{Y}_{R*}$ might turn out to be negative and hence might not be useful in practice. In case k belongs to $[,1,2]$, we can obtain the above result of Gupta by putting $s = 1$ in theorem 3.2.2.

Gupta (1971) has also considered the linear combination of product type estimators given by

$$\hat{Y}_{p*} = d\ \frac{\hat{Y}\ \hat{X}}{X} + (1-d)\ \hat{Y}\left(\frac{\hat{X}}{X}\right)^2 , \qquad (3.3.2)$$

and obtained the corresponding optimum product type estimator by minimising $M(\hat{Y}_{p*})$ with respect to d. It can be easily shown that the corresponding optimum product type estimator Y_{p*} will be meaningful only when k belongs to $[-2, -1]$. In case k belongs to $[-2, -1]$ the above result of Gupta follows directly by taking $s = -2$ in theorem 3.2.2.

3.4 Illustrative examples

The following nine populations are considered to illustrate the results of this chapter. The details of these populations are given in tables A1 to A6 of the appendix.

Population 1 : It consists of 80 factories in a region for which the data on number of workers $(\underline{\underline{X}})$ and output $(\underline{\underline{Y}})$ is available (p. 228, Murthy (1967)).

Population 2 : Here we have data on fixed capital $(\underline{\underline{X}})$ and output $(\underline{\underline{Y}})$ for 80 factories in a region (p. 228, Murthy (1967)).

Population 3 : Here the data consists of 1951 census village-wise data for a tehsil on geographic area $(\underline{\underline{X}})$ and total cultivated area $(\underline{\underline{Y}})$ for 128 villages (pp. 126-130, Murthy (1967)).

Population 4 : It consists of a specific group of 45 cities/urban agglomerations (u.a.'s) with population size 100,000 and above (Census of India 1971, Series-1, Paper-1 of 1971 Supplement, provisional population totals, issued by Registrar General and Census Commissioner, India) with data on female literacy rate $(\underline{\underline{X}})$ and female work participation rate $(\underline{\underline{Y}})$.

Population 5 : It consists of data for jute plants (Capsulanes) sown at Jute Agricultural Research Station Farm, Barrackpore, Calcutta, in the year 1962-63 on height of the jute plant $(\underline{\underline{X}})$ and yield of fibre from the plant $(\underline{\underline{Y}})$ for 50 jute plants.

Population 6 : Same as the above population 5 with $\underline{\underline{X}}$ being the base diameter of the plants.

For the 142 cities of India with population 100,000 and above we define populations 7 and 8 as

Population 7 : $\underset{=}{X}$: 1961 census population (Persons).

$\underset{=}{Y}$: 1971 census population (Persons).

Population 8 : $\underset{=}{X}$: 1971 census population (Persons).

$\underset{=}{Y}$: Workers in 1971 as per census.

Population 9 : Here we have data on number of workers ($\underset{=}{X}$) and number of absentees ($\underset{=}{Y}$) for 43 factories (p. 398, Murthy (1967)), where these 43 factories themselves constitute a simple random sample without replacement from a list of 325 factories situated in a district of India.

Values of some of the population parameters for the above 9 populations are **provided** in the summary table 3.4.1 below.

SUMMARY TABLE 3.4.1

Values of some of the population parameters

Pop. No.	$\bar{Y}$	$\bar{X}$	C_0	C_1	ρ	$k = \rho \frac{C_0}{C_1}$
1	5182.6375	285.1250	0.3542	0.9485	0.9150	0.3417
2	5182.6375	1126.4625	0.3542	0.7507	0.9412	0.4441
3	2303.2593	6.6280	0.5700	0.6200	0.8124	0.7469
4	6.6900	44.1400	0.5684	0.1944	-0.6501	-1.9008
5	5.6900	6.6500	0.2383	0.0920	0.7418	1.9214
6	5.6900	1.4200	0.2383	0.1127	0.5677	1.2004
7	4015.2200	2900.3900	2.1192	2.2049	0.9948	0.9561
8	1201.9900	4015.2200	2.3621	2.1192	0.9954	1.1095
9	9.6512	79.4651	0.6805	0.4590	0.6608	0.9797

The efficiencies of $\hat{Y}_R$ ($\hat{Y}_P$ when ρ is negative) and $\hat{Y}_k$ with respect to $\hat{Y}$ for the above 9 populations are presented in the following table.

TABLE 3.4.2

Percentage efficiencies of different estimators relative to $\hat{Y}$.

Pop. No.	$\hat{Y}$	$\hat{Y}_R$ or $\hat{Y}_P$	$\hat{Y}_k$
1	100	31	614
2	100	67	876
3	100	241	294
4	100	149	173
5	100	174	222
6	100	146	148
7	100	8026	9640
8	100	5311	10882
9	100	177	178

The efficiency of $\hat{Y}_R$ over $\hat{Y}$ or $\hat{Y}_k$ depends crucially on both ρ and k. In fact $\hat{Y}_R$ turns out to be less efficient than $\hat{Y}$ if $k < \frac{1}{2}$ (as can be seen from populations 1 and 2) and $\hat{Y}_R$ will be superior to $\hat{Y}$ and as efficient as $\hat{Y}_k$ if $k = 1$ (as is clear from population 9). If ρ is quite high, then a moderate deviation of k from 1 will result in $\hat{Y}_R$ to be much less efficient than $\hat{Y}_k$ (which can be observed from population 8). But if ρ is moderate, then a moderate departure of k from 1 may not reduce very much the efficiency of $\hat{Y}_R$ over $\hat{Y}_k$ (which is clear from population 6).

3.5 Order of the approximate variance

Assuming the population under consideration to be large and (Y, X) follows a bivariate Normal distribution, Sukhatme (1954) has investigated the errors involved in the approximate formulae for bias and m.s.e. of the ratio estimator obtained by taking the terms up to the fourth degree. On similar lines, Srivastava (1967) has studied the bias and m.s.e. of $\hat{Y}^{(k)}$ and compared them with those of the ratio estimator under the assumption $C_y = C_x$. In this section, we study the errors involved in the approximation of the bias and m.s.e. of $\hat{Y}^{(k)}$ and $\hat{Y}_x$ and compare them with those of $\hat{Y}_R(\hat{Y}_p$ if p is negative) without assuming that $C_y = C_x$.

If $\hat{Y}$ and $\hat{X}$ are unbiased estimators of the population totals Y and X respectively based on a simple random sample of size n, then from (3.2.5) and (3.2.6) it can be easily shown that, up to terms of order $\frac{1}{n^2}$,

$$B(\hat{Y}^{(a)}) = \frac{a\ Y\ C_x^2}{2n}(a+1-2k) + \frac{a\ Y\ C_x^4}{8n^2}(a+1)\ (a+2)\ (a+3-4k). \qquad (3.5.1)$$

$$B(\hat{Y}_\theta) = \frac{\theta\ Y\ C_x^2}{n}\ (\theta - k) + \frac{3\ \theta^3\ Y\ C_x^4}{n^2}\ (\theta - k) \qquad (3.5.2)$$

$$M(\hat{Y}^{(a)}) = \frac{Y^2}{n}[(C_y^2 - 2a\ k\ C_x^2 + a^2 C_x^2) + \frac{1}{n}\{a(3a+1)(1+2\rho^2)C_y^2 C_x^2$$

$$-a(a+1)(7a+2)k\ C_x^4 + \frac{1}{4}\ a^2(a+1)\ (7a+11)\ C_x^4\}] \qquad (3.5.3)$$

and

$$M(\hat{Y}_\theta) = \frac{Y^2}{n}[(C_y^2 - 2\theta k\ C_x^2 + \theta^2 C_x^2) + \frac{1}{n}\{3\theta^2(1+2\rho^2)C_y^2 C_x^2 + 3\theta^3(\theta - 4k)C_x^4\}]. \qquad (3.5.4)$$

putting $a = k$ in (3.5.1) and (3.5.3) we get

$$B(\hat{Y}^{(k)}) = \frac{(1-k)\ C_x^2\ Y}{n}\ [\ \frac{k}{2} + \frac{3}{8n}\ (k+1)\ (k+2)] \qquad (3.5.5)$$

d

$$\hat{Y}^{(k)}) = \frac{Y^2}{n}\ [C_y^2(1-\rho^2) + \frac{C_y^2\ C_x^2}{4n}\ \{4k(2k+1)\ -\ \rho^2(5k^2+10k-3)\}]. \qquad (3.5.6)$$

so, putting $\theta = k$ in (3.5.2) and (3.5.4) we obtain

$$B(\hat{Y}_k) = 0, \qquad (3.5.7)$$

nd

$$M(\hat{Y}_k) = \frac{Y^2}{n}\ [C_y^2(1-\rho^2)\ +\ \frac{3}{n}\ k^2(1-\rho^2)\ C_y^2\ C_x^2] \qquad (3.5.8)$$

imilar expressions for bias and m.s.e. of ratio and product estimators can be obtained by putting $a = 1$ and $a = -1$ in (3.5.1) and (3.5.3) respectively, viz.,

$$B(\hat{Y}_R) = \frac{Y}{n}(1-k)\ C_x^2\ (1 + \frac{3}{n}\ C_x^2) \qquad (3.5.9)$$

$$M(\hat{Y}_R) = \frac{Y^2}{n}[C_y^2(1-\rho^2)+C_x^2(1-k)^2+\frac{3}{n}\ \{C_y^2\ C_x^2(1-\rho^2)+3C_x^4(1-k)^2\}], \qquad (3.5.10)$$

$$B(\hat{Y}_P) = \frac{Y}{n}\ k\ C_x^2 \qquad (3.5.11)$$

and

$$M(\hat{Y}_P) = \frac{Y^2}{n}\ [C_y^2(1-\rho^2)\ +\ (1+k)^2\ C_x^2\ +\ \frac{1}{n}(1+2\rho^2)\ C_y^2\ C_x^2]. \qquad (3.5.12)$$

From (3.5.5) and (3.5.9) it is easy to see that

$$B(\hat{Y}^{(k)}) \leq B(\hat{Y}_R)$$

rovided $0 \leq k \leq 1$. Also from (3.5.6) and (3.5.10) we get

$$(\hat{Y}^{(k)})-M(\hat{Y}_R) = \frac{Y^2}{n}[-C_x^2(1-k)^2+\frac{C_y^2\,C_x^2}{4n}\{4k(2k+1)-\rho^2(5k^2+10k-3)$$

$$-12(1-\rho^2)\}-\frac{9C_x^4(1-k)^2}{n}]$$

$$\leq \frac{Y^2}{4n^2}C_y^2C_x^2[4(k-1)(2k+3)-5\rho^2(k-1)(k+3)-\frac{36\;\rho^2(1-k)^2}{k^2}]$$

$$\leq \frac{Y^2}{4n^2}C_y^2C_x^2(k-1)[8k+12-5\rho^2(k+3)+36\rho^2-36\rho^2k] \text{provided } 0<k\leq 1$$

$$= \frac{Y^2}{4n^2}C_y^2C_x^2(k-1)[8k(1-\rho^2)+21\rho^2(1-k)+12(1-\rho^2k)]$$

$$\leq 0.$$

rom (3.5.6) and (3.5.8) we obtain after simplification

$$(\hat{Y}_k)-M(\hat{Y}^{(k)}) = \frac{Y^2\;C_y^2\;C_x^2}{4n^2}(k-1)[(4-7\rho^2)k+3\rho^2]$$

$$= \frac{Y^2\;C_y^2\;C_x^2}{4n^2}(k-1)\;[4(1-\rho^2)k+3\rho^2(1-k)]$$

$$\leq 0, \text{ provided } 0<k\leq 1.$$

hus for $0 < k \leq 1$, we get

$$\left.\begin{array}{l} 0 = B(\hat{Y}_k) \leq B(\hat{Y}^{(k)}) \leq B(\hat{Y}_R) \\ \text{nd} \\ M(\hat{Y}_k) \leq M(\hat{Y}^{(k)}) \leq M(\hat{Y}_R). \end{array}\right\rangle \qquad (3.5.14)$$

From (3.5.6) and (3.5.12) we get that

$$M(\hat{Y}_p)-M(\hat{Y}^{(k)})= \frac{C_y^2C_x^2Y^2}{n^2}[(1+2\rho^2)-k(2k+1)+\frac{\rho^2}{4}(5k^2+10k-3)]+\frac{Y^2C_x^2(1+k)^2}{n}$$

$$\geq \frac{C_y^2C_x^2Y^2}{4n^2}\ [4(1-2k^2-k) + \rho^2 5(k+1)^2]$$

$$= \frac{C_y^2C_x^2Y^2}{4n^2}(k+1)\ [4(1-2k) + 5\rho^2(1+k)]$$

$$\geq 0, \quad \text{provided} \quad -1 \leq k < 0.$$

Also, from (3.5.13) we have

$$M(\hat{Y}_k)-M(\hat{Y}^{(k)}) = \frac{Y^2C_x^2C_y^2}{4n^2}\ (k-1)[(4-7\rho^2)\ k + 3\rho^2]$$

$$\leq \frac{Y^2C_y^2C_x^2}{4n^2}\ k(k-1)\ (4 - 10\rho^2) \quad \text{provided } -1 \leq k < 0$$

$$\leq\ 0, \quad \text{provided} \quad \rho^2 \geq 0.4.$$

Thus for $-1 \leq k < 0$, and $\rho^2 \geq 0.4$, we get

$$M(\hat{Y}_k) \leq M(\hat{Y}^{(k)}) \leq M(\hat{Y}_p). \tag{3.5.15}$$

From (3.5.3) we have

$$M(\hat{Y}_k) = \frac{Y^2}{n}\ [C_y^2(1 - \rho^2) + \frac{3}{n}\ C_y^2\ C_x^2\ k^2(1 - \rho^2)]$$

$$= V_1 + \frac{3}{n^2}\ \ Y^2\ C_y^2\ C_x^2\ (1 - \rho^2)\ k^2, \text{ say.}$$

Following Cochran (1963), section 6.4, we note that the use of V_1 understates the m.s.e. of $\hat{Y}_k$ by a proportion P_1 given by

$$P_1 = \frac{3}{n} k^2 C_x^2 ,$$

which is negligible in large samples. Similarly from (3.5.6) the proportion of understatement of $M(\hat{Y}^{(k)})$, denoted by P_2, is given by

$$P_2 = \frac{C_x^2}{4(1-\rho^2)n} \; [4k(2k+1) - \rho^2(5k^2+10k-3)].$$

From (3.5.14) it is easy to see that

$$P_1 \leq P_2 \quad \text{provided} \quad 0 < k \leq 1. \qquad (3.5.16)$$

From (3.5.10) we get P_3, the proportion of understatement of $M(\hat{Y}_R)$ to be

$$\frac{3C_x^2}{n} \quad \frac{C_y^2(1-\rho^2) + 3C_x^2(1-k)^2}{C_y^2(1-\rho^2) + C_x^2(1-k)^2} .$$

Now $P_2 - P_3$ reduces after simplification to

$$= \frac{C_x^2}{n}(k-1) \; [\frac{(2k+3) + 3\rho^2(k-1)(k^2-2k\rho^2+9\rho^2-8)}{\{k^2(1-\rho^2) + \rho^2(1-k)^2\}4(1-\rho^2)}]. \qquad (3.5.17)$$

It is easy to verify that for $0 < k \leq 1$, a sufficient condition for the above expression (3.5.17) to be less than or equal to zero is that $\rho^2 \leq \frac{12}{13}$, which will be satisfied in almost all practical situations. Thus we see that the proportion of understatement of $M(\hat{Y}^{(k)})$ is smaller than that of $M(\hat{Y}_R)$ for $0 < k \leq 1$ and $\rho^2 \leq \frac{12}{13}$.

From (3.5.12) the proportion of understatement of $M(\hat{Y}_P)$, denoted by P_4, can be seen to be

$$C_x^2 \left[\frac{(1+2\rho^2)\, C_y^2}{C_y^2(1-\rho^2)+C_x^2(1+k)^2} \right].$$

Now $P_1 - P_4$ simplifies to

$$= \frac{k^2 C_x^2}{n} \, [3k^2 + 6\rho^2 k - (1 - \rho^2)]$$

$$\leq 0, \quad \text{for } -1 \leq k \leq 0, \quad \text{provided } \rho^2 \geq 0.4.$$

For the populations numbered 1, 2, 3, 7 and 8 referred to in section 4 of this chapter, we present below a table of the proportions of understatement of the m.s.e. of different estimators. Here we are assuming that $(\underline{\underline{Y}}, \underline{\underline{X}})$ follows bivariate Normal distribution in the populations considered.

TABLE 3.5.1

Proportion of understatement of m.s.e. of different estimators, multiplied by the sample size.

Pop. No.	$M(\hat{Y}_R)$	$M(\hat{Y}^{(k)})$	$M(\hat{Y}_X)$
1	7.8282	2.0221	0.3151
2	4.8150	1.4863	0.3334
3	1.5736	0.7518	0.6433
7	19.4492	14.2099	13.9445
8	27.2850	20.3992	14.9483

the above table 3.5.1 it is clear that eventhough $M(Y_k)$ uniformly smaller proportion of understatement than those $M(\hat{Y}_R)$ and $M(\hat{Y}^{(k)})$ for all the five populations, but its rstatement for the populations numbered 7 and 8 will be l and negligible only in case of large samples. Another t to be noted here is that the proportions of understatement $M(\hat{Y}_R)$ and $M(\hat{Y}^{(k)})$ will be small and negligible for all five populations only in case of large samples.

CHAPTER IV

TRANSFORMED RATIO METHOD OF ESTIMATION

4.0 Summary

Let $\hat{Y}$ and $\hat{X}$ be unbiased estimators of the population totals Y and X of the study variate $\underline{\underline{Y}}$ and the auxiliary variate $\underline{\underline{X}}$ respectively, based on any sample design. Let

$$\hat{Y}_\theta = \frac{\hat{Y}\ X}{X + \theta(\hat{X}-X)}$$

be an alternative estimator of Y based on the supplementary information where θ is a scalar and we assume that $\left|\frac{\theta(\hat{X} - X)}{X}\right| < 1$. We have seen in the previous chapter III that the above estimator $\hat{Y}_\theta$ of Y is almost unbiased and has the m.s.e. equal to $V(\hat{Y})\ (1 - \rho^2)$ when $\theta = \rho \frac{C_Y}{C_X} = k$, where ρ is the correlation coefficient between $\hat{Y}$ and $\hat{X}$ and C_y and C_x denote the coefficients of variation of $\hat{Y}$ and $\hat{X}$ respectively. The justification for considering $\hat{Y}_k$ is provided by considering a modified ratio estimator with a suitable transformation on the auxiliary variate. Usually k will not be known but a good guess θ of k can be obtained either from a pilot study or past data or experience. So, the effects of the deviation of θ from k on the bias and m.s.e. have been studied and it is found that the absolute bias and m.s.e. of $\hat{Y}_\theta$ are smaller than those of the usual ratio estimator (or product estimator) for a fairly wide range of values of θ when k belongs to [0, 1] (or[-1, 0]). The estimator $\hat{Y}_{\theta}$, its bias and m.s.e. have been examined under a super-population model, and it is found that a ratio-type estimator with a suitable transformation on the auxiliary variate will result in a smaller absolute bias and smaller m.s.e. under fairly valid conditions. A multivariate extension of the estimator is suggested and a comparison of its bias has been made with that of Srivastava's (1971) generalised ratio estimator. An empirical example is also included for illustration.

[illegible].1 Introduction

Whenever some suitable auxiliary information is available on a supplementary variate $\underset{=}{X}$ highly correlated with the study variate $\underset{=}{Y}$ it is the usual practice in sample surveys to utilize this information in the method of estimation which increases efficiency. The ratio method of estimation (or the product method) gives a more efficient estimator than the simple unbiased estimator provided the correlation coefficient ρ between the study variate and the auxiliary variate has high positive value (or high negative value). Further, when a good guess of $K = \rho \frac{C_y}{C_x}$ is available, where C_y and C_x denote respectively the coefficients of variation of the estimators of the population totals Y and X, we have seen in chapter III that the estimator

$$\hat{Y}_k = \frac{\hat{Y}X}{X+k(\hat{X}-X)}$$

is $\underline{\underline{X}}$-optimum.

4.2 Justification for the use of $\hat{Y}_k$

It is well known that the use of ratio estimator is suggested whenever the correlation coefficient ρ between $\hat{Y}$ and $\hat{X}$ is positive and high and the use of product estimator is suggested when ρ is negative and high. For these two cases we separately present the justification for considering the estimator $\hat{Y}_k$.

Case (1) : $\rho_{\underline{\underline{YX}}}$, the correlation coefficient between $\hat{Y}$ and $\hat{X}$, is positive. It has been shown in corollary 3.2.1a that the usual ratio estimator is $\underline{\underline{X}}$-optimum if and only if $k = 1$. Hence, instead of the auxiliary variate $\underline{\underline{X}}$ if we consider a suitable transformed variate $\underline{\underline{X}}^+$ taking value X_i^+ on unit U_i, $i = 1,2,\ldots,N$, in such a way that

(A) : (1) $\rho_{\underline{\underline{Y}}\,\underline{\underline{X}}^+} = \rho_{\underline{\underline{Y}}\,\underline{\underline{X}}}$, (2) $\dfrac{\rho\, C_{\underline{\underline{Y}}}}{C_{\underline{\underline{X}}^+}} = 1$ and (3) $\hat{X}^+$ lies between 0 and $2X^+$ where $\hat{X}^+$ is an unbiased estimator of X^+,

then the transformed ratio estimator $\dfrac{\hat{Y}X^+}{\hat{X}^+}$ turns out to be $\underline{\underline{X}}^+$-optimum.

Now choose $\underline{\underline{X}}^+ = \underline{\underline{X}} + d\,\bar{X}$ where $d = \frac{1-k}{k}$ and

$$\hat{\bar{X}}^+ = \hat{\bar{X}} + d\,\bar{X}$$

$$= \frac{\bar{X} + k(\hat{\bar{X}} - \bar{X})}{k} .$$

Clearly $\bar{X}^+ = \bar{X} + d\,\bar{X} = \frac{\bar{X}}{k}$

and so $$\left| \frac{\hat{\bar{X}}^+ - \bar{X}^+}{\bar{X}^+} \right| = \left| \frac{(\hat{\bar{X}} - \bar{X})k}{\bar{X}} \right| . \qquad (4.2.1)$$

A sufficient condition for the above expression (4.2.1) to be less than 1 is that $k \leq 1$. Thus, if $k \leq 1$, then $\hat{\bar{X}}^+$ lies between 0 and $2\bar{X}^+$ and the other two conditions of (A) are trivially satisfied, hence $\frac{\hat{Y}\,X^+}{\hat{X}^+}$ will be $\underline{\underline{X}}^+$-optimum. Now

$$\frac{\hat{Y}\,X^+}{\hat{X}^+} = \frac{\hat{Y}\,X}{X + k(\hat{X} - X)}$$

$$= \hat{Y}_k ,$$

and since $\rho_{\underline{\underline{Y}}\,\underline{\underline{X}}^+} = \rho_{\underline{\underline{Y}}\,\underline{\underline{X}}}$, it follows that $\hat{Y}_k$ is $\underline{\underline{X}}$-optimum.

<u>Case (ii)</u> : $\rho_{\underline{\underline{Y}}\,\underline{\underline{X}}}$ is negative. From theorem 3.2.1 it easily follows that $\hat{Y}_{R'} = \frac{\hat{Y}\,X}{2X - \hat{X}}$ will be $\underline{\underline{X}}$-optimum if and only if $k = -1$. Hence instead of $\underline{\underline{X}}$ if we consider a suitable transformed variate $\underline{\underline{X}}^-$ taking value X_i^- on the unit U_i, $i = 1, 2, \ldots, N$, in such a way that

(B): (1) $\rho_{\underline{\underline{Y}}\,\underline{\underline{X}}^{-}} = \rho_{\underline{\underline{Y}}\,\underline{\underline{X}}}$, (2) $\frac{\rho\,\sigma_{\underline{X}}}{C_{\underline{X}^{-}}} = -1$ and $\hat{X}^{-}$ (an unbiased estimator of X^{-}) lies between 0 and $2X^{-}$ then the transformed ratio-type estimator $\frac{\hat{Y}X^{-}}{2X^{-}-\hat{X}^{-}}$ turns out to be $\underline{\underline{X}}^{-}$-optimum. Now choose $\underline{\underline{X}}^{-} = \underline{\underline{X}} - d\,\bar{X}$ where $d = \frac{1+k}{k}$ and

$$\hat{\bar{X}}^{-} = \hat{\bar{X}} - d\,\bar{X}$$

$$= \frac{k(\hat{\bar{X}}-\bar{X})-\bar{X}}{k}.$$

Then $\bar{X}^{-} = -\frac{\bar{X}}{k}$,

and $$\left|\frac{\hat{\bar{X}}^{-}-\bar{X}^{-}}{\bar{X}^{-}}\right| = \left|\frac{(\hat{\bar{X}}-\bar{X})\,(-k)}{\bar{X}}\right|. \qquad (4.2.2)$$

A sufficient condition for the above expression (4.2.2) to be less than 1 is that $|k| \leq 1$. Thus if $-1 \leq k < 0$, then $\hat{\bar{X}}$ lies between 0 and $2\bar{X}^{-}$ and the other conditions of (B) are trivially satisfied, hence $\frac{\hat{Y}X^{-}}{2X^{-}-\hat{X}^{-}}$ turns out to be $\underline{\underline{X}}^{-}$-optimum and so $\underline{\underline{X}}$-optimum.

$$\text{Now}\quad \frac{\hat{Y}X^{-}}{2X^{-}-\hat{X}^{-}} = \frac{\hat{Y}(-\bar{X})}{\bar{X}+k(\hat{\bar{X}}-\bar{X})}$$

$$= \hat{Y}_{k}\,.$$

4.3 Effect of the deviation of θ from k

Usually k will not be known but a good guess θ of k can be obtained either from a pilot study or from past data or experience. In this section we study the effect of the deviation of θ from k by comparing the bias and m.s.e. of $\hat{Y}_\theta$ with those of $\hat{Y}$, $\hat{Y}_R$ and $\hat{Y}_P$ for θ and k in $[-1, 1]$. We distinguish two cases.

Case (1) : $0 < k \leq 1$.

In this case it is reasonable to assume that $\theta \varepsilon (0, 1]$ and it is meaningful to make a comparison of the bias and m.s.e. of $\hat{Y}_\theta$ with those of $\hat{Y}_R$ and $\hat{Y}$. The bias and m.s.e. of $\hat{Y}_R$ to the second degree of approximation as well known to be

$$B(\hat{Y}_R) = Y\, C_x^2\, (1-k), \tag{4.3.1}$$

and

$$M(\hat{Y}_R) = Y^2\, [C_y^2 + (1-2k)\, C_x^2]. \tag{4.3.2}$$

Comparing (4.3.2) with (3.2.10) we find that

$$M(\hat{Y}_\theta) \leq M(\hat{Y}_R) \text{ if and only if } \theta \geq 2k - 1. \tag{4.3.3}$$

From (3.2.8) and (4.3.1) we find that

$$|B(\hat{Y}_\theta)| \leq |B(\hat{Y}_R)| \text{ if and only if } \theta \geq 2k - 1. \tag{4.3.4}$$

Also from (3.2.10) we find that

$$M(\hat{Y}_\theta) \leq V(\hat{Y}) \text{ if and only if } \theta \leq 2k. \tag{4.3.5}$$

It is known that

$$M(\hat{Y}_R) \lesseqgtr V(\hat{Y}) \text{ according as } k \gtreqless \frac{1}{2}. \qquad (4.3.6)$$

Thus for $k > \frac{1}{2}$ and for any θ in $(2k-1, 1)$ we have from (4.3.3) to (4.3.6) that

$$M(\hat{Y}_\theta) < M(\hat{Y}_R) < V(\hat{Y})$$

and

$$|B(\hat{Y}_\theta)| < |B(\hat{Y}_R)|.$$

Also, $M(\hat{Y}_\theta) \leq V(\hat{Y}) \leq M(\hat{Y}_R)$ for $k \leq \frac{1}{2}$ and for any θ in $[0, 2k]$. We present the above results diagrammatically in Fig. 4.3.1. Note that $\hat{Y}_\theta$ has also smaller absolute bias than $\hat{Y}_R$ in the region marked I in Fig. 4.3.1.

