ON A GENERAL CLASS OF UNBIASED RATIO, PRODUCT, RATIO-CUM-PRODUCT AND REGRESSION TYPE ESTIMATORS

N. S. SASTRY

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

Calcutta

1974

IN MEMORY OF
MY BROTHER
N. S. SARMA

ACKNOWLEDGEMENTS

I have great pleasure in expressing my deep gratitude to Prof. C.R. Rao, F.R.S., Secretary and Director of the Indian Statistical Institute; Dr. N.T. Mathew, U.N. Adviser, Department of Statistics, Malaysia (former Director, Central Statistical Organisation, India) and Mr. V.R. Rao, Director, Central Statistical Organisation, India, for providin facilities to carry out this research work. My sincere thanks are due to Prof. T.V. Hanurav, Dean of Studies, Research and Training School of the Indian Statistical Institute, for his valuable guidance and many critical comments that helped to considerably improve the presentation of the materia

I am grateful to Prof. B.V. Sukhatme, Professor of Statistics, Iowa State University, who initially set me to work in this field at the Institut of Agricultural Research Statistics, New Delhi. Thanks are due to Dr. T.J. Rao for participating in many valuable discussions on this work. Prof. J.N.K Rao has been kind enough to discuss the problem with me and offer valuable advice. I am happy to record my association with Dr. M.S. Avadhani, whicl has been very helpful to me.

Mr. P. Ravindra Kumar and Mr. C.H. Sastry have taken great pains in writing the programmes while organising the empirical work on H-400 Electronic Computer. To them are my sincere thanks. Mr. Pratap Narain has rendered valuable help in proof reading.

Finally, I wish to thank Mr. Prem Kumar Grover for his very patient and elegant typing of the thesis.

N. S. SASTRY

CONTENTS

CHAPTER I

INTRODUCTION AND SUMMARY

Probability sampling methods have extensive application in large-scale agricultural and socio-economic surveys. Sampling is also used in one or more stages of a census and sometimes as a substitute for a complete census. Over the past four decades, confidence in the use of sampling and appreciation for the technique have steadily grown thanks to the pioneering efforts of Neyman, Hansen, Hurwitz, Mahalanobis, Sukhatme, Cochran, Yates and others.

2. It is well known that the principal problem of the theory of sampling from finite populations consists in devising a suitable sample selection procedure for securing a representative sample from a given population and developing an appropriate estimator for the population parameter of interest such as the mean or total of a characteristic or ratio of totals of two characteristics, with a view to maximising the precision of the estimator within the available resources or minimising the cost for achieving a given level of precision.

3. Towards the solution of the principal problem of sampling from finite populations, the first significant attempt appears to be that of Neyman (1934) who laid down the principles governing stratified random sampling. Development of cluster sampling may be traced to the work of Hansen and Hurwitz (1942). Multistage sampling technique and their applications may be found in the contributions by Hansen and Hurwitz (1943), Sukhatme (1947) and others. Systematic sampling was

first investigated by Madow and Madow (1944). Use of auxiliary information in sample selection was suggested by Hansen and Hurwitz (1943). Cochran (1942) utilised auxiliary information in ratio and regression methods of estimation. The important concept of 'cost function' was introduced by Mahalanobis (1940). A number of reviews have appeared on the significant developments in theoretical and applied research on sampling from finite populations. Special mention may be made of the reviews by Yates (1946), Cochran (1947), Stephen (1948), Seng (1951), Sukhatme (1959, 1966), Dalenius (1962) and Murthy (1963).

4. Heuristic considerations seem to have played a major role in the earlier developments in devising representative methods of sample selection and, in general, efforts have been limited to construct some simple unbiased or biased estimators and study their relative efficiency. Though a number of sample designs and estimation methods have become available, the problem of choosing an optimum strategy (sampling design together with an estimator being called a strategy) has not become easier. Recognising the need to formulate the theory of sampling from finite populations in a more systematic manner, Horvitz and Thompson (1952) have initiated a fruitful line of research and defined three classes of linear estimators. Later Godambe (1955) has proposed a unified theory of sampling from finite populations with a view to discussing the fundamental problems and established his classic 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. Hanurav (1965) has, however, characterised some exceptions to Godambe's general result and obtained non-trivial designs (called unicluster designs) where a best estimator exists.

5. The major implication of Godambe's result is that, in order to arrive at an optimum choice, one has to conduct a search in the class of admissible estimators, where for any given sample design the concept of admissibility of an estimator T ensures that there does not exist another estimator T' which is better than T. Surely, the criterion of admissibility helps in the elimination of bad (inadmissible) estimators. Towards this end, important contributions have been made by Murthy (1957), Des Raj and Khamis (1958), Basu (1958), Roy and Chakravarty (1960) and others. Their results eliminate from consideration unbiased estimators which depend either on the order of units in the sample or on repetition of a unit in the sample. Godambe and Joshi (1965) and Joshi (1965 a, 1965 b) have considered admissibility removing the linearity restriction and later relaxing the criterion of unbiasedness. One may also mention the criteria of 'linear invariance' and 'regularity' studied by Roy and Chakravarty (1960). Introducing the concept of 'hyper admissibility', Hanurav (1965) has shown that for any sample design which is not a unicluster design, in the class of all polynomial unbiased estimators, the criterion of hyper admissibility restricts the choice of an estimator to the Horvitz and Thompson (1952) estimator.

6. Auxiliary information on a variate closely associated with the study variate, if available on all units of the population beforehand, may at

times be used to make some plausible guesses about the relative magnitudes of the values of the study variate. In some situations such guesses may be formulated in terms of an apriori distribution, at least partially. This concept, known as super-population concept in the theory of sampling from finite populations, has its origin in Bayesian inference. The concept was originally due to Cochran (1946). 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 apriori knowledge is available on the study variate. When information on a positive valued auxiliary 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 such that

i) every sample has the same number of distinct units,

ii) probability of including any unit in the sample is proportional to the auxiliary information on the unit,

and iii) the estimator used is the corresponding Horvitz and Thompson estimator,

is best in a well defined Bayesian sense. Sampling designs satisfying the condition (ii) are termed πPS designs by Hanurav (1965). The problem of construction of sampling designs, for which the conditions (i) and (ii) are satisfied, has naturally attracted the attention of many research workers. In this area, important contributions have been made by Goodman and Kish (1950), Horvitz and Thompson (1952), Yates and

Grundy (1953), Durbin (1953), Grundy (1954), Des Raj (1956), Hájek (1959), Hanurav (1962, 1965, 1967), Hartley and Rao (1962), Rao, Hartley and Cochran (1962), Felligi (1963), Stuart (1964), Vijayan (1967), Rao, T.J. (1967) and others.

7. Having given a brief account of the historical development of the theory of sampling from finite populations, we may now turn our attention to a specific area, namely the ratio, product and regression methods of estimation. Auxiliary information available on one or more variates positively correlated with the variate under study may be used in ratio or regression method of estimation. In product or regression method of estimation, auxiliary information on variates negatively associated with the study variate may be used. In the next chapter, important developments relating to the three methods of estimation are reviewed. Here it is convenient to concentrate on specific developments which have a bearing on the problems attempted in this thesis.

8. For simple random sampling without replacement, Mickey (1959) has formulated an interesting general class of unbiased estimators for estimating the finite population mean of the study variate utilising information on the population means of several auxiliary variates and obtained unbiased ratio and regression type estimators as special cases. This thesis contains the results of a critical study of Mickey's general class of estimators with particular reference to unbiased ratio, product, ratio-cum-product and regression type estimators which are derivable from the general class. It is useful to first review the present state of knowledge

on Mickey's estimators before giving a brief summary of the author's contributions in this area.

9. Let us consider a population of size N and suppose the vector $(y, x_1, x_2, \ldots, x_p)$ of variates is under observation, where y is the study variate and the x's are auxiliary variates. For the auxiliary variates, the population means $\bar{X}_i$ are assumed to be known. A simple random sample of size n is drawn without replacement from the population. Let Z_m denote any subset of m ($< n$) distinct sample elements, and $\bar{y}_m$ and $\bar{x}_{im}$ be the means of y and x_i based on Z_m. Further, let $a_i(Z_m)$ be some real valued functions defined on Z_m. Consider the derived sample of size $n-m$, obtained by the exclusion of Z_m from the total sample of size n, and define on it the means $\bar{y}_{n-m}$ and $\bar{x}_{i\,n-m}$. We may conceive the derived population of size $N-m$ by excluding Z_m from the total population of size N and define on it the means $\bar{Y}_{N-m}$ and $\bar{X}_{iN-m}$. For a given Z_m, it is easily seen that the derived sample of size $n-m$ is a simple random sample without replacement from the derived population of size $N-m$. Consequently, for a given Z_m, the statistic

$$U_m = \bar{y}_{n-m} - \sum_{i=1}^{p} a_i(Z_m)(\bar{x}_{i\,n-m} - \bar{X}_{i\,N-m})$$

is conditionally unbiased for $\bar{Y}_{N-m}$. It follows that, for a given Z_m, the estimator

$$G_m = \frac{1}{N}\left[\, m\,\bar{y}_m + (N-m)\,U_m \,\right]$$

is conditionally unbiased for the population mean ($\bar{\bar{Y}}$) of the study variate y. It is then obvious that G_m is also unconditionally unbiased for $\bar{\bar{Y}}$. The estimator G_m is Mickey's (1959) basic general unbiased estimator for estimating $\bar{\bar{Y}}$.

10. Mickey has pointed out that for a given sample one can choose Z_m in several ways and corresponding to each such choice a basic estimator G_m can be formed. For instance, treating the sample as an ordered sequence the order being that in which the units are drawn one by one, Z_m may be taken as the set of units occurring in the first m draws. Alternately, Z_m may be chosen as a sub-sample of size m from the sample of size n. It is then easy to think of classes of unbiased estimators which may be generated as weighted averages of basic estimators G_m. One interesting class given by Mickey depends on basic estimators constructed on portions of total sample. If Z_m comprises the units occurring in the first m draws of the sample and the corresponding estimator G_m is denoted by G (m, n), then for the choice of any k integers m_1, m_2, ..., m_k such that $0 \leq m_1 \leq m_2 \leq \ldots \leq m_k < m_{k+1} = n$, Mickey has shown that the k estimators $G(m_t, m_{t+1})$, t = 1 to k, are mutually un-correlated and, therefore, for the class of estimators

$$\bar{G}^{(k)} = \frac{1}{k} \sum_{t=1}^{k} G(m_t, m_{t+1})$$

a non-negative unbiased estimator of variance can be easily formulated.

11. Again when the units selected in the first m draws constitute the sub-set Z_m, for all possible orders of the sample one can construct

the basic estimators G_m and denote their simple average by G_m^*. As observed by Mickey, variance of G_m^* can never be greater than the variance of G_m. This follows readily from the results of Murthy (1957) and Basu (1958) on the relative efficiency of 'ordered' and the corresponding 'unordered' estimators. Williams (1961, 1962) has suggested the division of the sample at random into g groups each of size m (n = mg) and the computation of Mickey's basic estimator G_m treating each of the g groups in turn as Z_m and then taking the simple average, say $\bar{G}_m$, of the g such estimators G_m. Rao (1967) has also considered the splitting of the sample at random into g groups each of size m but omitted a particular group from the total sample to form a sub-set Z_{n-m} and computed the corresponding basic estimator G_{n-m}. In this process, g estimators G_{n-m} may be formed corresponding to the g possible sub-sets Z_{n-m} for a given split of the sample and Rao has suggested the use of their simple average $\bar{G}'_{n-m}$ for estimation purposes.

12. Mickey (1959) has derived unbiased ratio and regression type estimators from his general estimators G_m and G^*_m for specific choices of the coefficient functions $a_i\,(Z_m)$. For example, Hartley-Ross (1954) unbiased ratio type estimator $\hat{Y}_{HR}$ has been obtained from G^*_m using a single auxiliary variate x and the choice $a\,(Z_m) = \bar{r}_m = \frac{1}{m} \sum_{j \in Z_m} \frac{y_j}{x_j}$. For the choice $a\,(Z_m) = R_m = \frac{\bar{y}_m}{\bar{x}_m}$, unbiased ratio type estimators T_m and T^*_m have been derived from G_m and G^*_m. It may be noted that $T^*_1 = \hat{Y}_{HR}$. Using the choice $a(Z_m) = b_m$, the linear regression coefficient of y on x based on Z_m, unbiased regression type estimators D_m and D^*_m have been

shown to result from G_m and G_m^*. In view of the complex form of the estimators, Mickey has not investigated the variance of any one of his estimators. Even the problem of estimation of variance has been attempted by him only for the class of estimators $\bar{G}^{(k)}$

13. Robson (1957) has derived the exact variance of $\hat{\bar{Y}}_{HR}$ which is valid for any sample size. In large samples (i.e., when the variance or mean square error is considered only upto terms of order n^{-1}), Goodman and Hartley (1958) have shown that $\hat{\bar{Y}}_{HR}$ is more efficient than the classical biased ratio estimator $\hat{\bar{Y}}_R$ if and only if the slope of the population regression line of y on x is closer to $\bar{r}_N = \frac{1}{N} \sum_{j=1}^{N} \frac{y_j}{x_j}$ than to $R = \frac{\bar{Y}}{\bar{X}}$. In general, this condition is not satisfied and $\hat{\bar{Y}}_{HR}$ is found to be less efficient than $\hat{\bar{Y}}_R$ in large samples.

14. Splitting the sample at random into g groups each of size m, Rao (1967) has constructed the unbiased ratio type estimator $\bar{T}'_{n-m}$ as a particular case of $\bar{G}'_{n-m}$ with the choice $a(Z_{n-m}) = R_{n-m}$. He has investigated the relative precision of $\bar{T}'_{n-m}$, $\hat{\bar{Y}}_{HR}$ and $\hat{\bar{Y}}_R$ under two super-population models which were earlier formulated by Durbin (1959) in a study on ratio estimators. In Durbin's super-population model 1, it is assumed that (i) population is infinite, (ii) $y_j = \alpha + \beta x_j + u_j$, $E(u_j/x_j) = 0$, $E(u_j^2/x_j) = \delta$ (a constant), $E(u_j u_{j'}/x_j, x_{j'}) = 0$ for $j \neq j'$, and (iii) x is a normal variate, where $E(\,./x_j)$ denotes conditional expectation for given x_j, etc. Under this model, when the expected variance or mean square error is considered upto order n^{-3}, Rao (1967) has shown that $\bar{T}'_{n-1}$ is the optimum choice in the class of

unbiased ratio type estimators $\overline{T}'_{n-m}$ (for $1 \leq m \leq \frac{n}{2}$) and $\overline{T}'_{n-1}$ is slightly more efficient than $\hat{\overline{Y}}_R$. Durbin's super-population model 2 assumes that x is a gamma variate in addition to the conditions (i) and (ii) of his first model. Under this model Rao's (1967) investigation shows that $\overline{T}'_{n-1}$ is the optimum choice among the estimators $\overline{T}'_{n-m}$ (for $1 \leq m \leq \frac{n}{2}$) and it is considerably more efficient than $\hat{\overline{Y}}_{HR}$ ($n \geq 4$) but is only slightly better than $\hat{\overline{Y}}_R$ for $n \geq 8$. Similar results have been obtained by Rao and Beegle (1967) in a Monte Carlo study under Durbin's super-population model 1 for small and moderate size samples. In an empirical investigation using several sets of real data, Rao (1969) has again demonstrated that in small and moderate size samples $\overline{T}'_{n-1}$ is better than $\hat{\overline{Y}}_{HR}$ and compares favourably with $\hat{\overline{Y}}_R$, although in small samples $\hat{\overline{Y}}_R$ seems to be slightly more efficient than $\overline{T}'_{n-1}$. Other empirical studies with essentially similar observations are those of Hutchison (1971) and Rao and Kuzik (1973).

15. Using Mickey's estimators, Rao and Pereira (1968) have proposed two unbiased double ratio type estimators in sampling on two occasions for estimating the current population mean of the study variate. One of their estimators is a linear combination of three Hartley-Ross type ratio estimators and may be viewed as a special case of G_1^* with three auxiliary variates. For this unbiased double ratio type estimator, Rao and Pereira have given the large sample variance, using Goodman and Hartley's (1958) variance formula for $\hat{\overline{Y}}_{HR}$ and the covariance formula between two Hartley-Ross type ratio estimators due to Robson and Vithayasi (1961).

16. When information on the population mean of an auxiliary variate negatively correlated with the study variate is available, Robson (1957) has formulated an unbiased product type estimator $\hat{Y}_{RP}$ in simple random sampling without replacement and derived its exact variance. When several auxiliary variates of which some having positive and others negative correlation with the study variate are available, Srivastava (1965), Rao and Mudholkar (1967), and others have suggested an unbiased ratio-cum-product type estimator which is a weighted linear combination of Hartley-Ross ratio type and Robson's product type estimators.

17. Williams (1963) has investigated the precision of the class of unbiased regression type estimators $\bar{D}_m$ which is a particular case of the class of general unbiased estimators $\bar{G}_m$ for the choice $a(Z_m) = b_m$, the linear regression coefficient of y on x based on Z_m. His results show that, under Durbin's super-population model 1, the optimum number of groups (g) into which the sample is to be split is equal to $\sqrt{n/3}$ for the class of estimators $\bar{D}_m$ and that the estimators compare favourably (in the sense that the relative loss in precision is not much) with the classical linear regression estimator $\hat{Y}_r$ (which is also unbiased under the assumed super-population model) provided the group size $m \geq 3$. For certain non-normal x distributions, his asymptotic results suggest a similar conclusion provided the distribution of x is platykurtic. It may be noted that William's results hold good only in large samples.

18. Splitting the sample into n groups each of size 1, Rao (1969) has constructed the unbiased regression type estimator $\bar{D}'_{n-1}$ as a

particular case of $\bar{G}'_{n-1}$ and studied empirically its precision using several sets of real data. According to his results, in small and moderate size samples, $\bar{D}'_{n-1}$ is far inferior to the classical biased linear regression estimator $\hat{\bar{Y}}_r$ which itself is less efficient than the classical ratio estimator $\hat{\bar{Y}}_R$. In small samples, Sampford (1969) has also observed through an empirical investigation that $\hat{\bar{Y}}_r$ is less efficient than $\hat{\bar{Y}}_R$.

19. In the context of the above developments, the author's contributions are now presented in brief.

20. When the sample is split at random into groups of size m to construct the general unbiased estimators $\bar{G}_m$ and $\bar{G}'_{n-m}$, we note that the simple averages of $\bar{G}_m$ and $\bar{G}'_{n-m}$ over all possible splits of the sample are respectively G^*_m and G^*_{n-m}. The estimators $\bar{G}_m$ and $\bar{G}'_{n-m}$ are, therefore, inadmissible and not superior to G^*_m and G^*_{n-m} respectively. Consequently, our studies are mostly based on the estimator G^*_m. For investigating the variance of G^*_m, it is instructive to study first the variance of the basic estimator G_m. An exact expression for the variance of G_m is, therefore, derived. The approach adopted in the derivation is found useful in developing a non-negative unbiased estimator of variance of G_m, based on which an unbiased estimator of variance is formulated for any weighted average of the basic estimators G_m. Using this result it is possible to construct unbiased estimators of variance for $\bar{G}_m$, $\bar{G}'_{n-m}$ and G^*_m.

21. It seems difficult to obtain a meaningful expression for the exact variance of G^*_m for all choices of m in the range 1 to $n-1$. For $m = 1$,

however, the problem is satisfactorily solved. The exact variance of G_1^* has many interesting applications as may be seen from the results reported in the following. When the population size N is relatively large compared to the sample size n (i.e. $N \gg n$), the exact variance of G_1^* assumes a neat form. The expression for the exact variance of G_1^* is obtained in terms of population symmetric means and, therefore, an unbiased estimator of variance of G_1^* is readily derived by simply replacing the population symmetric means in the variance formula with the corresponding sample symmetric means. This unbiased estimator of variance for G_1^* is not the same as the one obtained by substituting $m = 1$ in the unbiased estimator of variance of G_m^* mentioned above.

22. While it has not been possible to obtain the exact variance of G_m^* for all choices of m and for any sample size n, a simple expression for the variance of G_m^* is obtained in large samples under two alternate sets of assumptions. When m is small in comparison with n, for any choice of the coefficient functions $a_i (Z_m)$, the large sample variance of G_m^* is derived. Alternatively, without making any assumption directly on the choice of m, but assuming that certain moment conditions are satisfied in respect of the coefficient functions $a_i (Z_m)$, the variance of G_m^* is obtained in large samples. It is interesting to note that the expression for the large sample variance of G_m^* is same under either of the two sets of assumptions. Using these results on the large sample variance of G_m^*, many interesting results are derived and reported in the following. Based on the large sample variance formula for G_m^*, an estimator of variance is also given.

23. One of the desirable properties of a multivariate estimator is that, whenever an additional auxiliary variate is introduced in the estimator, it should not result in a decrease in the efficiency of the estimator. For the multivariate unbiased estimator G^*_m, if $a_i(Z_m)$ is a coefficient function defined on y and the i-th auxiliary variate x_i only, then it is shown that the possibility of an increase in the variance of the estimator cannot be ruled out with the addition of a new auxiliary variate. To obviate this difficulty, an optimum method is suggested for introducing an additional auxiliary variate in G^*_m. The method has application in formulating unbiased ratio-cum-product type estimators.

24. Having examined Mickey's general unbiased estimators, our attention is then focussed on unbiased ratio type estimators which may be lerived from the general estimators for specific choices of the coefficient unctions $a_i(Z_m)$. With a single auxiliary variate x and the choice $(Z_m) = q\bar{y}_m / \bar{X}$, $q > 0$, it is shown that G^*_m reduces to an unbiased atio type estimator, say $S^*_{(q)}$, which is different from Mickey's unbiased atio type estimator T^*_m. Interestingly $S^*_{(q)}$ is the same for all values of in the range 1 to n-1. Noting that $S^*_{(q)}$ and Hartley-Ross unbiased atio estimator T^*_1 are special cases of G^*_1 for the choices $a(Z_1) = qy_j/\bar{X}$ d $a(Z_1) = y_j/x_j$ respectively, exact variances of these estimators are rived from the exact variance already obtained for G^*_1. It is verified that bson's (1957) exact variance formula for T^*_1 agrees with the one derived re. It is also noted that for one of the two unbiased double ratio type imators proposed by Rao and Pereira (1968), exact variance may be

written down from that of G_1^*.

25. For the choice $q = 1$, the estimator $S^*_{(q)}$ has the same large sample variance as the classical ratio estimator $\hat{Y}_R$. When the variance or mean square error is considered upto order n^{-2} and $N >> n$, $S^*_{(1)}$ is more efficient than $\hat{Y}_R$ if (i) the joint distribution of y and x is bivariate normal, (ii) the coefficients of variation of y and x are equal (i.e., $C_y = C_x$) and (iii) the correlation ρ between y and x does not exceed 0.78. When $N >> n$, for $S^*_{(q)}$ one can think of an optimum choice for q in the sense of minimising its variance. In this sense, the optimum choice q_o of q is obtained and the variance of $S^*_{(q_o)}$ is derived. It is further shown that the choice $q'_o = \rho C_y / C_x$ is a good approximation to the optimum choice q_o. Both $S^*(q_o)$ and $S^*(q'_o)$ have the same large sample variance as the classical linear regression estimator $\hat{Y}_r$. Information on q'_o available from a pilot survey or previous census or past experience may be used to compute $S^*_{(q'_o)}$ without disturbing its unbiased nature.