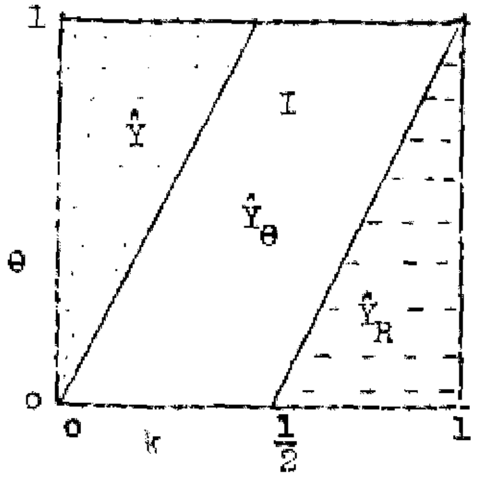


Fig.4.3.1. Configurational representation of the regions of preference for $\hat{Y}_\theta$, $\hat{Y}_R$ and $\hat{Y}$ based on m.s.e.

To get a clear idea of the comparison of the bias and m.s.e. of $\hat{Y}_\theta$ with those of $\hat{Y}_R$, we consider a simple example where $C_y = C_x$. Then $k = \rho$ and supposing $\rho = 0.65$, we find that the absolute bias and m.s.e. of $\hat{Y}_\theta$ are smaller than those of $\hat{Y}_R$ respectively for any θ in (0.3, 1) in this example.

Case (2) : $-1 \leq k \leq 0$.

In this case it is reasonable to assume that the guessed value $\theta \varepsilon [-1, 0)$ and it is meaningful to make a comparison among $\hat{Y}_\theta$, $\hat{Y}_P$ and $\hat{Y}$ with respect to absolute bias and m.s.e. The bias and m.s.e. of $\hat{Y}_P$ to the second degree of approximation are well known to be

$$B(\hat{Y}_P) = Y\ k\ C_x^2 \tag{4.3.7}$$

and
$$M(\hat{Y}_P) = Y^2[C_y^2 + (1 + 2k)\ C_x^2]. \tag{4.3.8}$$

Comparing the squared biases of $\hat{Y}_\theta$ and $\hat{Y}_P$ we find that

$$|B(\hat{Y}_\theta)| \leq |B(\hat{Y}_P)| \text{ if and only if } \theta \geq 2k. \tag{4.3.9}$$

By comparing (3.2.10) with (4.3.8) we find that

$$M(\hat{Y}_\theta) \leq M(\hat{Y}_P) \text{ if and only if } \theta \leq 2k + 1. \tag{4.3.10}$$

Also, it is known that

$$M(\hat{Y}_P) \lesseqgtr V(\hat{Y}) \text{ according as } k \lesseqgtr -\frac{1}{2}. \tag{4.3.11}$$

From (3.2.10) we have

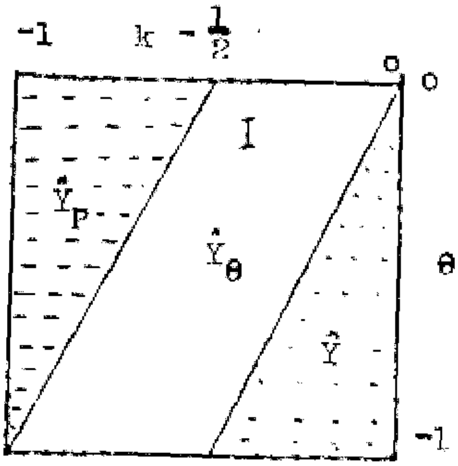
$$M(\hat{Y}_\theta) \leq V(\hat{Y}) \text{ if and only if } \theta \geq 2k. \qquad (4.3.12)$$

Thus from (4.3.5) and (4.3.12) we find that

$$M(\hat{Y}_\theta) < M(\hat{Y}_P) < V(\hat{Y}) \quad \text{for} \quad k < -\tfrac{1}{2} \text{ and for any } \theta \text{ in } (-1,\ 2k+1)$$

and $M(\hat{Y}_\theta) \leq V(\hat{Y}) \leq M(\hat{Y}_P)$ for $k \geq -\frac{1}{2}$ and for any θ in $[2k,\ 0]$.

Also, it is interesting to note that the absolute bias and m.s.e. of $\hat{Y}_\theta$ are smaller than those of $\hat{Y}_P$ respectively for any θ in $(2k,\ 2k+1)$. We present the above results diagrammatically in Fig. 4.3.2. Note that $\hat{Y}_\theta$ has also smaller absolute bias than $\hat{Y}_P$ in the region marked I in Fig. 4.3.2.



g.4.3.2. Configurational representation of the regions of preference for $\hat{Y}_\theta, \hat{Y}_P$ and $\hat{Y}$ based on m.s.e.

To get a clear idea of the comparison of the bias and m.s.e. of $\hat{Y}_\theta$ with those of $\hat{Y}_p$, we consider a simple example in which $k = \rho = -0.6$. For this example we find that the absolute bias and m.s.e. of $\hat{Y}_\theta$ are smaller than those of $\hat{Y}_p$ respectively for any θ in $(-1, -0.2)$.

4.4 The estimator and prior distributions

In this section we study the bias and m.s.e. of $\hat{Y}_\theta$ under a certain super-population set up. This concept was taken from Bayesian inference, where we assume that our prior knowledge about the variate $\underline{\underline{Y}}$ under study can be formulated in some sort of a prior distribution δ over $\underset{\sim}{Y} = (Y_1, Y_2, \ldots, Y_N)$. The role of the prior distribution δ is solely to choose between different estimators (and also designs) and has nothing to do with the final inference about $\underline{\underline{Y}}$, which will exclusively depend on the observed sample s and the variate values y_i, $i \in s$. This approach introduced by Cochran (1946) has been fruitfully exploited by Godambe (1955), Hajek (1959), Aggarwal (1959) and many others. Our sole criterion of judgment about an estimator in terms of a prior distribution would be the expected m.s.e. of the estimator under this prior distribution. Let us consider the class of all prior distributions Δ'_β of δ' for which

$$\left.\begin{aligned} &\mathcal{E}_{\delta'}(Y_i|X_i) = a + b\,X_i \\ &\mathcal{V}_{\delta'}(Y_i|X_i) = \sigma^2 X_i^g \\ \text{and}\quad &Cov_{\delta'}(Y_i, Y_j|X_i, X_j) = 0 \text{ for } i \neq j\,. \end{aligned}\right\} \quad \text{-- } (4.4.1)$$

from (3.2.8) we have

$$\mathcal{E}_{\delta'} B(\hat{Y}_\theta) = \frac{1}{X}\,[\theta^2\, V(\hat{X})\mathcal{E}_{\delta'}(R) - \theta\,\mathcal{E}_{\delta'} Cov(\hat{Y}, \hat{X})]$$

$$= 0 \quad \text{provided} \quad \theta = \theta_o = \frac{\mathcal{E}_{\delta'} Cov(\hat{Y}, \hat{X})}{V(\hat{X})\,\mathcal{E}_{\delta'}(R)}\,.$$

In the case of simple random sampling scheme

$$\theta_o = \bar{X}\;\frac{[\sum_{i=1}^{N} Y_i\, X_i - N\,\bar{Y}\,\bar{X}]}{N\,\sigma_x^2\,\mathcal{E}_{\delta'}(\bar{Y})}\,, \quad \text{where} \quad \sigma_x^2 = \frac{\sum_{i=1}^{N} (X_i - \bar{X})^2}{N}$$

$$= \frac{b\,\bar{X}}{a+b\,\bar{X}}\,.$$

Also it can be verified that, upto the second degree of approximation,

$$\mathcal{E}_{\delta'}(k) \doteq \theta_o + C_o^2(g)\,[\theta_o - \rho_{1g}\,\frac{C_g}{C_1}\,]$$

where $C_o^2(g) = \dfrac{\sigma^2 \sum_{i=1}^{N} X_i^g}{N^2(a+b\,\bar{X})^2}$, is the coefficient of variation of $\bar{Y}$ given X_i's, ρ_{1g} is the correlation coefficient between X^g and $\underset{=}{X}$ and C_g and C_1 are the population coefficients of variation of $\underline{\underline{X}}^{g}$ and $\underline{\underline{X}}$ respectively. So, if $C_o^2(g)$ is negligibly small, we get that.

$$\mathcal{E}_{\delta'}(R) \doteq \theta_o . \qquad (4.4.3)$$

From (3.2.10) we have that

$$\mathcal{E}_{\delta'} M(\hat{Y}_\theta) = \mathcal{E}_{\delta'} [V(\hat{Y}) - 2\theta R \, Cov(\hat{Y}, \hat{X}) + \theta^2 R^2 \, V(\hat{X})].$$

Equating to 0 the derivative of $\mathcal{E}_{\delta'} M(\hat{Y}_\theta)$ with respect to θ we obtain

$$\theta_1 = \frac{\mathcal{E}_{\delta'} R \, Cov(\hat{Y}, \hat{X})}{V(\hat{X}) \mathcal{E}_{\delta'}(R)}$$

as the value of θ which minimises $\mathcal{E}_{\delta'} M(\hat{Y}_\theta)$. In the case of simple random sampling scheme θ_1 simplies to

$$\theta_1 = \frac{\theta_o}{1 + C_o^2(g)} + \frac{C_o(g) \, \rho_{1g} \, C_g}{(1 + C_o^2(g))}$$

$$\doteq \quad \theta_o \text{ provided } C_o^2(g) \text{ is negligibly small.}$$

Thus we have

<u>Theorem 4.4.1</u> : Let $\hat{Y}$ and $\hat{X}$ be unbiased estimators of Y and X respectively based on simple random sampling, and let

$$\hat{Y}_\theta = \frac{\hat{Y} X}{X + \theta(\hat{X} - X)}$$

be an alternative estimator of Y. Then, under the super-population model Δ'_g, $\mathcal{E}_\delta B(\hat{Y}_\theta)$ is almost zero and $\mathcal{E}_\delta M(\hat{Y}_\theta)$ is minimum simultaneously at $\theta = \theta_o = \frac{b\bar{X}}{a + b\bar{X}}$ provided $C_o(g)$ the coefficient of variation of $\bar{Y}$ given X_i's is negligibly small.

Corollary 4.4.1 : The usual ratio estimator is almost unbiased (in fact unbiased Cochran (1963)) and has the minimum m.s.e. under the above super-population model Δ'_g provided $C_o(g)$ is small and the super-population parameter a of (4.4.1) is equal to zero.

We have from (4.4.2)

$$\theta_o = \frac{b\bar{X}}{a + b\bar{X}} = \frac{\bar{X}}{(\bar{X} + \frac{a}{b})} .$$

Now substituting $\theta = \theta_o$ in $\hat{Y}_\theta$ we obtain

$$\hat{Y}_{\theta_o} = \frac{\hat{Y}(\bar{X} + \frac{a}{b})}{(\hat{\bar{X}} + \frac{a}{b})}$$

$$= \frac{\hat{Y}\bar{X}'}{\hat{\bar{X}}'} \quad \text{where} \quad \bar{X}' = \bar{X} + \frac{a}{b} .$$

Thus we have

Theorem 4.3.2 : Let $\hat{Y}$ and $\hat{X}$ be unbiased estimators of Y and X respectively based on simple random sampling.

Then, under the super-population model Δ_g'' the transformed ratio estimator

$$\hat{Y}_{R_T} = \frac{\hat{Y}\, X'}{\hat{X}'}$$

where $\underset{=}{X}$ is transformed to $\underset{=}{X} + \frac{a}{b} = \underset{=}{X}'$, is almost unbiased and has the minimum m.s.e. provided $C_o(g)$, the coefficient of variation of $\bar{Y}$ given X_i's, is small.

Remark : The problem of simultaneously reducing the bias and m.s.e. of $\hat{Y}_R$ was considered by Mohanty and Das (1971) assuming that the relationship between $\underset{=}{Y}$ and $\underset{=}{X}$ is of the form $Y_i = a + b\, X_i$ for $i = 1,2,\ldots,N$. Clearly the results of Mohanty and Das follow as a special case of the above theorem 4.3.2.

4.5 Extension to multivariate transformed ratio estimator

We have already seen in chapter III that the usual ratio and product methods of estimation which use information on a single supplementary variate, will provide efficient estimators in certain situations commonly met with in practice. In many surveys we collect information on more than one auxiliary, variate and some of these may be highly correlated with the study variate under consideration. Olkin (1958) considered an estimator which utilizes information on several auxiliary variates which are positively correlated with the variate under study. He used a linear combination of ratio estimators based on each auxiliary variate separately. Since then several authors, particularly Srivastava (1965), Des Raj (1965), Singh (1967), Rao, P.S.R.S. and Mudholkar (1967), considered estimators which are linear combinations of several estimators based on each auxiliary variate separately. Further, Srivastava (1971) has obtained a class of generalised multivariate ratio estimators whose m.s.e. to the second degree of approximation is same as that of the multivariate regression estimator.

Here we study an extension of the suggested transformed ratio estimator of multi-supplementary variates (which may be either positively or negatively correlated with the study variate under consideration) defined as

$$\hat{Y}_{MR} = \hat{Y} \prod_{i=1}^{m} \frac{X_{(i)}}{X_{(i)} + \theta_i (\hat{X}_{(i)} - X_{(i)})} \qquad (4.5.1)$$

where $\hat{Y}$ is an unbiased estimator of Y, $\hat{X}_{(i)}$ is an unbiased estimator of $X_{(i)}$ (the total of i-th supplementary variate) for $i = 1, 2, \ldots, m$. We assume that $|\theta_i e_i| < 1$ for $i = 1, 2, \ldots, m$ where $e_i = \frac{\hat{X}_{(i)} - X_{(i)}}{X_{(i)}}$. Now we compare the bias of this estimator with that of Srivastava's (1971) generalised multivariate ratio estimator of the form

$$\hat{Y}_{GR} = \hat{Y} \prod_{i=1}^{m} \left(\frac{X_{(i)}}{\hat{X}_{(i)}} \right)^{r_i} . \qquad (4.5.2)$$

t is easy to establish that the $(\theta_1, \theta_2, \ldots, \theta_m)$ which minimizes $M(\hat{Y}_{MR})$ is same as the $(r_1, r_2, \ldots, r_m)$ which minimises $M(\hat{Y}_{GR})$ and

$$\text{minimum } M(\hat{Y}_{GR}) = \text{minimum } M(\hat{Y}_{MR}).$$

Also we can easily see that

$$B(\hat{Y}_{GR}) = B(\hat{Y}_{MR}) + Y \sum_{i=1}^{m} r_i(1-r_i)\, C_{x_{(i)}} \qquad (4.5.3)$$

where $C_{x_{(i)}}$ is the coefficient of variation of $\hat{X}_{(i)}$.

If ρ_{ij}, the correlation coefficient between $\hat{X}_{(i)}$ and $\hat{X}_{(j)}$ is zero for $i \neq j$, and $i, j = 1, 2, \ldots, m$, then

$$\text{optimum } \theta_i = \text{optimum } r_i = k_i, \text{ where } k_i = \frac{\rho_{0i}\, C_y}{C_{x(i)}},$$

ρ_{0i} is the correlation coefficient between $\hat{Y}$ and $\hat{X}_{(i)}$ and C_y and $C_{x(i)}$ respectively denote the coefficients of variation f $\hat{Y}$ and $\hat{X}_{(i)}$. Also then

$$B(\hat{Y}_{MR}) = 0. \qquad (4.5.4)$$

It then follows that

$$B(\hat{Y}_{MR}) = 0 \leq |B(\hat{Y}_{GR})|. \qquad (4.5.5)$$

ven if $\rho_{ij} \neq 0$, and is small, it is found in most of the situations met with in practice that

$$|B(\hat{Y}_{MR})| \leq |B(\hat{Y}_{GR})|.$$

This fact can be illustrated by the following empirical example which is considered by Singh (1965).

The data for all 61 blocks of Ahmedabad city ward No. I (Khadia I) taken from 1961 population census have been considered for the purpose of the present study. It is intended to find the total number of female workers (Y). The supplementary variates choosen for this purpose are

(i) $\underline{\underline{X}}_1$, the female population

and (ii) $\underline{\underline{X}}_2$, the female population in services (group IX of the population census).

The following population values were obtained by Singh (1965)

$\bar{Y} = 7.46$	$C_0^2 = 0.5046$	$\rho_{01} = 0.0388$
$\bar{X}_{(1)} = 265.54$	$C_1^2 = 0.0379$	$\rho_{02} = 0.7731$
$\bar{X}_{(2)} = 5.31$	$C_2^2 = 0.5737$	$\rho_{12} = -0.0474$

where $\bar{Y}$, $\bar{X}_{(1)}$ and $\bar{X}_{(2)}$ denote averages for the corresponding variates ($\underline{\underline{Y}}$, $\underline{\underline{X}}_{(1)}$ and $\underline{\underline{X}}_{(2)}$) and C_i^2, $i = 0, 1,$ and 2 stand for the square of the coefficients of variation of the respective variates and $\rho_{ij} (i \neq j,\ i, j = 0, 1, 2)$ for the corresponding correlation coefficients.

Let us suppose that a simple random sample of n blocks is selected from this population to estimate the total number of female workers Y. Then, we obtain opt. $\theta_1 = 0.2759$, opt. $\theta_2 = 0.7282$ and

$$B(\hat{Y}_{MR}) = (0.0014)\ Y,$$

whereas

$$B(\hat{Y}_{GR}) = (0.0620)\ Y.$$

Thus $\hat{Y}_{MR}$ has considerably smaller bias than $\hat{Y}_{GR}$.

CHAPTER V

A STUDY ON THE USE OF PRIOR KNOWLEDGE ON CERTAIN POPULATION PARAMETERS IN ESTIMATION

5.0 Summary

After a brief review of the different methods of estimation using <u>a priori</u> information on the population parameters R, the ratio of Y and X ; β the regression coefficient of $\underline{\underline{Y}}$ on $\underline{\underline{X}}$; k, the ratio of β and R ; and C_y, the coefficient of variation of the unbiassed estimator of Y, where Y and X are respectively the population totals of the study variate $\underline{\underline{Y}}$ and the auxiliary variate $\underline{\underline{X}}$; an analysis is carried out separately to find how far the <u>a priori</u> values of R, β, k and C_y can depart from their respective true values without incurring a substantial loss in precision. Also, the stability of these parameters over time and space has been examined through empirical examples. It is found not only that k is relatively more stable than the other parameters but also that the estimation procedure using the information on k is relatively more efficient. The relative advantages and disadvantages of these methods are discussed and the results are illustrated with empirical examples.

5.1 Introduction

Let $\hat{Y}$ and $\hat{X}$ be unbiased estimators of the population totals of the study variate $\underline{\underline{Y}}$ and the auxiliary variate $\underline{\underline{X}}$

respectively, based on any sample design. For improving $\hat{Y}$ in the case of simple random sampling design Searls (1965) has considered the estimator of the form $\lambda \hat{Y}$ and by minimising the m.s.e. of $\lambda \hat{Y}$ with respect to λ, he has obtained the optimum value of λ as equal to $(1 + C_y^2)^{-1}$ where C_y is the coefficient of variation of the estimator $\hat{Y}$.

When information on a supplementary variate X is available, considering the difference estimator $\hat{Y}_h = \hat{Y} + h(X - \hat{X})$, Des Raj (1965) has shown that the above estimator has the same efficiency as the usual ratio estimator when $h = R$. Also, it is well known that the above difference estimator has minimum variance when $h = \beta$, where $\beta = \frac{Cov(\hat{Y}, \hat{X})}{V(\hat{X})}$; the corresponding estimator is called the optimum difference estimator.

For improving the usual ratio and product estimators we have seen in chapters III and IV that the estimator

$$\frac{\hat{Y} X}{X + k(\hat{X} - X)}$$

is almost unbiased and has the same efficiency as the optimum difference estimator mentioned above.

5.2 Use of knowledge on C_y

Usually the *a priori* information on C_y (if available) is used to determine the sample size for a pre-assigned level of ecision (see Murthy (1967) section 4.3). Searls (1965,67) s proceeded a step further and shown that it can be used in stimation procedure as well, to obtain improved estimators. fact, if a guessed value of C_y is available he has shown hat

$$\hat{Y}^* = \frac{\hat{Y}}{1 + C_y^2}$$

as its m.s.e. equal to

$$\frac{V(\hat{Y})}{(1 + C_y^2)}.$$

Thus the estimator $\hat{Y}^*$ has an efficiency $1 + C_y^2$ over $\hat{Y}$ which is considerably high for populations with large values of C_y. Here the efficiency decreases as the sample size increases for all those designs for which $V(\hat{Y})$ is inversely proportional to the sample size. So, Searls has suggested that it is very much economical and profitable to use $\hat{Y}^*$ only in small samples. While suggesting this, Searls has ignored the magnitude of the bias and the necessity of a reasonably large sample size so that the absolute bias is small (or negligible) either relative to the standard deviation of the estimator or relative to Y. We study this in the following.

Consider Searl's estimator

$$\hat{Y}^* = \lambda\, \hat{Y}, \quad \text{where } \lambda = \frac{1}{1 + C_y^2}.$$

CHAPTER V

A STUDY ON THE USE OF PRIOR KNOWLEDGE ON CERTAIN POPULATION PARAMETERS IN ESTIMATION

0 Summary

After a brief review of the different methods of estima- n using *a priori* information on the population parameters the ratio of Y and X; β the regression coefficient of on $\underline{X}$; k, the ratio of β and R; and C_y, the coefficient variation of the unbiassed estimator of Y, where Y and are respectively the population totals of the study variate and the auxiliary variate $\underline{X}$; an analysis is carried out eparately to find how far the *a priori* values of R, β, k and y can depart from their respective true values without incurr- g a substantial loss in precision. Also, the stability of these arameters over time and space has been examined through empiri- cal examples. It is found not only that k is relatively more stable than the other parameters but also that the estimation procedure using the information on k is relatively more efficient. The relative advantages and disadvantages of these methods are discussed and the results are illustrated with empirical examples.

5.1 Introduction

Let $\hat{Y}$ and $\hat{X}$ be unbiased estimators of the population totals of the study variate $\underline{Y}$ and the auxiliary variate $\underline{X}$

figures in the paranthoses of the above table denote the centage efficiency of $\hat{Y}^*$ over $\hat{Y}$. Thus, we note here that a negligibly small bias in $\hat{Y}^*$ we require a fairly large ple size and if the sample size is fairly large, then the ficiency of $\hat{Y}^*$ over $\hat{Y}$ is negligibly small.

Let us now examine the assumption that the relative solute bias of $\hat{Y}^*$ is small, that is, $\frac{|B(\hat{Y}^*)|}{Y} \leq \alpha$, where is small we then obtain

$$C_y^2 \leq \frac{\alpha}{(1-\alpha)} \, . \qquad (5.2.8)$$

w for the designs for which $C_y^2 = \frac{C_o^2}{n}$, we require that

$$n \geq C_o^2 \frac{(1-\alpha)}{\alpha}$$

$$= 9 \, C_o^2 \quad \text{if} \quad \alpha = 0.1.$$

provide below in the table 5.2.2 the sample sizes required r different values of C_o and α.

TABLE 5.2.2

Sample sizes for different values of C_o and α

	α		
C_o	0.1	0.2	0.25
1	9(111.11)	4(125)	3(133)
2	36(")	16(")	12(")
3	81(")	36(")	27(")
4	144(")	64(")	48(")
5	225(")	100(")	75(")

The figures in parantheses denote the percentage efficiency of $\hat{Y}^*$ over $\hat{Y}$. Even though the efficiency of $\hat{Y}^*$ over $\hat{Y}$ is slightly more than that in the previous case, yet it is not high enough to suggest the use of $\hat{Y}^*$ over $\hat{Y}$.

Effect of the deviations of C_y

Usually C_y is not known, however an accurate guessed value C_y' of C_y may be available either from a pilot study, past data or experience and this can be used to calculate the near-optimum estimator $\lambda_o \hat{Y}$, where $\lambda_o = \frac{1}{1+C_y'^2}$. Let $\lambda_o = \lambda(1 + \varepsilon)$ where $\lambda = \frac{1}{1+C_y^2}$ and ε is the proportion of deviation of λ_o from λ. After some simplification it can be shown that

$$M(\lambda_o \hat{Y}) = M(\lambda \hat{Y}) \left[1 + \frac{\varepsilon^2}{C_y^2}\right]. \qquad (5.2.3)$$

Thus, for the proportional increase in m.s.e. of $\lambda_o \hat{Y}$ over $\lambda \hat{Y}$ to be small, say $\leq \alpha$, where α is small, we must have that

$$|\varepsilon| \leq C_y \sqrt{\alpha} . \qquad (5.2.4)$$

For example, if $C_y = 1$, the proportional increase in m.s.e. of $\lambda_o \hat{Y}$ over $\lambda \hat{Y}$ is less than 10 percent (that is $\alpha = 0.1$) provided $|\varepsilon| < \sqrt{0.1} = 0.3165$. Thus, the above expression (5.2.4) makes it clear that in order to ensure a small roportional increase in m.s.e., $|\varepsilon|$ must be close to zero if is considerably small but can depart substantially from ero if C_y is moderately high.

.3 Use of knowledge on R and β

For improving the usual ratio estimator $\hat{Y}_R = \frac{\hat{Y} X}{X}$, es Raj (1965) has considered the difference estimator

$$\hat{Y}_h = \hat{Y} + h(X - \hat{X})$$

and has shown that this estimator has the same efficiency as that of the ratio estimator when h = R. Thus, if a fairly accurate information on R is available in advance, it is possible to make use of this in the estimation procedure to get better estimators. The advantages of $\hat{Y}_h$ when h = R over the usual ratio estimator are that it is unbiased, the ariance expression is exact and an unbiased estimate of the variance could be easily obtained. Since the variance of $\hat{Y}_h$ is minimum when h = B, the estimator using information on R is usually less efficient than the optimum difference estimator.

Effect of the deviations of R and β

Let R_o and β_o be the guessed values of R and β respectively. Let $\beta_o = \beta(1 + \varepsilon)$. Then from Cochran (196 3), section 7.2, we obtain that

$$V(\hat{Y}_{\beta_o}) = V(\hat{Y})\ (1 - \rho^2)\ [1 + \frac{\varepsilon^2\ \rho^2}{(1- \rho^2)}]. \qquad (5.3.1)$$

Thus, the proportional increase in variance of $\hat{Y}_{\beta_0}$ over $\hat{Y}_\beta$ is less than α provided

$$|\varepsilon| < \sqrt{\frac{(1-\rho^2)\alpha}{\rho^2}} \qquad (5.3.2)$$

Let $h_0 = R_0$ be the guessed value of R that is used in the estimator $\hat{Y}_h$. Now comparing the variance of $\hat{Y}_{h_0}$ with .s.e. of $\hat{Y}_R$ (the ratio estimator) we obtain that

$$\begin{aligned} V(\hat{Y}_{h_0}) - M(\hat{Y}_R) &= [(R_0-\beta)^2 - (R-\beta)^2]\ V(\hat{X}) \\ &= (R_0 - R)(R_0 + R - 2\beta)\ V(\hat{X}) \qquad (5.3.3) \\ &\leq 0 \left\{ \begin{array}{l} \text{for } R_0 \text{ in } [2\beta-R, R] \text{ in case } \beta \leq R \\ \qquad\qquad \text{and} \\ \text{for } R_0 \text{ in } [R, 2\beta-R] \text{ in case } \beta \geq R. \end{array} \right. \end{aligned}$$

Thus, we see from (5.3.3) that even with a guessed value of R_0, it is possible to get better estimates than $\hat{Y}_R$ under certain conditions.

.4 Use of knowledge on k

We have seen in the chapters III and IV that the estimator

$$\hat{Y}_k = \frac{\hat{Y}\,X}{x+k(\hat{X}-X)}$$

is almost unbiased and has the m.s.e. equal to

$$V(\hat{Y})\ (1-\rho^2).$$

us, if information on k is available, it is possible to t an estimator whose efficiency to the second degree of pproximation is nearly same as that of the optimum difference stimator provided the guessed value k_o of k is fairly ccurate. Let $k_o = k(1 + \varepsilon)$ where ε is the proportion of eviation of k_o from the true value k. It is easy to verify hat

$$M(\hat{Y}_{k_o}) = V(\hat{Y})\ (1 - \rho^2)\ [1+\frac{\varepsilon^2\ \rho^2}{(1 - \rho^2)}] . \tag{5.4.1}$$

Thus the proportional increase in m.s.e. of $\hat{Y}_{k_o}$ over $\hat{Y}_k$ is $< \alpha$ if

$$|\varepsilon| < \sqrt{\frac{(1-\rho^2)\ \alpha}{\rho^2}} . \tag{5.4.2}$$

From (5.4.2) it is clear that in order to ensure a small proportional increase in m.s.e. of $\hat{Y}_{k_o}$, $|\varepsilon|$ must be close to zero if $|\rho|$ is very high but can depart substantially from zero if $|\rho|$ is only moderate.

Remark : By comparing the m.s.e. of $\hat{Y}_k$ with that of $\hat{Y}^*$ we obtain that

$$M(\hat{Y}^*) \leq M(\hat{Y}_k) \quad \text{provided} \quad C_y^2 \geq \frac{\rho^2}{(1-\rho^2)} ,$$

which shows that $\hat{Y}^*$ is better than $\hat{Y}_k$ only if C_y is quite high or $|\rho|$ is considerably small.

5 Illustrations of the efficiency of different procedures

Using data on productivity of jute and rice for a few tates in India, we examine below the efficiencies of different rocedures by comparing the m.s.e.'s of different estimators hich use knowledge respectively on R, β, k and C_0 (and sing a simple random sample of size 2 for the estimator $\hat{Y}^*$). he idea behind selecting a small size 2 for $\hat{Y}^*$ is that the fficiency of $\hat{Y}^*$ over $\hat{Y}$ decreases with increase in n n case of simple random sampling scheme (whereas the efficiencies of $\hat{Y}_R$, $\hat{Y}_k$ or $\hat{Y}_\beta$ over $\hat{Y}$ do not depend on n).

TABLE 5.5.1a

Percentage efficiency of different estimators over $\hat{Y}$ for the data on jute production in India over related time periods.

Year	$\hat{Y}$	$\hat{Y}^*$	$\hat{Y}_R$	$\hat{Y}_k$ (or $\hat{Y}_\beta$)
1951-52	100	152	1045	1362
1961-62	100	138	969	1254
1962-63	100	140	962	1110
1963-64	100	140	1082	1411
1964-65	100	137	718	1089
1970-71	100	137	1305	1644

he basic data for tables 5.5.1a and 5.5.1b is given in ables B1 and B2 of the appendix.

TABLE 5.5.1b

Percentage efficiency of different estimators over $\hat{Y}$ for the data on rice production in four states over two time periods.

States	Year	$\hat{Y}$	$\hat{Y}^*$	$\hat{Y}_R$	$\hat{Y}_k$ (or $\hat{Y}_\beta$)
Uttar Pradesh :	1954-55	100	155	522	533
	1964-65	100	130	2039	2058
Maharashtra :	1954-55	100	223	1357	1370
	1964-65	100	212	1250	1304
Tamilnadu :	1954-55	100	129	6439	6969
	1964-65	100	131	1564	1603
West Bengal :	1954-55	100	120	597	665
	1964-65	100	123	2579	2589

5.6 Stability of the population parameters

In the earlier sections we have discussed several estimation procedures which make use of knowledge on certain population parameters in order to obtain improved estimators of population total. This procedure will be quite valid if the population parameters under consideration are stable over time and space. In the cases of many repeated surveys and surveys which are based on multi-phase sampling techniques, information on the same variates is collected on several occasions. Hence, after gaining sufficient experience with the data, it might then be possible to predict accurately in advance the values of certain population parameters. We now examine this aspect of stability of the parameters R, β, k and C_0 using data on Indian agriculture. We utilise the district-wise data (output of jute ($\underline{\underline{Y}}$) and area under production of jute ($\underline{\underline{X}}$)) on production of jute in India for the years 1951-52, 1961-62 to 1964-65 and 1970-71 and district-wise production of rice (with output of rice as $\underline{\underline{Y}}$ and area under production of rice as $\underline{\underline{X}}$) for Uttar Pradesh, Maharashtra, West Bengal and Tamilnadu for the two time periods 1954-55 and 1964-65 to calculate the values of the population parameters R, β, k and C_o with a view to examining the stability of these parameters. The following tables give summary figures of the population values.

TABLE 5.6.1a

Summary table giving population parameters for the data on jute production in India over related time periods.

Year	R	β	k	C_o	ρ	C_1
1951-52	6.0191	5.2107	0.8657	1.0182	0.9626	1.1322
1961-62	6.9260	5.9780	0.8623	0.8759	0.9593	0.9744
1962-63	6.3565	5.6586	0.8992	0.8970	0.9539	0.9610
1963-64	7.0110	6.0855	0.8680	0.8968	0.9639	0.9959
1964-65	7.1518	6.1861	0.8650	0.8592	0.9277	0.9215
1970-71	6.5883	5.8303	0.8849	0.8652	0.9691	0.9438

TABLE 5.6.1b

ry table giving population parameters for the data on ction of rice in four states over two time periods.

te		Year	R	β	k	C_0	ρ	C_1
Pradesh	:	1954-55	0.5953	0.6386	1.0728	1.0507	0.9014	0.8828
		1964-65	0.7470	0.7266	0.9727	0.7676	0.9754	0.7697
shtra	:	1954-55	0.9960	1.0136	1.0177	1.5661	0.9628	1.4645
		1964-65	1.0617	1.1062	1.0419	1.4955	0.9609	1.3793
adu	:	1954-55	1.2877	1.2318	0.9566	0.7631	0.9928	0.7920
		1964-65	1.5369	1.6202	1.0542	0.7920	0.9683	0.7275
Bengal	:	1954-55	0.9692	0.7894	0.8144	0.6258	0.9218	0.7083
		1964-65	1.2334	1.2569	1.0191	0.6827	0.9805	0.6568

us from the above table 5.6.1a we see that, of all the parameters, k seems to be remarkably stable over time for the ta on jute production and furthermore from table 5.6.1b we te that k is stable not only over time but also over different regions.

We next use the data on jute production for the year 1951-52 to calculate the *a priori* values of the parameters β, k and λ denoted by β_0, k_0 and λ_0 respectively. Similarly for the data on production of rice we use 1954-55 data to calculate the *a priori* values of the parameters β_0, k_0 and λ_0 for different states. We provide below in the tables 5.6.2a and 5.6.2b the proportion of deviation of β_0, k_0 and λ_0 from the true values β, k and λ respectively for different years. The figures in the brackets denote the proportional rease in m.s.e. of the corresponding estimators due to the iation of the a priori values from the true values.

TABLE 5.6.2a

roportion of deviation of the a priori values from their rue values for the data on jute production for related ime periods.

Year	β	k	λ
1961-62	0.1275(0.1876)	0.0039(0.0002)	0.0472 (0.0058)
1962-63	0.0792(0.0634)	0.0275(0.0071)	0.0403 (0.0040)
1963-64	0.1438(0.2710)	0.0026(0.0001)	0.0403 (0.0040)
1964-65	0.1577(0.1536)	0.0008(0.0000)	0.0528 (0.0073)
1970-71	0.1063(0.1744)	0.0217(0.0073)	0.0407 (0.0044)

TABLE 5.6.2b

Proportion of deviation of the a priori values from their true values for the data on production of rice for different states.

State	β	k	λ
Uttar Pradesh	0.1211(0.2871)	0.1029(0.2073)	0.0937(0.0238)
Maharashtra	0.0837(0.0845)	0.0232(0.0065)	0.0199(0.0004)
Tamilnadu	0.2397(0.8634)	0.0926(0.1289)	0.0105(0.0004)
West Bengal	0.3719(3.4430)	0.2009(1.0047)	0.0217(0.0020)

Judging from the tables 5.5.1a, 5.5.1b, 5.6.1a, 5.6.1b, 5.6.2a and 5.6.2b, we note that the estimation procedure using information on k is superior to all others. Thus, this gives another justification for preferring the estimator

$$\frac{\hat{Y}\ X}{X+k(\hat{X}-X)}$$

to others.

CHAPTER VI

ICIENCY OF RATIO ESTIMATOR UNDER A SUPER POPULATION MODEL

Summary

Consider a finite population $\underline{\underline{U}} = (U_1, U_2, \ldots, U_N)$ of size Let $\underline{\underline{Y}}$ $(\underline{\underline{X}})$ be any study variate (auxilary variate) taking lue $Y_i(X_i)$ on the unit U_i, $i = 1,2,\ldots,N$. Suppose the lationship between $\underline{\underline{Y}}$ and $\underline{\underline{X}}$ is of the form

$$Y_i = a + bX_i + e_i, \quad i = 1,2,\ldots,N,$$

ere a and b are constants and e_i's are random variables lth $\mathcal{E}(e_i|X_i) = 0$, $V(e_i|X_i) = f(X_i)$ and $Cov(e_i, e_j|X_i, X_j) = 0$ r $i \neq j$; $i,j = 1,2,\ldots,N$. For estimating the population tal of $\underline{\underline{Y}}$ based on simple random sampling and using the available supplementary information on $\underline{\underline{X}}$, Des Raj (1954) has studied he efficiency of the conventional ratio estimator $\hat{Y}_R$ with spect to the usual unbiased estimator $\hat{Y}$ by comparing the xact variances of these estimators in the following special ases of the above model :

(1) $f(X_i) = 0$ for all i,

(2) $a = 0$ and $f(X_i) = \sigma^2$ for all i.

A comparison between $\hat{Y}_R$ and $\hat{Y}$ is made in this chapter by studying the expected variances of these estimators under the super-population model when $a = 0$ and $f(X_i) = \sigma^2 X_i^g$, and it is found that there exists a g' in (0,1) such that $\hat{Y}_R$

superior to $\hat{Y}$ for all $g \geq g'$. For the case when $g < g'$, simple condition is provided for $\hat{Y}_R$ to be superior to $\hat{Y}$.

The results are illustrated by an empirical example. lso, a comparison is made between combined and separate ratio stimators in stratified random sampling under the above superopulation model.

1 Introduction

Let $\underline{\underline{Y}}$ be any study variate taking value Y_i on the opulation unit U_i for $i = 1, 2, \ldots, N$. Whenever some suitable uxiliary information is available on a supplementary variate taking value X_i on the unit U_i, $i = 1, 2, \ldots, N$, a number of ethods have been proposed in sampling theory to improve upon the usual unbiased estimator of the population total $Y = \sum_{i=1}^{N} Y_i$. Of these, the ratio method of estimation has been widely used whenever the correlation coefficient ρ between $\underline{\underline{Y}}$ and $\underline{\underline{X}}$ is positive and high. To be specific, let us note that (see section 1 of chapter III)

$$M(\hat{Y}_R) \leq V(\hat{Y}) \text{ iff } k = \rho \frac{C_y}{C_x} \geq \frac{1}{2},$$

where $\hat{Y}_R$ and $\hat{Y}$ are respectively the ratio estimator and the usual unbiased estimator of Y based on any sample design, ρ is the coefficient of correlation between $\hat{Y}$ and $\hat{X}$; and C_y and C_x are respectively the coefficients of variation of nd $\hat{X}$. Thus $\hat{Y}_R$ is superior to $\hat{Y}$ if and only if

$$\rho \geq \frac{C_x}{2C_y} . \tag{6.1.1}$$

d on the above criterion (6.1.1) the sampler would tend to l that $\hat{Y}_R$ is preferable to $\hat{Y}$ if the correlation coeffi- nt between $\hat{Y}$ and $\hat{X}$ is quite high. With a view to examin- this aspect a little further Des Raj (1954) has compared efficiency of ratio estimator with the simple unbiased esti- tor and has obtained the following results.

orem 6.1.1 (Des Raj (1954) : Let $\hat{Y}$ and $\hat{X}$ be unbiased timators of Y and X respectively based on simple random pling without replacement design. Assume that there is a rfect linear relation between $\underline{\underline{Y}}$ and $\underline{\underline{X}}$, say $Y_i = a + bX_i$, r $i = 1,2,\ldots,N$, where a and b are constants. Then will be preferable to $\hat{Y}_R$ if

$$\frac{X^2 \; V(\frac{1}{\hat{X}})}{S_x^2} > \frac{b^2}{a^2} \; \frac{(N-n)}{Nn} ,$$

here $S_x^2 = \frac{1}{(N-1)} \sum\limits_{i=1}^{N} (X_i - \bar{X})^2 .$

heorem 6.1.2 (Des Raj (1954) : Let the relationship between and $\underline{\underline{X}}$ be such that $\mathcal{E}(Y_i|X_i) = a + bX_i$, $\mathcal{V}(Y_i|X_i) = \sigma^2$ $S_y^2(1 - \rho^2)$ and Cov $(Y_i, Y_j | X_i, X_j) = 0$ for $i \neq j$ and

= 1,2,...,N, where $S_y^2 = \frac{1}{(N-1)} \sum_{i=1}^{N} (Y_i - \bar{Y})^2$ and ρ is the relation coefficient between $\underset{=}{Y}$ and $\underset{=}{X}$. Then, in the case of ple random sampling without replacement design, $\hat{Y}_R$ will be ferable to $\hat{Y}$ if $\rho^2 > 1 - \frac{1}{X^2\, E(\frac{1}{\hat{X}^2})}$, where $\hat{Y}$ and $\hat{X}$ unbiased estimators of Y and X respectively.

From the above two theorems 6.1.1 and 6.1.2 it is clear that will be superior to $\hat{Y}_R$ if the regression line is far away m the origin. Even if the regression line passes through the gin the choice does not depend solely on ρ. This fact can be ed from the populations 1 and 2 mentioned in chapter III.

) Efficiency of ratio estimator

First, we prove the following lemma.

ma 6.2.1 : Suppose X_i, i = 1,2,...,N are non-negative. Then

$$m(g) = \frac{\sum_{i=1}^{N} X_i^g (X_i - \bar{X})}{\sum_{i=1}^{N} X_i^g}$$

an increasing function of g.

oof :

$$\frac{dm(g)}{dg} = \frac{\sum_{i=1}^{N} X_i^g \sum_{j=1}^{N} X_j^g (X_j - \bar{X}) \log X_j - \sum_{i=1}^{N} X_i^g (X_i - \bar{X}) \sum_{j=1}^{N} X_j^g \log X_j}{(\sum_{i=1}^{N} X_i^g)^2}$$

$$e.,\ (\sum_{i=1}^{N} X_i^g)^2 \frac{dm(g)}{dg} = \sum_{i=1}^{N} \sum_{j=1}^{N} X_i^g X_j^g (X_j - X_i) \log X_j$$

$$\frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} X_i^g X_j^g (X_j - X_i) (\log X_j - \log X_i)$$

$$> 0,$$

hich it follows that $m(g)$ is an increasing function of g.

m 6.2.1 : Let $\hat{Y}$ and $\hat{X}$ be unbiased estimators of Y I respectively based on simple random sampling without ement scheme of size n. Let $\hat{Y}_R$ be the usual ratio mator of Y. Then under the super-population model Δ_g exists a g' in $(0,1)$ such that

$$\mathcal{E}_\delta M(\hat{Y}_R) < \mathcal{E}_\delta V(\hat{Y})$$

ded either $g > g'$ or $g \leq g'$ and $\dfrac{\sigma^2 \sum_{i=1}^{N} X_i^g}{b^2 \quad X^2} < 1,$

Δ_g denotes the class of prior distributions δ ying

$$\mathcal{E}_\delta (Y_i | X_i) = b X_i,$$

$$\mathcal{V}_\delta (Y_i | X_i) = \sigma^2 X_i^g \qquad \text{and}$$

$$\text{Cov}_\delta (Y_i, Y_j | X_i, X_j) = 0 \quad \text{for} \quad i \neq j$$

: Under the super-population model Δ_g it can be shown

$$\mathcal{E}_\delta V(\hat{Y}) = \frac{(N-n)}{n} (N\ b^2\ S_x^2 + \sigma^2 \sum_{i=1}^{N} X_i^g), \tag{6.2.1}$$

$$\mathcal{E}_\delta M(\hat{Y}_R) = \frac{(N-n)}{Nn} \sigma^2 \sum_{i=1}^{N} X_i^g \{N + C_x^2(1-2k^{(g)})\}] \tag{6.2.2}$$

$$s_x^2 = \frac{1}{(N-1)} \sum_{i=1}^{N} (X_i-\bar{X})^2, \quad C_x^2 = \frac{S_x^2}{\bar{X}^2} \quad \text{and} \quad k^{(g)} = \frac{X \sum_{i=1}^{N} X_i^g (X_i-\bar{X})}{(N-1)\, S_x^2 \sum_{i=1}^{N} X_i^g} .$$

(.2.1) and (6.2.2) we get that

$$\xi_\delta V(\hat{Y}) - \xi_\delta M(\hat{Y}_R)$$

$$= \frac{(N-n)}{Nn} C_x^2 [b^2 X^2 + \sigma^2 \sum_{i=1}^{N} X_i^g (2k^{(g)} - 1)]$$

$$= \frac{(N-n)\, C_x^2}{Nn\, C^2(g)} \sigma^2 \sum_{i=1}^{N} X_i^g [1 + C^2(g) (2k^{(g)} - 1)] \qquad (6.2.3)$$

$C^2(g) = \frac{\sigma^2 \sum_{i=1}^{N} X_i^g}{b^2 X^2}$, is the square of the coefficient of

tion of $\bar{Y}$ given X_i's under the super-population model Δ_g.

we have $k^{(g)} = \frac{X\, m(g)}{(N-1)\, S_x^2} = \begin{cases} 0 & \text{at} \quad g = 0, \\ 1 & \text{at} \quad g = 1, \end{cases}$

from the lemma 6.2.1 it follows that $k^{(g)}$ is an increasing tion of g. Hence there exists a value g' in $(0, 1)$ such $2k^{(g')} = 1$. We distinguish two cases :

i) $\underline{g > g'}$:

Here, from (6.2.3) we get that

$$\xi_\delta V(\hat{Y}) > \xi_\delta M(\hat{Y}_R).$$

ii) $g \leq g'$:

In this case $(2k^{(g)} - 1) \geq -1$ and hence we obtain that

$$\mathscr{E}_\delta \, V(\hat{Y}) > \mathscr{E}_\delta \, M(\hat{Y}_R)$$

lded $C^2(g) < 1$. Hence the theorem follows.

lor $C^2(g) = \frac{\sigma^2 \sum_{i=1}^{N} X_i^g}{b^2 X^2} < 1$, for given X_i's and g we

ld have that

$$\frac{b^2}{\sigma^2} > \frac{\sum_{i=1}^{N} X_i^g}{X^2} . \tag{6.2.4}$$

g = 0, the above condition (6.2.4) reduces to

$$\frac{\rho^2 S_y^2}{S_x^2 S_y^2(1 - \rho^2)} > \frac{1}{N \bar{\bar{X}}^2}$$

$$\frac{\rho^2}{1-\rho^2} > \frac{C_X^2}{N} ,$$

it is easily satisfied when $\rho^2 > 0.5$ (since in this case

$> 1 > \frac{C_X^2}{N}$).

The efficiency of $\hat{Y}_R$ over $\hat{Y}$ (in the sense of expected

lance) is illustrated below by considering the data on 64

ge cities in United States (See Cochran (1963) p. 92), where

is the population total for the year 1930 and $\underline{\underline{X}}$ denotes he same for the year 1920. In case of simple random sampling he values

$$\frac{\sum_{g} V(\hat{Y})}{\sum_{g} M(\hat{Y}_R)}$$

r different values of g and $\frac{b^2}{\sigma^2}$ are presented in table 2.1.

TABLE 6.2.1

The efficiency of $\hat{Y}_R$ over $\hat{Y}$

	$\frac{b^2}{\sigma^2}$							
g	0.01	0.05	0.10	0.50	1.00	2.00	5.00	10.00
00	330	1646	3298	16455	32909	65817	164540	329079
10	194	965	1929	9640	19278	38555	96387	192774
20	114	564	1126	5636	11253	22505	56266	112520
30	66	328	655	3273	6544	13087	32716	65431
40	39	191	380	1896	3791	7581	18952	37902
50	23	110	220	1094	2188	4375	10934	21869
60	14	64	127	629	1258	2514	6284	12567
70	8	37	73	361	720	1439	3596	7191
80	5	22	42	206	411	821	2050	4098
90	3	13	24	117	234	466	1164	2326
00	2	8	14	67	132	264	658	1315

e also obtain

$$k^{(g)} = \begin{cases} 0.45824 & \text{at } g = 0.5 \\ 0.56163 & \text{at } g = 0.6 \end{cases}$$

nce g' lies between 0.5 and 0.6. The basic data for the table 2.1 is given in the table B' of the appendix.

ark (a) : Suppose $Y_i = a + b X_i$, and $Y_i, X_i \geq 0$, for $= 1,2,\ldots,N$. Let $\hat{Y}$ and $\hat{X}$ be unbiased estimators of Y d X respectively based on any sampling design and let the rrelation coefficient between $\hat{Y}$ and $\hat{X}$ be positive.

aking m.s.e. of $\hat{Y}_R$ up to the second degree of approximation e can obtain from (6.1.1) that $\hat{Y}_R$ will be less efficient han $\hat{Y}$ if

$$\rho < \frac{C_x}{2C_y}$$

$$= \frac{(a+b\,\bar{X})}{2b\,\bar{X}} \; . \qquad (6.2.5)$$

learly, the right hand expression of (6.2.5) will be > 1 if > b $\bar{X}$. Hence, it follows that $\hat{Y}_R$ will be less efficient han $\hat{Y}$ if a > b $\bar{X}$.

<u>emark (b)</u> : The comparison between $\hat{Y}_R$ and $\hat{Y}$ under the uper-population model Δ_g with $\mathcal{V}_{\delta}\,(Y_i|X_i) = \sigma^2$ (see Des Raj 1954)) follows easily from our theorem 6.2.1.

.3 Combined and separate ratio estimators in stratified sampling

Let Y_{ij} and X_{ij} denote respectively the $\underline{\underline{Y}}$ and $\underline{\underline{X}}$ alues of the j-th unit of the i-th stratum in the population or $j = 1, 2, \ldots, N_i$; $i = 1, 2, \ldots, h$. Let $\hat{Y}_i$ and $\hat{X}_i$ denote spectively the unbiased estimators of Y_i and X_i, the totals f $\underline{\underline{Y}}$ and $\underline{\underline{X}}$ variates for the i-th stratum, based on a sample f size n_i from the i-th stratum. Let $Y = \sum_{i=1}^{h} Y_i$ and $= \sum_{i=1}^{h} X_i$.

orem. 6.3.1 : Let $\hat{Y}_{R,C} = \dfrac{\sum_{i=1}^{h} \hat{Y}_i}{\sum_{i=1}^{h} \hat{X}_i} X$ and $\hat{Y}_{R,S} = \sum_{i=1}^{h} \dfrac{\hat{Y}_i}{\hat{X}_i} X_i$ the combined and separate ratio estimators of Y respectively sed on stratified sampling with any sample design in each ratum. Then, to the second degree of approximation

$$\left.\begin{array}{l} B(\hat{Y}_{R,C}) = 0 \leq |B(\hat{Y}_{R,S})| \\ M(\hat{Y}_{R,C}) = M(\hat{Y}_{R,S}) \end{array}\right\rangle \quad \text{provided } R_i\text{'s equal}$$

$$\left.\begin{array}{l} B(\hat{Y}_{R,S}) = 0 \leq |B(\hat{Y}_{R,C})| \\ M(\hat{Y}_{R,S}) \leq M(\hat{Y}_{R,C}) \end{array}\right\rangle \quad \text{provided } k_i = 1 \text{ for all } i$$

ere $k_i = \dfrac{Cov(\hat{Y}_i, \hat{X}_i)}{V(\hat{X}_i)} \dfrac{X_i}{Y_i}$.

The proof of this theorem is easy and hence is omitted. om the above theorem it is clear that $\hat{Y}_{R,C}$ will be prefer-le to $\hat{Y}_{R,S}$ whenever the strata ratios R_i's do not differ gnificantly from one another, whereas $\hat{Y}_{R,S}$ is preferable to $_C$ whenever the individual regression lines in each stratum ss through a point close to the origin.

orem 6.3.2 : Let $\hat{Y}_i$ and $\hat{X}_i$ be unbiased estimators of pulation totals Y_i and X_i of the i-th stratum respectively sed on a simple random sample (without replacement) of size n_i

drawn from the i-th stratum, $i = 1, 2, \ldots, h$. Let $\hat{Y}_{R,C} = \frac{\sum_{i=1}^{h} \hat{Y}_i}{\sum_{i=1}^{h} \hat{X}_i} X$

and $\hat{Y}_{R,S} = \sum_{i=1}^{h} \frac{\hat{Y}_i}{\hat{X}_i} X_i$ where $\sum_{i=1}^{h} X_i = X$. Then, under the super-population model Δ_g of (2.22) there exists a g_0 in (0,1) such that

$$\mathcal{E}_\delta\, M(\hat{Y}_{R,S}) < \mathcal{E}_\delta\, M(\hat{Y}_{R,C}) \quad \text{for} \quad g \geq g_0 .$$

Proof : It is well known (see Murthy (1967) pp. 376-377) that

$$M(\hat{Y}_{R,C}) = \sum_{i=1}^{h} \frac{N_i^2 f_i}{n_i} (S_{yi}^2 - 2RS_{yxi} + R^2 S_{xi}^2)$$

and

$$M(\hat{Y}_{R,S}) = \sum_{i=1}^{h} \frac{N_i^2 f_i}{n_i} (S_{yi}^2 - 2R_i\, S_{yxi} + R_i^2\, S_{xi}^2)$$

where $f_i = 1 - \frac{n_i}{N_i}$, $R_i = \frac{Y_i}{X_i}$, $R = \frac{Y}{X}$,

$S_{yxi} = \frac{1}{(N_i - 1)} \sum_{j=1}^{N_i} (Y_{ij} - \bar{Y}_i)(X_{ij} - \bar{X}_i)$, $S_{yi}^2 = S_{yyi}$ and $S_{xi}^2 = S_{xxi}$.

Under the super-population model Δ_g, we get

$$\mathcal{E}_\delta\, M(\hat{Y}_{R,S}) = \sigma^2 \sum_{i=1}^{h} \frac{N_i f_i}{n_i} \sum_{j=1}^{N_i} X_{ij}^g \left[1 + \frac{C_{xi}^2}{N_i} (1 - 2k_i^{(g)})\right] \tag{6.3.1}$$

where $C_{xi}^2 = \frac{S_{xi}^2}{\bar{X}_i^2}$ and $k_i^{(g)} = \frac{X_i \sum_{j=1}^{N_i} X_{ij}^g (X_{ij} - \bar{X}_i)}{(N_i - 1)\, S_{xi}^2 \sum_{j=1}^{N_i} X_{ij}^g}$.

$$r.\ \varepsilon\, M(\hat{Y}_{R,C}) = \sigma^2 \sum_{i=1}^{h} \frac{N_i f_i}{n_i} \sum_{j=1}^{N_i} X_{ij}^{g} \left[1 - \frac{2C_{xi}^2}{N_i}\frac{X_i}{X} + \frac{C_{xi}^2 X_i^2 \sum_{i=1}^{h} \sum_{j=1}^{N_i} X_{ij}^{g}}{N_i X^2 \sum_{j=1}^{N_i} X_{ij}^{g}}\right]$$

$$= \sigma^2 \sum_{i=1}^{h} N_i\ f_i \sum_{j=1}^{N_i} X_{ij}^{g} \left[1 - \frac{2C_{xi}^2 X_i}{N_i\ X} + \frac{C_{xi}^2 X_i^2}{N_i\ X^2}(1 + D_i(g))\right] \qquad (6.3.2)$$

ere $D_i(g) = \dfrac{\sum\limits_{\substack{i'\neq i\\=1}}^{h} \sum\limits_{j=1}^{N_{i'}} X_{i'j}^{g}}{\sum\limits_{j=1}^{N_i} X_{ij}^{g}} > 0.$

w, from (6.3.1) and (6.3.2), after simplification we get

$$M(\hat{Y}_{R,S}) - \xi_\delta\, M(\hat{Y}_{R,C}) < \sigma^2 \sum_{i=1}^{h} \frac{f_i}{n_i} C_{xi}^2 \sum_{j=1}^{N_i} X_{ij}^{g}\left(1 - \frac{X_i}{X}\right)\left(1 + \frac{X_i}{X} - 2k_i^{(g)}\right). \qquad (6.3.3)$$

om the lemma 6.2.1 it follows that there exists value g_i

n (0, 1) such that

$2\,k_i^{(g_i)} = 1 + \frac{X_i}{X}$, and $2\ k_i^{(g_i)} > 1 + \frac{X_i}{X}$ for $g > g_i$.

et $g_o = \underset{1\leq i\leq h}{\mathrm{Max}}\ g_i$. Then

$$\xi_\delta\, M(\hat{Y}_{R,S}) < \xi_\delta M(\hat{Y}_{R,C}) \quad \text{for} \quad g \geq g_o .$$

ence, the proof of the theorem.

Since g lies between 1 and 2 in many practical situations Jesson (1942), Mahalanobis (1944), Smith (1938)) we note that $_{R,S}$ is superior to $\hat{Y}_{R,C}$ (in the sense of expected variance) in ost of the cases.