26. It may be recalled that a simple expression for the variance of the general unbiased estimator G^*_m is derived in large samples <u>either</u> by assuming that m is small in relation to n <u>or</u> under certain moment conditions on $a_i (Z_m)$. When m is <u>not</u> small relative to n, it is now shown that the moment conditions are satisfied for the choice $a(Z_m) = R_m$. Consequently, making use of the large sample results on the variance of G^*_m, the large sample variance of Mickey's unbiased ratio type estimator T^*_m is obtained for any choice of m in the range 1 to n-1. In particular, for any large value of m, it is proved that T^*_m and the classical ratio estimator

$\hat{Y}_R$ have the same variance. The result implies that T^*_{n-1} and $\hat{Y}_R$ are equally efficient in large samples.

27. Ratio method of estimation is generally considered appropriate in populations where regression of y on x is linear. It is, therefore, of interest to know the behaviour of T^*_m in such populations. When the regression line of y on x passes through the origin, for all choices of m in the range 1 to n - 1, T^*_m has the same large sample variance which is also equal to the large sample variance of $\hat{Y}_R$. When the regression line does <u>not</u> pass through the origin but all values of x are of same sign, the choice m = n - 1 is the optimum choice for T^*_m in large samples, and T^*_{n-1} and $\hat{Y}_R$ are equally efficient.

28. Making use of the procedures developed for the estimation of variance of $\hat{\sigma}^*_m$, we derive for T^*_m and $S^*_{(q)}$ unbiased estimators of variance valid for any sample size and biased but simpler estimators of variance applicable in large samples. Using data of five live populations and samples of sizes 2, 3, 4 and 5 drawn from these populations, unbiased estimates of variance are computed for the estimators $N\,T^*_m (1 \leq m \leq n-1)$, $N\,S^*_{(1)}$ and $N\,S^*_{(q'_o)}$ For the same ratio type estimators biased estimates of large sample variance are also computed using samples of sizes 10, 15, 20 and 25 from a live population of size 142.

29. Continuing the study on unbiased ratio type estimators, we investigate the precision of T^*_m under two super-population models $\triangle$ and $\triangle'$. For the model $\triangle$ we assume: (i) population is infinite, (ii) $y_j = \alpha + \beta x_j + u_j$ $E(u_j/x_j) = 0$, (iii) $V(u_j/x_j) = \delta$ (a Constant),

and (iv) $Cov(u_j, u_{j'} / x_j, x_{j'}) = 0$ for $j \neq j'$. It may be noted that Durbin's (1959) super-population model 1 is a particular case of $\triangle$ with the further assumption that x is a normal variate. As our model $\triangle'$ we take Durbin's (1959) super-population model 2, which is also a special case of $\triangle$ with the additional assumption that x is a gamma variate with parameter, say h (a positive integer).

30. Under the model $\triangle$ we evaluate the expected variance of T^*_m and the expected mean square error of the classical ratio estimator $\hat{Y}_R$ upto order n^{-3} and show that T^*_{n-1} is the optimum choice estimator in the class of estimators T^*_m ($\frac{n}{2} \leq m \leq n-1$) and is also superior to $\hat{Y}_R$, provided the coefficient of variation of x is not less than the coefficient of skewness of x. In particular, the result holds for all symmetric or negatively skewed distributions of x. Using the result on T^*_m and the result;
$V(T^*_{n-m}) \leq V(\bar{T}'_{n-m})$, it is proved that $\bar{T}'_{n-1} = T^*_{n-1}$ is also the optimum estimator in the class of estimators $\bar{T}'_{n-m}$ ($1 \leq m \leq \frac{n}{2}$) under the model $\triangle$, provided the coefficient of variation of x is not less than the coefficient of skewness of x. Rao's (1967) result on $\bar{T}'_{n-m}$ under Durbin's (1959) super-population model 1 is, therefore, a special case of our result on $\bar{T}'_{n-m}$ under the model $\triangle$

31. We derive the expected variance of T^*_m under the super-population model $\triangle'$. It is applicable for any sample size n and any choice of m in the range 1 to $n-1$. Earlier Rao (1967) has obtained, under the model $\triangle'$, the expected variances for the estimators $\bar{T}'_{n-m}$($1 \leq m \leq \frac{n}{2}$) and T^*_1. It may be noted that the expected variances of T^*_1 and $\bar{T}'_{n-1} = T^*_{n-1}$ may

also be derived from our result on T^*_m. Using our result one can obtain a numerical comparison of the expected variance of T^*_m and the expected mean square error of $\hat{Y}_R$ under the model Δ' for specific choices of h, n and m. Such a comparison is made for the choices:

$$h = 3, \quad n = 2 \quad n = 3 \quad \text{and} \quad n = 4$$
$$m = 1, \quad m = 1, 2 \quad\quad m = 1, 2, 3 ;$$

and it is shown that (i) the precision of T^*_m increases as m increases from 1 to n - 1 for n = 3 and 4, (ii) T^*_{n-1} (for n = 3 and 4) is more efficient than $\hat{Y}_R$ and (iii) Hartley-Ross estimator T^*_1 is less efficient than $\hat{Y}_R$.

32. So far we have reported the results on ratio method of estimation in which the auxiliary variate x and the study variate y are positively correlated. We now present the results on product method of estimation which uses information on an auxiliary variate negatively associated with the study variate.

33. When a single auxiliary variate x negatively associated with y is available, we show that the general unbiased estimator G^*_m yields unbiased product type estimators :

$$S^*_{(q < o)} \text{ for the choice } a (Z_m) = \frac{q\,\bar{y}_m}{\bar{X}}, \; q < O \; ;$$

$$H^*_m \text{ for the choice } a (Z_m) = -\frac{P_m}{\bar{X}}, \; P_m = \frac{1}{m\,\bar{x}_m} \sum_{j \in Z_m} y_j x_j \; ;$$

and L^*_m for the choice $a (Z_m) = \dfrac{-\bar{y}_m \, \bar{x}_m}{\bar{X}}$.

In particular, it is noted that $S^*_{(-1)} = H^*_1 = L^*_1 = \hat{Y}_{RP}$, where $\hat{Y}_{RP}$ is the Robson's (1957) unbiased product type estimator. Thus Mickey's general class of estimators contain not only unbiased ratio and regression type estimators but also unbiased product type estimators. It may be noted that the product type estimator $S^*_{(q<o)}$ and the ratio type estimator $S^*_{(q)}$ earlier introduced have the same 'form' except that we take $q < 0$ for $S^*_{(q<o)}$ and $q > 0$ for $S^*_{(q)}$.

34. For the class of estimators L^*_m $(1 \leq m \leq n-1)$ we prove an interesting property which says that: given any two specific choices of m, say m' and m'', the entire class of estimators L^*_m may be generated as weighted linear combinations of $L^*_{m'}$ and $L^*_{m''}$. In particular, when the finite population correction is ignored and n is even, the estimator $L^*_{n/2}$ is of simple form; and then, for any m, L^*_m may be expressed as a weighted linear combination of $L^*_1 = \hat{Y}_{RP}$, and $L^*_{n/2}$.

35. We observe that the results obtained for the unbiased ratio type estimator $S^*_{(q)}$ on exact variance, optimum choice q_o of q and optimum variance are all applicable (without any change) for the unbiased product type estimator $S^*_{(q<o)}$. It may be noted that the optimum choice product type estimator $S^*_{(q_o<o)}$ is more efficient than Robson's (1957) unbiased product type estimator $\hat{Y}_{RP} = S^*_{(-1)}$. A good approximation to q_o is $q'_o = \rho\ C_y / C_x$. In large samples, $S^*_{(q_o<o)}$, $S^*_{(q'_o<o)}$ and the classical linear regression estimator $\hat{Y}_r$ are all equally efficient and superior to $\hat{Y}_{RP}$ and the conventional product estimator $\hat{Y}_P = \frac{\bar{y}\ \bar{x}}{\bar{X}}$.

36. Utilising the results on the large sample variance of G^*_m, we

derive in large samples the variances of H^*_m and L^*_m. It is proved that in large samples H^*_{n-1} is more or equally or less efficient compared to $H^*_I = \hat{\bar{Y}}_{RP}$, according as $|\rho \; C_y / C_x| \lesseqgtr (1 + \frac{1}{2} C_x^2)^{-1}$. We also show that the large sample variance of L^*_m decreases as m increases from 1 to n-1 when <u>either</u> $R > \bar{r}_N > |\beta|$ or $R < \bar{r}_N < |\beta|$, where $0 < \bar{r}_N = \frac{1}{N} \sum_{j=1}^{N} y_j / x_j$, $0 < R = \frac{\bar{Y}}{\bar{X}}$, and β is the population linear regression coefficient of y on x.

37. We develop estimators of variance for $S^*_{(q_0' < 0)}$, H^*_m and L^*_m on similar lines as for G^*_m and illustrate their computation for samples of sizes 2, 3, 4 and 5; and 9 and 12 from a population of size 45. Variance estimates computed for sample sizes 2 to 5 are unbiased while those for sizes 9 and 12 are biased.

38. When several auxiliary variates of which some having positive and others negative correlation with the study variate are available, unbiased ratio-cum-product type estimators may be developed as particular cases of Mickey's general unbiased estimator G_m. It may be recalled that we have earlier suggested an optimum method of introducing an additional auxiliary variate in G^*_m. This method is useful in formulating unbiased ratio-cum-product type estimators.

39. Assuming that the <u>first</u> p_1 of the p auxiliary variates $(x_1, x_2, \ldots, x_{p_1}, x_{p_1+1}, \ldots, x_p)$ have positive correlation with y and the remaining $p - p_1$ variates have negative correlation with y, for any weight vector $W = (W_1, W_2, \ldots, W_p)$ such that $\sum_{i=1}^{p} W_i = 1$, we show that the estimator G^*_m results in the unbiased ratio-cum-product type

estimators:

$\delta^*_{(\pm 1)}$ for the choice $\tilde{a}_i(Z_m) = \frac{W_i \bar{y}_m}{\bar{X}_i}$, (i = 1 to p_1)

$$= -\frac{W_i \bar{y}_m}{\bar{X}_i} \text{ , (i = } p_1+1 \text{ to p) ;}$$

TH^*_m for the choice $\tilde{a}_i(Z_m) = W_i R_{im}$, (i = 1 to p_1)

$$= -\frac{W_i P_{im}}{\bar{X}_i} \text{ , (i = } p_1+1 \text{ to p) ;}$$

and TL^*_m for the choice $\tilde{a}_i(Z_m) = W_i R_{im}$, (i = 1 to p_1)

$$= -\frac{W_i \bar{r}_{im} \bar{x}_{im}}{\bar{X}_i} \text{ , (i = } p_1+1 \text{ to p).}$$

In particular, $TH^*_1 = TL^*_1$ is a weighted combination of Hartley-Ross unbiased ratio type estimators $\hat{Y}_{HR}$ and Robson's unbiased product type estimators $\hat{Y}_{RP}$. Srivastava (1965), Rao and Mudholkar (1967) have earlier mentioned TH^*_1 without, however, determining the optimum choice of the weight vector W.

40. In the construction of the above unbiased ratio-cum-product type estimators, an important problem is the determination of the optimum choice for the weight vector W. We derive exact variances for $\delta^*_{(\pm 1)}$ and TH^*_1 and large sample variances for TH^*_m and TL^*_m, and making use of these variances we determine the optimum choices of W. We note that estimation of variance for the unbiased ratio-cum-product type estimators

$S^*_{(\pm 1)}$, TH^*_m and TL^*_m presents no new problem as it can be attempted directly using the results on G^*_m.

41. Having considered ratio and product methods of estimation, we now present the results on regression method of estimation in which auxiliary information on one or more variates, positively or negatively associated with the study variate, may be used. As already stated, Mickey (1959) has shown that G^*_m results in an unbiased regression type estimator D^*_m for a single auxiliary variate x and $a(Z_m) = b_m$, the linear regression coefficient of y on x based on Z_m. When $p\ (> 1)$ auxiliary variates are available, we may denote by D^*_{pm} the multivariate unbiased regression type estimator which results from G^*_m for the choice $a_i(Z_m) = b_{im}$, the partial regression coefficient of y on x_i based on Z_m.

42. We note that D^*_m and D^*_{pm} are both defined for $2 \leq m \leq n-1$. For $m = 1$, mainly to facilitate theoretical studies on exact variance of unbiased regression type estimators, we derive from G^*_1:

(a) the unbiased regression type estimator D^*_1, using a single auxiliary variate x and the choice $a(Z_1) = a_j = \dfrac{y_j\ (x_j - \bar{X})}{\sigma_x^2}$, where σ_x^2 is the population variance of x;

and (b) the multivariate unbiased regression type estimator D^*_{p1} for the choice of the column vector of coefficient functions

$$A_j = (a_{1j}, a_{2j}, \ldots, a_{pj})' = \frac{N}{(N-1)} S^{-1}_{XX}\ y_j\ (X_j - \underset{\sim}{\bar{X}}),$$

where $X_j = (x_{1j}, x_{2j}, \ldots, x_{pj})'$ $\underset{\sim}{\bar{X}} = (\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_p)'$, and

S_{XX} is the matrix of the population mean squares and products among the p auxiliary variates x_i .

43. We obtain the exact variance of D_1^* using the result on the exact variance of G_1^* . We also compare under two super-population models the expected exact variance of D_1^* and the expected mean square error (upto terms of order n^{-2}) of the classical linear regression estimator $\hat{Y}_r$. The first super-population model $\triangle_1$ assumes: (i) population is infinite, (ii) $y_j = a_1 + a_2 (x_j - \bar{X}) + a_3 (x_j - \bar{X})^2 + u_j$, $E (u_j / x_j) = 0$, $V(u_j/x_j) = \delta$ (a constant), and $Cov (u_j, u_{j'} / x_j , x_{j'}) = 0$ for $j \neq j'$; and (iii) x has a normal distribution. The second super-population model $\triangle_2$ assumes conditions (i) and (ii) of the first model and that x is a gamma variate. It is shown that the classical linear regression estimator $\hat{Y}_r$ is more efficient than the unbiased regression type estimator D_1^* under both the super-population models.

44. Making use of the results on the large sample variance of G_m^*, we derive in large samples the variances of D_1^* , D_m^* , D_{p1}^* and D_{pm}^* and observe that (a) the estimators D_1^* , D_m^* (for large m) and $\bar{Y}_r$ are all equally efficient, and (b) D_{p1}^* and D_{pm}^* (for large m) are as efficient as $\hat{Y}_{pr} = \bar{y} + \sum_{i=1}^{p} b_{in} (\bar{X}_i - \bar{x}_i)$, where b_{in} is the partial regression coefficient of y on x_i based on the sample of size n. It is also shown that, in large samples, n - 1 is an optimum choice for m in the range (2, n - 1) for D_m^* and D_{pm}^* ; and the estimators D_{n-1}^* and $D_{p\,n-1}^*$ have the same precision as $\hat{Y}_r$ and $\hat{Y}_{pr}$ respectively.

45. Unbiased estimators of exact variance and biased estimators of

large sample variance are developed for D_1^* and D_m^*, and numerical illustrations of their computation are provided for samples of sizes 2, 3, 4 and 5 in the former case and for samples of sizes 9, 10, 12, 15, 20 and 25 in the latter case. Estimation of variance for the multivariate estimators D_{p1}^* and D_{pm}^* may be attempted on similar lines.

46. We now consider the problem of determining optimum forms of the coefficient functions $a_i(Z_m)$ for the general unbiased estimator G_m^* and give a solution in large samples. It may be recalled that an expression for the large sample variance of G_m^* has been obtained when m is small relative to n or under certain moment conditions on $a_i(Z_m)$. We have also shown that the results on the large sample variance of G_m^* are applicable for the estimators G_m^* involving the coefficient functions $a_i(Z_m)$ such as:

$$\frac{q\,\bar{y}_m}{\bar{X}_i},\quad R_{im},\quad -\frac{P_{im}}{\bar{X}_i},\quad -\frac{\bar{r}_{im}\,\bar{x}_{im}}{\bar{X}_i},$$

linear regression coefficient of y on x_i based on Z_m, and

partial regression coefficient of y on x_i based on Z_m.

In this manner, the large sample variance has been derived for:

the unbiased ratio type estimators $S_{(q)}^*$, T_m^*;

the unbiased product type estimators $S_{(q<o)}^*$, H_m^*, L_m^*;

the unbiased ratio-cum-product type estimators $S_{(\pm 1)}^*$, TH_m^*, TL_m^*;

and the unbiased regression type estimators D_1^*, D_m^*, D_{p1}^*, D_{pm}^*.

It is now proved that the possible minimum for the large sample variance of G_m^* is equal to the large sample variance of the classical linear regression estimator $\hat{\bar{Y}}_r$ in the case of a single auxiliary variate and to that of the classical multivariate regression estimator $\hat{\bar{Y}}_{pr}$ for p auxiliary variates. From the results reported above we also note that, for a single auxiliary variate, estimators having the minimum possible large sample variance are:

the unbiased ratio type estimators $S^*_{(q_o)}$, $S^*_{(q'_o)}$;

the unbiased product type estimators $S^*_{(q_o < o)}$, $S^*_{(q'_o < o)}$; and

the unbiased regression type estimators D^*_1, D^*_m for large m.

For p auxiliary variates, such estimators are the unbiased regression type estimators D^*_{p1} and D^*_{pm} for large m.

47. We have reported earlier that unbiased estimates of variance are computed for unbiased ratio, product and regression type estimators in small samples of sizes 2, 3, 4 and 5 for purposes of illustration. As only a few samples of each size are used in those examples, the tentative observations that may be made from those examples on 'the relative efficiency of the unbiased estimators in small samples ' need further verification. A separate empirical study is, therefore, conducted on the relative efficiency of biased as well as unbiased estimators in small samples of sizes 2, 3, 4 and 5, following an approach in which a reasonable number of independent samples of same size (drawn from several sets of live populations) are used.

48. The estimators included in the empirical study on ratio and regression estimators are: the classical biased ratio estimator $\hat{\bar{Y}}_R$,

unbiased ratio type estimators T^*_m ($1 \leq m \leq n-1$), $S^*_{(1)}$ and $S^*_{(q'_o)}$; classical biased regression estimator $\hat{Y}_r$, and unbiased regression type estimators D^*_m ($2 \leq m \leq n-1$). In addition, by using the same number of independent Midzuno-Sen (1952) scheme samples, Lahiri's (1951) unbiased ratio estimator is also included. The tentative conclusions <u>in small samples</u> are: (i) T^*_{n-1} ($n > 2$) is, in general, more efficient than the other unbiased ratio type estimators based on simple random sampling scheme, (ii) T^*_{n-1} is <u>not</u> likely to be superior to $\hat{Y}_R$, (iii) the unbiased regression type estimators are less efficient than $\hat{Y}_r$, (iv) $\hat{Y}_R$ is superior to $\hat{Y}_r$, and (v) Lahiri's unbiased ratio estimator (based on Midzuno-Sen scheme) is, in general, more efficient than all other estimators.

49. The empirical study on product and regression estimators includes the classical biased product estimator $\hat{Y}_P$, unbiased product type estimators $\hat{Y}_{RP}$, $S^*_{(q'_o < o)}$, L^*_m and H^*_m ($2 \leq m \leq n-1$); classical biased regression estimator $\hat{Y}_r$, and unbiased regression type estimators D^*_m ($2 \leq m \leq n-1$). <u>In small samples</u>, the tentative conclusions from the study are: (i) $S^*_{(q'_o < o)}$ is the most efficient among the unbiased product estimators, (ii) $\hat{Y}_P$ is as efficient as $S^*_{(q'_o < o)}$, (iii) the unbiased regression type estimators are less efficient than $\hat{Y}_r$ and (iv) the regression estimators are inferior to the product estimators.

50. The thesis is divided into eight chapters. Chapter 1 (the present one) is introductory. It provides a brief introduction to 'sampling from finite populations', focusses attention on the specific developments having a bearing on the problems considered in this thesis and gives a

broad summary of the results obtained by the author. In Chapter II we explain the basic concepts and definitions used in this thesis and present a brief review of the literature on biased, almost unbiased and unbiased ratio, product and regression estimators. The rest of the thesis contains the author's contributions.

51. Chapter III is devoted to a systematic investigation of the variance of Mickey's general unbiased estimators. The problems attempted and the results obtained in this chapter are given above in paragraphs 20 to 23. This chapter lays the foundation for the investigations to be carried out in the subsequent chapters on unbiased ratio, product, ratio-cum-product and regression type estimators which are derivable from Mickey's general unbiased estimators.

52. Chapter IV deals with the study of the two classes of unbiased ratio type estimators $S^*_{(q)}$ ($q > 0$) and T^*_m ($1 \leq m \leq n-1$). Specific problems considered and the results obtained in this chapter are presented above in paragraphs 24 to 28. In Chapter V we investigate the behaviour of the expected variance of T^*_m under two super-population models $\triangle$ and $\triangle'$ and obtain a comparison between T^*_m and the classical ratio estimator $\hat{Y}_R$. Making use of the results on T^*_m under the model $\triangle$, similar results are derived for the estimator $\bar{T}'_{n-m}$. Details of the super-population models $\triangle$ and $\triangle'$ and the results are given above in paragraphs 29 to 31.

53. In Chapter VI we derive unbiased product and ratio-cum-product type estimators from the general estimator G^*_m and study their variance.

Results obtained in this chapter are described above in paragraphs 33 to 40. Chapter VII deals with two important problems. One problem refers to a detailed study of the variance and the efficiency of the unbiased regression type estimators D_1^*, D_m^* ($2 \leq m \leq n-1$), D_{p1}^* and D_{pm}^* ($2 \leq m \leq n-1$). The second problem is to determine the optimum forms of the coefficient functions $a_i(Z_m)$ for the general unbiased estimator G_m^* in large samples. Specific results obtained in this chapter are given above in paragraphs 42 to 46. Finally, Chapter VIII contains the results of an empirical study on the relative performance of biased and unbiased ratio and regression, and product and regression estimators in small samples of sizes 2, 3, 4 and 5. The tentative conclusions drawn from the study are presented above in paragraphs 48 and 49.

54. Mickey's unbiased estimators have been extended to multi-stage and stratified sampling designs by Rao (1964) and to multi-phase designs by Sastry (1964). Several results reported in this thesis can be extended to multi-stage, stratified and multi-phase designs. Such extensions are, however, not considered in this thesis.

55. In the concluding remarks we indicate some further lines of research on problems related to those considered in this thesis.

CHAPTER II

CONCEPTS, DEFINITIONS AND REVIEW OF LITERATURE

2.1. Introduction :

In this chapter the basic concepts and definitions used in this thesis are explained and a brief review of literature on biased, almost unbiased and unbiased ratio, product and regression estimators is given.

A 'finite population' $\mathcal{U}$ is a collection of a known number N of 'sampling units' $U_1, U_2, \ldots, U_N$ which are distinguishable. We represent the finite population by

$$\mathcal{U} : [U_1, U_2, \ldots, U_N] \qquad \ldots (2.1.1)$$

A list such as (2.1.1) is termed a 'sampling frame' and N is called the 'population size'.