CHAPTER VII

MODIFIED PPS METHOD OF ESTIMATION

mmary

nder certain circumstances it is well known that the ility proportional to size (PPS) method of estimation es highly efficient estimators. However, it was noted Raj (1954) that PPS method of estimation in case of scheme turns out to be inefficient compared to the simple ed estimator based on the same scheme if the regression f $\underline{\underline{Y}}$ on $\underline{\underline{X}}$ is far from the origin. In this case we t a PPS estimator with a transformation on the auxiliary e. The transformed auxiliary variate uses information population parameter $k = \frac{\beta}{R}$ where β is the regression cient of $\underline{\underline{Y}}$ on $\underline{\underline{X}}$. The suggested PPS estimator is found highly efficient in most of the situations met with in ce. The empirical efficiency of the suggested PPS estima- s compared with those of Horvitz-Thompson estimator and trized Des Raj estimator. The results are extended to the when $\underline{\underline{Y}}$ and $\underline{\underline{X}}$ are negatively correlated and are illustra- ith empirical examples. Finally, a comparison between the stimators which make use of information on k, one based on ithout replacement and the other based on PPS with replace- design, is made by assuming a linear finite-population model.

Introduction

Consider a finite population $\underline{\underline{U}} = (U_1, U_2, \ldots, U_N)$ of size et $\underline{\underline{Y}}$ be the study variate taking value Y_1 on the unit

, $i = 1,2,\ldots,N$. Suppose we are interested in estimating he population total $Y = \sum_{i=1}^{N} Y_i$. Whenever information on a pplementary variate $\underset{=}{X}$ taking the value X_i on the unit , $i = 1,2,\ldots,N$ (or prior information on certain population arameters) is available, there are a number of methods sugges-ed in the theory of sampling to improve upon the simple unbiased stimator of the population total. In fact the major uses of uch information may be classified as follows :

(a) in obtaining the estimation formulae

(b) for determining the probabilities of selection of the sampling units

(c) to suitably arrange the units of the population for the purpose of sample selection

(d) for stratifying the population and

(e) in allocation problems.

Combinations of (d) with any of (a) to (c) and (e) are also common in survey sampling. We have already discussed in detail the utilisation of auxiliary information to obtain the estimation formulae in chapters III to V. We discuss the methods of using the auxiliary information as in (b) above in this chapter.

7.2 PPS method of estimation

It is well known that the probability proportional to size (PPS) sampling consists of selecting units in the sample with probability proportional to a given measure of size, where the size measure is usually the value of an auxiliary variate $\underline{\underline{X}}$ which is presumably having high degree of correlation with the variate $\underline{\underline{Y}}$ under study. Suppose a sample $(u_1, u_2, \ldots, u_n)$ of size n is selected using PPS with replacement sampling scheme. Then an unbiased estimator of the population total $Y = \sum_{i=1}^{N} Y_i$ of the study variate $\underline{\underline{Y}}$ and its variance are given by

$$\left.\begin{aligned} \hat{Y}_{PPS} &= \frac{X}{n} \sum_{i=1}^{n} \frac{y_i}{x_i} \\ \text{and} \qquad V(\hat{Y}_{PPS}) &= \frac{1}{n} \left[\sum_{i=1}^{N} \frac{Y_i^2}{X_i} X - Y^2 \right] \end{aligned}\right\} \tag{7.2.1}$$

where $X = \sum_{i=1}^{N} X_i$. Also, it is well known that if the sample $u = (u_1, u_2, \ldots, u_n)$ is selected by using simple random sampling with replacement scheme, then

$$\left.\begin{aligned} \hat{Y} &= \frac{N}{n} \sum_{i=1}^{n} y_i \\ \text{and} \qquad V(\hat{Y}) &= \frac{N}{n} \left[\sum_{i=1}^{N} Y_i^2 - N \bar{Y}^2 \right] \end{aligned}\right\} \tag{7.2.2}$$

is clear from (7.2.1) that $V(\hat{Y}_{PPS})$ reduces to zero when Y_i
X_i are exactly proportional (which may not be the case
illy). Des Raj (1954) has stressed that the success of PPS
pling depends heavily on how appropriate the measure of size
In fact he has proved the following :

orem 7.2.1 : Let $\hat{Y} = \frac{N}{n} \sum_{i=1}^{n} y_i$ be the usual unbiased
imator of the population total Y based on a simple random
mple (with replacement) of size n. Similarly, let
$_{PS} = X \sum_{i=1}^{n} \frac{y_i}{x_i}$ be the usual unbiased estimator of the popula-
on total Y based on a PPS with replacement sample of size
Then

$$V(\hat{Y}_{PPS}) < V(\hat{Y}) \quad \text{provided} \quad \sum_{i=1}^{N} (X_i - \bar{X}) \frac{Y_i^2}{X_i} \geq 0. \qquad (7.2.3)$$

sing the above condition Des Raj has derived the following

heorem 7.2.2 : Let there be a linear relation between $\underline{\underline{Y}}$ and
say $Y_i = a + b X_i$, $i = 1, 2, \ldots, N$. Then the estimator of
$= \sum_{i=1}^{N} Y_i$ based on the PPS with replacement scheme will be
less efficient than the estimator based on the SRS with replace-
ment sampling scheme provided

$$\frac{\bar{X} - \tilde{X}}{\bar{X}\,\sigma_X^2} > \frac{b^2}{a^2} \qquad (7.2.4)$$

here $\tilde{X} = \frac{N}{\sum\limits_{i=1}^{N} \frac{1}{X_i}}$ and $\sigma_x^2 = \frac{1}{N} \sum\limits_{i=1}^{N} (X_i - \bar{X})^2$.

t is easy to see that the condition (7.2.2) is likely to be atisfied if a is sufficiently large (i.e., if the regression ine is very far from the origin). It was also noted by Des Raj hat the conditions (7.2.3) and (7.2.4) are difficult to verify n practice. Also, Des Raj (1958) has compared $\hat{Y}_{PPS}$ with $\hat{Y}$ nder a linear regression model.

3 PPS estimation with a transformed size measure

We now consider an alternative measure of size and develop formulae for estimation of the population total in various situations. Let $\underline{\underline{X}}' = \underline{\underline{X}} + \frac{d}{N}$ so that $X_i' = X_i + \frac{d}{N}$ for $i=1,2,\ldots,N$, here d is any non-negative scalar. Write $\bar{d} = \frac{d}{N}$.

Theorem 7.3.1 : Let $\hat{Y} = N\bar{y}$ be the usual unbiased estimator of Y based on a SRS with replacement scheme of size n. Let

$\hat{Y}_{PPS} = \frac{X}{n} \sum\limits_{i=1}^{n} \frac{y_i}{x_i}$ be the usual unbiased estimator of Y based n a PPS with replacement sample of size n and let

$\hat{Y}'_{PPS} = \frac{X'}{n} \sum\limits_{i=1}^{n} \frac{y_i}{(x_i+\bar{d})}$ be an alternative unbiased estimator of Y based on a PPS with replacement sample of size n, where size measure is the transformed variate X' and

$= \sum_{i=1}^{N} (X_i + \bar{d}) = X + d$. Then,

$$V(\hat{Y}'_{PPS}) < \frac{d}{(X+d)} V(\hat{Y}) + \frac{X}{(X+d)} V(\hat{Y}_{PPS}).$$

roof : It is well known that

$$V(\hat{Y}) = \frac{N}{n} \left(\sum_{i=1}^{N} Y_i^2 - N \bar{Y}^2 \right) \qquad (7.3.1)$$

nd

$$V(\hat{Y}_{PPS}) = \frac{1}{n} \left(\sum_{i=1}^{N} \frac{Y_i^2}{X_i} X - Y^2 \right). \qquad (7.3.2)$$

is also easy to show that

$$V(\hat{Y}'_{PPS}) = \frac{1}{n} \left(\sum_{i=1}^{N} \frac{Y_i^2 (X + d)}{(X_i + \bar{d})} - Y^2 \right). \qquad (7.3.3)$$

om (7.3.1) and (7.3.2) we have

$$V(\hat{Y}_{PPS}) - V(\hat{Y}) = \frac{N}{n} \sum_{i=1}^{N} \frac{Y_i^2 (\bar{X} - X_i)}{X_i} \quad , \qquad (7.3.4)$$

nd from (7.3.2) and (7.3.3) we obtain

$$\hat{Y}'_{PPS}) - V(\hat{Y}_{PPS}) = \frac{1}{n} \left[\sum_{i=1}^{N} Y_i^2 \left(\frac{X + d}{X_i + \bar{d}} - \frac{X}{X_i} \right) \right]$$

$$= \frac{d}{n} \sum_{i=1}^{N} \frac{Y_i^2 (X_i - \bar{X})}{X_i (X_i + \bar{d})}$$

$$= \frac{d}{n} \sum_{i=1}^{N} \frac{Y_i^2}{X_i} \frac{1}{(\bar{X}+\bar{d})} (X_i - \bar{X}) \left[1 - \frac{(X_i - \bar{X})}{(X_i + \bar{d})} \right]$$

$$= \frac{d}{n} \left[\sum_{i=1}^{N} \frac{Y_i^2 (X_i - \bar{X})}{X_i (\bar{X}+\bar{d})} - \sum_{i=1}^{N} \frac{Y_i^2 (X_i - \bar{X})^2}{X_i (X_i + \bar{d}) (\bar{X}+\bar{d})} \right]$$

$$< \frac{d}{n} \sum_{i=1}^{N} \frac{Y_i^2 (X_i - \bar{X})}{X_i (\bar{X} + \bar{d})}, \quad \text{unless all } X_i\text{'s are same.}$$

$$= \frac{d}{X+d} [V(\hat{Y}) - V(\hat{Y}_{PP3})]. \quad \text{(from (7.3.4))}$$

Consequently, $V(\hat{Y}'_{PPS}) < \frac{d}{X+d} V(\hat{Y}) + \frac{X}{X+d} V(\hat{Y}_{PPS})$ and the theorem is proved.

Thus from the above theorem we see that $\hat{Y}'_{PPS}$ is better than the worse of $\hat{Y}$ and $\hat{Y}_{PPS}$. Since d is any non-negative scalar, it is ideal to choose d such that $V(\hat{Y}'_{PPS})$ is minimum. Already we have noticed that there will be a considerable reduction in the variance of $\hat{Y}_{PPS}$ if the regression line of $\underline{\underline{Y}}$ on passes close to the origin. In case, this line does not s through the origin, choose d such that the regression e of Y on the transformed variate $X' = X + \overline{dX}$ passes ugh a point close to the origin. We have noticed in section for the ratio method of estimation that the choice $= \frac{(1-k)}{k}$, where $k = \frac{Cov(\hat{Y}, \hat{X})}{V(\hat{X})} \frac{X}{Y}$ gives efficient estimators. d to be ≥ 0, it is required that $0 < k \leq 1$. For the ice $d^* = \frac{(1-k)}{k}$, denote the corresponding estimator $\hat{Y}'_{PPS}$ Y by $\hat{Y}^*_{PPS}$. Clearly,

$$V(\hat{Y}^*_{PPS}) = \frac{1}{n} \sum_{i=1}^{N} \left(\frac{Y_i^2 \; X}{k X_i + (1-k) \bar{X}} - Y^2 \right) . \qquad (7.3.5)$$

, we have noted in section 5.6 that k is remarkably stable time and space. Hence, using the prior knowledge on k, w study the efficiency of $\hat{Y}^*_{PPS}$ compared to $\hat{Y}_{PPS}$ through ical examples.

Empirical efficiency of $\hat{Y}^*_{PPS}$

The empirical efficiency of $\hat{Y}^*_{PPS}$ is examined in this on by considering the populations 1 and 2 (referred to as 1 in section 3.5) and the following three populations (denoted) considered by Yates and Grundy (1953).

Unit number i	X_i	Population(Y_i) A	B	C
1	0.1	0.5	0.8	0.2
2	0.2	1.2	1.4	0.6
3	0.3	2.1	1.8	0.9
4	0.4	3.2	2.0	0.8

the above mentioned five populations we provide below in summary table, values of some of the population parameters.

SUMMARY TABLE 7.4.1

Values of some population parameters

Pop.	d^*	$n \times V(\hat{Y})$	$n \times V(\hat{Y}_{PPS})$	$n \times V(\hat{Y}^*_{PPS})$
1	1.9266	213×10^8	676×10^8	41×10^8
2	1.2515	213×10^8	287×10^8	30×10^2
A	-0.2000	16.360	1.000	0.198
B	0.5000	3.360	1.000	0.176
C	0.2000	1.150	0.250	0.282

Next we give in the table 7.4.2, the percentage efficiency $\hat{Y}^*_{PPS}$ over $\hat{Y}_{PPS}$ and $\hat{Y}$.

TABLE 7.4.2

Percentage efficiency of $\hat{Y}^*_{PPS}$ over $\hat{Y}_{PPS}$ and $\hat{Y}$

Pop.	$\hat{Y}_{PPS}$	$\hat{Y}$	$\hat{Y}^*_{PPS}$
1	100	318	1644
2	100	133	930
A	100	6	505
B	100	30	568
C	100	22	89

the above, we observe that $\hat{Y}$ is superior to $\hat{Y}_{PPS}$ for lations 1 and 2 because as noticed already, the regression is far away from the origin. But, it is interesting to that for all the above populations except population C, the iency of $\hat{Y}^*_{PPS}$ over $\hat{Y}_{PPS}$ and $\hat{Y}$ is considerably high.

Usually, the value of k will not be available to the ier before the survey. However, a fair and accurate value can be obtained from past experience, data or survey. We have noted in section 5.6 that k is remarkably stable time and space. Hence, a near optimum value of $d^* = \frac{(1-k)}{k}$ be obtained. Now, for various values of d we provide below $V(\hat{Y}^*_{PPS})$ for the first two populations considered in summary 7.4.1.

TABLE 7.4.3

Values of $n \times V(\hat{Y}'_{PPS}) \times 10^{-8}$ for different values of d

d	$n \times V(\hat{Y}'_{PPS}) \times 10^{-8}$	
	Pop. 1	Pop. 2
0	676	282
0.50	144	67
1.00	61	30
1.25*2	-	30
1.50	42	33
1.93*1	41	-
2.00	42	43
3.00	54	65
4.00	70	84
5.00	84	99

$*^1$ and $*^2$ denote respectively the optimum values d* for populations 1 and 2. Thus from the above table we note that $V(\hat{Y}'_{PPS})$ is not very much sensitive to the variations in d.

7.5 Empirical efficiency of different estimators

It is generally observed that sampling without replacement provides more efficient estimators than sampling with replacement, since the effective sample size is more in the former than in the latter. There has been tremendous development in the field of sampling with varying probabilities without replacement since 1950. Especially, the following three important estimators

Horvitz-Thompson estimator (1952), Symmetrized Des Raj mator (Murthy (1957)) and Rao-Hartley-Cochran estimator 2) have been studied at length in the literature and are known for their optimal properties.

itz-Thompson estimator : It is defined as

$$\hat{Y}_{HT} = \sum_{i \in s} \frac{y_i}{\pi_i} ,$$

e y_i is the measure of the variate $\underline{\underline{Y}}$ under study for i-th unit, $i = 1, 2, \ldots, N$, and π_i is the inclusion probabi-y of the i-th unit in the sample of size n.

metrized Des Raj estimator : We explain this estimator for special case of $n = 2$. Let the sample of size 2, taken probability porportional to $\underline{\underline{X}}$-values without replacement, tain the two units u_i and u_j. The estimator of $\underline{\underline{Y}}$ is ined as

$$\hat{Y}_{SD} = \frac{1}{(2-p_i-p_j)} [\frac{y_i}{p_i} (1 - p_j) + \frac{y_j}{p_j} (1 - p_i)]$$

re y_i and y_j denote respectively the $\underline{\underline{Y}}$-values of u_i u_j, and p_i and p_j denote respectively the probabilities selecting the units u_i and u_j in the sample.

-Hartley-Cochran estimator : This estimator can be explained efly as follows :

) Split the population at random into n groups of sizes

$_2,\ldots,N_n$ with

$$N_i = [\frac{N}{n}] + 1 \qquad \text{if} \quad i \leq r$$
$$= [\frac{N}{n}] \qquad \text{if} \quad i > r$$

where r is the remainder when N is divided by n.

From each group select exactly one unit with probability proportional to $\underline{\underline{X}}$-values of that group.

The estimator of Y is defined as

$$\hat{Y}_{RHC} = \sum_{i \epsilon s} \frac{y_i}{(x_i/\phi_i)}$$

pre $\phi_i = \sum_{j \epsilon S^{(i)}} X_j$; s being the sample selected and

$)$ the group to which i-th unit belongs.

far in the literature all these estimators have been studied sing information on $\underline{\underline{X}}$. Using information on $\underline{\underline{X}}^*$ instead $\underline{\underline{X}}$ where $X_i^* = X_i + \frac{(1-k)}{k} \bar{X}$ and $k = \frac{\beta}{\bar{X}}$, we study the pirical efficiency of these estimators compared to the estimator $_{PS}$ (which has been studied in the previous sections) since it ifficult to compare their efficiencies theoretically. For e efficiency comparisons we consider the three populations tes and Grundy (1953) widely used in the literature as being re extreme than situations usually met with in practice. se three populations have been denoted as A, B, C in the eding section of this chapter. The optimum values $d^* = \frac{(1-k)}{k}$ r the three populations A, B and C are respectively -0.2, 5 and 0.2. Using these optimum values in the modified size the variances of different estimators of the population total r a sample of size n = 2 together with their efficiencies pared to $\hat{Y}^*_{PPS}$ are presented below in the table 7.5.1.

TABLE 7.5.1

Efficiencies of different estimators

	Population A		Population B		Population C	
timator	variance	efficiency	variance	efficiency	variance	efficiency
$\hat{Y}^*_{PPS}$	0.099 (0.50)	100	0.088 (0.500)	100	0.141 (0.125)	100
$\hat{Y}^*_{HC}$	0.066 (0.333)	150	0.059 (0.333)	150	0.094 (0.083)	150
$\hat{Y}^*_{HT}$	0.273 (0.823)	36	0.114 (0.057)	77	0.107 (0.059)	132
$\hat{Y}^*_{SD}$	0.052 (0.333)	190	0.059 (0.333)	150	0.089 (0.083)	158

The figures in the parentheses denote the variance of the corresponding estimators using information on $\underline{\underline{X}}$.

will be observed here that all the estimators (which have been nsidered here) using information on $\underline{\underline{X}}^*$ are uniformly superior those using information on $\underline{\underline{X}}$ for population A where as it just the reverse in case of population C. Only Horvitz-mpson estimator which uses information on $\underline{\underline{X}}^*$ is inferior to he same based on $\underset{=}{X}$ for population B. Of all the estimators, ing information on $\underset{=}{X^*}$, the most efficient is the Symmetrized s Raj estimator for all the three populations which have been nsidered here. However, the empirical efficiency is of limited cope and more studies are needed to evaluate the comparative erits of various estimators utilising information on the modified ize $\underset{=}{X^*}$.

6. Comparison under a finite population model

We have discussed at length the various properties of $_k$ in chapters 3,4, and 5. Also, in the section 4 of this hapter we note that the PPS estimator utilizing information n k in the modified size $\underset{=}{X^*}$ is highly superior to the same ased on $\underset{=}{X}$. Further it is well known that Rao-Hartley-Cochran stimator is always superior to the conventional PPS estimator. o, we make a comparison of $\hat{Y}'_{RHC}$ with that of $\hat{Y}_k$ under the ollowing finite population model.

$$\left.\begin{array}{l} Y_i = a + b X_i + \varepsilon_i, \\ \sum_{i=1}^{N} \varepsilon_i = 0 \overset{\sum_{i=1}^{N} \epsilon_i X_i = 0}{\wedge} \text{and } \varepsilon_i^2 = \sigma^2 X_i^g, \end{array}\right\rangle i = 1,2,\ldots,N \qquad (7.6.1)$$

here $\hat{Y}'_{RHC}$ is based on a modified size $\underset{=}{X'}$.

case of SRSWOR sample of size n and from the above finite opulation model (7.6.1) we obtain that

$$M(\hat{Y}_k) = \frac{N^2\sigma^2}{(N-1)}\left(\frac{1}{n}-\frac{1}{N}\right)\sum_{i=1}^{N} X_i^g \qquad (7.6.2)$$

, using Rao-Hartley-Cochran estimator based on a sample of n with a modified size $\underline{\underline{X'}}$ and assuming N is a multiple n, we obtain from (7.6.1) that

$$\hat{Y}'_{RHC}) = \frac{N\sigma^2}{(N-1)}\left(\frac{1}{n}-\frac{1}{N}\right)\sum_{i=1}^{N}\frac{X_i^g\, X(1+d)}{(X_i + d\, X)}. \qquad (7.6.3)$$

It is easy to see that $M(\hat{Y}_k) - V(\hat{Y}'_{RHC})$ is proportional to

$$\sum_{i=1}^{N} X_i^g \left[1 - \frac{\bar{X}(1+d)}{(X_i+d\bar{X})} \right]$$

$$= \sum_{i=1}^{N}\frac{X_i^g(X_i-\bar{X})}{(X_i+d\bar{X})}$$

< 0, whenever $d \geq 0$ and $g \leq 1$ since

$$\frac{X_i^g}{(X_i+d\bar{X})} - \frac{X_j^g}{(X_j+d\bar{X})} = \frac{X_i\, X_j(X_j^{g-1} - X_i^{g-1}) - d\,\bar{X}(X_i^g - X_j^g)}{(X_i+d\,\bar{X})\,(X_j+d\,\bar{X})}$$

≥ 0 whenever $d \geq 0$, $g \leq 1$ and $X_i \leq X_j$.

note here that for the case $d = 0$, the result of Sukhatme Avadhani (1970) follows a special case of the above result.

.7 PPS method of estimation for the case ρ is negative

In the entire foregoing analysis of this chapter it is ssumed that $\underset{=}{Y}$ and $\underset{=}{X}$ are positively correlated. In the mplementary situation where the correlation between $\underset{=}{Y}$ and $\underset{=}{X}$ s negative, the conventional PPS method of estimation may lead erroneous results. In this section we provide a PPS method ith a transformed measure of size) of sampling which gives ficient estimators under certain circumstances when $\underset{=}{Y}$ and are negatively correlated. When $\underset{=}{Y}$ and $\underset{=}{X}$ are negatively rrelated, consider $X_i' = d\,\bar{X} - X_i$, $i = 1, 2, \ldots, N$, where d s such that $d\,\bar{X} > \underset{i}{\text{Max}}\, X_i$. Denote the transformed size by '. It is easy to verify that $\rho_{\underset{=}{YX'}} = -\rho_{\underset{=}{YX}}$. For the choice $d^* = \frac{(k-1)}{k}$ where $k = \frac{\beta}{R}$, we note that the regression line $\underset{=}{Y}$ on $\underset{=}{X^*}$ (where $X_i^* = d^*\,\bar{X} - X_i$) passes through the origin. let

$$\hat{Y}^*_{PPS} = \frac{1}{n} \sum_{i=1}^{n} \frac{y_i\, X}{k\, x_i + (1-k)\bar{X}} \qquad (7.7.1)$$

be the estimation of Y based on a PPSWR sample of size n tilizing the modified size $\underset{=}{X^*}$. It is easy to verify that

$$V(\hat{Y}^*_{PPS}) = \frac{1}{n} \sum_{i=1}^{N} \left(\frac{Y_i^2\, X}{k\, X_i + (1-k)\bar{X}} - Y^2 \right). \qquad (7.7.2)$$

For illustrating the efficiency of $\hat{Y}^*_{PPS}$ compared to $\hat{Y}$ we onsider the following four populations :

lation 1 : It is same as the population 4 referred to in ion 3.5. Populations A, B and C are same as the three s populations with size measure as $\frac{1}{X}$ instead of X.

he above mentioned 4 populations we provide in the summary e 7.7.1, the values of some of the population parameters.

SUMMARY TABLE 7.7.1

Value of some population parameters

ulation	d^*	$n \times V(\hat{Y})$	$n \times V(\hat{Y}^*_{PPS})$
1	1.5260	28633	16106
A	2.0877	16.3600	4.0080
B	2.8571	3.2600	0.1040
C	2.3382	1.1600	0.0567

e following table 7.7.2 we provide the percentage efficiency of over Y for the four populations mentioned above.

TABLE 7.7.2

Percentage efficiency of $\hat{Y}^*_{PPS}$ compared to $\hat{Y}$.

Population	percentage efficiency of $\hat{Y}$	 $\hat{Y}^*_{PPS}$
1	100	178
A	100	408
B	100	3251
C	100	2028

us we note that the efficiency of $\hat{Y}^*_{PPS}$ compared to $\bar{Y}$ is nsiderably high for all these four populations.

Since usually d^* is not known accurately, we provide low for different values of d the corresponding variances r a given sample size n to study the sensitivity of $\hat{Y}^*_{PPS}$) with respect to the variations of d^* for population 1.

TABLE 7.7.3

Values of $n \times V(\hat{Y}'_{PPS})$ for different values of d

Population 1

d	1.35	1.40	1.50	1.5230*	1.75
$\times V(\hat{Y}'_{PPS})$	38169	21217	16400	16108	16501

* indicates the optimum value d^*

us from the above table we note that $V(\hat{Y}'_{PPS})$ is not very ch sensitive to the variations of d.

For the modified size $\underline{\underline{X}}^*$ where $X_i^* = \frac{(k-1)}{k} \bar{X} - X_i$, = 1, 2, ..., N, we now compare the efficiency of $\hat{Y}^*_{PPS}$ th Horvitz-Thompson estimator, Symmetrized Des Raj estimar and Rao-Hartley-Cochran estimator. Since it is difficult mpare their efficiencies theoretically, the variances of e estimators are compared empirically in sampling two units the three hypothetical populations studied by Yates and ov (1953) as these are more extreme than situations usually with in practice. For our illustration here, we have idered $\frac{1}{X}$ as the size variate. For $n = 2$, the variances ifferent estimators together with their efficiencies compared Y^{*}_{PPS} are presented below in table 7.7.4.

TABLE 7.7.4

Efficiencies of different estimators

	Population					
	A		B		C	
ator	Variance	Efficiency	Variance	Efficiency	Variance	Efficiency
PS	1.483	100	0.050	100	0.028	100
[illegible]	0.955	153	0.033	152	0.019	156
[illegible]	1.128	131	0.074	67	0.025	115
[illegible]	0.833	178	0.032	157	0.016	177

from the above table it is interesting to note that Symme- nd Des Raj estimator is uniformly superior to all the other imators considered for all the three populations.

CHAPTER VIII

OPTIMUM POINTS OF STRATIFICATION

ummary

The problem of optimum demarcation of a finite population strata in the case of proportional and optimum allocations imple random sampling with replacement (SRSWR) or without ement (SRSWOR) in each stratum has been examined in this er. It is shown that in stratified SRSWR or SRSWOR with tional allocation it is necessary to arrange the $\underline{\underline{Y}}$-charac- increasing (or decreasing) order of magnitude for strati- ion. Optimum points of demarcation of the strata in the of proportional allocation have been examined. Further- it is shown that in stratified SRSWR or SRSWOR with opti- llocation the condition that the $\underline{\underline{Y}}$-character has to be ged in increasing (or decreasing) order of magnitude may be necessary. However, if the coefficient of variation of Y-character is the same in each stratum, the necessity of ging the $\underline{\underline{Y}}$-character in increasing (or decreasing) order agnitude for stratification is established.

Introduction

It can be easily seen that in the case of SRSWR or SRSWOR variance of the estimate of the population total depends on, t from the sample size, the variability of $\underline{\underline{Y}}$ (the variate r study) in the population. If the population is very heterogeneous and considerations of cost limit the size of the sample, it may be found impossible to get a sufficiently precise estimate by taking a sample using SRSWR or SRSWOR from the entire population. And, populations encountered in practice are generally very heterogeneous. In such situations, it might be economical in some cases to obtain a precise estimate by using stratified random sampling, which consists of classifying the population units into a number of groups called strata and then selecting the samples independently from each group or stratum. An appropriate estimator for the population total is obtained by adding the stratum-wise estimators of the strata totals of the variate Y, under study. The division of the population into strata is termed stratification and is usually done in such a way as to reduce the variability of the strata estimators.