A 'sample' s from $\mathcal{U}$ is a finite ordered sequence of units from $\mathcal{U}$ and is denoted by

$$s = (u_1, u_2, \ldots, u_{n_s}), \qquad \ldots (2.1.2)$$

where each u_i belongs to $\mathcal{U}$. We may also represent equivalently

$$s = (U_{i_1}, U_{i_2}, \ldots, U_{i_{n_s}})$$

where $1 \leq i_k \leq N$ for $1 \leq k \leq n_s$.

In (2.1.2) the units need not be distinct. We call n_s (or simply n if the context is clear) the 'sample size'. The number of distinct units in the sample is termed the 'effective sample size'.

A 'sample design' D over $\mathcal{U}$ is a collection S of samples s from $\mathcal{U}$ with a probability measure P defined on it (i.e. to each $s \in S$ is attached a probability P_s such that $P_s \geq 0$ and $\sum_{s \in S} P_s = 1$). We may denote

$$D = D(S, P). \quad \ldots (2.1.3)$$

The above definition of the sample design suggests an obvious method of selecting a sample which consists of listing down all possible samples and selecting a sample from the list with the corresponding probability. However, the obvious method is not practicable, especially for large values of N and n. Instead, a one-by-one 'drawing mechanism', in which units are drawn from $\mathcal{U}$ one-by-one with replacement with probabilities depending on the previous draws, is more convenient to operate. In this connection Hanurav (1962 a) has shown that for any given design D(S, P) there corresponds a unique drawing mechanism M such that sampling according to M results in the design D(S, P) and conversly. A rigorous definition of a 'drawing mechanism' is as follows.

A drawing mechanism is a function

$$q(u, k, s_{k-1}) \quad \ldots (2.1.4)$$

in $u \in \mathcal{U}$, k (a positive integer), and s_{k-1}, a sample of size k-1, such that

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

$$\sum_{u \in \mathcal{U}} q(u, k, s_{k-1}) = 1 \text{ for all } k \text{ and } s_{k-1}.$$

In the above definition, q denotes the probability of drawing the unit $u \in \mathcal{U}$ at the k-th draw, which depends on u, k and also on the outcome s_{k-1} of the previous (k-1) draws.

Corresponding to a given design D(S, P), the 'inclusion probabili of a unit U_i is defined as

$$\pi_i = \sum_{s \ni i} P_s \qquad \ldots (2.1.5)$$

where the summation is taken over all samples that contain U_i atleast once.

In finite populations the problem of estimation may be briefly described as follows. Let y be a real valued characteristic taking values Y_i on U_i of (2.1.1) for $i = 1, 2, \ldots, N$. Let us denote $\underset{\sim}{Y} = (Y_1, Y_2, \ldots, Y_N)$. Any function Θ of $\underset{\sim}{Y}$ is called a 'parametric function'. Any function T defined over a design D such that for any sample $s \in D$ the function T depends only on the values of y for the units belonging to s is termed a 'statistic'. A statistic T which is used to estimate a parametric function $\Theta(\underset{\sim}{Y})$ is called an '*estimator*' *of* Θ. *An estimator* T *of* Θ *is termed* an 'unbiased estimator' if and only if

$$E(T) = \sum_{s \in S} T_s P_s = \Theta(\underset{\sim}{Y}) \text{ for all values of } \underset{\sim}{Y}. \quad (2.1.6)$$

An estimator T which is not unbiased for Θ is called a 'biased estimator' of Θ.

In the estimation of Θ, the deviation of T_s from Θ is taken as the 'error' on the basis of the sample s. Any convex function $f(T_s - \Theta)$ is called a 'loss function', and $E(f)$ is called the 'expected loss'. A loss function most often used is $(T_s - \Theta)^2$ and the 'mean square error' is given by

$$MSE(T) = E(T_s - \Theta)^2. \qquad \ldots (2.1.7)$$

If T is unbiased for Θ, then $E(T_s - \Theta)^2$ is called the 'variance' of T which may also be written as

$$V(T) = \sum_{s \in S} T_s^2 \, P_s - \Theta^2 \qquad \ldots (2.1.8)$$

A design $D(S, P)$ together with an estimator T of Θ defined over D is called a 'sampling strategy' $H = H(S, P, T)$ for the estimation of Θ. This definition is due to Hájek (1958). The expectation, variance or mean square error of a strategy are defined as the expectation, variance or mean square error of the estimator T over D.

When T_1 and T_2 are both estimators of a parametric function $\Theta(\underset{\sim}{Y})$ and defined over a design $D(S, P)$, T_1 is said to be 'uniformly better' than T_2 if and only if

$$MSE(T_1) \leq MSE(T_2) \text{ for all } \underset{\sim}{Y}, \qquad \ldots (2.1.9)$$

with strict inequality occurring at least for one value of $\underset{\sim}{Y}$.

In a class $\mathcal{C}$ of estimators of $\Theta(\underset{\sim}{Y})$ defined over D, an estimator $T_1 \in \mathcal{C}$ is the 'best estimator' if and only if

$$MSE(T_1) \leq MSE(T_2) \text{ for all } \underset{\sim}{Y} \text{ and for all } T_2 \ldots (2.1.10)$$

different from T_1 and belonging to $\mathcal{C}$, with strict inequality occurring at least for one value of $\underset{\sim}{Y}$. In other words, T_1 is called the 'best' in $\mathcal{C}$ if it is 'uniformly better' than every other member of $\mathcal{C}$.

A strategy H_1 is said to be 'better' than another strategy H_2 if and only if

$$MSE(H_1) \leq MSE(H_2) \text{ for all } \underset{\sim}{Y}, \qquad \ldots (2.1.11)$$

with strict inequality at least once, and H_1 is called the 'best' in a class

$\mathcal{C}$ if and only if

$$MSE(H_1) \leq MSE(H_2) \text{ for all } \underset{\sim}{Y} \text{ and for all } H_2 \neq H_1 \quad \ldots (2.1.12)$$

and belonging to $\mathcal{C}$ with strict inequality at least once.

We say that an unbiased estimator T_1 of a parameteric function $\Theta(\underset{\sim}{Y})$ is 'admissible' if and only if *there does not* exist another unbiased estimator T_2 of $\Theta(\underset{\sim}{Y})$ which is better than T_1. In other words, given any other unbiased estimator $T_2 \neq T_1$, there exists at least one value of $\underset{\sim}{Y}^o$ such that

$$V(T_1)_{\text{at } \underset{\sim}{Y} = \underset{\sim}{Y}^o} < V(T_2)_{\text{at } \underset{\sim}{Y} = \underset{\sim}{Y}^o} \quad \ldots (2.1.13)$$

An estimator is said to be 'inadmissible' if it is <u>not</u> admissible.

In a class $\mathcal{C}$ of unbiased estimators of $\Theta(\underset{\sim}{Y})$, an estimator $T_1 \in \mathcal{C}$ is said to be 'hyper admissible', if given any other estimator $T_2 \in \mathcal{C}$, in every principal hyperplane of R^N (i.e. N - dimensional real space) there is at least one point at which

$$V(T_1) < V(T_2). \quad \ldots (2.1.14)$$

This definition is due to Hanurav (1965).

An estimator T is said to be possessing the property of 'linear invariance', if it is invariant under linear transformation of $\underset{\sim}{Y}$. An estimator T is called a 'regular estimator' if $V(T) = K\sigma^2$, where k is a constant and $\sigma^2 = \frac{1}{N}\left[\sum_{i=1}^{N} Y_i^2 - \frac{(\sum_{i=1}^{N} Y_i)^2}{N}\right]$.

A 'general linear estimator' in a design D (S, P) is given by

$$T_s = \sum_{i \in s} \beta_{si} Y_i, \quad \ldots (3.1.15)$$

where the coefficients β_{si} depend both on the sample s and the units to which they are attached but not on the variate values Y_i's. This definition is due to Godambe (1955).

A design D (S, P) is said to be a 'unicluster design' if for any two samples s_i and s_j of D, either both contain the same units of the population or they do not have any common unit.

If auxiliary information on a characteristic x closely related to the study variate y is available beforehand, it may sometimes be used to assume a reasonable a-priori distribution over $\underset{\sim}{Y}$. According to this 'super-population concept' (Cochran 1946), $\underset{\sim}{Y}$ is assumed to be a realisation of a *random N-vector with certain distribution depending on* $\underset{\sim}{X} = (X_1 , X_2 , \ldots , X_N)$ and unknown parameters. We can then talk of expectations, variances and covariances taken over this a-priori distribution. Comparison among estimators may then be attempted on the basis of their expected mean square errors.

Apart from the use mentioned above, it is a common practice for survey statisticians to make use of auxiliary information in P(design) or T (estimator) or both. Examples of such uses are stratification of the population, selection of sample units with varying probabilities and estimation procedures like ratio, product and regression methods. In the ratio method, an auxiliary variate x positively correlated with the variate under study y is chosen and, generally, the ratio (R) of the population totals Y and X of the variates y and x is estimated by the ratio ($\hat{R}$) of the unbiased estimators $\hat{Y}$ and $\hat{X}$ of Y and X in the specific sampling design adopted for the selection of the sample. When the population total Y is of interest, $\hat{R}$ is multiplied by the known population total X to obtain the ratio estimator, $\hat{Y}_R = \frac{\hat{Y}}{\hat{X}} X$. Product method essentially uses the information on an auxiliary variate x negatively correlated with y to provide an estimator $\hat{P} = \hat{Y}\hat{X}$ for the population product $P = YX$; and if Y is of interest $\hat{P}$ is divided by the known population total X giving the product estimator $\hat{Y}_P = \frac{\hat{Y}\hat{X}}{X}$. In the regression method, a supplementary variate x having either positive or negative correlation with y may be used; and the unbiased estimators $\hat{Y}$ and $\hat{X}$ and the known population total X are combined into a linear form with the use of an estimator $\hat{\beta}$ of the population regression coefficient β of $\hat{Y}$ on $\hat{X}$ to provide a regression estimator $\hat{Y}_r = \hat{Y} + \hat{\beta}(X - \hat{X})$ for Y. If, however, we consider linear estimators of the form $\hat{Y}_l = \hat{Y} + \hat{t}(X - \hat{X})$ where $\hat{t}$ is a real valued function defined on the sample, it is clear that the choices $\hat{t} = 0$, $\frac{\hat{Y}}{\hat{X}}$, $-\frac{\hat{Y}}{X}$, and $\hat{\beta}$ respectively result in the estimators $\hat{Y}$, $\hat{Y}_R$, $\hat{Y}_P$ and $\hat{Y}_r$.

2.2. Biased ratio estimators:

In general, for several of the commonly used sampling schemes, the ratio estimator $\hat{R}$ (or $\hat{Y}_R$) is biased. Hartley and Ross (1954) have shown that $\frac{|\text{bias in } \hat{R}|}{\sigma(\hat{R})} \leq C(\hat{X})$ where $\sigma(\hat{R})$ is the standard error of $\hat{R}$ and $C(\hat{X})$ is the coefficient of variation of $\hat{X}$. This result suggests that the bias in $\hat{R}$ is negligible in relation to the standard error if $C(\hat{X})$ is small, which is likely to be the case when the sample size is sufficiently large. The bias in $\hat{R}$ vanishes if the regression of $\hat{Y}$ on $\hat{X}$ is linear and passes through the origin, indicating that the ratio method is appropriate in situations where $\hat{Y}$ and $\hat{X}$ are approximately proportional. Writing $\hat{Y} = Y(1+e)$ and $\hat{X} = X(1+e')$, under the assumption $|e'| < 1$, it is easily seen that

$$\text{Bias in } \hat{R} = R\left[(v_{20} - v_{11}) + (v_{21} - v_{30}) + \ldots\right] \quad \ldots (2.2.1)$$

where $$v_{ij} = \frac{E(\hat{X} - X)^i (\hat{Y} - Y)^j}{X^i \; Y^j}$$

For a number of basic sampling schemes, v_{ij} takes the form $\frac{\theta_{ij}}{n^{i+j-1}}$ where θ_{ij} is a population parameter independent of the sample size n, so that the contribution to the bias from the higher order terms in (2.2.1) is expected to be negligible in large samples. The expression for the mean square error is given by

$$MSE(\hat{R}) = R^2\left[(v_{02} - 2v_{11} + v_{20}) - 2(v_{12} - 2v_{21} + v_{30}) + 3(v_{22} - 2v_{31} + v_{40}) + \ldots\right] \quad \ldots (2.2.2)$$

and the contribution from higher order terms may be neglected in large samples. If only the leading term in the mean square error is considered $\hat{Y}_R = \hat{R}X$ is a more efficient estimator of the population total Y than $\hat{Y}$ provided the correlation $\rho(\hat{X}, \hat{Y}) > \frac{1}{2} \frac{C(\hat{X})}{C(\hat{Y})}$ when R is positive.

The assumption $|e'| < 1$, under which (2.2.1) and (2.2.2) are derived, implies that all the possible estimates $\hat{X}$ of X lie in the range (0, 2X). It is likely to hold good only when the coefficient of variation of $\hat{X}$ is not large and to ensure that the sample size has to be fairly large. Assuming that all the x values are of same sign (which is true in most practical situations), Koop (1951) obtained in simple random sampling without replacement the expression for the bias in the ratio estimators upto the order n^{-4} which coincides with (2.2.1). Recently Sukhatme and David (1973) have examined Koop's approach and shown that the assumption $|e'| < 1$ is still necessary.

Application of the above large sample theory to the stratified sampling schemes suggests that, when the ratios of stratum totals of y and x vary much among the strata, it is difficult to choose a 'combined' ratio estimator in preference to a 'separate' ratio estimator built up from stratum level ratio estimators from the variance point of view, although the 'combined' estimator is likely to be less biased than the 'separate' estimator. At the same time when the number of strata is large resulting in small or moderate sample sizes within strata, the bias in the 'separate' ratio estimator relative to its standard error may not be negligible. Such situations have emphasised the need to explore the possibilities of either reducing or completely eliminating the bias in the classical ratio estimators and in the recent past considerable attention has been paid by several authors to develop the theory of approximately unbiased (i.e. estimators with a smaller bias than the classical ratio estimators) or unbiased ratio estimators.

2.3. Approximately unbiased ratio estimators:

In simple random sampling Quenouille (1956) has outlined a method of adjustment to a broad class of biased estimators whose bias is of the form $an^{-1} + O(n^{-2})$, 'a' being a constant independent of the sample size n, such that the adjusted estimator has bias of order at most n^{-2} and has variance differing from the variance of the unadjusted estimator only in terms of order n^{-2} and above. When applied to the classical ratio estimator $\hat{R}$, his method consists of dividing the sample at random into g groups each of size m (n = mg) and computing from the sample after omitting each of the g groups in turn the classical ratio estimator $R'_{n-m}(i)$ for i = 1, 2, ..., g and forming the estimator:

$$\hat{R}_{Q} = g\hat{R} - \frac{g-1}{g} \sum_{i=1}^{g} R'_{n-m}(i) \qquad \ldots (2.3.1)$$

Assuming g = 2, Durbin (1959) has studied the mean square error of $\hat{R}_{Q}$ in simple random sampling under two super-population models:

Model - 1: (i) population is infinite, $y_j = \alpha + \beta x_j + u_j$, $E(u_j / x_j) = 0$,

$E(u_j^2 / x_j) = \delta$ (a constant), $E(u_j u_{j'} / x_j, x_{j'}) = 0$ for $j \neq j'$ and

(ii) x is a normal variate ;

Model - 2 : (i) as in model 1 and

(ii) x is a gamma variate.

For model 1, $\hat{R}_{Q}$ has smaller mean square error than $\hat{R}$ in large samples; while for model 2 this result is true for any sample size provided the coefficient of variation of the sample mean of x is less than 0.25.
Rao (1965) and Rao and Webster (1966) have extended Durbin's results under the same super-population models and showed that g = n would be the

optimum choice for the number of groups. Rao and Rao (1971 a) have shown that when Durbin's model 2 is slightly modified by assuming that the finite population of size N is a sample from a super-population with the properties (i) $y_j = \alpha + \beta x_j + u_j$, ($j = 1, 2, \ldots, N$), $E(u_j / x_j) = 0$, $E(u_j^2 / x_j) = \delta$ (δ constant), $E(u_j u_{j'} / x_j, x_{j'}) = 0$ ($j \neq j'$) and (ii) x is a gamma variate, Durbin's result on the relative performance of $\hat{R}$ and $\hat{R}_Q$ (with g = 2) is no longer valid. Instead we may consider a slight modification of $\hat{R}_Q$ due to Jones (1963):

$$\hat{R}_{MQ} = W\hat{R} - (W-1)\frac{1}{g}\sum_{i=1}^{g} R'_{n-m}(i), \quad W = g-(g-1)\frac{n}{N} \ldots (2.3.1a)$$

Under the above model, $\hat{R}_{MQ}$ (with g = 2) has smaller mean square error than $\hat{R}$. It may be noted that the asymptotic bias of $\hat{R}_{MQ}$ does not contain terms of order n^{-1} and order N^{-1} while the asymptotic bias of $\hat{R}_Q$ contains terms of order N^{-1} Further, when n = N, $\hat{R}_{MQ}$ reduces to the population ratio $R = Y/X$ whereas $\hat{R}_Q$ does not have this property. Hence it is likely that for finite populations Jones' (1963) modification $\hat{R}_{MQ}$ may be of more advantage than $\hat{R}_Q$.

When the sample is split at random into a number of groups, another possibility is to compute the classical ratio estimator $R_m(i)$ for each of the groups of size m and an estimator similar to $\hat{R}_Q$ is obtained as:

$$\hat{R}_{Q'} = \frac{g}{g-1}\hat{R} - \frac{1}{g(g-1)}\sum_{i=1}^{g} R_m(i) \qquad (2.3.1b)$$

This estimator has been proposed by Tin (1965) and earlier to him by Hartley (see Pascual, 1961) with g = n. Clearly $\hat{R}_{Q'} = \hat{R}_Q$ for g = 2.

Bias in $\hat{R}_{Q'}$ does not contain terms of order n^{-1} but contains terms of order N^{-1}. Similar to Jones' (1963) modification of $\hat{R}_Q$ (2.3.1), a modification of $\hat{R}_{Q'}$ is:

$$\hat{R}_{MQ'} = W^{*}\hat{R} - (W^{*} - 1)\frac{1}{g}\sum_{j=1}^{g} R_m(i) \qquad \ldots (2.3.1c)$$

where $\quad W^{*} = (1 - \frac{n}{Ng})\frac{g}{g-1}$.

When $n = N$, $\hat{R}_{MQ'}$ reduces to $R = Y/X$. Bias of $\hat{R}_{MQ'}$ does not contain terms of order n^{-1} and of order N^{-1}

Along with $\hat{R}_{Q'}$, Tin (1965) has proposed another estimator:

$$\hat{R}_T = \frac{\bar{y}}{\bar{x}}\left[1 + \left(\frac{1}{n} - \frac{1}{N}\right)\left(\frac{s_{xy}}{\bar{x}\,\bar{y}} - \frac{s_x^2}{\bar{x}^2}\right)\right] \qquad \ldots (2.3.2)$$

and also included in his study the estimator due to Beale (1962):

$$\hat{R}_B = \frac{\bar{y}}{\bar{x}}\left[1 + \left(\frac{1}{n} - \frac{1}{N}\right)\frac{s_{xy}}{\bar{x}\,\bar{y}}\right] \Big/ \left[1 + \left(\frac{1}{n} - \frac{1}{N}\right)\frac{s_x^2}{\bar{x}^2}\right] \qquad \ldots (2.3.3)$$

where $\bar{y}$ and $\bar{x}$ are the sample means of y and x; $s_x^2 = \frac{1}{n-1}\sum_{j=1}^{n}(x_j - \bar{x})^2$ and $s_{xy} = \frac{1}{n-1}\sum_{j=1}^{n} x_j(y_j - \bar{y})$. The asymptotic biases of both $\hat{R}_T$ and $\hat{R}_B$ do not contain terms of order n^{-1} and of order N^{-1} In large samples, of the estimators $\hat{R}$, $\hat{R}_{Q'}$ $(g = 2)$, $\hat{R}_T$ and $\hat{R}_B$. Tin has shown that $\hat{R}_T$ and $\hat{R}_B$ are better than others with regard to bias, precision and approach to normality.

Again in simple random sampling without replacement, Pascual (1961) has approximately estimated the bias in $\hat{R}$ and used it to obtain an approximately unbiased ratio estimator:

$$\hat{R}_P = \hat{R} + \frac{(N-1)}{(n-1)\,\bar{X}} \left(\bar{y} - \bar{r}\,\bar{x} \right), \qquad \ldots (2.3.4)$$

where $\bar{r}$ is the sample mean of $r = y/x$. Making a rather restrictive assumption: $\left| \frac{x - \bar{X}}{\bar{X}} \right| < 1$, where $\bar{X}$ is the population mean of x, he has shown that $\hat{R}_P$, $\hat{R}_Q$, (2.3.1b) with $g = n$, and $\hat{R}$ have the same large sample variance.

Quenouille's technique of bias reduction in simple random sampling has been extended on similar lines to general sampling designs by Murthy and Nanjamma (1959) with the use of independent interpenetrating sub-samples. Thus when $[\hat{Y}_i]$ and $[\hat{X}_i]$, $(i = 1, 2, \ldots, k)$, are unbiased estimators of the population totals Y and X based on k independent (interpenetrating) sub-samples of the same size and selected according to the same sample design, the classical ratio estimators that may be computed are:

$$\hat{R}_1 = \frac{\hat{Y}}{\hat{X}}, \quad \left(\hat{Y} = \frac{1}{k} \sum_{i=1}^{k} \hat{Y}_i, \; \hat{X} = \frac{1}{k} \sum_{i=1}^{k} \hat{X}_i \right), \qquad \ldots (2.3.5)$$

$$\text{and} \quad \hat{R}_k = \frac{1}{k} \sum_{i=1}^{k} \frac{\hat{Y}_i}{\hat{X}_i}$$

Provided the sub-sample size is large, it can be shown that bias in $\hat{R}_k$ is approximately 'k' times the bias in $\hat{R}_1$. Using this result, Murthy and Nanjamma have suggested $(\hat{R}_k \hat{X} - \hat{Y}) / (k-1)\hat{X}$ as an approximate estimate of bias in $\hat{R}_1$; thereby getting an approximately unbiased ratio estimator:

$$\hat{R}_{MN} = \hat{R}_1 + \frac{(\hat{Y} - \hat{R}_k \hat{X})}{(k-1)\,\hat{X}} \qquad \ldots (2.3.7)$$

In simple random sampling, the similarity between $\hat{R}_{MN}$ with k = g and $\hat{R}_{Q'}$ (2.3.1b) may easily be recognised. Murthy and Nanjamma have investigated the mean square errors of $\hat{R}_{MN}$ and $\hat{R}_1$ upto some higher order terms and derived the conditions under which the corrected estimator $\hat{R}_{MN}$ is more efficient than the un-corrected estimator $\hat{R}_1$.

Pascual's (1961) *method of adjustment for bias in simple random* sampling (2.3.4) may also be extended to any general sampling design by taking $(\hat{R}_k \hat{X} - \hat{Y}) / (k-1) X$ as an approximate estimate of bias in $\hat{R}_1$ and thereby getting an approximately unbiased ratio estimator:

$$\hat{R}_{P'} = \hat{R}_1 + \frac{(\hat{Y} - \hat{R}_k \hat{X})}{(k-1) X} \qquad \ldots (2.3.8)$$

Clearly $\hat{R}_{MN}$ and $\hat{R}_{P'}$ differ only in the denominator of their second term (i.e. X is used in $\hat{R}_{P'}$ in place of $\hat{X}$ used in $\hat{R}_{MN}$).

2.4. Unbiased ratio estimators:

We now attempt a brief review of the development of ratio (or ratio type) estimators which are unbiased.

In simple random sampling without replacement, Hartley and Ross (1954) have corrected the bias in the mean of ratios estimator $\bar{r}$ and obtained an unbiased ratio type estimator:

$$\hat{R}_{HR} = \bar{r} + \frac{n(N-1)}{N(n-1)} \cdot \frac{1}{\bar{X}} (\bar{y} - \bar{r}\bar{x}) \qquad \ldots (2.4.1)$$

Robson (1957) has derived the exact variance of $\hat{R}_{HR}$ valid for any sample size. Goodman and Hartley (1958) have observed that in large samples $\hat{R}_{HR}$ is more efficient than $\hat{R} = \frac{\bar{y}}{\bar{x}}$ if and only if the slope of the population regression line of y on x is closer to $\frac{1}{N}\sum_{i=1}^{N}\frac{y_i}{x_i}$ than to $R = \bar{Y}/\bar{X}$. This condition, in general, is rather restrictive since one would ordinarily use a ratio estimator only when the regression coefficient is expected to be near the population ratio R. Consequently, $\hat{R}_{HR}$ is likely to be less efficient than $\hat{R}$ in large samples.

Rao, T.J. (1966) has considered in simple random sampling without replacement linear combinations of the two biased ratio estimators $\hat{R}$ and $\bar{r}$, where the weights attached are either constants or random variables, and obtained the conditions under which a linear combination yields an almost unbiased or unbiased ratio estimator. As examples of this approach he has derived the almost unbiased ratio estimators due to Murthy and Nanjamma (2.3.7) and Pascual (2.3.4); and the Hartley-Ross unbiased ratio estimator (2.4.1).

Mickey (1959) has outlined a general theory for the construction of unbiased estimators of population means or totals using auxiliary information in simple random sampling without replacement. In particular, he has obtained a class of unbiased ratio type estimators given by

$$t_m = R_m + \frac{n(N-m)}{N(n-m)} \cdot \frac{1}{\bar{X}} (\bar{y} - R_m \bar{x}), \qquad (2.4.2)$$

where R_m is the classical ratio estimator $\hat{R}$ computed on any sub-set (Z_m) of m sample elements ($1 \leqslant m \leqslant n-1$) chosen out of the total sample of n elements. Z_m may be taken as the set of m elements occurring in the first m draws of the sample from the population or as any sub-sample from the total sample, etc. Williams (1961, 1962) has suggested the division of the sample at random into g groups each of size m (n = mg) and computation of Mickey's estimator t_m treating each group in turn as Z_m, and then taking the average of the g such estimators due to Mickey as an unbiased estimator of the population ratio. This estimator may be represented as

$$\bar{t}_m = \bar{R}_m + \frac{n(N-m)}{N(n-m)} \cdot \frac{1}{\bar{X}} (\bar{y} - \bar{R}_m \bar{x}), \qquad (2.4.3)$$

where $\bar{R}_m = \frac{1}{g} \sum_{i=1}^{g} R_m(i)$, and $R_m(i)$ is the classical ratio estimator $\hat{R}$ computed on the i-th group of size m.

Rao (1967) has also considered the splitting of the sample at random into groups, but treated as Z_{n-m} a set of elements obtained by omitting a particular group of m elements from the total sample, and corresponding to the g possible sets Z_{n-m} thus obtained g Mickey type estimators t_{n-m} have been computed, and their average has been taken as an unbiased estimator of the population ratio. This estimator may be represented by

$$\bar{t}'_{n-m} = \bar{R}'_{n-m} + \frac{n(N-n+m)}{N m \bar{X}} (\bar{y} - \bar{R}'_{n-m} \bar{x}), \ldots \quad (2.4.4)$$

where $\bar{R}'_{n-m} = \frac{1}{g} \sum_{i=1}^{g} R'_{n-m}(i)$ and $R'_{n-m}(i)$ is the classical ratio estimator $\hat{R}$ computed on the set Z_{n-m} obtained by omitting i-th group of size m from the total sample.

For any choice of m in the range 1 to n - 1, clearly there are $\binom{n}{m}$ distinct sub-sets Z_m and one may take the average of the $\binom{n}{m}$ Mickey type estimators t_m which may be obtained corresponding to the $\binom{n}{m}$ subsets Z_m. Mickey has denoted this estimator by

$$t^*_m = R^*_m + \frac{n(N-m)}{N(n-m)} \cdot \frac{1}{\bar{X}} (\bar{y} - R^*_m \bar{x}), \qquad \ldots (2.4.5)$$

where $$R^*_m = \frac{1}{\binom{n}{m}} \sum_{Z_m}^{\binom{n}{m}} R_m .$$

It is easy to see that t^*_m is obtained by averaging $\bar{t}_m$ (2.4.3) over all possible splits of the sample into g groups each of size m. Similarly average of $\bar{t}'_{n-m}$ (2.4.4) over all possible splits of the sample into g groups each of size m results in t^*_{n-m} (obtained by changing m to n-m in (2.4.5)). It may also be noted that for any specific choice of Z_m the corresponding estimator t_m (2.4.2) or for any specific split of the sample into g groups the corresponding estimators $\bar{t}_m$ (2.4.3) and $\bar{t}'_{n-m}$ (2.4.4) are all examples of ordered estimators (i.e. estimators which depend upon the order of units in the sample), whereas t^*_m or t^*_{n-m} are the corresponding unordered estimators not depending on the order in the sample. Consequently, from the results of Murthy (1957) and Basu (1958) on ordered and the corresponding unordered estimators, it is noted that t^*_m is more efficient than t_m or $\bar{t}_m$, and similarly t^*_{n-m} is more efficient than $\bar{t}'_{n-m}$. For the choice m = 1, however, we have $\bar{t}_1 = t^*_1$ = Hartley-Ross ratio estimator $\hat{R}_{HR}$ (2.4.1); and $\bar{t}'_{n-1} = t^*_{n-1}$.

On the precision of Mickey's unbiased ratio estimators there have been a few studies by Rao (1967), Rao and Beegle (1967), Rao (1969), Rao and Rao (1971 b), Hutchison (1971), and Rao and Kuzik (1973). Assuming Durbin's (1959) super-population model 1 (see section 2.3), Rao (1967) has observed that in large and moderately large samples, for the estimators $\bar{t}'_{n-m}$ (2.4.4) and $\hat{R}_Q$(2.3.1), the optimum number of groups into which the sample is to be divided is equal to the size of the sample (i.e., $g = n$). Further, with $g = n$, the asymptotic variance of $\bar{t}'_{n-m}$ is slightly smaller than the mean square error of the classical ratio estimator $\hat{R}$ but is slightly larger than the mean square of $\hat{R}_Q$ with $g = n$. In the same study he has also shown that under Durbin's (1959) super-population model 2, for any sample size n, $g = n$ is again the optimum number of groups for $\bar{t}'_{n-m}$ and $\hat{R}_Q$. Further under this model (i) $\bar{t}'_{n-1}$ is considerably more efficient than Hartley-Ross unbiased ratio estimator $\hat{R}_{HR}$ (2.4.1) for $n > 2$; (ii) $\bar{t}'_{n-1}$ is slightly better than the classical ratio estimator $\hat{R}$ for $n \geq 8$; and (iii) $\hat{R}_Q$ with $g = n$ and Tin's estimator $\hat{R}_T$ (2.3.2) are better than $\hat{R}$, $\hat{R}_{HR}$, $\bar{t}'_{n-1}$, $\hat{R}_{Q'}$ (2.3.1b) and $\hat{R}_P$ (2.3.4). Rao and Beegle (1967) have conducted a Monte Carlo study of the eight ratio estimators $\hat{R}$, $\hat{R}_{HR}$, $\hat{R}_Q$, $\hat{R}_{Q'}$, $\hat{R}_T$, $\hat{R}_B$, $\bar{t}'_{n-m}$ and $\hat{R}_P$ for small and moderate sample sizes using Durbin's (1959) super-population model 1. Their study essentially corroborates the above theoretical results obtained under Durbin's super-population model 2.

Rao (1969) has also conducted an empirical investigation on the relative performance of nine ratio estimators $\hat{R}$, $\hat{R}_T$, $\hat{R}_B$, $\hat{R}_P$, $\hat{R}_{HR}$ and with $g = n$ the estimators $\hat{R}_Q$, $\hat{R}_{MQ}$ (2.3.1a), $\hat{R}_{MQ'}$ (2.3.1c) and $\bar{t}'_{n-m}$, along with some regression estimators and variance estimators, using small and moderate sample sizes and several sets of data representing real populations. From this study, he has observed that from the point of view of relative bias (in relation to mean square error) and mean square error the estimators $\hat{R}_{MQ}$ with $g = n$, $\hat{R}_T$ and $\hat{R}_B$ are likely to be promising. If, however, an unbiased estimator is needed, then $\bar{t}_{n-1}$ which is better than $\hat{R}_{HR}$ seems to be promising in the sense that it compares favourably with the classical estimator $\hat{R}$ in precision, although the latter appears slightly better (particularly for small n). Essentially similar conclusions have been arrived at in the other studies by Rao and Rao (1971 b), Hutchison (1971), and Rao and Kuzik (1973). At the same time, Rao (1969) has pointed out that the above observations are still tentative and it is necessary to continue both theoretical and empirical investigations to gain more understanding of these estimators.

The technique by which Hartley and Ross (1954) have corrected the bias in the ratio estimator $\bar{r}$ in simple random sampling can be easily extended to any general sampling design making use of independent interpenetrating sub - samples. Thus when the total sample is selected in terms of k independent sub - samples, with the definitions of $\hat{R}_1$ (2.3.5) and $\hat{R}_k$ (2.3.6), we note that the bias in $\hat{R}_k$ may be corrected and an unbiased ratio estimator is given by

$$\hat{R}_k' = \hat{R}_k + \frac{k}{k-1} \frac{(\hat{Y} - \hat{R}_k \hat{X})}{X} \qquad \ldots (2.4.6)$$

It can be easily shown that $\hat{R}_k'$, $\hat{R}_k$ and $\hat{R}_1$ are all equally efficient in large samples.

So far we have reviewed methods of getting unbiased ratio estimators in commonly adopted sampling schemes. We now consider a second approach in which the usual simple form of the ratio estimator is preserved but the sampling scheme is slightly modified so as to make the ratio estimator unbiased. Lahiri (1951) has suggested that the biased classical ratio estimator $\hat{R} = \bar{y}/\bar{x}$ in simple random sampling without replacement would become unbiased for the population ratio, if the sampling design is so modified that the probability of selecting a sample is proportional to the sample total (or mean) of the characteristic x appearing in the denominator of the ratio. This can be achieved by selecting one unit with probability proportional to its x value and the rest of the sample with the original sampling design, i.e., with equal probability and without replacement, from the remaining units of the population [Midzuno (1952), Sen (1952)].

The above technique of changing the selection procedure so as to obtain unbiased ratio estimators has been further explored by Des Raj (1954) and by Nanjamma, Murthy and Sethi (1959), who have given the actual modifications needed in several of the commonly adopted schemes such as equal probability sampling, varying probability sampling, systematic sampling, stratified and multistage sampling etc. In many of these situations the modification is obtained by selecting the first unit with probability proportional to its value of the characteristic occurring in the denominator of the ratio and then choosing the remaining units according to the original scheme of sampling. In large samples the unbiased ratio estimator and the corresponding biased ratio estimator in the original design are of equal precision. When, however, higher order terms in the approximations for the variance/mean square error are retained upto fourth degree, in large or moderately large samples the unbiased ratio estimator is more efficient than the corresponding biased ratio estimator in the original scheme, if the estimators of the numerator and the denominator of the ratio are distributed in the Bivariate Normal form for the original sampling scheme.

For Lahiri's unbiased ratio estimator in Midzuno - Sen sampling scheme, Rao, P.S.R.S. (1968) has shown that its expected variance is less than the expected mean square error of the classical biased ratio estimator in simple random sampling without replacement for any sample size under the following super - population model:

$$y_j = \alpha + \beta x_j + u_j \ , \ (j = 1, 2, \ldots, N), \ E (u_j / x_j) = 0,$$

$$E (u_j u_{j'} / x_j , x_{j'}) = 0 (j \neq j'), \ E (u_j^2 / x_j) = \delta x_j^{\ell} \ (0 \leq \ell \leq 2),$$

and x is a gamma variate,

We now give a brief account of some developments in product method of estimation.

2.5. Biased product estimators :

As an estimator of the population total Y, the product estimator $\hat{Y}_P = \frac{\hat{Y}\hat{X}}{X}$ defined in section - 2.1 is biased and its bias: $\frac{\text{Cov}(\hat{X}, \hat{Y})}{X}$ decreases for many of the commonly adopted sampling schemes when the sample size is increased. It can be easily shown that the exact mean square error of the product estimator is given by

$$MSE(\hat{Y}_p) = Y^2 \left[(v_{02} + 2v_{11} + v_{20}) + 2(v_{12} + v_{21}) + v_{22} \right], \ldots (2.5.1)$$

where v_{ij} is as defined at (2.2.1). For many of the usual sampling schemes, v_{ij} is of the form Θ_{ij} / n^{i+j-1} where Θ_{ij} is a population parameter independent of the sample size n. Hence in large samples the contribution to the mean square error from the terms involving moments of order greater than two may be neglected. In that case, we note that $MSE(\hat{Y}_p)$ is less than $V(\hat{Y})$ provided the correlation $\rho(\hat{X}, \hat{Y}) < -\frac{1}{2} \frac{C(\hat{X})}{C(\hat{Y})}$ when $P = YX$ is positive. Combining this result with the similar result in ratio method, one may decide whether to use the supplementary variate x for obtaining a ratio estimator or product estimator according as $\rho(\hat{X}, \hat{Y})$ is greater than $\frac{C(\hat{X})}{2C(\hat{Y})}$ or less than $-\frac{C(\hat{X})}{2C(\hat{Y})}$. The variance and variance estimators of product of estimators have been studied by Goodman (1960). Murthy (1964) and Singh (1965, 1967) have discussed in detail the technique of product method of estimation.

2.6. Unbiased product estimators:

The motivation for the development of unbiased product estimators is the same as in the case of ratio estimators. In simple random sampling without replacement, Robson (1957) has corrected the bias in the product estimator $\hat{Y}_P = \frac{N \bar{y} \bar{x}}{\bar{X}}$ and obtained an unbiased product type estimator:

$$\hat{Y}_{RP} = \frac{N\bar{y}\bar{x}}{\bar{X}} - \frac{N-n}{N} \cdot \frac{s_{xy}}{\bar{X}} \qquad \ldots (2.6.1)$$

He has also derived the exact variance of $\hat{Y}_{RP}$.

For a general sampling design, Murthy (1964) has given an unbiased product estimator based on independent interpenetrating sub-samples. The method is similar to the method of obtaining almost unbiased ratio estimators (2.3.7). In this case, however, the total bias in the product estimator is eliminated by the type of correction suggested. Using the same definitions for $\hat{Y}_i$, $\hat{X}_i$, $\hat{Y}$ and $\hat{X}$ as at (2.3.5), Murthy's unbiased product estimator is given by

$$\hat{Y}_{MP} = \frac{k}{k-1} \frac{\hat{Y}\hat{X}}{X} - \frac{1}{k(k-1)} \sum_{i=1}^{k} \frac{\hat{Y}_i \hat{X}_i}{X} \qquad \ldots (2.6.2)$$

where k is the number of independent interpenetrating sub-samples. In simple random sampling with $k = n$ and sub-sample size one, the estimator $\hat{Y}_{MP}$ reduces to

$$\hat{Y}_{MP}\ (SRS) = \frac{N\bar{y}\bar{x}}{\bar{X}} - \frac{N}{n} \frac{s_{xy}}{\bar{X}}, \qquad \ldots (2.6.3)$$

which is easily seen to be the 'with replacement' version of Robson's 'without replacement' estimator $\hat{Y}_{RP}$ (2.6.1). The conditions under which $\hat{Y}_{MP}$ (2.6.2) is more efficient than $\hat{Y}_P = \frac{\hat{Y}\hat{X}}{X}$ are the same as those given by Murthy and Nanjamma (1959) in the case of obtaining an almost unbiased ratio estimator.

2.7. Biased regression estimators:

The classical regression estimator: $\hat{Y}_r = \hat{Y} + \hat{\beta}(X - \hat{X})$, mentioned in section 2.1, is biased in estimating the population total Y ; and the bias: $-\text{Cov}(\hat{X}, \hat{\beta})$ usually decreases for many of the commonly used sampling schemes as the sample size increases. When the sample size is large, to the first order of approximation,

$$MSE(\hat{Y}_r) = V(\hat{Y}) [1 - \rho^2(\hat{X}, \hat{Y})]. \qquad \ldots (2.7.1)$$

Hence in large samples $\hat{Y}_r$ is more efficient than the conventional estimator $\hat{Y}$ if $\rho(\hat{X}, \hat{Y})$ is non-zero, a condition less restricted than that for the classical ratio or product estimator to be more efficient than $\hat{Y}$ In large samples it also follows that the classical regression estimator is more efficient than either the classical ratio or product estimator. When, however, the line of regression of $\hat{Y}$ on $\hat{X}$ passes through the origin, the regression and the corresponding ratio estimators are equally efficient.

2.8 Almost unbiased regression estimators:

In a general sampling design, Murthy (1962) has considered the question of obtaining almost unbiased (i.e., unbiased only to the second degree of approximation) estimators of non-linear parametric functions making use of independent interpenetrating sub-samples. As an illustration of this technique, an almost unbiased regression estimator of the population total Y has been given by him. Based on i-th sub-sample, $(i=1, 2, \ldots, k)$, let $\hat{Y}_i$, $\hat{X}_i$, $\hat{Y}$ and $\hat{X}$ be as defined at (2.3.5). Further, let $\widehat{\text{Cov}}(\hat{Y}_i, \hat{X}_i)$, and $\hat{V}(\hat{X}_i)$ be unbiased estimators of $\text{Cov}(\hat{Y}_i, \hat{X}_i)$ and $V(\hat{X}_i)$; and denote

$$\hat{\beta}_i = \frac{\hat{Cov}(\hat{Y}_i, \hat{X}_i)}{\hat{V}(\hat{X}_i)} \quad \text{and} \quad \hat{\beta}' = \frac{\sum_{i=1}^{k} \hat{Cov}(\hat{Y}_i, \hat{X}_i)}{\sum_{i=1}^{k} \hat{V}(\hat{X}_i)} \qquad \ldots (2.8.1)$$

Then two biased regression estimators that may be computed are:

$$\hat{Y}_{1r} = \hat{Y} + \frac{1}{k} \sum_{i=1}^{k} \hat{\beta}_i (X - \hat{X}_i),$$

$$\ldots (2.8.2)$$

$$\text{and} \quad \hat{Y}_{kr} = \hat{Y} + \hat{\beta}' (X - \hat{X}).$$

With this set-up, Murthy's (1962) almost unbiased regression estimator is given by

$$\hat{Y}_{cr} = \frac{k\hat{Y}_{kr} - \hat{Y}_{1r}}{k-1} \qquad \ldots (2.8.3)$$

The conditions under which the corrected estimator $\hat{Y}_{cr}$ is more efficient than the uncorrected estimator $\hat{Y}_{kr}$ have also been given by Murthy (1962).

2.9. Unbiased regression estimators:

In simple random sampling without replacement, unbiased ratio estimators, obtained as particular cases of Mickey's (1959) general class of unbiased estimators, have been reviewed in section 2.4. Here we briefly review unbiased regression estimators which may also be derived as special cases of Mickey's general class of estimators. For the population total Y, Mickey's basic unbiased regression estimator is of the form:

$$d_m = N\left[\bar{y} + b_m(\bar{X} - \bar{x})\right] + \frac{m(N-n)}{(n-m)}\left[\bar{y} - \{\bar{y}_m + b_m(\bar{x} - \bar{x}_m)\}\right],$$

$$\ldots (2.9.1)$$

where $\bar{y}_m$, $\bar{x}_m$ and b_m are respectively the mean of y, mean of x, and the linear regression coefficient of y on x, computed on a sub-set Z_m of m sample elements ($2 \leq m \leq n-1$) chosen out of the total sample of size n.

If the sample is split at random into g groups each of size m ($n = mg$) and the estimators d_m (2.9.1) are computed treating each of the g groups in turn as Z_m, and the average of the g such estimators is taken; we obtain William's (1961) unbiased regression estimator. This estimator may be represented as

$$\bar{d}_m = N \left[\bar{y} + \bar{b}_m (\bar{X} - \bar{x}) \right] + \frac{(N-n)}{g(g-1)} \left[\sum_{i=1}^{g} b_m^{(i)} \bar{x}_m^{(i)} - g \, \bar{b}_m \bar{x} \right], \quad \ldots (2.9.2)$$

where $\bar{x}_m^{(i)}$ and $b_m^{(i)}$ correspond to the i-th group ($i = 1, 2, \ldots, g$) of size m; and $\bar{b}_m = \frac{1}{g} \sum_{i=1}^{g} b_m^{(i)}$.

When the sample is split into groups, alternately one may treat a set of n - m sample elements obtained by omitting a particular i-th group of size m from the total sample as Z_{n-m}, and compute Mickey's unbiased regression estimator (2.9.1) corresponding to each of the g possible sets Z_{n-m} and take the average of g such estimators. This estimator, after some simplification, would be

$$\bar{d}'_{n-m} = N \left[\bar{y} + \bar{b}'_{n-m} (\bar{X} - \bar{x}) \right] - \frac{(N-n)}{g} \left[\sum_{i=1}^{g} b'^{(i)}_{n-m} \bar{x}_m^{(i)} - g \, \bar{b}'_{n-m} \bar{x} \right], \quad \ldots (2.9.3)$$

where $b'^{(i)}_{n-m}$ is the linear regression coefficient computed on a set Z_{n-m} obtained by the omission of i-th group of size m from the sample, and

$\bar{b}'_{n-m} = \frac{1}{g} \sum_{i=1}^{g} b'^{(i)}_{n-m}$. Rao (1969) has included $\bar{d}'_{n-m}$ with $m = 1$ (i.e. $g = n$) in his empirical study on ratio and regression estimators.

For the $\binom{n}{m}$ possible selections of sub-sets Z_m out of the sample, the estimator d_m (2.9.1) may be computed and the average of these $\binom{n}{m}$ estimators may be represented by

$$d^*_m = N \left[\bar{y} + b^*_m (\bar{X} - \bar{x}) \right] + \frac{m(N-n)}{(n-m)} \cdot \frac{1}{\binom{n}{m}} \left[\sum_{Z_m} b_m \bar{x}_m - \binom{n}{m} b^*_m \bar{x} \right], \quad \ldots (2.9.4)$$

where $b^*_m = \frac{1}{\binom{n}{m}} \sum_{Z_m}^{\binom{n}{m}} b_m$.