Suppose a population of N units is divided into h strata. Let N_i be the number of units in the i-th stratum and Y_{ij} be the value of the study variate for the j-th unit in the i-th stratum. The population total Y is $\sum_{i=1}^{h} Y_i$ where $Y_i = \sum_{j=1}^{N_i} Y_{ij}$ is the total of Y-variate for the i-th stratum. An unbiased estimator of Y can be obtained by estimating unbiasedly the stratum totals $\{Y_i\}$ on the basis of probability samples drawn independently from each stratum with sampling schemes which need not be the same for all strata. Let $\hat{Y}_i$ be an unbiased imator of Y_i based on a sample of size n_i drawn by using sampling scheme, then an unbiased estimator of Y is given

$$\hat{Y} = \sum_{i=1}^{h} \hat{Y}_i \qquad (8.1.1)$$

its variance is given by

$$V(\hat{Y}) = \sum_{i=1}^{h} V(\hat{Y}_i) . \qquad (8.1.2)$$

ppose b_i is the cost of surveying (or collecting information m) one unit in the i-th stratum. Let

$$C = a_o + \sum_{i=1}^{h} b_i \, n_i \qquad (8.1.3)$$

the total cost of the sample where a_o is the overhead cost. optimum allocation of the total sample $n = \sum_{i=1}^{h} n_i$ to differ- strata is obtained by minimising (8.1.2) with respect to the xed cost C .

If SRSWR is used in each stratum, an unbiased estimator Y is given by

$$\hat{Y}_{st} = \sum_{i=1}^{h} \frac{N_i}{n_i} \sum_{j=1}^{n_i} y_{ij} , \qquad (8.1.4)$$

ere y_{ij} is the value of the j-th sample unit in the i-th ratum.

so, $$V(\hat{Y}_{st}) = \sum_{i=1}^{h} \frac{1}{n_i} \sum_{\substack{j=1 \\ j \neq j'}}^{N_i} \sum_{j'=1}^{N_i} (Y_{ij} - Y_{ij'})^2 . \qquad (8.1.5)$$

hus from (8.1.5) we see that in order to reduce $V(\hat{Y}_{st})$ it is
ecessary to reduce the value of the terms $(Y_{ij} - Y_{ij'})^2$, $j \neq j'$.
his would be achieved by grouping the units to form the strata
such a way that the units belonging to the same stratum are
similar or homogeneous as possible with respect to a strati-
ication variable $\underline{\underline{X}}$, highly (positively or negatively) correla-
ed with $\underline{\underline{Y}}$, since, in practice, the values of $\underline{Y}$ would not be
vailable for stratification purposes.

For a specified allocation of the sample size n, the
ptimum demarcation of strata has been studied earlier by, among
thers, Dalenius (1950, 1952), Dalenius and Gurney (1951),
alenius and Hodges (1957), Sethi (1963), Taga (1967), Isii and
aga (1969) and Singh and Sukhatme (1969, 71, 73). Mahalanobis
1952), Dalenius and Hodges (1957) and Ekman (1959) have sugges-
ted approximations to the theoretical solutions which are easier
apply in practice.

.2 Optimum demarcation of strata-proportional allocation

If SRSWR is used in each stratum, then from (8.1.2) the
riance of the unbiased estimator of the population total,
${}_{st} = \sum_{i=1}^{h} N_i \bar{y}_i$, is given by

$$V(\hat{Y}_{st}) = \sum_{i=1}^{h} \frac{N_i^2 \sigma_i^2}{n_i} \qquad (8.2.1)$$

here $\sigma_i^2 = \frac{1}{N_i} \sum_{j=1}^{N_i} (Y_{ij} - \overline{Y}_i)^2$.

If b_i, the cost of surveying (or collecting information from) ne unit in the i-th stratum is the same for all i, then the ptimum sample size in the i-th stratum is

$$n_{i,opt.} = \frac{n N_i \sigma_i}{\sum_{i=1}^{h} N_i \sigma_i} . \qquad (8.2.2)$$

If σ_i's are equal, then $n_{i,opt.}$ reduces to

$$n_{i,prop.} = \frac{n N_i}{N} . \qquad (8.2.3)$$

In the following summations, i runs over the strata 1,2,...,h and j over the population units 1,2,...,N_i of the i-th stratum unless otherwise specified. In the case of proportional allocation i.e., when $n_i = \frac{n N_i}{N}$, the expression (8.2.1) reduces to

$$\frac{N}{n} \sum_i N_i \sigma_i^2 .$$

$$= \frac{N}{n} [\sum_i \sum_j Y_{ij}^2 - \sum_i \frac{Y_i^2}{N_i}] \qquad (8.2.4)$$

Therefore, minimisation of (8.2.4) is same as maximisation of $\sum_i \frac{Y_i^2}{N_i}$.

In the following we denote by S any stratificatification of a population of size N into h strata with sizes

$N_2,\ldots,N_h$. And for each S we define a gain function as

$$G(S) = \sum_i \frac{Y_i^2}{N_i}$$

d a loss function as

$$L(S) = \sum_i N_i S_i^2$$

ere $$S_i^2 = \frac{1}{(N_i-1)} \sum_j (Y_{ij} - \bar{Y}_i)^2.$$

finition 8.2.1 : We call S a mis-stratification if at least e of the strata contains a $\underline{\underline{Y}}$-value which lies between the nimum and the maximum of $\underline{\underline{Y}}$ in another stratum.

Now we prove the following.

eorem 8.2.1 : For any stratification S of the population th SRSWR in each stratum and with proportional allocation of e sample size n, it is necessary for the $\underline{\underline{Y}}$-character to be rranged in increasing (or decreasing) order of magnitude so that $(\hat{Y}_{st})$ is minimum.

oof : In the case of SRSWR in each stratum and with proportional allocation of the sample size n, we have from (8.1.5)

$$V(\hat{Y}_{st}) = \frac{N}{n} \sum_i N_i \sigma_i^2$$

$$= \frac{N}{n} [\sum_i \sum_j Y_{ij}^2 - G(S)] \ . \qquad (8.2.5)$$

nce, minimisation of $V(\hat{Y}_{st})$ is same as maximisation of $G(S)$. hen $N_1, N_2, \ldots, N_h$ are given, let $S_1, S_2, \ldots, S_M$ be all the possible stratifications, where $M = \frac{N\,!}{N_1!\,N_2!\,\ldots\,N_n!}$. Among these, let S^* be one which maximises $G(S)$. Suppose, if possible, there exists a mis-stratification between r-th and s-th strata of S^*, say y_s (Y-value of an element of s-th stratum) lies between y_{r1} and y_{r2}, the minimum and the maximum of the Y-values in the r-th stratum respectively.

We distinguish two cases :

(i) $\bar{Y}_r \geq \bar{Y}_s$:

In this case exchange y_s with y_{r1} and denote the resulting stratification by $S^{*\prime}$. After some simplification, we obtain that

$$G(S^{*\prime})-G(S^*) = 2(y_s-y_{r1})(\bar{Y}_r-\bar{Y}_s)+(y_s-y_{r1})^2\left(\frac{1}{N_r}+\frac{1}{N_s}\right)$$
$$> 0.$$

(ii) $\bar{Y}_r < \bar{Y}_s$:

In this case exchange y_s with y_{r2} and denote the resulting stratification by $S^{*\prime\prime}$. After simplification, we obtain

$$G(S^{*\prime\prime})-G(S^*) = 2(y_s-y_{r2})\,(\bar{Y}_r-\bar{Y}_s) + (y_s-y_{r2})^2\left(\frac{1}{N_r}+\frac{1}{N_s}\right)$$
$$> 0.$$

Therefore, in both the cases we arrive at a contradiction to the assumption of S^*. Hence, the theorem follows.

ollary 8.2.1 : For any stratification S of the population th stratum sizes $N_1, N_2, \ldots, N_h$ and which is mis-stratified ere exists a stratification S' with the same stratum sizes $N_2, \ldots, N_h$ in which the Y-character is arranged in increasing der of magnitude such that

$$G(S') > G(S).$$

ark 8.2.1 : In the case of SRSWOR in each stratum and with portional allocation of the sample size n, it can be easily ified that

$$V(\hat{Y}_{st}) = \frac{(N-n)}{n} \sum_i N_i S_i^2 = \frac{(N-n)}{n} L(S)$$

$$= \frac{(N-n)}{n} \sum_i \frac{N_i}{(N_i-1)} (\sum_j Y_{ij}^2 - N_i \bar{Y}_i^2)$$

$$\doteq \frac{(N-n)}{n} [\sum_i \sum_j Y_{ij}^2 - G(S)],$$

vided $\frac{N_i}{(N_i-1)}$ are approximately equal. Hence, a result ilar to the above theorem 8.2.1 works out for SRSWOR case o. However, we give below an independent proof of this which es not assume that $\frac{N_i}{(N_i-1)}$ are approximately equal.

Assume that S* is a stratification which minimises L(S). pose, if possible, there exists a mis-stratification between th and s-th strata of S*, say y_s (Y-value of an element of e s-th stratum) lies between y_{r1} and y_{r2}, the minimum and ximum of the Y-values in the r-th stratum respectively. $\overline{Y}'_s$ denotes the mean of the (N_s-1) units other than y_s e s-th stratum. Similarly, $\overline{Y}'_{r1}$ $(\overline{Y}'_{r2})$ denotes the mean (N_r-1) units other than y_{r1} $(y_{r2}$ respectively) of the stratum.

We now distinguish two cases :

(i) $\underline{\overline{Y}'_s \geq \overline{Y}_r}$:

's case exchange y_{r2} for y_s and denote the resulting ification by $S^{*\prime}$. After some simplification, we obtain

$$L(S^{*\prime}) - L(S^*) = 2(y_s - y_{r2})\,(\overline{Y}'_s - \overline{Y}'_{r2})$$

$$< 0, \quad \text{since } y_s < y_{r2} \text{ and } \overline{Y}_r > \overline{Y}'_{r2}.$$

(ii) $\underline{\overline{Y}'_s < \overline{Y}_r}$:

is case exchange y_{r1} for y_s and denote the resulting ification by $S^{*\prime\prime}$. Once again, simplifying we obtain

$$L(S^{*\prime\prime}) - L(S^*) = 2(y_s - y_{r1})\,(\overline{Y}'_s - \overline{Y}'_{r1})$$

$$< 0, \quad \text{since } y_s > y_{r1} \text{ and } \overline{Y}_r < \overline{Y}'_{r1}.$$

fore, in both the cases we arrive at a contradiction to ssumption of S^*. Hence, the required result follows.

In general if $N_1, N_2, \ldots, N_h$ are all different from each r, there exist $h!$ ways of stratification for which $\underline{\underline{Y}}$-ster is arranged in increasing (or decreasing) order of itude. Denote them by $S_1, S_2, \ldots, S_{h!}$ respectively. Then rly the maximum of $G(S)$ is attained by one of these.

k 8.2.2 : For any stratification S of the population stratum sizes $N_1, N_2, \ldots, N_h$ and which is mis-stratified, e exists a stratification S' with the same stratum sizes $_2, \ldots, N_h$ in which the $\underline{\underline{Y}}$-character is arranged in increasing decreasing) order of magnitude such that

$$L(S') < L(S).$$

rk 8.2.3 : We note here that for any stratification S, though it is necessary for the $\underline{\underline{Y}}$-character to be arranged increasing (or decreasing) order of magnitude for $G(S)$ to be mum, this condition is not sufficient. In order to show s, we consider the following example :

$h = 2$, $N_1 = 2$, $N_2 = 3$, the population values of $\underline{\underline{Y}}$ being 1,2,3,4 and 6.

this population, it is easy to show that

$$G(S) = 60.83,$$

e S is the stratification with $Y_{11} = 1$, $Y_{12} = 2$; $Y_{21} = 3$, $= 4$ and $Y_{23} = 6$.

$G(S') = 62$,

; S' is the stratification with $Y_{11} = 1$, $Y_{12} = 2$, : 3 ; $Y_{21} = 4$ and $Y_{22} = 6$.

s, even though for both S and S' the $\underline{\underline{Y}}$-character is nged in increasing order of magnitude, still G(S) is not mum.

We, now provide a set of necessary and sufficient conditions $G(S) = \sum\limits_i \frac{Y_i^2}{N_i}$ to be maximum. We have already seen that for to be maximum, it is necessary that the $\underline{\underline{Y}}$-character has e arranged in increasing (or decreasing) order of magnitude. the units be arranged in increasing order of $\underline{\underline{Y}}$ and let the ts of demarcation be $(y_{(1)}, y_{(2)}, \dots, y_{(h-1)})$, i.e., the t stratum consists of all units for which $\underline{\underline{Y}} < y_{(1)}$ the nd stratum consists of all units for which $y_{(1)} \leq \underline{\underline{Y}} < y_{(2)}$, , the h-th stratum consists of all units for which $\underline{\underline{Y}} > y_{(h-1)}$. ming that all the points of demarcation except $y_{(s)}$ are d, we now show that the value of $y_{(s)}$ which maximises G(S) iven by

$$y_{(s)} \doteq \frac{\overline{Y}_s + \overline{Y}_{s+1}}{2} . \qquad (3.2.6)$$

owing Murthy (1967) we note that the value of $y_{(s)}$ would ct only two terms in $\sum\limits_i \frac{Y_i^2}{N_i}$ namely $\frac{Y_s^2}{N_s}$ and

For $y_{(s)}$ to be optimum, it is necessary and sufficient the following two conditions are satisfied :

$$\frac{Y_s^2}{N_s} + \frac{Y_{s+1}^2}{N_{s+1}} > \frac{(Y_s - y_s')^2}{(N_s-1)} + \frac{(Y_{s+1} + y_s')^2}{(N_{s+1}+1)} \qquad (8.2.7)$$

$$\frac{Y_s^2}{N_s} + \frac{Y_{s+1}^2}{N_{s+1}} > \frac{(Y_s + y_s')^2}{(N_s + 1)} + \frac{(Y_{s+1} - y_s''\ ^2}{(N_{s+1}-1)}, \qquad (8.2.8)$$

e y_s' and y_s'' are the values of the units just preceding following the optimum value $y_{(s)}$. Simplifying (8.2.7) we in that

$$_{+1}-\bar{Y}_s)\ (\bar{Y}_{s+1}+\bar{Y}_s-2y_s') - \frac{(y_s'-\bar{Y}_s)^2}{(N_s-1)} - \frac{(y_s' - \bar{Y}_{s+1})^2}{(N_{s+1}+1)} > 0, \qquad (8.2.9)$$

simplifying (8.2.8) we obtain that

$$_{+1}-\bar{Y}_s)\ (\bar{Y}_{s+1}+\bar{Y}_s-2y_s'') + \frac{(y_s''- \bar{Y}_s)^2}{(N_s+1)} + \frac{(y_s'' - \bar{Y}_{s+1})^2}{(N_{s+1}-1)} < 0. \qquad (8.2.10)$$

e, from (8.2.9) and (8.2.10) it follows that

$$y_s' < \frac{\bar{Y}_s + \bar{Y}_{s+1}}{2} < y_s'' .$$

e, $y_{(s)}$ also lies between y_s' and y_s'' it follows that optimum $y_{(s)}$ satisfies

$$y_{(s)} \doteq \frac{\bar{Y}_s + \bar{Y}_{s+1}}{2} .$$

s, simplifying (8.2.9) and (8.2.10) we obtain that

$$\frac{(N_s-1)\,N_{s+1}}{N_s(N_{s+1}+1)} > \frac{(y_s' - \bar{Y}_s)^2}{(y_s' - \bar{Y}_{s+1})^2} \tag{8.2.11}$$

d

$$\frac{(N_s+1)\,N_{s+1}}{N_s(N_{s+1}-1)} < \frac{(y_s'' - \bar{Y}_s)^2}{(y_s'' - \bar{Y}_{s+1})^2} \tag{8.2.12}$$

s, we obtain the set of necessary and sufficient conditions G(S) to be maximum as follows :

(i) The Y-character should be arranged in increasing (or decreasing) order of magnitude

(ii) the s-th optimum demarcation point should satisfy that

$$y_{(s)} \doteq \frac{\bar{Y}_s + \bar{Y}_{s+1}}{2}$$

$$\frac{(N_s-1)\,N_{s+1}}{N_s(N_{s+1}+1)} > \frac{(y_s' - \bar{Y}_s)^2}{(y_s' - \bar{Y}_{s+1})^2}$$

$$\frac{(N_s+1)\,N_{s+1}}{N_s(N_{s+1}-1)} < \frac{(y_s'' - \bar{Y}_s)^2}{(y_s'' - \bar{Y}_{s+1})^2} ,$$

re y_s' and y_s'' are the values of the units just preceding following the optimum value $y_{(s)}$.

ark : Since the Y-values are not available at the time of survey, the usual practice is to use X-variate which is

ghly correlated with $\underline{\underline{Y}}$ for stratification purposes. We ovide the justification for using the $\underline{\underline{X}}$-variate for strati-
cation in the chapter IX.

3 Optimum demarcation of strata-optimum allocation

If SRSWR is used in each stratum and with optimum alloca-
ion of the sample size n, we obtain from (8.2.1)

$$V_{opt.}(\hat{Y}_{st}) = \frac{1}{n}\left(\sum_i N_i \sigma_i\right)^2. \tag{8.3.1}$$

e now, show that it is not necessary to arrange the $\underline{\underline{Y}}$-character n increasing (or decreasing) order of magnitude for (8.3.1) to e minimum. In order to show this, we consider the following ample :

h=2, N_1=4, N_2=3 and the values of $\underline{\underline{Y}}$ being 1,2,2,2,2,2 and 3. r this population it is easy to show that $(\sum_i N_i \sigma_i)^2$ is inimum for the stratification S with $Y_{1j} = 2$ for j = 1,2,3,4 ; $Y_{21} = 1$, $Y_{22} = 2$ and $Y_{23} = 3$. Clearly, <u>S is mis-stratification</u>.

However, when C_i, the coefficient of variation of the variate in the i-th stratum is the same for all i, then 8.3.1) reduces to

$$\frac{Y^2 \sum_i N_i \sigma_i^2}{n \sum_i N_i \bar{Y}_i^2} = \frac{Y^2}{n}\left[\frac{\sum_i \sum_j Y_{ij}^2}{G(S)} - 1\right].$$

us we prove the following.

orem 8.3.1 : For any stratification S of the population th strata sizes $N_1, N_2, \ldots, N_h$, if

(i) SRSWR is used in each stratum,

(ii) $n_{i,opt.} = \dfrac{n N_i \sigma_i}{\sum\limits_i N_i \sigma_i}$, for $i = 1, 2, \ldots, h$

d (iii) the coefficient of variation C_i (with respect to the $\underline{\underline{Y}}$-variate) is the same in all the strata

hen, for $V_{opt.}(\hat{Y}_{st})$ to be minimum it is necessary for the -character to be arranged in increasing (or decreasing) order f magnitude.

ark : A result similar to the above holds good in the case f SRSWOR in each stratum also.

CHAPTER IX

STRATIFIED RANDOM SAMPLING AND PRIOR DISTRIBUTIONS

Summary

The problem of optimum stratification with proportional optimum allocations in the case of simple random sampling h or without replacement) has been examined in the light n appropriate super-population model and a formal proof been provided here for arranging the auxiliary character ncreasing (or decreasing) order of magnitude for stratifi- on in the case of proportional allocation. It is also n here that the same may not be necessary in the case of mum allocation. However if the coefficient of variation respect to the auxiliary variate is the same in each tum the necessity of arranging the auxiliary character in reasing (or decreasing) order of magnitude for stratification stablished. The results are illustrated by empirical mples. Also, some comparisons among different estimators e been made under the super-population model.

Introduction

Let a finite population of size N be divided into h ata of sizes N_i, $i = 1,2,\ldots,h$. Let Y_{ij} and X_{ij} be values of the characteristics $\underset{\sim}{Y}$ (variate under study) and (auxiliary variate closely related to $\underset{\sim}{Y}$) respectively for j-th unit of the i-th stratum in the population. Let a ple random (without replacement) sample of size n_i be taken m the i-th stratum such that $\sum_{i=1}^{h} n_i = n$. As an estimator of population total $\sum_{i=1}^{h} \sum_{j=1}^{N_i} Y_{ij}$, consider the conventional mator

$$\hat{Y}_{st} = \sum_{i=1}^{h} N_i \bar{y}_i \ , \quad \text{where} \quad \bar{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij} \ .$$

know that

$$V(\hat{Y}_{st}) = \sum_{i=1}^{h} \frac{N_i^2 \ f_i \ S_i^2}{n_i} \qquad (9.1.1)$$

ere $f_i = 1 - \frac{n_i}{N_i}$, $S_i^2 = \frac{1}{(N_i - 1)} \sum_{j=1}^{N_i} (Y_{ij} - \bar{Y}_i)^2$ and $\bar{Y}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} Y_{ij}$.

is well known that (9.1.1) is minimum for a fixed sample size

if $n_i = n_{i,opt.} = \frac{n \ N_i \ S_i}{\sum_{i=1}^{h} N_i \ S_i}$ (Neyman's allocation)

ich reduces to $n_i = \frac{n \ N_i}{N}$ (proportional allocation or wley's allocation), when all S_i's are equal. If the number of strata and the sample sizes $n_1, n_2, \ldots, n_h$ are specified en the optimum stratification in stratified simple random pling consists of finding (h-1) points of demarcation such at (9.1.1) is minimized.

As a first step in finding out the optimum points of arcation of strata in stratified simple random sampling as been quite freely assumed in the literature that the aracter which is presumably having high correlation with Y-character (the variate under study), should be arranged ncreasing order of magnitude, but till now we are not aware ny formal proof or valid justification for doing the same. this chapter optimum demarcation of strata for stratified ple random sampling with proportional and optimum allocations e been examined in the light of an appropriate super-population model. In the former case it has been proved that the haracter should be arranged in increasing order of magnitude d in the latter case it has been shown that the same is necessary if the coefficient of variation with respect to the auxiliary riate is the same in each stratum.

In case of Neyman's optimum allocation of sample size to rat we have $n_{i,opt.} = \frac{n N_i S_i}{\sum\limits_{i=1}^{h} N_i S_i}$. For the computation of $_{i,opt.}$'s, one requires the knowledge of at least the proportion- te values of S_i^2's which are unknown prior to sampling. So in ractice, for S_i^2's, we substitute some estimates α_i^2's which are sually nothing but the S_i^2's of some auxiliary variate which is losely related to the study variable under consideration. The stification for the assumption that the unknown proportionate lues of S_i^2's are usually not quite different from the portionate values of the known α_i^2's was examined in the light an *a priori* distribution by Hanurav (1965), which was further udied by Rao, T.J. (1968) and Vijayan, K. (1971) among others d this is pursued further in this chapter.

2 Demarcation of strata and prior distributions

Under Λ'_{K} of (8.23) we have that

$$(S_i^2|X_{i1},\dots,X_{iN_i})=\frac{\sum_{j=1}^{N_i}(a+bX_{ij})^2-(aN_i+bX_i)^2+\sigma^2\sum_{j=1}^{N_i}X_{ij}^{g}(1-\frac{1}{N_i})}{(N_i-1)}$$

$$=b^2\alpha_i^2+\frac{\sigma^2}{N_i}\sum_{j=1}^{N_i}X_{ij}^{g} \qquad (9.2.1)$$

here $\alpha_i^2=\frac{1}{(N_i-1)}\sum_{j=1}^{N_i}(X_{ij}-\overline{X}_i)^2$ and $\overline{X}_i=\frac{1}{N_i}\sum_{j=1}^{N_i}X_{ij}$.

ence, we have that

$$(V(\hat{Y}_{st})|X_{11},\dots,X_{1N_1},\dots,X_{hN_h})$$

$$=\sum_{i=1}^{h}\frac{N_i^2f_i}{n_i}(b^2\alpha_i^2+\frac{\sigma^2}{N_i}\sum_{j=1}^{N_i}X_{ij}^{g}) \; . \qquad (9.2.2)$$

n arriving at the above expression (9.2.2) it is assumed that $_i$'s are given constants. We note that in the following summa-tions i runs over strata 1,2,...,h and j over population nits 1,2,...,N_i of the i-th stratum unless otherwise specified.

he case of proportional allocation, i.e., when $n_i = \frac{n\, N_i}{N}$
expression (9.2.2) reduces to

$$\frac{(N-n)}{n} \left(b^2 \sum_i N_i \alpha_i^2 + \sigma^2 \sum_i \sum_j X_{ij}^2 \right) . \qquad (9.2.3)$$

te the above expression by V_1. Note that the last term
9.2.3) is independent of stratification, therefore, in the
of proportional allocation, minimization of V_1 reduces
inimization of $\sum_i N_i \alpha_i^2$. We provide here the justification
rranging the auxiliary character $\underline{\underline{X}}$ in the increasing order
gnitude for demarcation of strata, in the case of propor-
al allocation.

In the following we denote by S any stratification of
ulation of size N into h strata of sizes $N_1, N_2, \ldots, N_h$
for each S we define a loss function as

$$L(S) = \sum_i N_i \alpha_i^2 ,$$

gain function, as

$$G(S) = \sum_i \frac{X_i^2}{N_i} .$$

ition : We call S a mis-stratification (with respect to
if at least one of the strata contains an $\underline{\underline{X}}$-value which
between the minimum and the maximum of $\underline{\underline{X}}$ in another
tum.

<u>Theorem 9.2.1</u> : For any stratification S of the population with simple random sampling without replacement in each stratum and with proportional allocation of sample size n, it is necessary that the X-character is arranged in increasing (or decreasing) order of magnitude for

$$\mathcal{E}_\delta\,(V(\hat{Y}_{st})\,|\,X_{11},\ldots,X_{1N_1},\ldots,X_{h\,N_h})\quad \text{to be minimum.}$$

<u>Proof</u> : In case of SRSWOR in each stratum and with proportional allocation of sample size n, we know from (9.2.3) that

$$\mathcal{E}_\delta\,(V(\hat{Y}_{st})\,|\,X_{11},\ldots,X_{h\,N_h}) = \frac{(N-n)}{n}\,(b^2\,L(S) + \sigma^2 \sum_i \sum_j X_{ij}^g).$$

Now the theorem follows easily from remark 8.2.1.

<u>Remark 9.2.1</u> : In the case of SRSWR in each stratum and with proportional allocation of sample size n, it can be easily verified that

$$\mathcal{E}_\delta(V(\hat{Y}_{st})\,|\,X_{11},\ldots,X_{hN_h}) = \frac{N}{n}\,[\sum_i \sum_j \{(X_{ij}^2 - \frac{G(S)}{N})b^2 + \frac{\sigma^2(N_i-1)}{N_i}X_{ij}^g\}]$$

$$\doteq \frac{N}{n}\,(b^2\,L(S) + \sigma^2 \sum_i \sum_j X_{ij}^g)$$

provided $\frac{(N_i-1)}{N_i}$ are approximately equal to 1. Hence, a result similar to the above theorem 9.2.1 works out for SRSWR case also.

We have already shown in (9.2.2) that

$$E_{\xi}(V(\hat{Y}_{st})|X_{11},\ldots,X_{hN_h}) = \sum_i \frac{N_i^2 f_i}{n_i}(b^2 \alpha_i^2 + \frac{\sigma^2}{N_i} \sum_j X_{ij}^g).$$

Following Rao, T.J.(1968), we have that the Δ_g-optimum n_i will be proportional to $N_i \alpha_i$ if $\sum_j \frac{X_{ij}^g}{N_i}$ is proportional to α_i^2 and if this condition is satisfied, we have

$$n_{i,\ \Delta_g\text{-opt.}} = \frac{n\ N_i\ \alpha_i}{\sum_i N_i\ \alpha_i},$$

and the expected variance corresponding to this Δ_g-optimum allocation of sample size n is given by

$$V_2 = (b^2 + \sigma^2 C)\ [\frac{1}{n}(\sum_i N_i\ \alpha_i)^2 - \sum_i N_i\ \alpha_i^2] \qquad (9.2.4)$$

where $C = (\sum_i \sum_j X_{ij}^g / \sum_i N_i\ \alpha_i^2)$.

Now as in section 8.3, it can be proved that in the case of simple random sampling (WOR) with optimum allocation, the arrangement of the auxiliary character in increasing (or decreasing) order of magnitude for stratification may not be necessary.

When $g = 2$, $n_{i,\ \Delta_g\text{-opt.}}$ given above will be optimum when $\frac{1}{N_i} \sum_j X_{ij}^2$ is proportional to α_i^2, i.e., the coefficient of variation C_i (with respect to the auxiliary variate) are the same. This condition is usually satisfied when each N_i is sufficiently large. In this case we prove the following.

Theorem 9.2.2 : For any stratification S of the population with

(1) SRSWOR in each stratum,

(2) $n_i = \dfrac{n\, N_i\, \alpha_i}{\sum\limits_{i=1}^{h} N_i\, \alpha_i}$, for $i = 1, 2, \ldots, h$,

(3) the same coefficient of variation C_i (with respect to the auxiliary variate) in all the strata

(4) neglecting terms of order $\left(\frac{1}{N_i} - \frac{1}{N_i'} \right)$,

it is necessary that the X-character should be arranged in increasing (or decreasing) order of magnitude for

$$\mathcal{E}_{\delta}(V(\hat{Y}_{st}) \mid X_{11}, \ldots, X_{1N_1}, \ldots, X_{hN_h})$$

to be minimum, when the parameter g of the super-population model is 2.

Proof : Under the above assumptions when $g = 2$, from (9.2.4) it can be shown that

$$\mathcal{E}_{\delta}(V(\hat{Y}_{st}) | X_{11}, \ldots, X_{hN_h}) = [b^2\, L(S) + \sigma^2 \sum_i \sum_j X_{ij}^2][\frac{X^2}{n\, C(S)} - 1]$$

$$= F(S) \text{ say.}$$

When $N_1, \ldots, N_h$ are given, let $S_1, \ldots, S_M$ be the possible stratifications where $M = \frac{N!}{N_1! \ldots N_h!}$. Among these, let S^* be the one which minimizes $F(S)$. Suppose if possible, there sts a mis-stratification between r-th and s-th strata of S^*, x_s (an element of the s-th stratum) lies between x_{r1} (the nimum of elements of the r-th stratum) and x_{r2} (the maximum those). Let $\overline{X}'_s$ denote the mean of the (N_s-1) units of the th stratum other than x_s. Similarly $\overline{X}'_{r1}$ and $\overline{X}'_{r2}$ denote means of the (N_r-1) units other than x_{r1} and x_{r2} respectively of the r-th stratum.

We distinguish two cases

(1) $\overline{X}'_s \geq \overline{X}_r$.

this case exchange x_{r2} for x_s and denote the resulting ratification by $S^{*\prime}$. After some simplifications, we obtain

$$L(S^{*\prime}) - L(S^*) = 2(x_s - x_{r2})(\overline{X}'_s - \overline{X}'_{r2}) < 0,$$

i

$$\begin{aligned} G(S^{*\prime})-G(S^*) &= (x_s-x_{r2})[2(\overline{X}_r-\overline{X}_s)+(x_s-x_{r2})(\frac{1}{N_r}+\frac{1}{N_s})] \\ &= (x_s-x_{r2})[2(\overline{X}_r-\overline{X}'_s)(1-\frac{1}{N_s})+\frac{2(\overline{X}_r-x_s)}{N_s}+(x_s-x_{r2})(\frac{1}{N_r}+\frac{1}{N_s})] \\ &= (x_s-x_{r2})[2(\overline{X}_r-\overline{X}'_s)(1-\frac{1}{N_s})+\frac{2(\overline{X}_r-x_{r2})}{N_s}+(x_s-x_{r2})(\frac{1}{N_r}-\frac{1}{N_s})] \\ &\geq (x_s-x_{r2})[2(\overline{X}_r-\overline{X}'_s)(1-\frac{1}{N_s})+\frac{2(\overline{X}_r-x_{r2})}{N_s}] \\ &\geq 0. \end{aligned}$$

nce, $F(S^{*\prime}) < F(S^*)$.

(ii) $\bar{X}'_s < \bar{X}_r$.

this case exchange x_{r1} for x_s and denote the resulting tratification by $S^{*''}$. Now, it is easy to show that

$$(S^{*''})-L(S^*)=2(x_s-x_{r1})(\bar{X}'_s - \bar{X}'_{r1}) < 0,$$

nd

$$\begin{aligned}(S^{*''})-G(S^*) \quad & (x_s-x_{r1})[2(\bar{X}_r-\bar{X}_s)+(x_s-x_{r1})(\frac{1}{N_r} + \frac{1}{N_s})] \\ & \doteq (x_s-x_{r1})[2(\bar{X}_r-\bar{X}'_s)(1-\frac{1}{N_s})+\frac{2(\bar{X}_r-x_{r1})}{N_s}+(x_s-x_{r1})(\frac{1}{N_r} + \frac{1}{N_s})] \\ & \doteq (x_s-x_{r1})[2(\bar{X}_r-\bar{X}'_s)(1 - \frac{1}{N_s}) + \frac{2(\bar{X}_r-x_{r1})}{N_s}] \\ & > 0.\end{aligned}$$

ence, $F(S^{*''}) < F(S^*)$.

herefore, in both the cases we arrive at a contradiction to he assumption of S^*. Hence, the theorem follows.

emark. A result similar to the above theorem 9.2.2 will work out also in the case of SRSWR in each stratum if $\frac{(N_i-1)}{N_i}$ are pproximately equal to 1.

We note here that the assumption that the coefficient of ariation C_i (with respect to the auxiliary variate $\underset{=}{X}$) is pproximately the same in all the strata irrespective of any stratification may look a little restrictive. But in practice, we do come across such populations and three such empirical examples are illustrated in the following.

9.3 Empirical examples

We now illustrate the results of section 9.2 using live data of the following three different populations.

(1) Population I consists of 1961 census village-wise population totals for two neighbouring taluks, viz. Hatkangale and Shirol of Kolhapur district (excluding two highly populated villages) of Maharashtra state of India.

(2) Population II consists of village-wise data on literates and educated persons per thousand for two neighbouring police station zones, viz. Kulti and Asansol of Asansol sub-division derived from the Burdwan district 1961 census book published by the Government of India.

(3) Population III consists of district-wise totals of the production of Jowar crop for the year 1958-59 for the two neighbouring states of Maharashtra and Gujarat of India.

The raw data for these three populations is given in tables C1 to C3 of the appendix.

Here we have considered two types of stratification. One is based on regions, i.e., each taluk in case of Population I, each police station zone in case of Population II and each state in case of Population III is considered as a stratum.

The other stratification is based on the supplementary variate, i.e., with respect to the corresponding auxiliary information in each population the units are arranged in increasing order of magnitude and the first N_1 are considered as stratum 1 and the remaining N_2 as stratum 2. For the purpose of illustration the respective stratum sizes are same in both the types of stratification and for all the three populations.

TABLE 9.3.1

Basic data for the stratification based on regions.

	Pop. I			Pop. II			Pop. III		
Stratum i	N_i	$N_i \alpha'_i$	$N_i \alpha'^2_i$	N_i	$N_i \alpha'_i$	$N_i \alpha'^2_i$	N_i	$N_i \alpha'_i$	$N_i \alpha'^2_i$
1	51	92973	16949×10^4	39	5659	821052	16	267023	4456×10^6
2	45	70240	10964×10^4	40	6280	985840	22	1914450	166596×10^6
	96	163213	27913×10^4	79	11939	1806892	38	2181473	171052×10^6

TABLE 9.3.2

Basic data for the stratification based on supplementay variate

	Pop. I			Pop. II			Pop. III		
Stratum i	N_j	$N_i \alpha_j$	$N_i \alpha^2_i$	N_i	$N_i \alpha_i$	$N_i \alpha^2_i$	N_i	$N_i \alpha_i$	$N_i \alpha^2_i$
1	51	28296	1570×10^4	39	1987	101275	16	187430	2196×10^4
2	45	62414	8637×10^4	40	5325	709037	22	1812723	149362×10^4
	96	90710	10227×10^4	79	7312	810312	38	2000153	151558×10^4

TABLE 9.3.3

Coefficient of variation within the strata for both the stratifications.

	Pop. I		Pop. II		Pop. III	
Stratum i	C_i'	C_i	C_i'	C_i	C_i'	C_i
1	0.5925	0.3540	0.5956	0.3780	0.6540	0.5601
2	0.6022	0.3223	0.6265	0.3733	0.5903	0.5451

TABLE 9.3.4

Approximate efficiency of the estimator in the sense of expected variance when stratification is based on supplementary variate over that when stratification is based on regions, for proportional and optimum allocations.

	Percentage	efficiency	
Allocation	Pop. I	Pop. II	Pop. III
Proportional	273	223	113
Optimum	324	267	119

We note here that the above approximate efficiencies in the sense of expected variance for proportional and optimum allocations are calculated from

$$(\sum_i N_i \alpha_i'^2/\sum_i N_i \alpha_i^2) \quad \text{and} \quad (\sum_i N_i \alpha_i')^2/(\sum_i N_i \alpha_i)^2$$

respectively. The actual efficiencies can be obtained provided the values of the super-population parameters h, g, and σ^2 are known.

9.4 Comparisons of different estimators

As the value of g is usually not available at the time of stratification, it is not possible to stratify the population in such a way that $\sum_j \frac{X_{ij}^g}{N_i \alpha_i^2}$ is the same for all i. So, we relax this assumption and still consider the allocation (may be sub-optimum)

$$n_i = \frac{n N_i \alpha_i}{\sum_i N_i \alpha_i} .$$

Under the above allocation $\mathcal{E}\,(V(\hat{Y}_{st})|X_{11},\ldots,X_{hN_h})$ reduces to

$$V_3 = \sum_i (b^2 N_i \alpha_i^2 + \sigma^2 \sum_j X_{ij}^g)\left(\frac{\sum_i N_i \alpha_i}{n \alpha_i} - 1 \right)$$

$$= b^2 \left[\frac{(\sum_i N_i \alpha_i)^2}{n} - \sum_i N_i \alpha_i^2 \right] + \sigma^2 \left(\sum_i N_i \alpha_i \sum_i \sum_j \frac{X_{ij}^g}{n \alpha_i} - \sum_i \sum_j X_{ij}^g \right) .$$

Now consider

$$V_1 - V_3 = \frac{N}{n} (b^2 L + \sigma^2 L') \qquad (9.4.1)$$

where $L = \sum_i N_i \alpha_i^2 - \frac{1}{N}(\sum_i N_i \alpha_i)^2,$

and $L' = \sum_i \sum_j X_{ij}^g \left[1 - (\sum_i N_i \alpha_i) \frac{1}{N \alpha_i} \right].$

We first prove two lemmas.

Lemma 9.4.1 : Let $0 < p_1 \leq p_2 < \dots \leq p_N < 1$ and not all p_i's are equal and let $a_1 < a_2 \leq \dots \leq a_N$ with $a_N > 0$. If $F(g) = \sum_{i=1}^{N} a_i \, p_i^g \leq 0$ at $g = g'$, then $F(g)$ increases with g at that point.

Proof : Let $a_i \leq 0$ for $i = 1, 2, \dots, m$ and $a_i > 0$ for $i = m+1, \dots, N$.

Now

$$
\begin{aligned}
\frac{d\,F(g)}{d\,g} &= \sum_{i=1}^{N} a_i \, p_i^g \log p_i \\
&= \sum_{i=1}^{m} a_i \, p_i^g \log p_i + \sum_{i=m+1}^{N} a_i \, p_i^g \log p_i \\
&\geq \Big(\sum_{i=1}^{m} a_i \, p_i^g\Big) \log p_m + \sum_{i=m+1}^{N} a_i \, p_i^g \log p_i \\
&= \Big(\sum_{i=1}^{N} a_i \, p_i^g\Big) \log p_m + \sum_{i=m+1}^{N} a_i \, p_i^g (\log p_i - \log p_m) \\
&\geq 0 + \sum_{i=m+1}^{N} a_i \, p_i^g \, (\log p_i - \log p_m) \text{ at } g = g' \\
&\geq 0.
\end{aligned}
$$

Hence, it follows that $F(g)$ increases with g at $g = g'$.

Note : The above lemma is a slight extension of a lemma due to Vijayan (1966).

Lemma 9.4.2 : For any stratification S of the population with $0 < X_{11} \leq X_{12} \leq \dots \leq X_{1N_1} \leq \dots \leq X_{hN_h}$, we prove the following :

(1) $f(g) = \sum_i \sum_j (1 - \frac{\bar{X}}{\bar{X}_i}) X_{ij}^g \geq 0$ for $g \geq 1$,

and

(2) $f'(g) = \sum_i \sum_j (1 - \frac{1}{N\sqrt{\bar{X}_i}} \sum_i N_i \sqrt{\bar{X}_i}) X_{ij}^g \geq 0$ for $g \geq g'$

where g' lies between 0 and 1.

Proof : $$f(1) = \sum_i \sum_j (1 - \frac{\bar{X}}{\bar{X}_i}) X_{ij}$$
$$= X - \bar{X} \sum_i N_i$$
$$= X - \bar{X} N$$
$$= 0.$$

From the condition that $0 < X_{11} \leq X_{12} \leq \dots \leq X_{1N_1} \leq \dots \leq X_{hN_h}$, it easily follows that $(1 - \frac{\bar{X}}{\bar{X}_i})$'s are non-decreasing. Hence, from lemma 9.4.1 it follows that $f(g) \geq 0$ for $g \geq 1$.

Now consider $$f'(0) = \sum_i N_i - \frac{1}{N} \sum_i N_i \sqrt{\bar{X}_i} \sum_i \frac{N_i}{\sqrt{\bar{X}_i}}$$
$$\leq N - \frac{1}{N}(\sum_i N_i)^2 \text{ by Cauchy-Schwartz inequality}$$
$$= 0.$$

and
$$f'(1) = \sum_i X_i - \frac{1}{N}\left(\sum_i N_i\sqrt{\overline{X}_i}\right)^2$$
$$\geq \frac{1}{N}\left[\left(\sum_i N_i\sqrt{\overline{X}_i}\right)^2 - \left(\sum_i N_i\sqrt{\overline{X}_i}\right)^2\right]$$
$$= 0.$$

Also, it is clear that $\left(1 - \frac{1}{N\sqrt{\overline{X}_j}}\sum_i N_i\sqrt{\overline{X}_i}\right)$'s are non-decreasing. Once again applying the lemma 9.4.1 it follows that there exists g' in $[0,1]$ such that

$$f'(g) = 0 \quad \text{for} \quad g = g'$$
$$\geq 0 \quad \text{for} \quad g > g'.$$

Hence, the proof of the lemma.

Theorem 9.4.1 : Under the super-population model Δ'_g of (2.23) the estimator $\hat{Y}_{st}$ with the allocation $n_i = \frac{n\,N_i\,\alpha_i}{\sum_i N_i\,\alpha_i}$ is superior to $\hat{Y}_{st}$ with proportional allocation in the sense of expected variance if

(1) $0 < X_{11} \leq X_{12} \leq \ldots \leq X_{1N_1} \leq \ldots, \leq X_{hN_h}$,

(2) either C_i (the coefficient of variation of $\underset{=}{X}$ in the i-th stratum) or $C_i\sqrt{\overline{X}_i}$ is same for $i = 1, 2, \ldots, h$

and

(3) $g \geq 1$.

Proof : Let V_1 denote the expected variance of $\hat{Y}_{st}$ with proportional allocation under the super-population model A'_g, and V_3 denote the expected variance of $\hat{X}_{st}$ with $n_i = \frac{n}{\sum_i N_i \alpha_i} N_i \alpha_i$ for $i = 1, 2, \ldots, h$, under the same super-population model A'_g. From (9.4.1) we have that

$$V_1 - V_3 = \frac{N}{n} (b^2 L + \sigma^2 L')$$

where $L = \sum_i N_i \alpha_i^2 - \frac{1}{N} (\sum_i N_i \alpha_i)^2 \geq 0$,

and $L' = \sum_i \sum_j X_{ij}^g (1 - \frac{1}{N \alpha_i} \sum_i N_i \alpha_i)$.

We consider two cases

Case (i) : $C_i = C$ for all i.

Under this assumption the expression L' reduces to

$$\sum_i \sum_j X_{ij}^g (1 - \frac{X}{N \bar{X}_i})$$

≥ 0, from the first part of lemma 9.4.2.

Case (ii) : $C_i \sqrt{\bar{X}_i} = C'$ for all i.

Under this assumption the expression L' reduces to

$$\sum_i \sum_j X_{ij}^g (1 - \frac{1}{N \sqrt{\bar{X}_i}} \sum_i N_i \sqrt{\bar{X}_i})$$

≥ 0, from the second part of lemma 9.4.2.

nce the proof of the theorem.

In case of PPS with replacement sampling the estimator f the population total is given by

$$\hat{Y}_{PPS} = \frac{X}{n} \sum_{i=1}^{n} \frac{y_i}{x_i} .$$

t can be easily verified that

$${}_{\delta}(V(\hat{Y}_{PPS})|X_{11},\ldots,X_{hN_h}) = \frac{1}{n} [N^2a^2 (\frac{\bar{X}}{\tilde{X}} -1) + \sigma^2 \sum_{i}\sum_{j} X_{ij}^{g-1}(X-X_{ij})]$$

$$= V_4 \text{ say,}$$

here $\tilde{X}$ donote for the harmonic mean of X_{ij}'s.

ow consider

$-V_4$

$$= \frac{(N-n)}{n} b^2\sum_{i} N_i\alpha_i^2 - \frac{N^2a^2}{n} (\frac{\bar{X}}{X} -1) + \frac{\sigma^2}{n} [((N-n+1)\sum_{i}\sum_{j} X_{ij}^{g-1}(X_{ij}-X)]$$

$$\doteq \frac{(N-n)}{n} b^2\sum_{i} N_i\alpha_i^2 - \frac{N^2a^2}{n} (\frac{\bar{X}}{X} -1) + \frac{\sigma^2}{n} N[\sum_{i}\sum_{j} X_{ij}^{g-1}(X_{ij} - \tilde{X})]$$

rovided $\frac{N}{(N-n+1)} \doteq 1.$

learly, $\sum_{i}\sum_{j} X_{ij}^{g-1}(X_{ij}-\tilde{X}) = \sum_{i}\sum_{j} (X_{ij} - \bar{X}_i) (T_{ij} - \bar{T}_i)$

$$\geq 0 \text{ for } g \geq 1,$$

here $T_{ij} = X_{ij}^{g-1}$ and $\bar{T}_i = \frac{1}{N_i} \sum_{j} X_{ij}^{g-1}$.

Therefore, it follows that

$$V \geq V_1$$

provided a is negligibly small and $g \geq 1$.

Thus we have the following

<u>Theorem 9.4.2</u> : In the sense of expected variance the estimator $\hat{Y}_{st}$ with proportional allocation and with simple random sampling without replacement in each stratum is inferior to $\hat{Y}_{PPS}$ provided a and g of (2.23) satisfy the conditions : a is negligibly small and $g \geq 1$.

We note here that Des Raj (1958) has obtained the above theorem under the following special case of the super-population model (2.23) :

$$\mathcal{E}_\delta (Y_{ij} | X_{ij}) = a + b\,\bar{X}_i$$

$$\mathcal{V}_\delta (Y_{ij} | X_{ij}) = \sigma^2\, \bar{X}_i^g$$

and $\mathrm{Cov}_\delta (Y_{ij}, Y_{i'j'} | X_{ij}, X_{i'j'}) = 0$ for $i \neq i'$, $j \neq j'$.

$$j = 1, 2, \ldots, N_i \;;\quad i = 1, 2, \ldots, h.$$

CHAPTER X

A COMPARISON BETWEEN STRATIFIED AND UNSTRATIFIED RANDOM SAMPLING

10.0 Summary

Consider a finite population of size N divided into h strata of sizes N_i, $i = 1,2,\ldots,h$. Let $Y_{ij}(X_{ij})$ denote the value of the $\underline{\underline{Y}}$ ($\underline{\underline{X}}$) characteristic for j-th unit of the i-th stratum.

For estimating the population total $Y = \sum\limits_{i=1}^{h} \sum\limits_{j=1}^{N_i} Y_{ij}$, it is shown in several books on sampling (eg. Cochran (1963)) that stratified random sampling with proportional allocation is superior to unstratified random sampling provided the finite population correction factor (fpc) in each stratum is ignored. It is shown here that the above result is true even without ignoring the fpc, if the $\underline{\underline{Y}}$-character satisfies the following property $\underline{\underline{P}}$:

$$\operatorname*{Max}_{j} Y_{ij} \leq \operatorname*{Min}_{j} Y_{i+1,j} , \quad i = 1,2,\ldots,h-1.$$

Since the above condition cannot be verified in practice, the comparison between stratified and unstratified random sampling has been examined in the light of an appropriate super-population model and it is found that stratified random sampling with proportional allocation is always superior to unstratified random sampling (in the sense of expected variance) provided the $\underline{\underline{X}}$-character satisfies the property $\underline{\underline{P}}$.

The comparison between stratified and unstratified random sampling with respect to ratio, product and regression estimators has been examined. The results are illustrated with empirical examples.

10.1 Introduction

Let a population of size N be divided into h strata of sizes N_i, $i = 1,2,\ldots,h$. Let Y_{ij} and X_{ij} be the values of the characteristics $\underline{\underline{Y}}$ (the variate under study) and $\underline{\underline{X}}$ (the auxiliary variate closely related to $\underline{\underline{Y}}$) respectively for the j-th unit of the i-th stratum in the population. Let a simple random sample of size n_i be taken from the i-th stratum so that $\sum_{i=1}^{h} n_i = n$. As an estimator of the population total $Y = \sum_{i=1}^{h} \sum_{j=1}^{N_i} Y_{ij}$, consider the following conventional estimators :

$$\hat{Y}_{st} = \sum_{i=1}^{h} N_i \bar{y}_i, \quad \hat{Y}_{R,C} = \frac{\hat{Y}_{st}\, X}{\hat{X}_{st}}, \quad \hat{Y}_{P,C} = \frac{\hat{Y}_{st}\, \hat{X}_{st}}{X}$$

and $$\hat{Y}_{Reg,C} = \hat{Y}_{st} + \hat{b}_c\,(X - \hat{X}_{st})$$

where $\hat{X}_{st} = \sum_{i=1}^{h} N_i \bar{x}_i$, $\bar{y}_i$ and $\bar{x}_i$ denote the sample averages for the characteristics $\underline{\underline{Y}}$ and $\underline{\underline{X}}$ respectively in the i-th stratum and $\hat{b}_o = \frac{\hat{Cov}(\hat{Y}_{st}, \hat{X}_{st})}{\hat{V}(\hat{X}_{st})}$. Let $\hat{Y}$, $\hat{Y}_R$, $\hat{Y}_P$ and $\hat{Y}_{Reg}$ be the corresponding estimators of the population total Y based on the unstratified random sampling of size n. In the following summations, i runs over the strata 1,2,....,h and j runs over the population units $1,2,\ldots,N_i$ of the i-th stratum unless otherwise specified. It is well known that

$$V(\hat{Y}_{st}) = \sum_i \frac{N_i^2 f_i S_i^2}{n_i} ,$$

where $f_i = 1 - \frac{n_i}{N_i}$, $S_i^2 = \frac{1}{(N_i-1)} \sum_j (Y_{ij} - \bar{Y}_i)^2$ and

$\bar{Y}_i = \frac{1}{N_i} \sum_j Y_{ij}$.

Also, it is known that $V(\hat{Y}_{st})$ is minimum for a fixed sample size n if $n_i = n_{i,opt.} = \frac{n N_i S_i}{\sum_i N_i S_i}$ (Neyman's allocation) which reduces to $n_i = \frac{n N_i}{N}$ (proportional allocation) when all S_i's are equal. For the computation of n_i's in Neyman's allocation, one requires the knowledge of at least the proportionate values of S_i's which are unknown prior to sampling. One method of overcoming this limitation suggested by Sukhatme (1935) is to estimate S_i^2's from a priliminary large sample of size n'.

However, if more than one character are to be estimated from a sample survey, then optimum allocation of sample into different strata on the basis of any one character may lead to loss in precision on other characters as compared to the method of proportional allocation. Hence, one might prefer proportional allocation to optimum allocation if the study of several characters is the aim or the size of the preliminary sample to obtain the estimates of S_i^2's is found to be quite large. In light of these, a comparison is made in this chapter between stratified random sampling with proportional allocation and unstratified random sampling with respect to mean per unit, ratio, product and regression estimators without ignoring the fpc.

10.2 Comparisons

Firstly, we prove the following lemma,

Lemma 10.2.1 : If $\underline{Y}$ is any non-negative character with the property $\underline{P}$, i.e.,

$$\underset{j}{\text{Max}}\ Y_{ij} \leq \underset{j}{\text{Min}}\ Y_{i+1,j}\ ; \quad i = 1, 2, \ldots, k-1,$$

and $\underline{X}$ is any non-negative character, then

$$N \sum_i N_i (\bar{Y}_i - \bar{Y})\ (\bar{X}_i - \bar{X}) \underset{<}{\geq} \sum_i (N-N_i)\ S_{yxi}$$

according as $Y_{ij} \geq Y_{i',j'} \Rightarrow X_{ij} \underset{<}{\geq} X_{i',j'}$

re $$S_{yxi} = \sum_{j} \frac{(Y_{ij}-\overline{Y}_i)(X_{ij}-\overline{X}_i)}{(N_i-1)} .$$

of : Consider $N \sum_{i} N_i(\overline{Y}_i - \overline{Y})\ (\overline{X}_i - \overline{X})$

$$= N \sum_{i} N_i\ \overline{Y}_i\ \overline{X}_i - Y X$$

$$= \sum_{\substack{i=1 \\ i < i'}}^{h} \sum_{i'=1}^{h} \frac{(N_{i'}Y_i - N_iY_{i'})\ (N_{i'}X_i - N_i\ X_{i'})}{N_i\ N_{i'}} . \qquad (10.2.1)$$

ting that $N_{i'}Y_i - N_iY_{i'} = \sum_{j} A_j$

ere $A_j = (Y_{ij} - Y_{i'1}) + (Y_{ij} - Y_{i'2}) + \ldots + (Y_{ij} - Y_{i'N_{i'}})$

d writing $Y_{ij}-Y_{i'j'}$ as $(Y_{ij}-Y_{iN_i})+(Y_{iN_i}-Y_{i'1}) + (Y_{i'1}-Y_{i'j'})$

r $j' = 1, 2, \ldots, N_{i'}$, it follows that

$$A_j = N_{i'}(Y_{ij}-Y_{iN_i})+N_{i'}(Y_{iN_i}-Y_{i'1})+ \sum_{J'=1}^{N_{i'}} (Y_{i'1}-Y_{i'j'}).$$

en, $(N_{i'}Y_i - N_iY_{i'}) = \sum_{j} A_j$, which simplifies to

$$N_{i'} \sum_{j}(Y_{ij}-Y_{iN_i})+N_iN_{i'}(Y_{iN_i}-Y_{i'1})+N_i \sum_{j'=1}^{N_{i'}} (Y_{i'1} - Y_{i'j'}).$$

milarly, $(N_{i'}X_i-N_iX_{i'})=N_{i'} \sum_{j}(X_{ij}-X_{iN_i})+N_iN_{i'}(X_{iN_i}-X_{i'1})$

$$+ N_i \sum_{j'=1}^{N_{i'}} (X_{i'1}-X_{i'j'}).$$

nce, $\underset{=}{Y}$ is non-negative and satisfies the property $\underset{=}{P}$, we tain

$$(N_{i'} Y_i - N_i Y_{i'})(N_{i'} X_i - N_i X_{i'})$$

$$N_{i'}^2 \sum_j (Y_{ij} - Y_{iN_i}) \sum_j (X_{ij} - X_{iN_i}) + N_i^2 \sum_{j'=1}^{N_{i'}} (Y_{i'1} - Y_{i'j'}) \sum_{j'=1}^{N_{i'}} (X_{i'1} - X_{i'j'}) \qquad (10.2.2)$$

ccording as $Y_{ij} \geq Y_{i'j'} \Rightarrow X_{ij} \gtrless X_{i'j'}$.

ow,

$$\sum_j (Y_{ij} - Y_{iN_i}) \sum_j (X_{ij} - X_{iN_i}) \gtrless \sum_j (Y_{ij} - Y_{iN_i})(X_{ij} - X_{iN_i})$$

$$\gtrless \sum_{j=1}^{(N_i-1)} (Y_{ij} - Y_{i(N_i-1)})(X_{ij} - X_{i(N_i-1)}) \gtrless \ldots \gtrless (Y_{i1} - Y_{i2})(X_{i1} - X_{i2}).$$

So, we get that

$$(N_i - 1) \sum_j (Y_{ij} - Y_{iN_i})(X_{ij} - X_{iN_i}) \gtrless \sum_{\substack{j=1 \\ j < j'}}^{N_i} \sum_{j'=1}^{N_i} (Y_{ij} - Y_{ij'})(X_{ij} - X_{ij'})$$

$$= N_i(N_i - 1)\, S_{yxi}. \qquad (10.2.3)$$

Similarly,

$$(N_{i'} - 1) \sum_{j=1}^{N_{i'}} (Y_{i'1} - Y_{i'j})(X_{i'1} - X_{i'j}) \gtrless \sum_{\substack{j=1 \\ j < j'}}^{N_{i'}} \sum_{j'=1}^{N_{i'}} (Y_{i'j} - Y_{i'j'})(X_{i'j} - X_{i'j'})$$

$$= N_{i'}(N_{i'} - 1) S_{yxi'} \qquad (10.2.4)$$

From (10.2.2) to (10.2.4) it follows that

$$\frac{(N_{i'}Y_i - N_iY_{i'})\,(N_{i'}X_i - N_iX_{i'})}{N_iN_{i'}} \gtrless N_{i'}\,S_{yxi} + N_i\,S_{yxi'}\ ,$$

and from this the required lemma follows immediately.