As stated in section 2.4, by averaging $\bar{d}_m$ (2.9.2) and $\bar{d}'_{n-m}$ (2.9.3) over all possible splits of the sample into groups of size m, we get respectively d^*_m (2.9.4) and d^*_{n-m} (obtained by changing m to n-m in (2.9.4)). Consequently, the class of unbiased regression estimators d^*_m is superior to the class $\bar{d}_m$ as well as the class $\bar{d}'_{n-m}$ from the point of view of efficiency. For $m = 1$, however, we note $\bar{d}'_{n-1} = d^*_{n-1}$.

Under Durbin's (1959) super-population model 1 (see section 2.3), Williams (1963) has shown that for the class of estimators $\bar{d}_m$ (2.9.2) the optimum number of groups into which the sample is to be split is equal to $\sqrt{n/3}$; and that the unbiased estimators compare favourably (in the sense that the relative loss in precision is not much) with the classical regression estimator $\hat{Y}_r$ (which is also unbiased under the assumed super-population model), provided the group size $m > 3$. For certain non-normal x distributions, his asymptotic results suggested a similar conclusion provided the distribution of x is platykurtic. It must, however, be noted that

Williams illustrated these results only with large samples.

No theoretical results are available either on $\bar{d}'_{n-m}$ of (2.9.3) or on d^*_m of (2.9.4). However, in Rao's (1969) empirical study on ratio and regression estimators using several sets of real data with small and moderate sample sizes, the tentative observations made are: (i) the unbiased regression estimator $\bar{d}'_{n-m}$ with m = 1 is far inferior to the classical biased regression estimator $\hat{Y}_r$, and (ii) the classical regression estimator $\hat{Y}_r$ itself is inferior to ratio estimators. Sampford (1969) has conducted an independent empirical study using small sample sizes and observed that 'the least expected result was the excessively bad performance of the classical regression estimator, both in the stratified and, to a lesser extent, in the un-stratified sample'. Sukhatme (1966) has also drawn attention to Johnson's (1950) work in which Johnson has shown that in small samples, and with certain types of populations, the classical ratio estimator is occasionally more efficient than the classical regression estimator.

Finally, we note that Mickey's method of constructing unbiased regression estimators, in simple random sampling without replacement may be extended easily to general sampling designs with the use of independent interpenetrating sub-samples. Thus, with the set-up described in section 2.8 and further defining $\hat{\beta} = \frac{1}{k} \sum_{i=1}^{k} \hat{\beta}_i$, an unbiased regression estimator corresponding to the biased regression estimator $\hat{Y}_r = \hat{Y} + \hat{\beta} (X - \hat{X})$ is given by

$$\hat{Y}_{ur} = \hat{Y} + \hat{\beta}(X - \hat{X}) + \frac{1}{k(k-1)} \sum_{i=1}^{k} \hat{\beta}_i (\hat{X}_i - \hat{X}), \quad \ldots (2.9.5)$$

when the total sample is selected in terms of k independent interpenetrating sub-samples.

CHAPTER III

ON THE VARIANCE OF MICKEY'S GENERAL CLASS OF UNBIASED ESTIMATORS

3.0 Summary

Mickey (1959) has introduced a general class of unbiased estimators for the population mean of a variate utilising information on the population means of several auxiliary variates in simple random sampling without replacement. He has not investigated the variance of these estimators. However, for a sub-class of his general class of estimators he has given an unbiased method of estimating the variance. In this chapter we present a systematic investigation of the variance of Mickey's general class of unbiased estimators. The problem of unbiased estimation of the variance of any estimator belonging to Mickey's general class is also solved. Further this chapter lays the foundation for the investigations in subsequent chapters on the precision of unbiased ratio, product, ratio-cum-product, and regression type estimators which belong to Mickey's general class of unbiased estimators.

The chapter is divided into six sections. Section 1 gives a review of Mickey's (1959) method of estimating the population mean of the variate under study. Essentially, Mickey's 'basic' general unbiased estimator (G_m) is based on (i) the choice of a sub-set Z_m of m ($1 \leq m < n$) distinct sample elements from the total sample of size n, and (ii) a set of p (equal to the number of auxiliary variates) coefficient functions $a_i(Z_m)$, $i = 1, 2, \ldots, p$ which are some real valued functions defined on Z_m. For a given sample there are several possible ways of choosing Z_m and corresponding to each such choice of Z_m

a basic estimator G_m can be formed. It is then easy to think of classes of unbiased estimators that may be generated as weighted averages of Mickey's basic estimators G_m. In section 2, four such important classes of unbiased estimators are described and their relative efficiency is examined using the results due to Murthy (1957) and Basu (1958) on 'ordered' and 'unordered' estimators. Of all such classes of unbiased estimators, the most efficient is the class of estimators, say, G^*_m, which is obtained as the average of the $\binom{n}{m}$ basic estimators, G_m, corresponding to the $\binom{n}{m}$ possible choices of distinct Z_m from the sample.

The variance of Mickey's basic estimator G_m is derived in section 3. The approach adopted to derive the variance of G_m is useful for developing later in section 6 an unbiased estimator for the variance of any estimator belonging to Mickey's general class of estimators. In section 4, the variance of Mickey's averaged general unbiased estimator G^*_m is investigated. In the first instance, the exact variance of G^*_1 (i.e., for the choice m=1) is derived using multivariate symmetric functions on similar lines as Robson (1957). The variance expression of G^*_1 assumes a neat form when the population size N is relatively large compared to the sample size n (i.e., $N \gg n$). Secondly a simplified expression for the variance of G^*_m is obtained in large samples (i.e., variance upto order n^{-1}) under certain assumptions on $a_i(Z_m)$ which are usually satisfied for several 'forms' of these coefficient functions. By another result, the same simplified expression is obtained for the large sample variance of G^* when m is small in relation to the sample size n.

One of the desirable properties of a multivariate estimator is that whenever an additional auxiliary variate is introduced in the estimator, it should not result in a decrease in the efficiency of the estimator. In section 5, the unbiased estimator G_m^* is examined from this angle and an optimum method of introducing an additional auxiliary variate in Mickey's method of estimation is suggested. Finally in section 6, we consider the problem of variance estimation and present three results. Using the first result, the variance of any member of Mickey's general class of unbiased estimators including the estimator G_m^* can be estimated unbiasedly. The second result gives a different unbiased estimator of the variance of G_1^* by making use of sample symmetric means. The third result provides an estimator of the large sample variance of G_m^*.

3.1 Introduction

Mickey (1959) has formulated a general class of unbiased estimators for the finite population mean of a variate utilising information on the population means of several auxiliary characters in simple random sampling without replacement. To describe his estimators, suppose in a population of size N the vector ($y, x_1, x_2, \ldots, x_p$) of variables is under observation. We assume the population means $\overline{X}_i$ ($i = 1, 2, \ldots, p$) of the auxiliary variates x_i are all known. To estimate the population mean $\overline{Y}$ of y, let a simple random sample of size n be drawn without replacement from the population. Let Z_m denote any sub-set of m ($<$ n) distinct sample elements, and $\overline{y}_m$ and $\overline{x}_{im}$ be the means of y and x_i defined on Z_m. Further let $a_i(Z_m)$, ($i = 1, 2, \ldots, p$) be some real valued functions defined on Z_m. In the derived sample of size n - m obtained by the exclusion of Z_m from the total sample, let $\bar{y}_{n-m}$ and $\bar{x}_{i\,n-m}$ denote the means of y and x_i. Similarly in the derived population of size $N-m$ conceived by the exclusion of Z_m from the total population, $\bar{Y}_{N-m}$ and $\bar{X}_{i\,N-m}$ represent the means of y and x_i. Now define the statistics

$$U_m = \bar{y}_{n-m} - \sum_{i=1}^{p} a_i (Z_m) (\bar{x}_{i\,n-m} - \bar{X}_{i\,N-m}) \tag{3.1.1}$$

and

$$G_m = \frac{m\, \bar{y}_m + (N-m)\, U_m}{N} . \tag{3.1.2}$$

For a given Z_m, we note that the derived sample of size $n-m$ is a simple random sample without replacement from the derived population of size $N-m$. Hence if $E_m(\cdot)$ denotes the conditional expectation for given Z_m, we have from (3.1.1) and (3.1.2)

$$E_m (U_m) = \bar{Y}_{N-m} , \tag{3.1.3}$$

and

$$E_m (G_m) = \bar{Y} \tag{3.1.4}$$

It then follows that G_m is also unconditionally unbiased for $\bar{Y}$.

From now onwards, we call G_m Mickey's basic general unbiased estimator or simply basic estimator if the context is clear. By the word 'general' we mean the choice of the coefficient functions $a_i (Z_m)$, ($i = 1, 2, \ldots, p$) is unspecified. For particular specifications of $a_i (Z_m)$ we obtain unbiased ratio, product, ratio-cum-product and regression type estimators and these are discussed in detail in later chapters. We call G_m a 'basic' estimator because there are several possible ways of choosing a sub-set Z_m of sample elements from the total sample and one can think of classes of unbiased estimators that may be generated as weighted averages of the estimators of the form G_m. In particular we now describe some important classes of unbiased estimators that may be formulated using the basic estimators G_m.

3.2 Some important classes of unbiased estimators

When the sample is split at random into g groups each of size m ($n = mg$), treating each of the g groups in turn as Z_m, g estimators of the form G_m may be constructed and the average of these estimators may be denoted by $\bar{G}_m$. $\bar{G}_m$ is the class of unbiased estimators suggested by Williams (1961, 1962).

Again in the above split-sample set-up, a sub-set of n-m elements derived by omitting a group of m elements from the total sample may be taken as Z_{n-m} and the corresponding basic estimator may be computed. The average of the g such estimators, corresponding to the g possible sub-sets Z_{n-m}, may be denoted by $\bar{G}'_{n-m}$. Some unbiased ratio and regression estimators falling in the class $\bar{G}'_{n-m}$ have been studied by Rao (1967), Rao and Beegle (1967) and Rao (1969).

Given a sample of size n, for any particular choice of m ($< n$) there are $\binom{n}{m}$ distinct sub-sets Z_m and the average of the corresponding $\binom{n}{m}$ basic estimators G_m may be represented by G^*_m. It is also easy to see that $\bar{G}_m$ averaged over all possible splits of the sample into groups of size m results in G^*_m. Similarly average of $\bar{G}'_{n-m}$ over all possible splits of the sample is equivalent to G^*_{n-m}. Consequently from the results of Murthy (1957) and Basu (1958) on ordered and the corresponding unordered estimators it follows that

$$\begin{aligned} &\text{Variance of } G^*_m \leq \text{Variance of } G_m, \\ &\text{Variance of } G^*_m \leq \text{Variance of } \bar{G}_m, \qquad (3.2.1)\\ \text{and } &\text{Variance of } G^*_{n-m} \leq \text{Variance of } \bar{G}'_{n-m} \end{aligned}$$

An interesting class of unbiased estimators depending on basic estimators constructed on portions of the total sample is as follows. Consider the order in which the sample is drawn from the population and let $G(m, n)$ denote the basic estimator (3.1.2) depending on this ordered sample of size n and the sub-set Z_m comprising the elements occurring in the first m draws of the sample. Choose k integers $m_1, m_2, \ldots, m_k$ such that

$$0 < m_1 < m_2 < \ldots < m_k < m_{k+1} = n.$$

Then Mickey (1959) has shown that the k unbiased basic estimators $G(m_t, m_{t+1})$, $(t = 1, 2, \ldots, k)$ are mutually un-correlated. It may be noted that $G(m_t, m_{t+1})$ uses only that portion of the ordered sample made up by the elements occurring in the first m_{t+1} draws of the sample. Clearly then, for the unbiased estimato

$$\bar{G}^{(k)} = \frac{1}{k}\sum_{t=1}^{k} G(m_t, m_{t+1}), \qquad (3.2.2)$$

a non-negative unbiased estimator of variance is given by

$$\hat{V}(\bar{G}^{(k)}) = \frac{1}{k(k-1)} \sum_{t=1}^{k} \left[G(m_t, m_{t+1}) - \bar{G}^{(k)} \right]^2 \quad (3.2.3)$$

In view of the considerations mentioned at (3.2.1), in this thesis we confine our attention mainly to the class of unbiased estimators G^*_m. Befo we undertake the study of the variance of estimators G^*_m, it is, however, bo instructive and useful to examine the variance of the basic unbiased estimator G_m (3.1.2). Such an attempt is likely to throw light on the complexities involved in the study of the variance function of classes of estimators built upon the basic estimators.

3.3 Variance of Mickey's basic general unbiased estimator

In this section we obtain the exact variance of Mickey's basic general unbiased estimator G_m (3.1.2). We first introduce some notations which are useful in representing the results of this section.

At (3.1.3) we let $E_m(\cdot)$ to denote the conditional expectation for a given sub-set Z_m. Similarly let $V_m(\cdot)$ denote the conditional variance for given Z_m. For a unit $j \notin Z_m$, define

$$y'_j = y_j - \sum_{i=1}^{p} a_i(Z_m)\, x_{ij} \quad (3.3.1)$$

For a given Z_m, in section 3.1, we have introduced the concepts of derived sample and derived population. With the same concepts, let $\bar{y}'_{n-m}$ and $\bar{Y}'_{N-m}$ be the means of y'_j in the derived sample and derived population respectivel

Then we note

$$E_m(\bar{y}'_{n-m}) = \bar{Y}'_{N-m} \tag{3.3.2}$$

and $$V_m(\bar{y}'_{n-m}) = \frac{(N-n)}{(N-m)(n-m)} S^2_{y'} , \tag{3.3.3}$$

where $$S^2_{y'} = \frac{1}{(N-m-1)} \sum_{j \notin Z_m}^{N-m} (y'_j - \bar{Y}'_{N-m})^2 \tag{3.3.4}$$

Given a pair of distinct units (j, j') not contained in Z_m, we observe that Z_m may be treated as a simple random sample without replacement of size m from the N - 2 units of the population obtained by exclusion of the pair (j,j') from the total population. Hence define

$$E[a_i(Z_m)/Z_m \not\ni (j,j')] = \frac{1}{\binom{N-2}{m}} \sum_{Z_m \not\ni (j,j')}^{\binom{N-2}{m}} a_i(Z_m)$$

$$= (a_i)_{jj'} , \tag{3.3.5}$$

and $$E[a_i(Z_m) a_{i'}(Z_m)/Z_m \not\ni (j,j')] = \frac{1}{\binom{N-2}{m}} \sum_{Z_m \not\ni (j,j')}^{\binom{N-2}{m}} a_i(Z_m) a_{i'}(Z_m)$$

$$= (a_i a_{i'})_{jj'} \tag{3.3.6}$$

We now prove a lemma.

Lemma 3.3.1 : For a given Z_m the conditional variance of U_m(3.1.1) is given by

$$V_m(U_m) = \frac{(N-n)}{(N-m)(n-m)} S^2_{y'} \qquad (3.3.7)$$

where $S^2_{y'}$ is as defined at (3.3.4).

Proof : From the definition of U_m(3.1.1) and $\bar{y}'_{n-m}$ we note

$$U_m = \bar{y}'_{n-m} + \sum_{i=1}^{p} a_i(Z_m)\,\bar{X}_{i\,N-m}. \qquad (3.3.8)$$

Since the second term in (3.3.8) is constant for given Z_m,

$$V_m(U_m) = V_m(\bar{y}'_{n-m}) \qquad (3.3.9)$$

Now by using (3.3.3) the lemma follows.

We now state and prove the theorem on the exact variance of G_m(3.1.2)

Theorem 3.3.1 : The variance of Mickey's basic general unbiased estimator G_m(3.1.2) is given by

$$V(G_m) = \frac{(1-f_n)(1-f_m)}{(n-m)} \cdot \frac{1}{2N(N-1)} \sum_{j \neq j'}^{N}\sum^{N} \left[(y_j - y_{j'})^2 - 2\sum_{i=1}^{p}(y_j - y_{j'})(x_{ij} - x_{ij'})(a_i)_{jj'} \right.$$

$$\left. + \sum_{i=1}^{p}\sum_{i'=1}^{p}(x_{ij} - x_{ij'})(x_{i'j} - x_{i'j'})(a_i a_{i'})_{jj'} \right], \qquad (3.3.10)$$

where $f_n = \frac{n}{N}$, $f_m = \frac{m}{N}$, and $(a_i)_{jj'}$ and $(a_i a_{i'})_{jj'}$ are as defined at (3.3.5) and (3.3.6).

Proof : We have

$$V(G_m) = E\left[V_m(G_m) \right] + V\left[E_m(G_m) \right]$$

$$= E\left[V_m(G_m) \right] \qquad E_m(G_m) = \bar{Y} \text{ from } (3.1.4),$$

$$= E\left[V_m\left(\frac{m\,\bar{y}_m + (N-m)\,U_m}{N} \right) \right] \text{ from } (3.1.2),$$

$$= \frac{(N-m)^2}{N^2}\, E\left[V_m(U_m) \right],$$

$$= \frac{(N-n)(N-m)}{N^2(n-m)}\, E\left[S^2_{y'} \right] \text{ from lemma } 3.3.1,$$

$$= \frac{(1-f_n)(1-f_m)}{(n-m)} \; \frac{1}{2(N-m-1)(N-m)}\, E\left[\sum\sum^{(N-m)}_{(j \neq j') \notin Z_m} (y'_j - y'_{j'})^2 \right]$$

by re-writing $S^2_{y'}$ (3.3.4) in terms of pairs of units not belonging to Z_m,

$$= \frac{(1-f_m)(1-f_n)}{(n-m)} \; \frac{1}{2(N-m-1)(N-m)\binom{N}{m}} \sum_{Z_m}^{\binom{N}{m}} \left[\sum\sum^{(N-m)}_{(j \neq j') \notin Z_m} (y'_j - y'_{j'})^2 \right]. \tag{3.3.11}$$

Now fixing a pair of units ($j \neq j'$) in the population and choosing Z_m from the remaining N - 2 units of the population, the summations in (3.3.11) may be rearranged as under:

$$V(G_m) = \frac{(1-f_n)(1-f_m)}{(n-m)} \cdot \frac{1}{2N(N-1)\binom{N-2}{m}} \sum_{j \neq j'}^{N} \sum^{N} \left[\sum_{Z_m \not\ni (j,j')}^{\binom{N-2}{m}} (y'_j - y'_{j'})^2 \right] \tag{3.3.12}$$

From the definition (3.3.1) of y'_j we may write

$$\sum_{Z_m \not\ni (j,j')}^{\binom{N-2}{m}} (y'_j - y'_{j'})^2 = \sum_{Z_m \not\ni (j,j')}^{\binom{N-2}{m}} \left[(y_j - y_{j'}) - \sum_{i=1}^{p} a_i(Z_m)(x_{ij} - x_{ij'}) \right]^2$$

$$= \binom{N-2}{m} \left[(y_j - y_{j'})^2 - 2\sum_{i=1}^{p} (y_j - y_{j'})(x_{ij} - x_{ij'})(a_i)_{jj'} \right.$$

$$\left. + \sum_{i=1}^{p} \sum_{i'=1}^{p} (x_{ij} - x_{ij'})(x_{i'j} - x_{i'j'})(a_i a_{i'})_{jj'} \right], \tag{3.3.13}$$

using the definitions (3.3.5) and (3.3.6) for $(a_i)_{jj'}$ and $(a_i a_{i'})_{jj'}$.

We now substitute (3.3.13) in (3.3.12) and obtain the expression (3.3.10).

Hence the theorem is proved.

<u>Remark</u>: Using theorem 3.3.1, one can formulate the variance of the class of unbiased estimators $\bar{G}^{(k)}$ (3.2.2).

<u>Theorem 3.3.2</u> The variance of the unbiased estimator $\bar{G}^{(k)}$ (3.2.2) may be formulated as

$$V(\bar{G}^{(k)}) = \frac{1}{k^2} \sum_{t=1}^{k} V\left[G(m_t, m_{t+1}) \right], \text{ where}$$

$V[G(m_t, m_{t+1})]$ is given by (3.3.10) in which m_t and m_{t+1} are substituted for m and n respectively.

Proof : Follows from the fact that the k estimators $G(m_t, m_{t+1})$, $(t = 1, 2, \ldots, k)$ are mutually un-correlated.

3.4 Variance of Mickey's averaged general unbiased estimator.

As stated in section 3.2, the class of estimators G_m^* is obtained by taking the average of the $\binom{n}{m}$ basic estimators G_m(3.1.2) corresponding to the $\binom{n}{m}$ possible sub-sets Z_m that may be generated from a given sample of size n. Alternately the $\binom{n}{m}$ possible sub-sets Z_m may be treated as the $\binom{n}{m}$ possible sub-samples of size m when a simple random sample of size m is drawn without replacement from the total sample of size n. Consequently G_m^* may be represented as $E_n(G_m)$ where G_m is a basic estimator using a sub-sample of size m as the sub-set Z_m and $E_n(\cdot)$ denotes the conditional expectation for a given total sample over all possible sub-samples of size m. This kind of representation for G_m^* is useful in investigating its variance, for the problem can in effect be treated as a two-phase sampling problem using the basic estimator G_m in the two-phase set-up.

Mickey (1959) has given three alternate but equivalent forms for the basic estimator G_m (3.1.2). One of these forms, in our notation, is

$$G_m = \left[\bar{y} - \sum_{i=1}^{p} a_i(Z_m)(\bar{x}_i - \bar{X}_i) \right] - \frac{m(N-n)}{(n-m)N} \left[(\bar{y}_m - \bar{y}) - \sum_{i=1}^{p} a_i(Z_m)(\bar{x}_{im} - \bar{x}_i) \right] \tag{3.4.1}$$

where $\bar{y}$ and $\bar{x}_i$ $(i = 1, 2, \ldots, p)$ are the means defined on the total sample and the others are as defined earlier. We use this form for writing down $G_m^* = E_n(G_m)$. Noting $E_n(\bar{y}_m) = \bar{y}$, and writing $E_n[a_i(Z_m)] = a_i^*(Z_m)$,

and
$$E_n[a_i(Z_m)(\bar{x}_{im} - \bar{x}_i)] = Cov_n[a_i(Z_m), \bar{x}_{im}] \quad (3.4.2)$$

we have from (3.4.1)

$$E_n(G_m) = G_m^* = [\bar{y} - \sum_{i=1}^{p} a_i^*(Z_m)(\bar{x}_i - \bar{X}_i)] + \frac{m(N-n)}{(n-m)N} \sum_{i=1}^{p} Cov_n[a_i(Z_m), \bar{x}_{im}] \quad (3.4.3)$$

In particular for the choice $m = 1$, Z_m contains only one element, say j-th element. Then we may write $a_i(Z_m) = a_{ij}$ and $a_i^*(Z_m) = E_n[a_{ij}] = \bar{a}_i$. Further in this case, $Cov_n[a_i(Z_m), \bar{x}_{im}] = Cov_n[a_{ij}, x_{ij}] = \frac{n-1}{n} s_{a_i x_i}$

where $s_{a_i x_i} = \frac{1}{n-1} \sum_{j=1}^{n} a_{ij}(x_{ij} - \bar{x}_i)$. Using these in (3.4.3), we obtain

$$G_1^* = \bar{y} - \sum_{i=1}^{p} [\bar{a}_i(\bar{x}_i - \bar{X}_i) - (\frac{1}{n} - \frac{1}{N}) s_{a_i x_i}]. \quad (3.4.4)$$

The problem of finding a meaningful expression for the exact variance of G_m^* (3.4.3) valid for any m ($1 \leq m \leq n-1$) is rather complicated. Our attempts in this direction have resulted in too unwieldy expressions and it has been felt difficult to give meaningful interpretation to the terms appearing in those expressions. For the choice $m=1$, however, the problem, though still complicated, has been solved satisfactorily. The exact variance obtained in this section for G_1^* (3.4.4) has many interesting particular cases as will be evident from the results reported in later chapters. We also give a solution to the variance of G_m^* (3.4.3) under some asymptotic conditions, the special cases of which are also many and discussed in the subsequent chapters.