<u>Corollary 10.2.1</u> : If $\underset{=}{Y}$ is non-negative and satisfies the property $\underset{=}{\flat}$, then

$$N \sum_i N_i(\bar{Y}_i-\bar{Y})^2 \geq \sum_i (N-N_i)\, S^2_{yi}.$$

From Cochran (1963, p.99) we have that

$$V(\hat{Y}) - V_{prop}(\hat{Y}_{st})$$

$$= \frac{(N-n)}{Nn(N-1)} [N \sum_i N_i(\bar{Y}_i - \bar{Y})^2 - \sum_i(N - N_i)\, S^2_{yi}] \ . \qquad (10.2.5)$$

Thus $V(\hat{Y}) \geq V_{prop}(\hat{Y}_{st})$ if $\underset{=}{Y}$ is non-negative and satisfies the property $\underset{=}{\flat}$.

The comparison between unstratified and stratified random sampling with proportional allocation in the case of ratio, product and regression estimators can be obtained by replacing Y_{ij} by $Z_{ij} = Y_{ij} + \theta_o\ (\bar{X} - X_{ij})$ in (10.2.5) and suitably choosing θ_o in each case. In fact for $\theta_o = R$, we obtain to the second degree of approximation

$$M(\hat{Y}_R)-M(\tilde{Y}_{R,C}) = \frac{N-n}{Nn(N-1)} [N \sum_i N_i(R_i-R)^2 \bar{X}_i^2 - \sum_i (N-N_i) S_{zi}^2]$$

$$< 0, \quad \text{if} \quad R_i = R \quad \text{for all} \quad i. \qquad (10.2.6)$$

Similarly, putting $\theta_o = b_o$ we obtain that

$$M(\tilde{Y}_{Reg})-M(\tilde{Y}_{Reg,C}) = \frac{(N-n)}{Nn(N-1)} [N \sum_i N_i(R_i-b_o)^2 \bar{X}_i^2-(R-b_o)^2 X^2$$

$$- \sum_i (N-N_i) S_{zi}^2]$$

$$< 0, \text{ provided } R_i = b_o \text{ for all } i. \quad (10.2.7)$$

Putting $\theta_o = -R$ we note that $\underline{\underline{Z}}$ satisfies the property $\underline{\underline{P}}$ if $\underline{\underline{Y}}$ satisfies the property $\underline{\underline{P}}$ and $Y_{ij} \geq Y_{i'j'} \Rightarrow X_{ij} \leq X_{i'j'}$. Hence, if this condition is satisfied we obtain to the second degree of approximation

$$M(\tilde{Y}_P) > M(\hat{Y}_{P,C}).$$

10.3 Related remarks

For the interesting comparisons between the stratified random sampling with proportional and optimum allocations and the unstratified random sampling, the reader is referred to Armitage (1947) and Evans (1951). In fact, Armitage had shown that stratified random sampling even with optimum allocation is inferior to unstratified random sampling provided $\bar{Y}_i$'s are equal and S_i's are equal. On the contrary if the stratification is done by using an auxiliary variate $\underline{\underline{X}}$ which is closely related with $\underline{\underline{Y}}$, then not only $\overline{Y}_i$'s will be different but they will be either in increasing or decreasing order. Moreover, in this case it is likely that the $\underline{\underline{Y}}$-character satisfies the property and hence, stratified random sampling with proportional allocation is superior to unstratified random sampling. Also, from remark (8.2.1) it follows that $\underline{\underline{Y}}$-character should satisfy the property $\underline{\underline{p}}$ for optimum demarcation of strata in the case of proportional allocation. Also, from remark 8.2.2 for estimating the population total Y we note that in case of SRSWOR and with proportional allocation, among all the stratifications of sizes $N_1, N_2, \ldots, N_h$, the stratification in which the $\underline{\underline{Y}}$-character satisfies the property $\underline{\underline{p}}$ will have smaller variance. But, the condition that the $\underline{\underline{Y}}$-variate satisfies the property $\underline{\underline{p}}$ cannot be verified in practice, since it depends on the unknown $\underline{\underline{Y}}$-values. Since $\underline{\underline{X}}$ is the stratification variable, it satisfies the property $\underline{\underline{p}}$. We now examine in the following, the comparison between stratified and unstratified random sampling under the superpopulation model Δ'_g of (2.23).

From (9.2.1) and (9.2.3) it is easy to see that under Δ'_g

$$\xi_\delta V(\hat{Y}) - \xi_\delta V_{prop}(\hat{Y}_{st}) \text{ is proportional to}$$

$$[N \sum_i N_i(\overline{X}_i - \overline{X})^2 - \sum_i (N-N_i)\, S^2_{xi}]$$

$$> 0, \text{ since } \underline{\underline{X}} \text{ satisfies the property } \underline{\underline{p}}.$$

Thus, in the case of simple random sampling $\hat{Y}_{st}$ with proportional allocation is superior to $\hat{Y}$ (in the sense of expected variance).

From (10.2.6) and (10.2.7) we note that stratified ratio and regression estimators with proportional allocation can result in large gain over their corresponding estimators in unstratified random sampling only when R_i's differ significantly.

Since, usually the product method of estimation is suggested when $\underset{=}{Y}$ and $\underset{=}{X}$ are negatively correlated, it is reasonable to assume that R_i's differ significantly and then stratification with product method of estimation is expected to have an edge over unstratified sampling.

The optimum allocation of sample size to strata in the case of ratio or regression methods of estimation is not considered for the simple reason that optimum n_i's will involve knowledge of several population parameters and also one can easily derive the conditions under which stratified estimators with optimum allocation will be inferior to the corresponding estimators based on unstratified sampling.

10.4 Illustrations

In this section we illustrate the above results with the help of the following two populations.

Population I consists of 64 large cities in the United States (see Cochran (1963, p.92)). The number of inhabitants in 1930 is considered as $\underline{\underline{Y}}$ and the same in 1920 as $\underline{\underline{X}}$. The cities are stratified into two strata on the basis of $\underline{\underline{X}}$, the first stratum containing the 16 large cities and the second the remaining 48 cities. Population II consists of 364 villages in the Etawah subdivision (India). The villages are stratified into two strata on the basis of agricultural area ($\underline{\underline{X}}$), the first stratum containing the 319 smallest villages and second the remaining 45 villages. The basic data about this population is taken from Sukhatme (1954, pp. 161-165).

The mean square errors (m.s.e.) of different estimators of the population total based on a random sample of size n are obtained for both the populations and the efficiencies of these estimators over $\hat{Y}$ are presented in the following table. Proportional allocation is used in the stratified random sampling.

TABLE 10.4.1

Efficiencies of different estimators

Estimator	Pop. I		Pop. II	
	Unstra.	Stra.	Unstra.	Stra.
Mean per unit	100	297	100	170
Ratio	853	817	330	338
Regression	855	821	336	338

From the above table it is clear that for population I, stratified random sampling with ratio and regression estimators and with proportional allocation is inferior to the corresponding estimators based on unstratified sampling since R_1's are approximately equal ($R_1 = 1.206$ and $R_2 = 1.196$). In the case of the population II there is marginal gain in using stratification in the case of ratio and regression estimators because of the fact that $R_1 = 0.308$ and $R_2 = 0.265$.

APPENDIX

DATA FOR POPULATIONS 1 AND 2

TABLE A1

Data on number of workers (x_1), fixed capital (x_2) and output (y) for 80 factories in a region.

Sr. no.	x_1	x_2	y
(1)	(2)	(3)	(4)
1	51	106	1350
2	51	162	1176
3	52	235	1841
4	52	325	2606
5	53	244	2656
6	54	214	2546
7	57	310	2911
8	60	370	3280
9	65	385	3425
10	67	390	3416
11	68	367	3390
12	70	412	3395
13	71	407	3417
14	73	430	3290
15	74	435	3481
16	76	450	3520
17	78	463	3570
18	80	520	3740
19	81	470	3520
20	85	469	3601
21	87	510	3717
22	88	600	3750
23	92	584	3730
24	93	605	3767
25	97	590	3821
26	100	534	3886
27	107	618	3972
28	110	625	4065
29	113	630	4109
30	116	641	4216
31	119	720	4950
32	121	755	4302
33	125	663	4385
34	127	695	4426
35	127	680	4530
36	131	700	4689
37	134	750	5386
38	135	745	4961
39	139	732	4822
40	144	782	5097

Contd.

Table A1(Contd.)

Sr. no. (1)	x_1 (2)	x_2 (3)	y (4)
41	152	760	5124
42	160	837	5286
43	162	774	5113
44	166	781	5230
45	173	820	5330
46	185	672	4762
47	192	815	5420
48	198	870	5562
49	211	885	5630
50	242	948	5582
51	253	770	5684
52	285	1086	5790
53	291	1073	5839
54	314	1160	5920
55	335	1320	6315
56	352	1465	6510
57	375	1530	6510
58	387	1690	6719
59	425	1670	6752
60	443	1720	6660
61	452	1780	6854
62	466	1705	6760
63	481	1650	6825
64	495	1775	6940
65	528	2370	7295
66	544	1930	7070
67	563	2290	7152
68	585	2065	7186
69	598	2170	7215
70	644	2355	7288
71	667	2500	7540
72	705	2498	7516
73	750	2500	7610
74	775	2780	7894
75	824	2750	8063
76	870	2695	8180
77	913	2880	8315
78	951	2970	8576
79	980	3000	8675
80	1095	3485	9250

x_2 and y in (000) rupees.

DATA FOR POPULATION 3

TABLE A2

Village-wise complete anumeration data obtained in 1951 census for a Tehsil.

Sr. no.	area in sq. miles	cultivated area (in acres)
(1)	(2)	(3)
1	6.97	2544
2	1.55	428
3	3.51	1177
4	9.96	4567
5	8.46	2618
6	9.71	4113
7	6.16	1869
8	12.06	2713
9	9.42	2237
10	1.21	600
11	15.11	3420
12	12.86	4012
13	7.73	1949
14	2.13	695
15	6.03	1569
16	12.74	4562
17	6.90	2221
18	7.80	2423
19	1.63	608
20	3.03	1124
21	1.29	527
22	9.09	2767
23	7.40	2770
24	2.56	719
25	2.95	607
26	2.46	482
27	4.03	1527
28	3.74	1367
29	1.95	767
30	6.44	1648
31	11.33	2440
32	9.28	2434
33	4.90	1638
34	0.18	61
35	13.25	4505
36	5.37	1751
37	5.28	2622
38	5.86	2848
39	6.91	3013
40	4.98	1599
41	8.33	2549
42	13.23	2641
43	4.38	1959
44	4.28	1371
45	7.70	3290
46	5.85	2526
47	4.97	2935
48	2.71	1109
49	14.30	2821
50	13.96	3678
51	6.40	811
52	9.13	1453
53	3.86	1665
54	8.07	2350
55	1.35	564
56	9.06	2487
57	4.60	904
58	10.67	2040
59	3.57	1314
60	6.52	1506
61	6.17	1657
62	5.01	1053
63	5.57	2071
64	2.53	872
65	5.64	1718
66	6.98	316
67	1.78	653
68	3.80	2357
69	6.35	3258
70	8.97	4051
71	3.47	1209
72	2.55	1658

Contd.

Table A2 (Contd.)

Sr. no.	area in sq. miles	cultivated area (in acres)
(1)	(2)	(3)
73	6.27	2608
74	3.61	1289
75	4.59	599
76	7.21	2573
77	4.39	1414
78	3.31	980
79	4.65	1543
80	10.15	3060
81	9.27	2600
82	2.32	1210
83	6.07	2937
84	3.08	1867
85	2.77	1337
86	4.66	1031
87	4.58	1930
88	2.24	1333
89	2.58	1509
90	0.94	509
91	8.47	4424
92	5.56	1881
93	10.87	4139
94	7.35	4072
95	1.20	612
96	16.36	5507
97	11.29	4634
98	3.05	1667
99	3.43	2013
100	0.80	156
101	3.67	1425
102	5.17	2566
103	4.60	2394
104	3.09	1356
105	0.82	610
106	0.96	603
107	1.23	631
108	1.88	1074
109	3.03	1959
110	4.57	2366
111	4.32	2618
112	1.77	428
113	7.18	2075
114	5.56	2296
115	4.66	1870
116	3.56	1328
117	3.61	1612
118	3.25	1653
119	1.91	933
120	8.15	2698
121	1.44	730
122	5.72	2128
123	2.79	1753
124	2.75	772
125	4.03	2096
126	8.51	2862
127	6.56	2377
128	4.77	1318

DATA FOR POPULATION 4

TABLE A3

Female literacy rate and work participation rate (1971) in 45 selected cities and urban agglomerations of India with population 100,000 and above.

Name of the city	Female literacy rate	Female work participation rate
(1)	(2)	(3)
izayawada	45.79	6.07
ntur	37.42	15.81
urnool	36.49	14.75
izamabad	26.40	16.12
auhati	41.87	8.26
anchi	49.39	4.83
okaro Steel city	16.25	12.23
ihar	24.82	9.14
medabad	49.76	4.96
ajkot	51.41	4.27
havanagar	45.54	5.51
amnagar	43.91	5.39
adiad	52.63	5.42
htak	47.15	3.19
mbala Cantt.	46.60	3.88
ammu	49.53	4.39
alicut	58.38	6.84
ndore	46.60	5.59
atlam	44.23	4.88
holapur	33.24	11.93
olhapur	48.46	5.40
alegaon	31.73	9.83
lhasnagar	48.07	4.26
Belgaum	50.35	5.10
Bellary	35.77	8.55
Cuttack	46.90	4.64
Berhampur	37.94	7.48
Bhubaneswar	50.46	5.08
Amritsar	51.50	3.71
Ludiana	50.66	2.88
Jullundur	50.98	2.78
Patiala	51.92	3.12
Jodhpur	34.69	7.38
Ajmer	45.95	5.42
Dindigul	47.44	5.99
Kumbakonam	48.36	4.52
Kanpur	41.21	3.71
Lucknow	43.38	3.90
Varanasi	31.92	10.36
Gorakhpur	41.42	3.81
Dehradun	54.97	4.40
Calcutta	50.48	4.75
Asansol	48.95	4.38
Delhi	50.88	5.13
Imphal	43.53	18.97

DATA FOR POPULATIONS 5 AND 6

TABLE A4

Data on height (x_1), base diameter (x_2) and fibre weight for 50 jute plants sown at Jute Agricultural Research Institute Farm, Barrackpore in 1962-63.

Sr. no. (1)	x_1 (2)	x_2 (3)	y (4)
1	7.08	1.3	5.0
2	7.00	1.3	5.5
3	7.08	1.4	5.5
4	5.67	1.3	3.5
5	5.75	1.2	4.0
6	6.08	1.3	6.0
7	6.83	1.2	4.5
8	7.58	1.4	6.5
9	6.33	1.5	5.0
10	6.17	1.3	4.5
11	6.75	1.4	5.5
12	6.25	1.3	4.5
13	5.92	1.3	3.5
14	6.00	1.3	4.5
15	7.25	1.5	7.0
16	5.50	1.3	4.5
17	6.83	1.2	5.5
18	6.67	1.2	4.5
19	6.83	1.2	5.5
20	7.33	1.5	7.5
21	5.25	1.4	3.5
22	6.17	1.4	4.5
23	7.25	1.7	9.5
24	7.17	1.6	6.5
25	7.08	1.2	5.0
26	6.83	1.5	7.0
27	7.17	1.7	7.5
28	6.83	1.4	6.5
29	7.50	1.8	9.0
30	6.25	1.2	5.0
31	6.67	1.5	6.0
32	6.42	1.4	4.5
33	6.25	1.7	4.5
34	6.58	1.3	5.0
35	7.25	1.2	6.0
36	7.33	1.3	6.5
37	5.33	1.5	4.5
38	6.92	1.4	6.0
39	6.75	1.4	5.0
40	7.25	1.5	7.0
41	5.58	1.6	6.5
42	7.42	1.5	7.5
43	7.08	1.4	7.0
44	7.00	1.7	6.5
45	7.33	1.4	7.0
46	7.58	1.6	7.5
47	6.42	1.7	5.5
48	6.83	1.6	7.0
49	6.00	1.5	4.0
50	6.17	1.5	4.5

x_1 in feet, x_2 in cms. and y in gms.

DATA FOR POPULATIONS 7 AND 8

TABLE A5

1961 population, 1971 population and workers in 1971 of 142 cities and urban agglomerations with population 100,000 and above of India.

(Figures in '00s)

Name of the city	1961 Pop.	1971 Pop.	Workers in 1971	Name of the city	1961 Pop.	1971 Pop.	Workers in 1971
(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Hyderabad	12490	17989	4369	Ahmedabad	11499	15884	4525
Visakhapatnam	2112	3623	1000	Surat	2880	4718	1492
Vizayawada	2344	3437	1018	Baroda	2984	4674	1305
Guntur	1871	2699	888	Rajkot	1941	3002	768
Warangal	1561	2071	566	Bhavnagar	1765	2261	578
Rajahmundry	1300	1888	575	Jamnagar	1486	2149	572
Kakinada	1229	1642	458	Nadiad	790	1083	291
Kurnool	1008	1367	433	Rohtak	882	1248	301
Nellore	1068	1336	397	Ambala Cantt.	1055	1025	264
Eluru	1083	1270	391	Srinagar	2853	4036	1083
Nizamabad	791	1148	397	Jammu	1027	1552	438
Machilipatnam	1014	1126	311	Cochin	2007	4384	1195
Tenali	785	1029	309	Trivandrum	2398	4098	1154
Gauhati	1007	1230	441	Calicut	1925	3340	823
Patna	3646	4903	1373	Alleppey	1388	1601	410
Jamshedpur	3280	4652	1332	Quilon	910	1241	330
Dhanbad	2006	4331	1597	Indore	3949	5726	1479
Ranchi	1402	2560	608	Jabalpur	3670	5338	1541
Gaya	1511	1798	463	Gwalior	3006	4068	1069
Bhagalpur	1438	1727	447	Bhopal	2229	3921	1127
Darbhanga	1030	1321	341	Durg-Bhilaynagar	1332	2453	751
Mazafferpur	1091	1270	403	Ujjain	1442	2091	540
Bokaro Steel city	751*	[illegible]	471	Raipur	1398	2059	607
Monghyr	898	1025	241	Sagar	[illegible]	[illegible]	[illegible]
Bihar	786	1001	293	Bilaspur	867	1308	391

Contd.

Table A5 (Contd.)

Name of the city	1961 Pop.	1971 Pop.	Workers in 1971
(1)	(2)	(3)	(4)
Ratlam	875	1186	298
Burhanpur	821	1053	287
Greater Bombay	41520	59685	21862
Nagpur	6436	8661	2350
Poona	5976	8532	2495
Sholapur	3376	3981	1159
Kolhapur	1874	2591	684
Amravati	1379	1936	511
Malegaon	1214	1918	539
Nasik	1311	1762	498
Thana	1011	1702	575
Akola	1158	1685	456
Ulhasnagar	1078	1681	473
Aurangabad	876	1505	401
Dhulia	989	1371	347
Nanded	811	1264	327
Ahmednagar	970	1173	327
Sangli	738	1151	319
Jalgaon	804	1067	278
Bangalore	11526	16482	4884
Hubli-Dharwar	2485	3796	1055
Mysore	2539	3556	937
Mangalore	1745	2141	771
Belgaum	1468	2138	585
Gulbarga	971	1456	383
Bellary	857	1251	361
Devonagere	781	1210	366
Bijapur	789	1033	251
Simoga	638	1027	299
Bhadravathi	658	1013	271
Cuttak	1463	1940	529
Rourkela	903	1725	584
Berhampur	769	1176	322
Bhubaneswar	382	1055	339
Amritsar	3763	4327	1297
Ludhiana	2440	4011	1208
Jullundur	2296	2961	804
Patiala	1252	1519	425
Jaipur	4034	6131	1655
Jodhpur	2248	3189	854
Ajmer	2312	2625	661
Kota	1203	2130	618
Bikaner	1506	1886	461
Udaipur	1111	1699	494
Alwar	727	1008	263
Madras city	17292	24703	7155
Madurai	4248	5483	1559
Coimbatore	2863	3535	1102
Salem	2492	3083	987
Tiruchirapalli	2499	3062	887

Contd.

Table A5 (Contd.)

Name of the city	1961 Pop.	1971 Pop.	Workers in 1971	Name of the	1961 Pop.	1971 Pop.	Workers in 1971
(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Tuticorin	1242	1548	441	Gorakhpur	1802	2307	617
Nagercoil	1062	1412	385	Saharanpur	1852	2257	618
Thanjavur	1111	1405	368	Dehradun	1563	1994	591
Vellore	1137	1382	412	Jhansi	1697	1981	497
Dindigul	929	1274	353	Rampur	1354	1618	453
Singanallur	247	1134	385	Shahjahanpur	1177	1441	384
Tiruppur	798	1132	410	Mathura	1253	1405	377
Kumbakonam	926	1130	305	Firozabad	986	1339	361
Kanchipuram	927	1105	343	Ghaziabad	704	1280	368
Tirunelveli	880	1085	306	Muzaffarnagar	876	1149	279
Erode	738	1037	344	Farrukhabad-cum-Fatehgarh	946	1114	322
Cuddalore	792	1018	271	Faizabad	883	1098	311
Kanpur	9710	12730	3866	Mirzapur-cum-Vindhyachal	1001	1059	308
Lucknow	6557	8262	2390	Calcutta	57369	70054	22826
Agra	5087	6378	1695	Durgapur	417	2072	660
Varanasi	4898	5829	1823	Kharagpur	1473	1619	401
Allahabad	4307	5140	1499	Asansol	1034	1574	428
Meerut	2840	3678	1058	Burdwan	1082	1450	387
Bareilly	2728	3261	888	Chandigarh	993	2330	774
Moradabad	1918	2724	771	Delhi	23593	36298	11151
Aligarh	1850	2540	668	Imphal	677	1006	286

* Bokaro Steel city came into existence only after 1961. To ensure uniformity, a hyphothetical 1961 population is estimated based on growth rate in other cities of Bihar state in which Bokaro is situated.

X

DATA FOR POPULATION 9

TABLE A6

Data on number of workers (x) and number of absentees (y) for the 43 sample factories.

Sr. no.	x	y	Sr. no.	x	y
(1)	(2)	(3)	(1)	(2)	(3)
1	95	9	23	75	6
2	79	7	24	69	8
3	30	3	25	63	5
4	45	2	26	83	7
5	28	3	27	124	13
6	142	8	28	31	2
7	125	9	29	96	23
8	81	10	30	42	13
9	43	6	31	85	18
10	53	2	32	91	14
11	148	16	33	73	7
12	89	4	34	159	18
13	57	5	35	54	13
14	132	13	36	69	14
15	47	4	37	61	1
16	43	9	38	164	35
17	116	12	39	132	21
18	65	8	40	82	5
19	103	9	41	33	4
20	52	8	42	86	11
21	67	14	43	41	10
22	64	6			

DATA FOR PRODUCTION OF JUTE

TABLE B1

District-wise area (x) and production (y) of Jute in India.

Name of the district	1951-52		1961-62		1962-63		1963-64		1964-65		1970-71	
	x	y	x	y	x	y	x	y	x	y	x	y
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Darrang	1700	11431	1712	13578	1681	8163	1665	11993	1534	11539	2522	15904
Goalpara	3440	20996	3764	26646	3387	16578	3438	28216	3492	26637	3500	30294
Kamrup	2833	20391	2999	19701	3053	15116	3123	25299	3371	19871	2960	18188
Lakhimpur	243	1835	234	1829	266	1379	224	1706	225	1510	356	2579
Nowgong	4694	25371	4563	41795	4246	24403	4268	30431	4275	27885	3602	26715
Champaran	956	3770	1227	7138	1157	8554	955	6768	837	6210	672	4007
Darbhanga	203	459	654	3303	562	4322	507	3631	514	3219	263	1573
Purnea	15598	78882	15328	83799	13830	70012	14754	80294	12208	62139	10594	60919
Saharsa	1341	6349	4608	27490	3464	17969	3198	20809	2460	11179	2460	11179
Garo Hills	486	3699	641	5005	647	3350	676	3847	688	3753	607	4903
Balasore	809	4838	405	2686	405	2688	448	2976	695	44621	695	4462
Cuttack	5009	29943	3586	23800	3586	23815	4451	29561	3998	27925	3209	25066
Bahraich	208	786	243	1974	243	1824	813	6100	239	2032	204	1588
Kheri	886	3344	1107	9004	942	6816	308	3837	748	6365	542	4228
Sitapur	519	1958	532	4327	424	3068	1582	13205	451	3837	268	2091
Burdwan	1429	11562	1453	10241	1433	12912	5399	34695	1708	18970	1280	11610
Cooch-Behar	3679	31701	6204	46357	4929	34251	243	1824	5342	37255	5650	36880
Darjeeling	146	958	417	4788	380	2560	255	1411	291	3689	333	1830
Hooghly	4108	33616	4314	42547	4520	43686	3954	37517	4051	42577	2680	29370
Jalpaiguri	1692	15210	4715	41569	4039	26762	4282	24635	3877	28103	4210	27070
Malda	3092	15926	2894	15059	2424	12983	2703	17166	2388	11179	2340	12000
Midnapore	1785	14374	1550	9193	1659	10453	1534	9667	1740	20543	1180	10610
Murshidabad	6324	27679	7078	46720	7025	46367	7301	57102	7272	57062	6910	42010
Nadia	5253	26339	5638	41982	5229	40108	5666	48685	6621	58372	6270	38560
24-Parganas	5051	40108	5026	40057	4844	39694	5229	44284	5395	48776	3200	23200
West Bengal	1938	10392	6176	32447	5930	32638	5917	35370	6265	31983	6140	31070

x in 10 hectares and y in bales of 1800 kgs.

DATA FOR PRODUCTION OF RICE FOR FOUR STATES

TABLE B2

District-wise information on area in hectares (x) and production of rice in tonnes (y) for 1954-55 and 1964-65.

Uttar Pradesh

of the trict	1954-55		1964-65	
	x	y	x	y
1)	(2)	(3)	(4)	(5)
adun	1388	903	1268	1287
ranpur	7044	4720	8358	8696
ffarnagar	2980	1845	3332	3020
at	1345	828	2137	2030
ndshahar	446	282	621	562
arh	401	267	819	623
ura	10	9	85	59
	9	6	89	70
puri	2902	1943	5483	4616
	1087	722	2323	1788
illy	7977	4768	10359	7903
ur	7387	3098	8350	7080
un	2323	1284	3847	3037
dabad	6085	2592	7860	5433
jhanpur	6711	3998	9161	8613
ibhit	6780	2638	9309	5970
pur	3324	1644	4659	3770
ukhabad	2306	1898	3069	2994
ah	3188	2611	5739	5521
ur	4454	3773	7071	6463
pur	5862	4845	7114	5396
habad	11938	6291	14566	9333
si	1017	429	2085	1075
un	42	18	557	291
rpur	296	123	642	330
a	5454	3312	9236	6795
masi	10915	2249	15283	13212
spur	10065	3899	14899	10648
pur	7282	2111	10289	6700
dpur	8681	2057	9605	5550
lla	7268	2508	8065	4369
i	30042	20789	24114	18776

Contd.

Table B2 (Contd.)

Uttar Pradesh (Contd.)

Name of the district	1954-55		1964-65	
	x	y	x	y
(1)	(2)	(3)	(4)	(5)
Gorakhpur	22930	16128	18947	12198
Deoria	16877	10317	32163	25805
Azamgarh	18390	6497	19640	12458
Nainital	5234	3531	6630	6246
Lucknow	2897	2071	4182	2932
Unnao	5073	3734	7089	5736
Rae-Bareli	9132	6135	11239	8280
Sitapur	8867	5708	12646	8080
Hardoi	5099	3690	7083	5274
Kheri	10673	6584	14013	8838
Faizabad	12395	7225	14257	10177
Gonda	23036	12899	25003	19446
Bahreich	15822	8432	17380	9283
Sultanpur	11387	6227	14138	10292
Pratapgarh	6305	3454	8588	7174
Bara-Banki	10702	7771	12872	10542
Hill districts	16074	20608	8557	7569

Area (x) in 10 hectares and production of rice (y) in 10 tonnes.