3.4a <u>Exact Variance of G_1^*</u>.

From the formula (3.4.4) of G_1^* it is clear that to evaluate $V(G_1^*)$ we need to evaluate

$$\text{Cov} \left[\bar{y},\ \bar{a}_i (\bar{x}_i - \bar{X}_i) - \left(\frac{1}{n} - \frac{1}{N} \right) s_{a_i x_i} \right] \tag{3.4.5}$$

and $$\text{Cov}\left[\bar{a}_i(\bar{x}_i - \bar{X}_i) - \left(\frac{1}{n} - \frac{1}{N}\right) s_{a_i x_i},\ \bar{a}_{i'} (\bar{x}_{i'} - \bar{X}_{i'}) - \left(\frac{1}{n} - \frac{1}{N}\right) s_{a_{i'} x_{i'}} \right] \tag{3.4.6}$$

Writing $\bar{a}_i (\bar{x}_i - \bar{X}_i) - \left(\frac{1}{n} - \frac{1}{N} \right) s_{a_i x_i} = \bar{a}_i(\bar{x}_i - \bar{X}_i) - \left(\frac{1}{n} - \frac{1}{N}\right) \left[\frac{\sum_{j=1}^{n} a_{ij} x_{ij} - n \bar{a}_i \bar{x}_i}{n-1} \right]$

$$= \frac{n(N-1)}{N(n-1)}\ \bar{a}_i\, \bar{x}_i - \frac{(N-n)}{N(n-1)\,n} \sum_{j=1}^{n} a_{ij} x_{ij} - \bar{a}_i\, \bar{X}_i, \tag{3.4.7}$$

It may be noted that to evaluate the covariances (3.4.5) and (3.4.6) we need the expansions of $\text{Cov}(\bar{h}_1, \bar{h}_2)$, $\text{Cov}(\bar{h}_2, \bar{h}_3 \bar{h}_4)$ and $\text{Cov}(\bar{h}_1 \bar{h}_2, \bar{h}_3 \bar{h}_4)$, where $\bar{h}_i$ $(i = 1, 2, 3, 4)$ are the sample means of four random variables h_i. Making use of multivariate symmetric functions as illustrated by Robson(1957), we first obtain $\text{Cov}(\bar{h}_2, \bar{h}_3 \bar{h}_4)$ and $\text{Cov}(\bar{h}_1 \bar{h}_2, \bar{h}_3 \bar{h}_4)$.

<u>Definition 3.4.1</u> : For a vector $(h_1, h_2, h_3, \ldots, h_p)$ of p variates a sample symmetric function of order k is defined as

$$\left[(a_{11}\,a_{21}\ldots a_{p1}),(a_{12}\,a_{22}\ldots a_{p2}),\ldots,(a_{1k}\,a_{2k}\ldots a_{pk})\right]$$

$$=\frac{1}{(n)_k}\sum_{(k)}^{n}(h_{1j_1}^{a_{11}}h_{2j_1}^{a_{21}}\ldots h_{pj_1}^{a_{p1}})(h_{1j_2}^{a_{12}}h_{2j_2}^{a_{22}}\ldots h_{pj_2}^{a_{p2}})\ldots(h_{1j_k}^{a_{1k}}h_{2j_k}^{a_{2k}}\ldots h_{pj_k}^{a_{pk}}), \tag{3.4.8}$$

where $(n)_k = n(n-1)\ldots(n-k+1), 1 \leq k < n$; and $\sum_{(k)}^{n} = \sum_{j_1\neq}^{n}\sum_{j_2\neq}^{n}\ldots\sum_{\neq j_k}^{n}$

Definition 3.4.2 : The population symmetric function corresponding to the k-th order sample symmetric function (3.4.8) is defined by

$$\left[(a_{11}\,a_{21}\ldots a_{p1}),(a_{12}\,a_{22}\ldots a_{p2}),\ldots(a_{1k}\,a_{2k}\ldots a_{pk})\right]'$$

$$=\frac{1}{(N)_k}\sum_{(k)}^{N}(h_{1j_1}^{a_{11}}h_{2j_1}^{a_{21}}\ldots h_{pj_1}^{a_{p1}})(h_{1j_2}^{a_{12}}h_{2j_2}^{a_{22}}\ldots h_{pj_2}^{a_{p2}})\ldots(h_{1j_k}^{a_{1k}}h_{2j_k}^{a_{2k}}\ldots h_{pj_k}^{a_{pk}}), \tag{3.4.9}$$

where $(N)_k = N(N-1)(N-2)\ldots(N-k+1)$, $1 \leq k < n$, and $\sum_{(k)}^{N} = \sum_{j_1\neq}^{N}\sum_{j_2\neq}^{N}\ldots\sum_{\neq j_k}^{N}$.

With the above definitions, we have the following well-known lemma.

Lemma 3.4.1 : For a simple random sample of size n drawn without replacement from a population of size N, the expected value of the k-th order ($1 \leq k < n$) sample symmetric function (3.4.8) is the corresponding k-th order population symmetric function (3.4.9).

Lemma 3.4.2 :

$$
\begin{aligned}
\operatorname{Cov}(\bar{h}_1\,\bar{h}_2,\ \bar{h}_3\bar{h}_4) = \frac{1}{n^2}\left(\frac{1}{n}-\frac{1}{N}\right)\Big[\,& [(1111)]' - [(1100),(0011)]' \\
& + \frac{(n-1)(N+n-1)}{N-1}\Big\{[(1010),(0101)]' + [(1001),(0110)]'\Big\} \\
& + (n-1)\Big\{[(1000),(0111)]' + [(0100),(1011)]' + [(0010),(1101)]' \\
& \qquad + [(0001),(1110)]' - 2[(1100),(0010),(0001)]' \\
& \qquad - 2[(1000),(0100),(0011)]'\Big\} \\
& + (n-1)\Big\{(n-2) - \frac{n}{(N-1)}\Big\}\Big\{[(1010),(0100),(0001)]' \\
& \qquad + [(1001),(0100),(0010)]' \\
& \qquad + [(1000),(0110),(0001)]' \\
& \qquad + [(1000),(0101),(0010)]'\Big\} \\
& + \frac{2(n-1)}{(N-1)}(3N+3n-2Nn-3)[(1000),(0100),(0010),(0001)]'\,\Big]
\end{aligned}
\tag{3.4.10}
$$

Proof; Following Robson (1957) we have

$$\bar{h}_1\bar{h}_2\bar{h}_3\bar{h}_4 = [(1000)][(0100)][(0010)][(0001)]$$

$$\bar{h}_1 \bar{h}_2 \bar{h}_3 \bar{h}_4 = \frac{1}{n^3}\Big[[(1111)] + (n-1)\{ [(1100),(0011)] + [(1010),(0101)]$$

$$+ [(1001),(0110)] + [(1000),(0111)] + [(0100),(1011)]$$

$$+ [(0010),(1101)] + [(0001),(1110)]\}$$

$$+ (n-1)(n-2)\{ [(1100),(0010),(0001)] + [(1010),(0100),(0001)]$$

$$+ [(1001),(0100),(0010)] + [(0110),(1000),(0001)]$$

$$+ [(0101),(1000),(0010)] + [(0011),(1000),(0100)]\}$$

$$+ (n-1)(n-2)(n-3) [(1000),(0100),(0010),(0001)]\Big] \qquad (3\ 4\ 11)$$

Also

$$E(\bar{h}_1\bar{h}_2)E(\bar{h}_3\bar{h}_4) = \frac{1}{n^2}\Big[[(1100)]' + (n-1)[(1000),(0100)]' \Big]\Big[[(0011)]'$$

$$+ (n-1)[(0010),(0001)]' \Big]$$

$$= \frac{1}{Nn^2}\Big[[(1111)]' + (N-1)[(1100),(0011)]' + (n-1)\{ [(1011),(0100)]'$$

$$+ [(0111),(1000)]' + [(1110),(0001)]' + [(1101),(0010)]'$$

$$+ (N-2)[(1000),(0100),(0011)]'$$

$$+ (N-2)[(0010),(0001),(1100)]'\} + \frac{(n-1)^2}{(N-1)}\{ [(1010),(0101)]'$$

$$+ [(1001),(0110)]'\} + \frac{(n-1)^2(N-2)}{(N-1)}\{ [(1010),(0100),(0001)]'$$

$$+ [(1001),(0100),(0010)]' + [(1000),(0110),(0001)]'$$

contd...

$$+[(1000),(0010),(0101)]' + (N-3)[(1000),(0100),(0010),(0001)]'\}]$$

(3.4.12)

By definition $Cov(\bar{h}_1\bar{h}_2, \bar{h}_3\bar{h}_4) = E(\bar{h}_1\bar{h}_2\bar{h}_3\bar{h}_4) - E(\bar{h}_1\bar{h}_2)E(\bar{h}_3\bar{h}_4)$.

(3.4.13)

From lemma 3.4.1 we note that $E(\bar{h}_1\bar{h}_2\bar{h}_3\bar{h}_4)$ is obtained from (3.4.11) just by replacing each of the sample symmetric functions by the corresponding population symmetric functions. Thus using (3.4.11) with this change and (3.4.12) in (3.4.13), we obtain, after some simplification, the result (3.4.10) of lemma 3.4.2.

Lemma 3.4.3 : $Cov(\bar{h}_2, \bar{h}_3\bar{h}_4) = \frac{1}{n}(\frac{1}{n} - \frac{1}{N})[[(0111)]' - [(0100),(0011)]' +$

$$+ (n-1)\{[(0110),(0001)]' + [(0010),(0101)]'$$

$$- 2[(0100),(0010),(0001)]'\}]$$

(3.4.14)

Proof : This may be derived from lemma 3.4.2 by writing $h_1 \equiv 1$. It implies that in the formula (3.4.10) we write $\alpha_{1i} = 0$ in each of the brackets $(\alpha_{1i}\alpha_{2i}\alpha_{3i}\alpha_{4i})$, and by this substitution whenever $(\alpha_{1i}\alpha_{2i}\alpha_{3i}\alpha_{4i}) = (0000)$ the order of the particular symmetric function gets reduced by one.

Remark (a) : Bohrnstedt and Goldberger (1969) have given formulae for the exact covariances of products of random variables. Their formulae relate to $Cov(h_1h_2, h_3h_4)$ and $Cov(h_2, h_3h_4)$ and for our present purpose need further amplification. The results above are obtained directly with the use of symmetric functions.

We now proceed with the derivation of $V(G_1^*)$. For the sake of simplicity in presentation, in what follows we use j, j', j'' and j''' to denote distinct units instead of j_1, j_2, j_3, and j_4.

Lemma 3.4.4 :

$$\mathrm{Cov}\left[\bar{y}, \bar{a}_i(\bar{x}_i - \bar{X}_i) - \left(\frac{1}{n} - \frac{1}{N}\right)s_{a_i x_i}\right] = \left(\frac{1}{n} - \frac{1}{N}\right)\frac{1}{N(N-1)}\left[\sum_{(2)}^{N} y_j x_{ij} a_{ij'} - \frac{1}{(N-2)}\sum_{(3)}^{N} y_j x_{ij'} a_{ij''}\right]$$

(3.4.15)

Proof: Let us write $h_1 = 1$, $h_2 = y$, $h_3 = x_i$ and $h_4 = a_i$. From (3.4.7) we have

$$\mathrm{Cov}\left[\bar{y}, \bar{a}_i(\bar{x}_i - \bar{X}_i) - \left(\frac{1}{n} - \frac{1}{N}\right)s_{a_i x_i}\right] = \mathrm{Cov}\left[\bar{y}, \frac{n(N-1)}{N(n-1)}\bar{a}_i\bar{x}_i - \frac{(N-n)}{N(n-1)}\cdot\frac{1}{n}\sum_{j=1}^{n} a_{ij}x_{ij} - \bar{a}_i\bar{X}_i\right]$$

(3.4.16)

In (3.4.16) consider

$$\mathrm{Cov}\left(\bar{y}, \frac{1}{n}\sum_{j=1}^{n} a_{ij}x_{ij}\right) = E\left[[(0100)][(0011)]\right] - [(0100)]'[(0011)]'$$

$$= \left(\frac{1}{n} - \frac{1}{N}\right)\left[[(0111)]' - [(0100),(0011)]'\right]$$

(3.4.17)

Similar to (3.4.17) we write

$$\text{Cov}(\bar{y}, \bar{a}_i) = \left(\frac{1}{n} - \frac{1}{N}\right)[(0010)]' \,[\,[(0101)]' - [(0100), (0001)]'\,]$$

$$= \left(\frac{1}{n} - \frac{1}{N}\right) \cdot \frac{1}{N} \,[\,[(0111)]' + (N-1)[(0010), (0101)]'$$

$$- [(0110), (0001)]' - [(0100), (0011)]'$$

$$-(N-2)[(0100), (0010), (0001)]'\,] \qquad (3.4.18)$$

w from (3.4.16) we note, by using lemma 3.4.3 and the results (3.4.17) and 4.18),

$$v[\bar{y}, \bar{a}_i(\bar{x}_i - \bar{X}_i) - (\frac{1}{n} - \frac{1}{N}) s_{a_i x_i}] = \frac{n(N-1)}{N(n-1)}(3.4.14) - \frac{(N-n)}{N(n-1)}(3.4.17) - (3.4.18).$$

(3.4.19)

(3.4.19) we have to collect together the coefficients of each of the population mmetric functions $[(ijkl)]'$ which may be denoted by $Q[(ijkl)]'$. Thus may be seen that

$$Q[(0111)]' \quad \left(\frac{1}{n} - \frac{1}{N}\right)\left[\frac{(N-1)}{N(n-1)} - \frac{(N-n)}{N(n-1)} - \frac{1}{N}\right] = 0 ,$$

$$Q[(0100), (0011)]' = \left(\frac{1}{n} - \frac{1}{N}\right)\left[-\frac{(N-1)}{N(n-1)} + \frac{(N-n)}{N(n-1)} + \frac{1}{N}\right] = 0 ,$$

$$Q[(0110), (0001)]' = \left(\frac{1}{n} - \frac{1}{N}\right)\left[-\frac{(N-1)}{N} + \frac{1}{N}\right] = \left(\frac{1}{n} - \frac{1}{N}\right),$$

$$Q[(0101), (0010)]' = \left(\frac{1}{n} - \frac{1}{N}\right)\left[\frac{(N-1)}{N} - \frac{(N-1)}{N}\right] = 0 ,$$

$\angle\overline{(0100), (0010), (0001)}\overline{\rule{0pt}{0pt}}\!\!/'= (\frac{1}{n}-\frac{1}{N})\angle\overline{\ } - \frac{2(N-1)}{N} + \frac{(N-2)}{N}\overline{/} = -(\frac{1}{n}-\frac{1}{N})$.

ence $Cov \angle\overline{\ } \bar{y}, \bar{a}_i(\bar{x}_i - \bar{X}_i) - (\frac{1}{n}-\frac{1}{N}) s_{a_i x_i} \overline{/} = (\frac{1}{n}-\frac{1}{N}) \Big/ \angle\overline{(0110), (0001)}\overline{/}'$

$- \angle\overline{(0100), (0010), (0001)}\overline{/}' \Big/$

= expression (3.4.15).

is completes the proof of the lemma.

We now proceed to evaluate the covariance (3.4.6), which is rather

thus.

mma 3.4.5 : $Cov\angle\overline{\ } \bar{a}_i(\bar{x}_i - \bar{X}_i) - (\frac{1}{n}-\frac{1}{N}) s_{a_i x_i}, \bar{a}_{i'}(\bar{x}_{i'} - \bar{X}_{i'}) - (\frac{1}{n}-\frac{1}{N}) s_{a_{i'} x_{i'}} \overline{/}$

$$= \frac{1}{N^2(n-1)}(\frac{1}{n}-\frac{1}{N})\angle\overline{\ } \sum_{(2)}^{N} x_{ij} a_{ij'} x_{i'j} a_{i'j'} + \frac{(N-n)}{(N-1)} \sum_{(2)}^{N} x_{ij} a_{ij'} x_{i'j'} a_{i'j}$$

$$+ \frac{(Nn-2N+1)}{(N-1)(N-2)} \sum_{(3)}^{N} x_{ij} a_{ij'} x_{i'j} a_{i'j''} - \frac{(N-n)}{(N-1)(N-2)} \sum_{(3)}^{N} x_{ij} a_{ij'} x_{i'j''} a_{i'j}$$

$$\frac{(N-n)}{(N-1)(N-2)} \sum_{(3)}^{N} x_{ij} a_{ij'} x_{i'j'} a_{i'j''} - \frac{1}{(N-2)} \sum_{(3)}^{N} x_{ij} a_{ij'} x_{i'j''} a_{i'j'}$$

$$\frac{(Nn-3N+n+1)}{(N-1)(N-2)(N-3)} \sum_{(4)}^{N} x_{ij} a_{ij'} x_{i'j''} a_{i'j'''} \overline{/} \qquad (3.4.20)$$

Let us write $x_i = h_1$, $a_i = h_2$, $x_{i'} = h_3$ and $a_{i'} = h_4$ to prove lemma. Using (3.4.7) we write

$$\text{ov}\,[\,\bar{a}_i(\bar{x}_i - \bar{X}_i) - (\frac{1}{n} - \frac{1}{N}) s_{a_i x_i},\ \bar{a}_{i'}(\bar{x}_{i'} - \bar{X}_{i'}) - (\frac{1}{n} - \frac{1}{N}) s_{a_{i'} x_{i'}}\,]$$

$$\text{Cov}[\,\bar{a}_i(\bar{x}_i - \bar{X}_i) - (\frac{1}{n} - \frac{1}{N}) s_{a_i x_i},\ \frac{n(N-1)}{N(n-1)} \bar{a}_{i'} \bar{x}_{i'} - \frac{(N-n)}{N(n-1)} \cdot \frac{1}{n} \sum_{j=1}^{n} a_{i'j} x_{i'j} - \bar{a}_{i'} \bar{X}_{i'}\,]$$

(3.4.21)

(3.4.21) we note $\text{Cov}[\,\bar{a}_i(\bar{x}_i - \bar{X}_i) - (\frac{1}{n} - \frac{1}{N}) s_{a_i x_i},\ \frac{1}{n}\sum_{j=1}^{n} a_{i'j} x_{i'j}\,]$

$$= (\frac{1}{n} - \frac{1}{N}) \frac{1}{N(N-1)} [\,\sum_{(2)}^{N} x_{ij} a_{ij'} x_{i'j} a_{i'j} - \frac{1}{(N-2)} \sum_{(3)}^{N} x_{ij} a_{ij'} x_{i'j''} a_{i'j''}\,]$$

(follows on similar lines as lemma 3.4.4)

$$= (\frac{1}{n} - \frac{1}{N}) [\,[(1011), (0100)]' - [(1000), (0100), (0011)]'\,] \qquad (3.4.22)$$

on the same lines as in lemma 3.4.4, we observe

$$[\,\bar{a}_i(\bar{x}_i - \bar{X}_i) - (\frac{1}{n} - \frac{1}{N}) s_{a_i x_i},\ \bar{a}_{i'} \bar{X}_{i'}\,]$$

$$= (\frac{1}{n} - \frac{1}{N}) \frac{\bar{X}_{i'}}{N(N-1)} [\,\sum_{(2)}^{N} x_{ij} a_{ij'} a_{i'j} - \frac{1}{(N-2)} \sum_{(3)}^{N} x_{ij} a_{ij'} a_{i'j''}\,]$$

$$= (\frac{1}{n} - \frac{1}{N}) [(0010)]' [\,[(1001), (0100)]' - [(1000), (0100), (0001)]'\,]$$

$$= \frac{1}{N}(\frac{1}{n} - \frac{1}{N}) [\,[(1011), (0100)]' + [(1001), (0110)]' + (N-2)[(1001), (0100), (0010)]'$$

$$- [(1010), (0100), (0001)]' - [(1000), (0110), (0001)]'$$

$$- [(1000), (0100), (0011)]' - (N-3)[(1000), (0100), (0001), (0010)]'\,]$$

(3.4.23)

To evaluate the covariance (3.4.21), apart from (3.4.22) and (3.4.23), we should also obtain

$$\text{Cov}\left[\bar{a}_i(\bar{x}_i - \bar{X}_i) - \left(\frac{1}{n} - \frac{1}{N}\right) s_{a_i x_i}, \ \bar{a}_i, \bar{x}_i\right]$$

$$= \text{Cov}\left[\frac{n(N-1)}{N(n-1)}\bar{a}_i \bar{x}_i - \frac{(N-n)}{N(n-1)} \cdot \frac{1}{n}\sum_{j=1}^{n} a_{ij} x_{ij} - \bar{a}_i \bar{X}_i, \bar{a}_i, \bar{x}_i\right]$$ on using form (3.4.7). (3.4.24)

In (3.4.24) we have, by using lemma 3.4.3,

$$\text{Cov}\left[\frac{1}{n}\sum_{j=1}^{n} a_{ij} x_{ij}, \bar{a}_i, \bar{x}_i\right] = \frac{1}{n}\left(\frac{1}{n} - \frac{1}{N}\right)\Big[[(1111)]' - [(1100),(0011)]' + (n-1)\left\{[(1110),(0001)]' + [(1101),(0010)]' - 2[(1100),(0010),(0001)]'\right\}\Big] \quad (3.4.25)$$

Again by using lemma 3.4.3, we obtain,

$$(\bar{a}_i \bar{X}_i, \bar{a}_i, \bar{x}_i) = \left(\frac{1}{n} - \frac{1}{N}\right)\frac{1}{n}[(1000)]\Big[[(0111)]' - [(0100),(0011)]' + (n-1)\left\{[(0110),(0001)]' + [(0101),(0010)]' - 2[(0100),(0010),(0001)]'\right\}\Big]$$

$$= \frac{1}{Nn}\left(\frac{1}{n} - \frac{1}{N}\right)\Big[[(1111)]' + (N-1)[(1000),(0111)]' - [(1100),(0011)]' - [(0100),(1011)]' - (N-2)[(1000),(0100),(0011)]'$$

contd .

$$+ (n-1)\Big\{ [(1110),(0001)]' + [(0110),(1001)]' + (N-2)[(1000),(0110),(0001)]'$$

$$+ [(1101),(0010)] + [(0101),(1010)]' + (N-2)[(1000),(0101),(0010)]'$$

$$- 2[(1100),(0010),(0001)]' - 2[(0100),(1010),(0001)]$$

$$- 2[(0100),(0010),(1001)]' - 2\,(N-2)[(1000),(0100),(0010),(0001)]'\Big\}\Big]$$

(3.4.26)

Finally we note that $\mathrm{Cov}\,(\bar{a}_i\,\bar{x}_i,\ \bar{a}_{i'}\,\bar{x}_{i'})$ is given by lemma 3.4.2. Now we are in a position to formulate the covariance (3.4.24) as under:

$$\mathrm{Cov}\Big[\bar{a}_i(\bar{x}_i - \bar{X}_i) - \Big(\frac{1}{n} - \frac{1}{N}\Big)s_{a_i x_i},\ \bar{a}_{i'}\bar{x}_{i'}\Big] = \frac{n(N-1)}{N(n-1)}(3.4.10) - \frac{(N-n)}{N(n-1)}(3.4.25) - (3.4.26)$$

(3.4.27)

Using (3.4.22), (3.4.23) and (3.4.27) in (3.4.21) it may be seen that

$$\Big[\bar{a}_i(\bar{x}_i - \bar{X}_i) - \Big(\frac{1}{n} - \frac{1}{N}\Big)s_{a_i x_i},\ \bar{a}_{i'}(\bar{x}_{i'} - \bar{X}_{i'}) - \Big(\frac{1}{n} - \frac{1}{N}\Big)s_{a_{i'} x_{i'}}\Big]$$

$$\frac{n^2(N-1)^2}{N^2(n-1)^2}(3.4.10) - \frac{n(N-1)(N-n)}{N^2(n-1)^2}(3.4.25) - \frac{n(N-1)}{N(n-1)}(3.4.26)$$

$$- \frac{(N-n)}{N(n-1)}(3.4.22) - (3.4.23). \qquad (3.4.28)$$

In the above formulation we have to collect the coefficient $Q[(ijkl)]'$ h of the population symmetric functions $[(ijkl)]'$. Thus it can be seen that

$$Q[(1111)]' = \left(\frac{1}{n} - \frac{1}{N}\right) \frac{(N-1)}{N^2(n-1)^2} \left[(N-1) - (N-n) - (n-1)\right] = 0$$

$$Q[(1100),(0011)] = \left(\frac{1}{n} - \frac{1}{N}\right) \frac{(N-1)}{N^2(n-1)^2} \left[-(N-1) + (N-n) + (n-1)\right] = 0$$

$$Q[(1010),(0101)] = \left(\frac{1}{n} - \frac{1}{N}\right) \frac{(N-1)}{N^2(n-1)} \left[N + n - 1 - (n-1)\right] = \left(\frac{1}{n} - \frac{1}{N}\right) \frac{(N-1)}{N(n-1)}$$

$$Q[(1001),(0110)]' = \left(\frac{1}{n} - \frac{1}{N}\right) \left[\frac{(N-1)}{N^2(n-1)} \left[N + n - 1 - n + 1\right] - \frac{1}{N}\right] = \left(\frac{1}{n} - \frac{1}{N}\right)^2 \frac{n}{(n-1)}$$

$$Q[(1000),(0111)]' = \left(\frac{1}{n} - \frac{1}{N}\right) \left[\frac{(N-1)^2}{N^2(n-1)} - \frac{(N-1)^2}{N^2(n-1)}\right] = 0$$

$$Q[(0100),(1011)]' = \left(\frac{1}{n} - \frac{1}{N}\right) \left[\frac{(N-1)}{N^2(n-1)} (N-1+1) - \frac{(N-n)}{N(n-1)} - \frac{1}{N}\right] = 0$$

$$Q[(0010),(1101)]' = Q[(0001),(1100)] = \left(\frac{1}{n} - \frac{1}{N}\right) \left[\frac{(N-1)}{N^2(n-1)} \left[(N-1) - (N-n) - (n-1)\right]\right] = 0$$

$$Q[(1100),(0010),(0001)]' = 2\left(\frac{1}{n} - \frac{1}{N}\right) \left[\frac{(N-1)}{N^2(n-1)} \left[-(N-1) + (N-n) + (n-1)\right]\right] = 0$$

$$Q[(1010),(0100),(0001)]' = \left(\frac{1}{n} - \frac{1}{N}\right) \left[\frac{(N-1)}{N^2(n-1)} \left[(n-2)(N-1) - n + 2(n-1)\right] + \frac{1}{N}\right]$$

$$= \left(\frac{1}{n} - \frac{1}{N}\right) \left[\frac{(N-1)(n-2)}{N(n-1)} + \frac{1}{N}\right] = \left(\frac{1}{n} - \frac{1}{N}\right) \frac{(Nn - 2N + 1)}{N(n-1)}$$

$$Q[(1001),(0100),(0010)]' = \left(\frac{1}{n}-\frac{1}{N}\right)\left[\frac{(N-1)}{N^2(n-1)}\left[(n-2)(N-1)-n+2(n-1)\right]-\frac{(N-2)}{N}\right]$$

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

$$Q[(1000),(0101),(0010)]' = \left(\frac{1}{n}-\frac{1}{N}\right)\left[\frac{(N-1)}{N^2(n-1)}\left[(n-2)(N-1)-n-(N-2)(n-1)\right]\right]$$

$$= -\left(\frac{1}{n}-\frac{1}{N}\right)\frac{(N-1)}{N(n-1)}$$

$$Q[(1000),(0110),(0001)]' = \left(\frac{1}{n}-\frac{1}{N}\right)\left[\frac{(N-1)}{N^2(n-1)}\left[(n-2)(N-1)-n-(N-2)(n-1)\right]+\frac{1}{N}\right]$$

$$= \left(\frac{1}{n}-\frac{1}{N}\right)\left[\frac{-(N-1)}{N(n-1)}+\frac{1}{N}\right] = -\left(\frac{1}{n}-\frac{1}{N}\right)\frac{(N-n)}{N(n-1)}$$

$$Q[(1000),(0100),(0011)]' = \left(\frac{1}{n}-\frac{1}{N}\right)\left[\frac{(N-1)}{N^2(n-1)}\left[-2(N-1)+(N-2)\right]+\frac{(N-n)}{N(n-1)}+\frac{1}{N}\right]$$

$$= \left(\frac{1}{n}-\frac{1}{N}\right)\left[\frac{-(N-1)}{N(n-1)}+\frac{(N-1)}{N(n-1)}\right] = 0$$

$$Q[(1000),(0100),(0001),(0010)]' = \left(\frac{1}{n}-\frac{1}{N}\right)\left[-\frac{2(N-1)}{N^2(n-1)}(3N+3n-2Nn-3)\right.$$

$$\left. + \frac{2(N-1)(N-3)}{N^2}+\frac{(N-3)}{N}\right]$$

$$= \left(\frac{1}{n} - \frac{1}{N}\right) \left[\frac{-2(N-1)(-Nn+2N)}{N^2(n-1)} + \frac{(N-3)}{N} \right] = -\left(\frac{1}{n} - \frac{1}{N}\right) \frac{(Nn - 3N + n + 1)}{N(n-1)} .$$

Hence $\text{Cov}\left[\bar{a}_i(\bar{x}_i - \bar{X}_i) - \left(\frac{1}{n} - \frac{1}{N}\right) s_{a_i x_i},\ \bar{a}_{i'}(\bar{x}_{i'} - \bar{X}_{i'}) - \left(\frac{1}{n} - \frac{1}{N}\right) s_{a_{i'} x_{i'}} \right]$

$$= \frac{1}{(n-1)} \left(\frac{1}{n} - \frac{1}{N}\right) \Bigg[\frac{(N-1)}{N} \left[(1010), (0101)\right]' + \frac{(N-n)}{N} \left[(1001), (0110)\right]'$$

$$+ \frac{(Nn-2N+1)}{N} \left[(1010), (0100), (0001)\right]' - \frac{(N-n)}{N} \left[(1001), (0100), (0010)\right]'$$

$$- \frac{(N-n)}{N} \left[(1000), (0110), (0001)\right]' - \frac{(N-1)}{N} \left[(1000), (0101), (0010)\right]'$$

$$- \frac{(Nn-3N+n+1)}{N} \left[(1000), (0100), (0010), (0001)\right]' \Bigg] \qquad (3.4.29)$$

= expression (3.4.20).

Hence the lemma is proved.

We now have the following theorem on $V(G_1^*)$.

<u>Theorem 3.4.1</u>: The exact variance of Mickey's averaged general unbiased estimator G_1^* (3.4.4) is given by

$$V(G_1^*) = \left(\frac{1}{n} - \frac{1}{N}\right) \Bigg[\frac{1}{N-1} \sum_{(1)}^{N} (y_j - \bar{Y})^2 - \frac{2}{N(N-1)} \sum_{i=1}^{p} \Big[\sum_{(2)}^{N} y_j x_{ij} a_{ij'} - \frac{1}{(N-2)} \sum_{(3)}^{N} y_j x_{ij} a_{ij''} \Big]$$

contd...

$$+ \frac{1}{N^2(n-1)} \sum_{i=1}^{p} \sum_{i'=1}^{p} \Big[\sum_{(2)}^{N} x_{ij}\, a_{ij'}\, x_{i'j}\, a_{i'j'} + \frac{(N-n)}{(N-1)} \sum_{(2)}^{N} x_{ij}\, a_{ij'}\, x_{i'j'}\, a_{i'j}$$

$$+ \frac{(Nn-2N+1)}{(N-1)(N-2)} \sum_{(3)}^{N} x_{ij}\, a_{ij'}\, x_{i'j}\, a_{i'j''} - \frac{(N-n)}{(N-1)(N-2)} \sum_{(3)}^{N} x_{ij}\, a_{ij'}\, x_{i'j''}\, a_{i'j}$$

$$- \frac{(N-n)}{(N-1)(N-2)} \sum_{(3)}^{N} x_{ij}\, a_{ij'}\, x_{i'j'}\, a_{i'j''} - \frac{1}{(N-2)} \sum_{(3)}^{N} x_{ij}\, a_{ij'}\, x_{i'j''}\, a_{i'j'}$$

$$- \frac{(Nn-3N+n+1)}{(N-1)(N-2)(N-3)} \sum_{(4)}^{N} x_{ij}\, a_{ij'}\, x_{i'j''}\, a_{i'j'''} \Big] \qquad (3.4.30)$$

Proof : From the formula (3.4.4) of G_1^*, we get

$$V(G_1^*) = V(\bar{y}) - 2\sum_{i=1}^{p} \mathrm{Cov}\Big[\bar{y},\ \bar{a}_i(\bar{x}_i - \bar{X}_i) - \Big(\frac{1}{n} - \frac{1}{N}\Big) s_{a_i x_i}\Big]$$

$$+ \sum_{i=1}^{p} \sum_{i'=1}^{p} \mathrm{Cov}\Big[\bar{a}_i(\bar{x}_i - \bar{X}_i) - \Big(\frac{1}{n} - \frac{1}{N}\Big) s_{a_i x_i},\ \bar{a}_{i'}(\bar{x}_{i'} - \bar{X}_{i'}) - \Big(\frac{1}{n} - \frac{1}{N}\Big) s_{a_{i'} x_{i'}}\Big]$$

$$(3.4.31)$$

We note $V(\bar{y}) = \left(\frac{1}{n} - \frac{1}{N}\right) \frac{1}{(N-1)} \sum_{(1)}^{N} (y_j - \bar{Y})^2$. Using this and the results of lemmas 3.4.4 and 3.4.5 in (3.4.31), we obtain the result (3.4.30) for $V(G_1^*)$.

Remark (b) : The population symmetric functions of higher orders (orders greater than one) appearing in the variance (3.4.30) may be expressed in terms of symmetric functions of order one to facilitate numerical computation of the variance. This exercise is, however, not attempted here.

Corollary 3.4.1a : When the population size (N) is large and the finite population correction is negligible ($N \gg n$), $V(G_1^*)$ takes the simpler form:

$$V(G_1^*) = \frac{1}{n} \Big[V(y) - 2 \sum_{i=1}^{p} E(a_i) Cov(y, x_i) + \sum_{i=1}^{p} \sum_{i'=1}^{p} E(a_i) E(a_{i'}) Cov(x_i, x_{i'}) \Big]$$

$$+ \frac{1}{n(n-1)} \Big[\sum_{i=1}^{p} \sum_{i'=1}^{p} Cov(x_i, x_{i'}) Cov(a_i, a_{i'}) + Cov(x_i, a_{i'}) Cov(x_{i'}, a_i) \Big] \qquad (3.4.32)$$

Proof : It may be noted that when the population is infinite,

$$\Big[(a_{11} a_{21} \ldots a_{p1}), (a_{12} a_{22} \ldots a_{p2}), \ldots (a_{1k} a_{2k} \ldots a_{pk}) \Big]'$$

$$= \Big[(a_{11} a_{21} \ldots a_{p1}) \Big]' \Big[(a_{12} a_{22} \ldots a_{p2}) \Big]' \ldots \Big[(a_{1k} a_{2k} \ldots a_{pk}) \Big]'. \qquad (3.4.33)$$

In large populations this equality may be assumed and if, further, $N \gg n$ we obtain from (3.4.30)

$$V(G_1^*) = \frac{1}{n} \Big[V(y) - 2 \sum_{i=1}^{p} \Big[E(y x_i) E(a_i) - E(y) E(x_i) E(a_i) \Big]$$

$$+ \frac{1}{(n-1)} \sum_{i=1}^{p} \sum_{i'=1}^{p} \Big[E(x_i x_{i'}) E(a_i a_{i'}) + E(x_i a_{i'}) E(a_i x_{i'}) + (n-2) E(x_i x_{i'}) E(a_i) E(a_{i'})$$

$$- E(x_i a_{i'}) E(a_i) E(x_{i'}) - E(x_i) E(a_i x_{i'}) E(a_{i'}) - E(x_i) E(a_i a_{i'}) E(x_{i'})$$

$$- (n-3) E(x_i) E(a_i) E(x_{i'}) E(a_{i'}) \Big] \Big] = \text{expression (3.4.32)}.$$

Corollary 3.4.1b: With a single auxiliary variate x, the exact variance of G_1^* (3.4.4) is given by

$$V(G_1^*)=\left(\frac{1}{n}-\frac{1}{N}\right)\Big[\frac{1}{(N-1)}\sum_{(1)}^{N}(y_j-\bar{Y})^2-\frac{2}{N(N-1)}\Big[\sum_{(2)}^{N}y_jx_ja_{j'}-\frac{1}{(N-2)}\sum_{(3)}^{N}y_jx_{j'}a_{j''}\Big]$$

$$+\frac{1}{N^2(n-1)}\Big[\sum_{(2)}^{N}x_j^2a_{j'}^2+\frac{(N-n)}{(N-1)}\sum_{(2)}^{N}x_ja_jx_{j'}a_{j'}+\frac{(Nn-2N+1)}{(N-1)(N-2)}\sum_{(3)}^{N}x_j^2a_{j'}a_{j''}$$

$$-\frac{2(N-n)}{(N-1)(N-2)}\sum_{(3)}^{N}x_ja_jx_{j'}a_{j''}-\frac{1}{(N-2)}\sum_{(3)}^{N}x_ja_{j'}^2x_{j''}$$

$$-\frac{(Nn-3N+n+1)}{(N-1)(N-2)(N-3)}\sum_{(4)}^{N}x_ja_{j'}x_{j''}a_{j'''}\Big]\Big] \qquad (3.4.34)$$

Proof: This follows from (3.4.30) by putting $i = i'$ and $p = 1$.

Corollary 3.4.1c: In large populations with the finite population correction negligible, variance of G_1^* (3.4.4) based on a single auxiliary variate x has the form:

$$V(G_1^*)=\frac{1}{n}\Big[V(y)-2E(a)\,Cov(y,x)+E^2(a)V(x)\Big]+\frac{1}{n(n-1)}\Big[V(x)V(a)+Cov^2(x,a)\Big]. \qquad (3.4.35)$$

Proof: This follows from (3.4.32) by putting $i = i'$ and $p = 1$.

3.4b. Large sample variance of G_m^*:

In this sub-section we prove two theorems on the variance of G_m^* (3.4.3) under some assumptions on the coefficient functions $a_i(Z_m)$, $(i=1,2,\ldots,p)$. In the subsequent chapters it will be shown that for several forms of the coefficient functions $a_i(Z_m)$ resulting in unbiased ratio, product, and regression type estimators, these assumptions (under which the variance is deduced here) hold good in large samples. It will thus be possible to derive the large sample variance of a number of unbiased estimators from the theorems proved here.

In stating the theorems we use the following definitions:

$$S_Y^2 = \frac{1}{(N-1)}\sum_{j=1}^{N}(y_j-\bar{Y})^2,\quad S_{x_iy} = \frac{1}{(N-1)}\sum_{j=1}^{N}x_{ij}(y_j-\bar{Y}),\ \text{and}\ S_{x_ix_{i'}} = \frac{1}{(N-1)}\sum_{j=1}^{N}x_{ij}(x_{i'j}-\bar{X}_{i'})$$

(3.4.36)

Also we denote by μ_2 and μ_4 the second and fourth central moments.

<u>Theorem 3.4.2</u> : If the choice of coefficient functions $a_i(Z_m)$, $(i=1,2,\ldots,p)$, is such that

$$\mu_2\left[a_i^*(Z_m)\right] = O(n^{-v_i}),\quad \mu_4\left[a_i^*(Z_m)\right] = O(n^{-s_i})\ \text{and}$$

$$\mu_2\left[\mathrm{Cov}_n\left[a_i(Z_m),\ \bar{x}_{i\,n-m}\right]\right] = O(n^{-k_i}),\qquad (3.4.37)$$

where $v_i > 0$, $s_i > 0$ and $k_i > 1$; then in large samples the unbiased estimator G_m^* (3.4.3) has the variance:

$$V(G_m^*) = \left(\frac{1}{n}-\frac{1}{N}\right)\left[S_y - 2\sum_{i=1}^{p}E\left[a_i(Z_m)\right]S_{x_iy} + \sum_{i=1}^{p}\sum_{i'=1}^{p}E\left[a_i(Z_m)\right]E\left[a_{i'}(Z_m)\right]S_{x_ix_{i'}}\right]$$

(3.4.38)

We note $V(Q) = O(n^{-1})$ and for any two random variables h_1, h_2 having finite second moments $|\,Cov(h_1, h_2)\,| \leq [V(h_1)V(h_2)]^{\frac{1}{2}}$. Consequently from (3.4.43) it follows that σ_1^2 will be of smaller order than n^{-1} if $V(\Theta)$ and $V(\phi)$ are of smaller orders than n^{-1}. Again, by the same argument, from the definitions (3.4.41) and (3.4.42) of Θ and ϕ we infer that $V(\Theta)$ and $V(\phi)$ will be of smaller orders than n^{-1} if $V(\Theta_i)$ and $V(\phi_i)$ have that property. Now it may be noted that the last one of our conditions (3.4.37) says that $V(\phi_i)$ is of smaller order than n^{-1}. It, therefore, remains to show that $V(\Theta_i)$ is of smaller order than n^{-1} under the conditions (3.4.37).

Noting $E[a_i(Z_m)] = E[E_n[a_i(Z_m)]] = E[a_i^*(Z_m)]$, we have from the definition (3.4.41)

$$V(\Theta_1) = E\left[[a_i^*(Z_m) - E\{a_i^*(Z_m)\}]^2(\bar{x}_i - \bar{X}_i)^2\right] - Cov^2[a_i^*(Z_m), \bar{x}_i]. \tag{3.4.44}$$

Using the conditions (3.4.37) we observe:

$$\text{first term in (3.4.44)} \leq \left[\mu_4[a_i^*(Z_m)]\,\mu_4(\bar{x}_i)\right]^{\frac{1}{2}}$$
$$= O\left(n^{-\frac{(s_i+2)}{2}}\right), \tag{3.4.45}$$

and

$$Cov^2[a_i^*(Z_m), \bar{x}_i] \leq \mu_2[a_i^*(Z_m)]\,\mu_2(\bar{x}_i)$$
$$= O\left(n^{-(v_i+1)}\right). \tag{3.4.46}$$

ince $s_i > 0$ and $v_i > 0$, it now follows from (3.4.44), (3.4.45) and (3.4.46) hat $V(\Theta_1)$ is of smaller order than n^{-1}

This completes the proof of the theorem.

We now prove another theorem which uses a different set of ssumptions but gives the same expression (3.4.38) for the large sample ariance of G_m^* (3.4.3).

Theorem 3.4.3 : In large samples, when m is small in comparison with the sample size n, the unbiased estimator G_m^* (3.4.3) has the variance given by expression (3.4.38).

Proof : When n is large and m is small in comparison with n,

$$\frac{m(N-n)}{(n-m)N} = \left(\frac{N-m}{N}\right)\left(\frac{n}{n-m}\right) - 1$$

$$\doteq 1 - 1 = 0.$$

ence in this case G_m^* (3.4.3) reduces to

$$G_m^* \doteq \bar{y} - \sum_{i=1}^{p} a_i^*(Z_m)(\bar{x}_i - \bar{X}_i). \qquad (3.4.47)$$

Let us now write

$$a_i^*(Z_m) = E_n[a_i(Z_m)] = E[a_i(Z_m)] + e_{in},$$

$$\bar{x}_i = \bar{X}_i + e_{ix},$$

$$\text{and} \qquad = \quad + e_y$$

Then $E(e_{ia}) = E(e_{ix}) = E(e_y) = 0$ and from (3.4.47)

$$(G_m^* - \bar{Y}) = e_y - \sum_{i=1}^{p} e_{ix} \left[e_{ia} + E\left\{a_i(Z_m)\right\} \right].$$

Hence in large samples, when m is small in comparison with n,

$$V(G_m^*) = E\left[e_y - \sum_{i=1}^{p} e_{ix} \left[e_{ia} + E\left\{a_i(Z_m)\right\} \right] \right]^2. \quad (3.4.48)$$

As the sample size is large, in (3.4.48) we may neglect the contribution of terms involving the powers in e's higher than second and obtain

$$V(G_m^*) = E\left[e_y^2 - 2\sum_{i=1}^{p} E\left[a_i(Z_m)\right] e_y e_{ix} + \sum_{i=1}^{p} \sum_{i'=1}^{p} E\left[a_i(Z_m)\right] E\left[a_{i'}(Z_m)\right] e_{ix} e_{i'x} \right]$$

$$= \text{expression } (3.4.38).$$

3.5. Effect of introducing an additional auxiliary variate:

One of the desirable properties of a multivariate estimator is that whenever an additional auxiliary variate is introduced in the estimator, it does not result in a decrease in the efficiency of the estimator. In this section we examine the multivariate unbiased estimator G_m^* (3.4.3) from this angle.

To start with suppose $a_1(Z_m)$ is a coefficient function defined on y and x_1 only. Let us denote by G_{1m}^* the unbiased estimator of the form (3.4.3) using the single auxiliary variate x_1 and the coefficient function $a_1(Z_m)$.

Formally

$$G^*_{im} = \bar{y} - a^*_i(Z_m)(\bar{x}_i - \bar{X}_i) + \frac{m(N-n)}{N(n-m)} \operatorname{Cov}_n[a_i(Z_m), \bar{x}_{im}] \quad (3.5.1)$$

Corresponding to the p auxiliary variates we may define p such estimators G^*_{im}, each depending on a single auxiliary variate.

Now using the <u>same</u> p coefficient functions $a_i(Z_m)$, $(i=1,2,\ldots,p)$, suppose we construct the multivariate estimator G^*_m (3.4.3). Then we note

$$G^*_m = \bar{y} - \sum_{i=1}^{p} a_i(Z_m)(\bar{x}_i - \bar{X}_i) + \frac{m(N-n)}{N(n-m)} \sum_{i=1}^{p} \operatorname{Cov}_n[a_i(Z_m), \bar{x}_{im}] \quad (3.5.2)$$

$$= \sum_{i=1}^{p} G^*_{im} - (p-1)\,\bar{y}, \quad \text{where } G^*_{im} \text{ is given by (3.5.1).} \quad (3.5.3)$$

From (3.5.3) it is clear that the multivariate estimator G^*_m (3.5.2) <u>need not have less variance than</u> an estimator G^*_{im} based on a single auxiliary variate.