Maharashtra

Greater Bombay	114	171	210	346
Thana	14549	21733	15091	25208
Kolba	13041	16784	13504	18502
Ratnagiri	12974	13168	13152	15464
Nasik	3130	2927	3909	4288
Dhulia	2813	1286	2958	1433
Jalgaon	773	534	360	325
Ahmednagar	807	346	1081	1108
Poona	4763	3951	5293	5111
Satara	2142	1754	2594	2652
Sangli	1125	978	1504	1311

Contd.

Table B2 (Contd.)

Maharashtra (Contd.)

Name of the district	1954-55		1964-65	
	x	y	x	y
(1)	(2)	(3)	(4)	(5)
Sholapur	893	362	935	539
Kolhapur	7394	8029	10481	9968
Aurangabad	393	272	510	386
Parbhani	773	454	1704	833
Bhir	551	429	967	681
Nanded	1422	1451	3509	2865
Osmanabad	2164	1284	3408	2174
Buldhana	182	92	441	295
Akola	618	295	830	559
Amaravati	346	163	720	488
Yeotmal	851	366	1218	823
Wardah	89	41	405	303
Nagpur	1770	1661	2566	2337
Bhandara	24768	20809	27256	28378
Chanda	17283	15922	22497	19081

x in hectares and y in tonnes.

Tamilnadu

Chingleput	27364	28762	33923	44654
South Arcot	24000	32139	30515	48043
North Arcot	20385	26591	30549	47973
Salem	11232	15667	13524	24639
Coimbatore	5212	7275	11230	20475
Tiruchirapalli	21057	28483	23508	34874
Thanjavur	55161	69714	60186	104033
Madurai	13948	19874	15507	28343
Ramanadhpuram	17409	21854	24048	20719
Tirunalveli	14622	20680	13606	19428
Nilgiris	335	453	355	329
Kanyakumari	5181	6535	5663	10105

Area (x) in 10 hectares and production of rice (y) in 10 tonnes.

Table B2 (Contd.)

West Bengal

Name of the district	1954-55		1964-65	
	x	y	x	y
(1)	(2)	(3)	(4)	(5)
24 Parganas	60335	53552	59936	76270
Nadia	19559	17606	19589	22242
Murshidabad	29162	31116	28572	33885
Burdwan	43933	57069	46202	69423
Birbum	30279	36604	31471	44801
Bankura	30999	30519	34549	47902
Midnapore	80350	56441	84410	100315
Hoogly	17871	30185	18521	24572
Howrah	8511	8452	8579	13750
Jalpaiguri	17677	17366	21910	24310
Darjeeling	2962	4103	3750	3780
Malda	15451	11150	20016	18653
West Dinajpur	23860	23438	40674	42101
Cooch Behar	16969	13988	23458	24710
Purulia	27067	30291	25444	32360

Area (x) in 10 hectares and production of rice (y) in 10 tonnes.

TABLE B'

Sizes of 64 cities (in 1000's) in 1920 and 1930.

1920 size				1930 size			
Stratum				Stratum			
h = 1	2			1	2		
797	314	172	121	900	364	209	113
773	298	172	120	822	317	183	115
748	296	163	119	781	328	163	123
734	258	162	118	805	302	253	154
588	256	161	118	670	288	232	140
577	243	159	116	1238	291	260	119
507	238	153	116	573	253	201	130
507	237	144	113	634	291	147	127
457	235	138	113	578	308	292	100
438	235	138	110	487	27[illegible]	164	107
415	216	138	110	442	284	143	114
401	208	138	108	451	255	169	111
387	201	136	106	459	270	139	163
381	192	132	104	464	214	170	116
324	180	130	101	400	195	150	122
315	179	126	100	366	260	143	134

Cities are arranged in the same order in both years

DATA FOR POPULATION I

TABLE C1

1961 census village-wise population tables for Hatkangale and Shirol Taluks of Kolhar district of Maharashtra state of India.

Hathangale Taluk

Name of village (1)	Population (2)
Gavre	1702
Ghunaki	4211
Pargaon	5332
Kini	3814
Talsande	2280
Khochi	2866
Bhadole	4458
Vathar	2157
Bhendawade	1885
Latawadi	2604
Padali	1466
Ambap	3802
Vathar T. Udgaon	1619
Vadgaon Kasba	569
Manpadele	1673
Savarde	3690
Narande	2723
Minche	2415
Tasgaon	869
Top	4349
Hingangaon	1092
Kumbhoj	7336
Nej	1923
Majale	1510
Alte	6036
Mauje Vadgaon	1567
Nagaon	2505
Male	1069
Here	5539
Mudshingi	1240
Chokak	1355
Aligre	1584
Hatkangale	5053
Tardal	4091
Shiroli	5966
Halondi	1438
Rukadi	7096
Korochi	2216
Tilawani	1030
Sajani	2147
Mangaon	4803
Kabnur	4454
Rui	3710
Ingali	3101
Chandur	3788
Rattam-Kodoli	7366
Rangoli	2931
Talandage	2263
Rendal	5354
Jangamwadi	408
Yelnd	2470

We have omitted the village Hupari which is highly populated.

Contd.

Table C1 (Contd.)

Shirol Taluk

Name of village (1)	Population (2)
Kavathesar	1931
Kothali	3591
Danoli	5745
Umalwad	2015
Udgaon	4485
Chinchawad	1923
Jainapur	1126
Chipri	3198
Nimshirgaon	2216
Temadalge	1153
Kondigre	663
Ariunwad	2491
Ghalwad	1649
Kutwad	1149
Kanwad	1659
Hasur	1296
Shirati	2670
Kavatheguland	1591
Shedshal	2676
Ganeshwadi	2625
Gurwad	1266
Aurwad	1995
Alas	4819
Dharangutti	1474
Nandani	6916
Haroli	1606
Jambhali	2041
Yadrav	1944
Shahapur	1189
Takavade	3290
Shirdhon	4016
Lat	6868
Shiradwad	3486
Kurundwad	170
Terwad	816
Majarewad	1126
Bastawad	1218
Herwad	3017
Akivat	3893
Rajapur	2524
Khidrapur	1767
Ghosarwad	3336
Takali	4886
Dattawad	4219
Danwad	2910

DATA FOR POPULATION II

TABLE C2

Village-wise data on literates and educated persons per 1000 persons (in 1961) in Kulti and Assansol police-stations of Assansol sub-division of Burdwan district of West Bengal, India.

Kulti police-station

[illegible]me of the village	literates per 1000	Name of the village	literates per 1000
(1)	(2)	(1)	(2)
Debipur	132	Chungari	228
Duburdi	123	Hatinal	136
Damagaria	157	Parra	529
Champtaria	84	Jasaidi	188
Digari	84	Shitalpur	214
Sabanpur	267	Chhota Dhemua	117
Baria	142	Sodepur	352
Lalbazar	246	Radhanagar	222
Ramnagar	298	Asanbani	295
Manbenia	328	Sitarampur	374
Balitara	150	Belrui	377
Retana	538	Lachhipur	89
Lachhmanpur	191	Bamandiha	240
Rampur	857	Aldihi	220
Chalbalpur	203	Methani	340
Kultara	187	Kamalpur	241
Mabutdi	130	Henreigaria	109
Shibpur	208	Bejdihi	204
Gangutia	191	Paidi	132
Haydi	600	China Kurni	300

The villages with literacy rate less than 25 are not considered in Kulti police-station zone.

Contd.

Table C2 (Contd.)

Assansol police-station

Name of the village	literates per 1000
(1)	(2)
Sarkdi	563
Nadiha	656
Ganrui	246
Gopalpur	283
Narasamuda	302
Kumarpur	422
Gobindpur	244
Shitala	198
Palasdiha	179
Barsarakdi	248
Mahujuri	128
Barapukhuriya	92
Garparira	269
Uttar dhadka	171
Dakshin dhadka	494
Kalla	304
Satpukhurjia	106
Kankhaya	318
Banbishnupur	108
Nischinta	111
Keshabganj	47
Chak Keshabganj	27
Assansol	457
Mohishila	353
Kotaldihi	167
Kalipapari	158
Ghoshik	194
Damra	161
Hatgarui	449
Gopalpur	421
Marichkata	256
Sudi	116
Raghunathbati	224
Ramjibanpur	41
Barachak	187
Phatepur	238
Bara Dhemo	145
Jagatdi	168
Bartaria	250

DATA FOR POPULATION III

TABLE C3

District-wise totals of Jowar production in Maharashtra and Gujarat of India in 1958-59.

Maharashtra

Name of the district (1)	Production of Jowar (in tonnes) (2)
Ahmednagar	286,661
East Khandesh	112,436
Kolhapur	57,634
North Satara	114,840
Nasik	24,202
Poona	167,680
Sholapur	364,617
South Satara	193,110
West Khandesh	100,210
Aurangabad	237,130
Bhir	194,668
Nanded	170,546
Osmanbad	245,061
Parbhani	227,811
Akola	126,245
Amraoti	103,311
Bhandara	14,730
Buldana	125,980
Chanda	59,567
Nagpur	105,897
Wardah	73,748
Yeotmal	139,464

Gujarat

Name of the district (1)	Production of Jowar (in tonnes) (2)
Amreli	8,010
Bhavnagar	28,629
Jamnagar	23,327
Junagadh	17,521
Kutch	20,462
Rajkot	20,684
Surendranagar	25,362
Ahmedabad	20,753
Banaskantha	23,286
Baroda	41,317
Broach	47,102
Kaira	7,623
Mehsana	56,995
Panch Mahals	4,402
Sabarkantha	7,170
Surat	55,637

REFERENCES

Aggarwal, O.P. (1959) : Bayes and minimax procedures in sampling from finite and infinite populations-I, *Ann. Math.* Stat., 30, 206-218.

Armitage, P. (1947) : A comparison of stratified with unrestricted random sampling from a finite population, *Biometrika*, 34, 273-280.

Avadhani, M.S. and Sukhatme, B.V. (1970) : Comparison of two sampling procedures with application to successive sampling, *Appl. Stat.*, 19, 251-259.

Basu, D. (1958) : On sampling with and without replacement, *Sankhyā*, 20, 287-294.

Bowley, A.L. (1926) : Measurement of the precision attained in sampling, *Bull. Int. Stat. Inst.*, 22(1), 1-62.

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

Cochran, W.G. (1942) : Sampling theory when the sampling units are of unequal sizes, Jour. *Amer. Stat. Assn.*, 37, 199-212.

Cochran, W.G. (1946) : Relative accuracy of systematic and stratified random samples for a certain class of populations, *Ann. Math. Stat.*, 17, 164-167.

Cochran, W.G. (1947) : Recent developments in sampling theory in the United States, *Rev. Int. Stat. Inst.*, 3, 40-66.

Cochran, W.G. (1963) : *Sampling Techniques*, Wiley, New York, Ed. 2.

Dalenius, T. (1950) : The problem of optimum stratification I, *Skand. Akt.*, 33, [illegible], 203-213.

Dalenius, T. (1952) : The problem of optimum stratification in a special type of design, Skand. Akt., 35, 61-70.

Dalenius, T. and Gurney, M. (1951) : The problem of optimum stratification-II, Skand. Akt., 34, 133-148.

Dalenius, T. and Hodges, J.L.(Jr.) (1957) : The choice of stratification points, Skand. Akt., 40, 198-203.

Dalenius, T. and Hodges, J.L.(Jr.) (1959): Minimum variance stratification, Jour. Amer. Stat. Assn., 54, 88-101

Dalenius, T. (1962) : Recent advances in sample survey theory and methods, Ann. Math. Stat., 33, 325-349.

Das, M.N. and Mohanty, S. (1973) : On PPS sampling without replacement ensuring selection probabilities exactly proportional to sizes, Aust. Jour. Stat., 15 (2), 87-94.

Des Raj (1954) : Ratio estimation in sampling with equal and unequal probability, Jour. Ind. Soc. Agr. Stat., 6, 127-138.

Des Raj (1956) : Some estimators in sampling with varying probabilities without replacement, Jour. Amer. Stat. Assn., 51, 269-284.

Des Raj (1958) : On the relative accuracy of some sampling techniques, Jour. Amer. Stat. Assn., 53, 98-101.

Des Raj (1963) : Some apparently unconnected problems encountered in sampling work, Contributions to Statistics, 67-72, Permagon Press, London, and Statistical Publishing Society, Calcutta.

Des Raj (1965) : On a method of using multi-auxiliary information in sample surveys, Jour. Amer. Stat. Assn., 60, 270-277.

Des Raj and Khamis, S. H. (1958) : Some remarks on sampling with replacement, Ann. Math. Stat., 29, 550-557.

Dodds, D. J. and Fryer, J. G. (1971) : Some families of selection probabilities for sampling with probability proportional to size, Jour. Roy. Stat. Soc. B, 33, 263-274.

Durbin, J. (1953) : Some results in sampling theory when the units are selected with unequal probabilities, Jour. Roy. Stat. Soc., Sr. B,

Durbin, J. (1959) : A note on the application of Quenouille's method of bias reduction to the estimation of ratios, Biometrika, 46, 477-480.

Ekman, G. (1959) : An approximation useful in univariate stratification, Ann. Math. Stat., 30, 219-229.

Evans, W. D. (1951a) : On stratification and optimum allocation, Jour. Amer. Stat. Assn., 46, 95-104.

Fellegi, I. M. (1963) : Sampling with varying probabilities without replacement : Rotating and non-rotating samples, Jour. Amer. Stat. Assn., 58, 183-201.

Foreman, E. K. and Brewer, K.R.W. (1971) : The use of supplementary information in standard sampling procedures, Jour. Roy. Stat. Soc., B, 33, 391-400.

Godambe, V. P. (1955) : A unified theory of sampling from finite populations, Jour. Roy. Stat. Soc., B, 17, 269-278.

Godambe, V. P. and Joshi, V. M. (1965) : Admissibility and Bayes estimation in sampling finite populations-I, Ann. Math. Stat., 36, 1707-1722.

Goodman, L. A. (1960) : On the exact variance of products, Jour. Amer. Stat. Assn., 55, 708-713.

Goodman, L. A. and Kish, L. (1950) : Controlled selection - a technique in probability sampling, Jour. Amer. Stat. Assn., 45, 350-372.

Grundy, P. M. (1954) : A method of sampling with probability exactly proportional to size, Jour. Roy. Stat. Soc., B, 16, 236-238.

Gupta, P. C. (1971) : On some ratio and product type estimators, Jour. Ind. Soc. Agr. Stat., 23, 111-112.

Hajek, J. (1958) : Some contributions to the theory of probability sampling, Bull. Int. Stat. Inst., 36(3), 127-134.

Hajek, J. (1959) : Optimum strategy and other problems in probability sampling, Casipis Pest. Math., 84, 387-423.

Hansen, M. H. and Hurwitz, W. N. (1942) : Relative efficiencies of various sampling units in population inquiries, Jour. Amer. Stat. Assn., 37, 89-94.

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

Hansen, M. H., Hurwitz, W. N. and Madow, W. G. (1953) : Sample Surveys Methods and Theory, Wiley, New York.

Hanurav, T. V. (1962a) : An existence theorem in sampling, Sankhyā, A, 24, 327-330.

Hanurav, T.V. (1965) : Optimum sampling strategies and some related problems, Ph.D. Thesis submitted to the Indian Statistical Institute.

Hanurav, T. V. (1966) : *Some aspects of unified sampling theory*, Sankhyā, A, 28, 175-204.

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

Hartley, H.O. and Ross, A. (1954) : *Unbiased ratio estimators*, Nature, 174, 270-271.

Hayashi, C. and Maruyama, F. (1948) : On some method for stratification, Research Memoir, Inst. Statist. Math., 4, 399-411.

Hayashi, C., Maruyama, F. and Isida, M.D. (1951) : *On some criteria for stratification*, Ann. Inst. Statist. Math., 2, 77-86.

Horvitz, D.G. and Thompson, D.J. (1952): a generalisation of sampling without replacement from a finite universe, Jour. Amer. Stat. Assn., 47, 663-665.

Isii, K. and Taga, Y. (1969) : On optimal stratification for multivariate distributions, Skand. Akt., 1-2, 24-38.

Jessen, R. J. (1942) : *Statistical investigation of a sample survey for obtaining farm facts*, Iowa Agricultural Experimental Station Research Bulletin, No. 304.

Joshi, V. M. (1965a) : Admissibility and Bayes estimation on sampling finite populations II, Ann. Math. Stat., 36, 1723-1729.

Joshi, V. M. (1965b) : Admissibility and Bayes estimation in sampling finite populations III, Ann. Math. Stat., 36, 1730-1742.

Koop, J.C. (1951) : A note on the bias of the ratio estimate, Bull. Int. Stat. Inst., 33(1), 141-146.

Koop, J.C. (1963) : On the axioms of sample formation and their bearing on the construction of linear estimators in sampling theory for finite universes, Parts, I, II and III, Metrika, 7, 81-114, 165-204.

Koop, J.C. (1964) : On an identity for the variance of a ratio of two random variables, Jour. Roy. Stat. Soc., B, 26, 484-486.

Lahiri, D.B. (1951) : A method of sample selection providing unbiased ratio estimators, Bull. Int. Stat. Inst., 33(2), 133-140.

Lahiri, D.B. (1954) : On the question of bias of systematic sampling, Proceedings of World Population Conference, 6, 349-362.

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

Madow, W.G. and Madow, L.H. (1944) : On the theory of systematic sampling, Ann. Math. Stat., 15, 1-24.

Mahalanobis, P.C. (1940) : A sample survey of the acreage under Jute in Bengal, Sankhyā, 4, 511-530.

Mahalanobis, P.C. (1944) : On large scale sample surveys, Roy. Soc. Phil. Trans., B 231, 329-451.

Mahalanobis, P.C. (1946) : Recent experiments in statistical sampling in the Indian Statistical Institute, Jour. Roy. Stat. Soc., 109, 325-378.

Mahalanobis, P.C. (1952) : Some aspects of the design of sample surveys, Sankhyā, 12, 1-7.

Mickey, M.R. (1959) : Some finite population unbiased ratio and regression estimators, Jour. Amer. Stat. Assn., 54, 594-612.

Midjuno, H. (1951) : On the sampling system with probability proportional to the sum of sizes, Ann. Inst. Stat. Math., 3, 99-108.

Mohanty, S. and Das, M. N. (1971) : Use of transformation in sampling, *Jour. Ind. Soc. Agri. Stat.*, 23,(2), 83-87.

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

Murthy, M. N. (1963) : Some recent advances in sampling theory, *Jour. Amer. Stat. Assn.*, 58, 737-755.

Murthy, M. N. (1963) : Generalised unbiased estimation in sampling from finite populations, *Sankhyā*, B, 25, 245-262.

Murthy, M. N. (1964) : Product method of estimation, *Sankhyā*, A, 26, 69-74.

Murthy, M. N. (1967) : *Sampling Theory and Methods*, Statistical Publishing Society, Calcutta, India.

Murthy, M. N. and Nanjamma, N.S. (1959) : Almost unbiased ratio estimates based on interpenetrating sub-sample estimates, *Sankhyā*, 21, 381-392.

Murthy, M. N., Nanjamma, N. S. and Sethi, V. K. (1959) : Some sampling systems providing unbiased ratio estimators, *Sankhyā*, 21, 299-304.

Neyman, J. (1934) : On two different aspects of the representative method, the method of stratified sampling and the method of purposive selection, *Jour. Roy. Stat. Soc.*, 97, 558-606.

Neyman, J. (1938) : Contributions to the theory of sampling human populations, *Jour. Amer. Stat. Assn.*, 33, 101-116.

Olkin, I. (1958) : Multivariate ratio-estimation for finite populations, *Biometrika*, 45, 154-165.

Pascual, J. N. (1961) : Unbiased ratio estimators in stratified sampling, *Jour. Amer. Stat. Assn.*, 56, 70-87.

Pathak, P. K. (1961) : *Some contributions to the theory of sampling*, Thesis submitted to the Indian Statistical Institute.

Pathak, P. K. (1964) : Sufficiency in sampling theory, *Ann. Math. Stat.*, 35, 795-808.

Quenouille, M. H. (1956) : Notes on bias in estimation, *Biometrika*, 43, 353-360.

Ramakrishnan, M. K. (1971) : *Optimum estimators and strategies in survey sampling*, Ph.D. Thesis submitted to the Indian Statistical Institute.

Rao, J.N.K. (1964) : Unbiased ratio and regression estimators in multi-stage sampling, *Jour.Ind. Soc. Agr. Stat.*, 16, 175-188.

Rao, J.N.K. (1966a) : Alternative estimators in pps sampling for multiple characteristics, *Sankhyā*, A, 28, 47-60.

Rao, J.N.K. (1966b) : On the relative efficiency of some estimators in pps sampling for multiple characteristics, *Sankhyā*, A, 28, 67-70.

Rao, J.N.K. Hartley, H.O. and Cochran, W. G. (1962) : A simple procedure of unequal probability sampling without replacement, *Jour. Roy. Stat. Soc.*, B, 24, 482-491.

Rao, P.S.R.S. (1969) : Comparison of four ratio type estimators, *Jour. Amer. Stat. Assn.*, 64, 574-580.

Rao, P.S.R.S. and Mudholkar, G. S. (1967) : Generalised multivariate estimator for the mean of finite populations, *Jour. Amer. Stat. Assn.*, 62, 1009-1012.

Rao, P.S.R.S. and Rao, J.N.K. (1971) : On the jack-knife estimator in finite populations, *Proc. of 38-th session of Int. Stat. Inst.*, 325-329.

Rao, T. J. (1966) : On certain unbiased ratio estimators, *Ann. Inst. Stat. Math.*, 18, 117-121.

Rao, T. J. (1967) : *Contributions to the theory of sampling strategies*, Ph. D. Thesis submitted to the Indian Statistical Institute.

Rao, T. J. (1968) : On the allocation of sample size in stratified sampling, *Ann. Inst. Stat. Math.*, 18, 117-121.

Reddy, V. N. (1973) : On ratio and product methods of estimation, *Sankhyā*, B, 35, 307-316.

Reddy, V. N. (1974a) : On a transformed ratio method of estimation, *Sankhyā*, C, 36, 59-70.

Reddy, V. N. (1974b) : On stratified random sampling from finite populations, *presented in the International Symposium on recent trends of research in Statistics*, Calcutta, 1974.

Reddy, V. N. (1975a) : Stratified simple random sampling and prior distributions, To appear in *Ann. Inst. Stat. Math.*

Reddy, V. N. (1975b) : A note on the efficiency of ratio estimator, *Sankhyā*, C, 37, 95-99.

Robson, D. S. (1957) : Applications of multivariate polykays to the theory of unbiased ratio-type estimators, *Jour. Amer. Stat. Assn.*, 52, 511-522.

Robson, D. S. and Vitheyasai, C. (1961) : Unbiased component-wise ratio estimation, *Jour. Amer. Stat. Assn.*, 56, 350-358.

Roy, J. and Chakravarti, I. M. (1960) : Estimating the mean of finite population, *Ann. Math. Stat.*, 31, 392-398.

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

Sastry, N. S. (1974) : *On a general class of unbiased ratio, product, ratio-cum-product and regression estimators*, Ph.D.Thesis submitted to the Indian Statistical Institute.

Searls, D. T. (1965) : The utilization of a known coefficient of variatio in the estimation procedure, *Jour. Amer. Stat. Assn.*, 59, 1225-1226.

Searls, D. T. (1967) : A note on the use of an approximately known coefficient of variation, *American Statistician*, 21(3), 20-21.

Sen, A. R. (1952) : Present status of probability sampling and its use in estimation of farm characteristics (abstract), *Econometrica*, 20, 103.

Seng, Y. P. (1951) : Historical survey of the development of sampling theories and practice, *Jour. Roy. Stat. Soc.*, A, 114, 214-231.

Sethi, V. K. (1963) : A note on optimum stratification of populations for estimating the population means, *Aust. Jour. Stat.*, 5, 20-35.

Singh, M. P. (1965) : On the estimation of ratio and product of populatio parameters, *Sankhyā*, B, 27, 321-328.

Singh, M. P. (1967a) : Multivariate product method of estimation for finite populations, *Jour. Ind. Soc. Agr. Stat.*, 19, 1-10.

Singh, M. P. (1967b) : Ratio-cum-product method of estimation, *Metrika*, 12, 34-42.

Singh, M. P. (1970) : *Some aspects of estimation in sampling from finite populations*, Ph. D. Thesis submitted to the Indian Statistical Institute.

Singh, M. P. and Roy, A. S. (1968) : On the utilization of an apriori value of the parameter in the estimation procedure, *Tech. Paper No. 30, NSS-ISI*.

Singh, R. and Sukhatme, B. V. (1969) : Optimum stratification, *Ann. Inst. Stat. Math.*, 21, 515-528.

Singh, R. and Sukhatme, B. V. (1971) : Optimum stratification in sampling with varying probabilities, *Ann. Inst. Stat. Math.*, 24, 485-494.

Singh, R. and Sukhatme, B. V. (1972) : A note on optimum stratification, *Jour. Ind. Soc. Agr. Stat.*, 26(2), 91-98.

Singh, R. and Sukhatme, B. V. (1973) : Optimum stratification with ratio and regression methods of estimation, *Ann. Inst. Stat. Math.*, 25, 627-633.

Smith, F. H. (1938) : An empirical law governing soil heterogeneity, *Jour. Agr. Sci.*, 28, 1-23.

Srivastava, S. K. (1965) : An estimate of the mean of a finite population using several auxiliary variables. *Jour. Ind. Stat. Assn.*, 3, 189-194.

Srivastava, S. K . (1967) : An estimator using auxiliary information in sample surveys, Cal. Stat. Assn. Bull., 16, 121-132.

Srivastava, S. K. (1971) : Generalised estimator for mean of a finite population using multiauxiliary information, Jour. Amer. Stat. Assn., 66, 404-407.

Stephan, F. F. (1948) : History of the use of modern sampling procedures, Jour. Amer. Stat. Assn., 43, 12-39.

Stuart, A. (1964) : Multistage sampling with preliminary random stratification of first stage units, Rev. Int. Stat. Inst., 32, 193-201.

Sukhatme, B. V. and David, I. P. (1973) : A note on Koop's approach for finding the bias of the ratio estimate, Jour. Amer. Stat. Assn., 68, 405-407.

Sukhatme, P. V. (1935) : Contribution to the theory of the representative method, Jour. Roy. Stat. Soc., suppl., 2, 253-268.

Sukhatme, P. V. (1945) : Random sampling for estimating rice yields in Madras province, Ind. Jour. Agr. Sci., 15, 308-318.

Sukhatme, P. V. (1946) : Bias in the use of small size plots in sample surveys for yield, Current Science, 15, 119-120, Nature, 157, 630.

Sukhatme, P. V. (1954) : Sampling theory of surveys with applications, Ames, Iowa.

Sukhatme, P. V. (1959) : Major developments in the theory and application of sampling during the last twenty-five years, Estadistica, 17, 652-679.

Sukhatme, P. V. (1966) : Major developments in sampling theory and practice, Research Papers in Statistics, Festschrift for J. Neyman, Wiley, New York, 367-409.

Taga, Y. (1967) : On optimum stratification for the objective variable based on concomitant variables using prior information, Ann. Inst. Stat. Math., 19, 101-129.

Vijayan, K. (1966) : On Horvitz-Thompson and Des Raj estimator, Sankhyā, A, 28, 87-92.

Vijayan, K. (1967) : On an exact πPS sampling (abstract), Sankhyā, A, 29, 113.

Vijayan, K. (1971) : Some contributions to sampling theory, Ph. D. Thesis submitted to Indian Statistical Institute.

Vos, J.W.E. (1974) : Steekproeftrekking met ongelijke Kansen, Statistica Neerlandica, 28; Part I, 11-49, Part II, 69-108.

Walsh, J. E. (1970) : Generalisation of ratio estimate for population total, Sankhyā, A, 32, 99-106.

Yates, F. (1946) : A review of recent statistical developments in sampling and sampling surveys, Jour. Roy. Stat. Soc., 109, 12-43.

Yates, F. (1947) : The influence of agricultural research statistics on the development of sampling theory, Proceedings of International Statistical Conference, 3 (A), 27-39.

Yates, F. and Grundy, P. M. (1953) : Selections without replacement from within strata with probability proportional to size, Jour. Roy. Stat. Soc., B, 15, 253-261.

INDIAN STATISTICAL INSTITUTE LIBRARY
No. T 088
Date 26 NOV 1984
CALCUTTA

RESTRICTED COLLECTION