As an alternative, consider the multivariate estimator G^*_m based on the coefficient functions $a'_i(Z_m) = W_i\, a_i(Z_m)$, where $W = (W_1, W_2, \ldots, W_p)$ is a <u>known</u> weight vector $(\sum_{i=1}^{p} W_i = 1)$. Then

$$G^*_m = \bar{y} - \sum_{i=1}^{p} W_i a'_i(Z_m)(\bar{x}_i - \bar{X}_i) + \frac{m(N-n)}{N(n-m)} \sum_{i=1}^{p} W_i \operatorname{Cov}_n[a_i(Z_m), \bar{x}_{im}] \quad (3.5.4)$$

$$\sum_{i=1}^{p} W_i\, G^*_{im}, \quad \text{where } G^*_{im} \text{ is given by (3.5.1).} \quad (3.5.5)$$

Now it is possible to choose the weight vector W in such a way that G^*_m (3.5.5) does not loose precision in comparison with G^*_{im}. The optimum choice for W will naturally depend on the variance-covariance matrix of the vector (G^*_{1m}, G^*_{2m}, ..., G^*_{pm}) of estimators. The important point to be noted here is that the weighted estimator $\sum_{i=1}^{p} W_i G^*_{im}$ with a __non-random__ weight vector W can be accomodated in the class of unbiased estimators G^*_m (3.4.5) with an appropriate definition of the coefficient functions. This enables us to apply to $G^*_m = \sum_{i=1}^{p} W_i G^*_{im}$ the theorems on variance proved in section 3.4 and determine the optimum choice of the weight vector W.

The above discussion will be of use when we formulate some biased ratio-cum-product estimators in chapter - VI.

__Estimation of variance:__

In this section we consider the problem of variance estimation for different classes ($\bar{G}_m$, $\bar{G}'_{n-m}$, G^*_m and $\bar{G}^{(k)}$) of unbiased estimators described in section 3.3. For the class of estimators $\bar{G}^{(k)}$ (3.3.2) we have already mentioned the solution (3.3.3) suggested by Mickey (1959) which rests on the fact that $\bar{G}^{(k)}$ is the average of k mutually un-correlated unbiased estimators. Such a solution is not available for the classes $\bar{G}_m$, $\bar{G}'_{n-m}$ and G^*_m because the estimators used for their construction are all mutually correlated. We give here a common solution to the problem that is applicable to any class of estimators involving Mickey's basic unbiased estimators G_m (3.1.3). Besides, for the sub-class of estimators G^*_1 (3.4.4) or which an exact variance formula has been derived in section 3.4a, we may frame an alternate unbiased estimator of variance different from the ne obtainable through the common solution. Also, in large samples, for the class of estimators G^*_m for which the variance formula (3.4.38) is pplicable we can have a simpler estimator of variance. In the subsequent hapters, for specific choices of coefficient functions $a_i\,(Z_m)$ resulting in nbiased ratio, product, and regression type estimators, these methods of ariance estimation are illustrated with numerical examples.

We now formally present the methods of variance estimation.

In developing a method which is applicable to all classes of estimators $\bar{G}_m$, $\bar{G}'_{n-m}$, G^*_m and $\bar{G}^{(k)}$, our approach is to first obtain an nbiased estimator for the variance of the basic estimator G_m (3.1.2). For his we recall our discussion on the variance of G_m in section 3.3. Using the notations adopted in that section, we now define the sample version of $\sigma^2_{y'}$ (3.3.4) as

$$s^2_{y'} = \frac{1}{(n-m-1)} \sum_{j \notin Z_m}^{n-m} (y'_j - \bar{y}'_{n-m})^2 . \qquad (3.6.1)$$

We note that $s^2_{y'}$ is defined only for $1 \leq m \leq n-2$. When $m = n-1$, there is only one element, say j-th element, not belonging to Z_{n-1}. In this case we rite U_m (3.1.1) as

$$U_{n-1} = y_j - \sum_{i=1}^{p} a_i\,(Z_{n-1})\,(x_{ij} - \bar{X}_{i\ N-n+1}) . \qquad (3.6.2)$$

We also denote $s_y^2 = \frac{1}{(n-1)} \sum_{j=1}^{n} (y_j - \bar{y})^2$. (3.6.3)

In terms of the above notations we have the following theorem.

Theorem 3.6.1 : For the basic unbiased estimator G_m (3.1.2) an unbiased estimator of variance is given by

$$\hat{V}(G_m) = \frac{(1-f_m)(1-f_n)}{(n-m)} s_{y'}^2, \quad \text{for } 1 \leq m \leq n-2, \qquad (3.6.4)$$

$$= (1-f_{n-1}) \left[(1-f_{n-1})(U_{n-1}^2 - y_j^2) + (1-f_n) s_y^2 \right] \qquad (3.6.5)$$

for $m = n-1$;

ere $f_m = \frac{m}{N}$, $f_{n-1} = \frac{n-1}{N}$, and $f_n = \frac{n}{N}$

oof : While proving the theorem 3.3.1 on the variance of G_m, we have erved

$$V(G_m) = \left(\frac{N-m}{N}\right)^2 E\left[V_m(U_m) \right] \qquad (3.6.6)$$

$$= \frac{(1-f_m)(1-f_n)}{(n-m)} E(S_{y'}^2), \qquad (3.6.7)$$

re $S_{y'}^2$ is given by (3.3.4). When $1 \leq m \leq n-2$, we note that for a given , $s_{y'}^2$ (3.6.1) is an unbiased estimator of $S_{y'}^2$ (3.3.4). Using this in .7) we obtain the first part (3.6.4) of the theorem.

When $m = n-1$, assuming that the j-th sample element is not belonging $_{n-1}$, we re-write (3.6.6) as

$$V(G_{n-1}) = (1 - f_{n-1})^2 E\left[V_{n-1}(U_{n-1}) - V_{n-1}(y_j) + V_{n-1}(y_j) \right]. \quad (3.6.8)$$

Now it may be noted $E_{n-1}(U_{n-1}) = \bar{Y}_{N-n+1} = E_{n-1}(y_j)$,

$$V_{n-1}(y_j) = \frac{(N-n)}{(N-n+1)} S^2_{y,N-n+1}, \text{ where } S^2_{y,N-n+1} = \frac{1}{(N-n)} \sum_{j \notin Z_{n-1}}^{N-n+1} (y_j - \bar{Y}_{N-n+1})$$

Using these in (3.6.8), we observe

$$V(G_{n-1}) = (1-f_{n-1})^2 E\left[E_{n-1}(U^2_{n-1}) - E_{n-1}(y_j^2) + \frac{(N-n)}{(N-n+1)} S^2_{y,N-n+1} \right]$$

$$=(1-f_{n-1})\left[(1-f_{n-1})E(U^2_{n-1} - y_j^2) + (1-f_n) S^2_y\right]. \quad (3.6.9)$$

where $S^2_y = \frac{1}{(N-1)} \sum_{j=1}^{N} (y_j - \bar{Y})^2$

ow since s^2_y (3.6.3) is unbiased for S^2_y, we obtain from (3.6.9) the second art (3.6.5) of the theorem.

This completes the proof of the theorem.

emark (a) : In the above theorem the suggested estimator of variance is on-negative for $1 \leq m \leq n-2$, while for $m = n-1$ it may admit negative values. ...r in a number of empirical studies reported in subsequent chapters, it is ...t that the estimator of variance for $m = n-1$, whenever it is positive, nsistently overestimated the variance. As such the solution for $m = n-1$ not satisfactory.

Remark (b) : For an unbiased estimator belonging to any of the four classes of unbiased estimators: $\bar{G}_m$, $\bar{G}'_{n-m}$, G^*_m and $\bar{G}^{(k)}$, an unbiased estimator of variance may be formulated using theorem 3.6.1 and the following lemma.

Lemma 3.6.1 : If $G(1), G(2), \ldots, G(h)$ are h basic unbiased estimators of the form G_m (3.1.2), then for the weighted average $G_w = \sum_{t=1}^{h} W_t G(t)$, $(\sum_{t=1}^{h} W_t = 1)$, an unbiased estimator of variance may be computed as

$$\hat{V}(G_w) = \sum_{t=1}^{h} W_t \left[\hat{V}[G(t)] - [G(t) - G_w]^2 \right], \quad (3.6.10)$$

where $\hat{V}[G(t)]$ is an unbiased estimator of variance of $G(t)$.

Proof : We note G_w is also unbiased for the population mean $\bar{Y}$. Now writing

$$\hat{V}(G_w) = \sum_{t=1}^{h} W_t \left[\hat{V}[G(t)] - [G(t) - \bar{Y} + \bar{Y} - G_w]^2 \right],$$

$$= \sum_{t=1}^{h} W_t \left[\hat{V}[G(t)] - [G(t) - \bar{Y}]^2 \right] + (G_w - \bar{Y})^2,$$

we have

$$E[\hat{V}(G_w)] = \sum_{t=1}^{h} W_t \left[V[G(t)] - V[G(t)] \right] + V(G_w),$$

$$= V(G_w).$$

Remark (c) : $\hat{V}(G_w)$ given by (3.6.10) may assume negative values.

We now formulate the result which will be used repeatedly in the later chapters.

Corollary 3.6.1: For the averaged general unbiased estimator G_m^* (3.4.3) unbiased estimator of variance is provided by

$$\hat{V}(G_m^*) = \frac{1}{\binom{n}{m}} \sum_{Z_m}^{\binom{n}{m}} \left[\hat{V}(G_m) - (G_m - G_m^*)^2 \right], \qquad (3.6.11)$$

re $\hat{V}(G_m)$ is given by theorem 3.6.1.

oof: Follows from theorem 3.6.1 and lemma 3.6.1.

Another method of variance estimation is available for the sub-class unbiased estimators G_1^* (3.4.4) by using the result (3.4.30) on the exact iance of G_1^*. This is contained in the following theorem.

orem 3.6.2: The variance of the unbiased estimator G_1^* (3.4.4) is biasedly estimated by

$$\hat{V}(G_1^*) = \left(\frac{1}{n} - \frac{1}{N}\right) \Bigg\{ \Bigg[\frac{1}{(n-1)} \sum_{(1)}^{n} (y_j - \bar{y})^2 - \frac{2}{n(n-1)} \sum_{i=1}^{p} \Big[\sum_{(2)}^{n} y_j x_{ij} a_{ij'}$$

$$- \frac{1}{(n-2)} \sum_{(3)}^{n} y_j x_{ij'} a_{ij''} \Big] + \frac{1}{Nn(n-1)} \sum_{i=1}^{p} \sum_{i'=1}^{p} \Big[\frac{(N-1)}{(n-1)} \sum_{(2)}^{n} x_{ij} a_{ij} x_{i'j} a_{i'j'}$$

$$+ \frac{(N-n)}{(n-1)} \sum_{(2)}^{n} x_{ij} a_{ij'} x_{i'j'} a_{i'j} + \frac{(Nn-2N+1)}{(n-1)(n-2)} \sum_{(3)}^{n} x_{ij} a_{ij'} x_{i'j} a_{i'j''}$$

$$- \frac{(N-n)}{(n-1)(n-2)} \sum_{(3)}^{n} x_{ij} a_{ij'} x_{i'j''} a_{i'j} - \frac{(N-n)}{(n-1)(n-2)} \sum_{(3)}^{n} x_{ij} a_{ij'} x_{i'j'} a_{i'j''}$$

$$- \frac{(N-1)}{(n-1)(n-2)} \sum_{(3)}^{n} x_{ij} a_{ij'} x_{i'j''} a_{i'j'} - \frac{(Nn-3N+n+1)}{(n-1)(n-2)(n-3)} \sum_{(4)}^{n} x_{ij} a_{ij'} x_{i'j''} a_{i'j'''} \Big] \Bigg] \Bigg\}. \qquad 3.6.12$$

Proof: We use lemma 3.4.1 and theorem 3.4.1 to obtain this result.

Remark (d) : Est. $V(G_1^*)$ may also assume negative values. It is not the same as $\hat{V}(G_1^*)$ given by (3.6.11), the obvious difference being the use of information on the population means $\bar{X}_i$, $(i = 1, 2, \ldots, p)$, in $\hat{V}(G_1^*)$ while Est. $V(G_1^*)$ is completely based on sample data only. It may be relatively easy to compute $\hat{V}(G_1^*)$ than Est. $V(G_1^*)$.

Finally we observe that in large samples if the variance formula (3.4.38) is applicable to G_m^* (3.4.3), then we may use the following result to compute an estimate for the variance of G_m^*.

Theorem 3.6.3 : In large samples, when the variance of G_m^* (3.4.3) is given by (3.4.38), an estimator of variance of G_m^* is :

$$\hat{V}(G_m^*) = \left(\frac{1}{n} - \frac{1}{N}\right)\left[s_y^2 - 2\sum_{i=1}^{p} a_i^*(Z_m)\, s_{x_i y} + \sum_{i=1}^{p}\sum_{i'=1}^{p} a_i^*(Z_m)\, a_{i'}^*(Z_m)\, s_{x_i x_{i'}} \right], \tag{3.6.13}$$

where $s_y^2 = \frac{1}{(n-1)}\sum_{j=1}^{n} (y_j - \bar{y})^2$, $s_{x_i y} = \frac{1}{(n-1)}\sum_{j=1}^{n} y_j (x_{ij} - \bar{x}_i)$ and

$$s_{x_i x_{i'}} = \frac{1}{(n-1)}\sum_{j=1}^{n} x_{ij}\,(x_{i'j} - \bar{x}_{i'}).$$

Proof : Since $E[a_i(Z_m)] = E[a_i^*(Z_m)]$, $E(s_y^2) = S_y^2$, etc.;

(3.6.13) is an estimator of (3.4.38).

CHAPTER IV

UNBIASED RATIO TYPE ESTIMATORS

4.0 Summary

In this chapter we deal with the study of two classes of unbiased ratio type estimators which may be obtained from Mickey's averaged general unbiased estimator G_m^* for specific choices of the coefficient functions $a_i(Z_m)$. The chapter is organised into five sections. The results reported in the individual sections are briefly presented in the following.

Mickey (1959) has shown that the estimator G_m^* results in a ratio type estimator, say T_m^*, if a single auxiliary variate x positively correlated with the study variate y is used and $a(Z_m)$ is chosen as $R_m = \bar{y}_m / \bar{x}_m$, where $\bar{y}_m$ and $\bar{x}_m$ are the means based on Z_m. He has pointed out that Hartley-Ross (1954) unbiased ratio type estimator $\hat{\bar{Y}}_{HR}$ is a special case of T_m^* for the choice m=1. Alternately, he has also arrived at $\hat{\bar{Y}}_{HR}$ by choosing $a(Z_m) = \bar{r}_m = \frac{1}{m} \sum_{j \in Z_m} y_j/x_j$ in G_m^* for any m in the range 1 to n-1. These aspects have been reviewed in the introductory section.

In section 2, we show that for the choice $a(Z_m) = \frac{q\,\bar{y}_m}{\bar{X}}$, where $q > 0$ and $\bar{X}$ is the population mean of x, the estimator G_m^* reduces to a new unbiased ratio type estimator, say $S_{(q)}^*$.

Interestingly $S^*_{(q)}$ is the same for all values of m in the range 1 to n-1. Also $S^*_{(q)}$ is a linear combination of the sample mean $\bar{y}$ and Robson's (1957) unbiased product type estimator $\hat{\bar{Y}}_{RP}$ and is given by $S^*_{(q)} = (q+1)\,\bar{y} - q\,\hat{\bar{Y}}_{RP}$. By itself, the unbiased product type estimator $\hat{\bar{Y}}_{RP}$ is recommended for use when x is negatively correlated with y. However, here when x is positively associated with y, $\hat{\bar{Y}}_{RP}$ appears in $S^*_{(q)}$ with a negative coefficient since q is taken to be a positive constant.

Both the Hartley-Ross ratio type estimator $\hat{\bar{Y}}_{HR}$ and the new ratio type estimator $S^*_{(q)}$ may be considered as special cases of G^*_1 for the choices $a(Z_1) = y_j / x_j$ and $a(Z_1) = q\,y_j / \bar{X}$ respectively. Using this fact, in section 3 we derive the exact variances of $\hat{\bar{Y}}_{HR}$ and $S^*_{(q)}$ from the exact variance formula for G^*_1 obtained in chapter III. It may, however, be noted that Robson (1957) has earlier independently obtained the exact variance of $\hat{\bar{Y}}_{HR}$. Rao and Pereira (1968) have proposed two unbiased double ratio type estimators in sampling on two occasions for estimating the current population mean of the study variate. In section 3 we show that one of their estimators is a special case of G^*_1 with three auxiliary variates and appropriate choices of $a_i(Z_1)$ and, consequently, its exact variance is also derivable from that of G^*_1. When

$N >> n$, the exact variance of the unbiased double ratio type estimator assumes a simpler form. This simplified version of the variance has earlier been obtained by Rao and Pereira (1968) by making use of (i) Goodman and Hartley's (1958) variance formula for $\hat{Y}_{HR}$ and (ii) the formula for the covariance between two $\hat{Y}_{HR}$ type estimators as given by Robson and Vithayasai (1961). In our approach, the exact variance of the unbiased double ratio type estimator can be derived directly from that of G_1^* with appropriate choices of a_i (Z_i).

When $N >> n$, the variance of $S^*_{(q)}$ is further studied in detail in section 3. For instance, we show that $S^*_{(1)}$ (i.e., for the choice $q = 1$) has the same large sample variance as the classical biased ratio estimator $\hat{Y}_R = (\bar{y} / \bar{x}) \bar{X}$. When the variance or mean square error is considered upto order n^{-2}, $S^*_{(1)}$ is more efficient than $\hat{Y}_R$ if (i) the joint distribution of (y, x) is bivariate normal, (ii) the coefficients of variation of y and x are equal (i.e., $C_y = C_x$), and (iii) the correlation ρ between y and x does not exceed 0.78. For $S^*_{(q)}$ one can also think of the optimum choice for q in the sense of minimising the variance of $S^*_{(q)}$. In this sense, the optimum choice q_o of q is obtained and the variance of $S^*_{(q_o)}$ is derived. Further, it is shown that the choice $q_o' = \rho\, C_y / C_x$ is a good approximation to the optimum choice q_o. It is also interesting to note that both $S^*_{(q_o)}$ and $S^*_{(q'_o)}$ have the same large sample variance as the classical biased linear regression estimator $\hat{Y}_r = \bar{y} + b\ (\bar{X} - \bar{x})$, where b is the sample linear regression coefficient of y on x. Information on q'_o available from a pilot survey or previous census or past experience may be used to construct $S^*_{(q'_o)}$ without disturbing its unbiased nature.

It may be recalled that in chapter III, a simplified expression has been derived for the large sample variance of G^*_m either under certain assumptions on $a_i\ (Z_m)$ or by assuming that m is small in comparison with n. In section 4 of this chapter, we show that for the choice $a\ (Z_m) = R_m$ those assumptions are satisfied provided n is large and m is not small relative to n. Consequently, we use the results on the large sample variance of G^*_m and obtain the variance of Mickey's unbiased ratio type estimators T^*_m in large samples for all values of m in the range 1 to n-1. Further, we prove that T^*_m has the same large sample variance as the classical ratio estimator $\hat{Y}_R$ for any large value of m; and, in particular, T^*_{n-1} and $\hat{Y}_R$ equally efficient in large samples.

Since ratio method of estimation is generally considered appropriate in populations where regression of y on x is linear, it is of interest to know the behaviour of T^*_m in such populations for various choices of m in the range (1, n-1). In this direction, two results of section 4 may be mentioned. When the regression of y on x is linear and passes through the origin, in large samples, all choices of m result in the same variance for T^*_m which is also equal to the variance of $\hat{Y}_R$. When the regression line does not pass through the origin but all values of x in the population are of same sign, we show that the choice m = n - 1 is the optimum choice for T^*_m; and since T^*_{n-1} and $\hat{Y}_R$ have the same large sample variance, it follows that none of the estimators T^*_m is more efficient than $\hat{Y}_R$ in large samples.

In section 5, we consider the problem of estimating the variance of the estimators T^*_m and $S^*_{(q)}$. Unbiased estimators of variance valid for sample size $n \geq 2$ and biased but simpler estimators of variance applicable in large samples are derived by making use of the procedures developed in chapter III for the estimation of variance of G^*_m. For purposes of numerical illustration, we consider five live populations (data given in section 5) and from each of them select one sample each of sizes 2, 3, 4 and 5 and compute unbiased estimates of variance for the estimators $N\bar{y}$, $N T^*_m$ $(1 \leq m < n)$, $N S^*_{(1)}$, and $N S^*_{(q'_o)}$. These estimates are presented in Tables - 4.5 to 4.9 and the tentative conclusions in small samples are as follows. All the unbiased ratio type estimators performed better than the simple estimator $N \bar{y}$. Among the ratio type estimators, in four of the five populations, the estimators $N T^*_m$ appear to be more efficient than the estimators $N S^*_{(1)}$ and $N S^*_{(q_o')}$. Among the estimators $N T^*_m$ (for $1 \leq m \leq n$), there is a visible trend of increasing efficiency as m increases from 1 to $n - 1$. Following a different approach, in chapter VIII we examine more thoroughly in small samples of sizes 2, 3, 4 and 5 the relative efficiencies of (i) the unbiased ratio type estimators considered in this chapter, (ii) some unbiased regression type estimators to be studied in chapter VII, and (iii) the classical biased ratio and regression estimators, using a reasonable number of independent samples drawn from each of the five populations used in this chapter and also four additional populations.

In addition to the illustrations in small samples, in section 5 of this chapter estimates of large sample variance are also computed for the estimators $N T^*_m$, $N S^*_{(1)}$, $N S^*_{(q_o')}$ and $N \hat{Y}_R$ using samples of four different sizes: 10, 15, 20, and 25 drawn from a population of size 142. For each sample size, estimates of variance are computed on more than one sample. Thus 15, 10, 8, and 5 independent samples are used respectively.

for the four sizes 10, 15, 20 and 25. The estimates of variance are presented in tables 4.15 to 4.18. The important conclusions are as follows. The precision of the estimator $N\,T^*_m$ has clearly improved as m increases from 1 to n-1. The estimators $N\,T^*_{n-1}$ and $N\,S^*_{(1)}$ are of equal efficiency as the classical estimator $N\,\hat{\bar{Y}}_R$. The estimator $N\,S^*_{(q'_o)}$, which uses information on $\rho\,C_y/C_x$, is the most efficient among the estimators compared. These empirical results are consistent with our theoretical results in large samples.

4.1. Introduction

In the previous chapter we have considered Mickey's general unbiased estimators without specifying the form of the coefficient functions $a_i(Z_m)$, $(i = 1, 2, \ldots, p)$. In this chapter we study unbiased ratio type estimators which result from the general estimators for specific forms of the coefficient functions $a_i(Z_m)$.

When information on a single auxiliary variate x positively correlated with y is available, for the choice $a(Z_m) = R_m = \frac{\bar{y}_m}{\bar{x}_m}$ Mickey has shown that the unbiased estimators G_m (3.4.1) and G^*_m (3.4.3) result in the unbiased ratio type estimators:

$$T_m = R_m \bar{X} + \frac{(N-m)\,n}{N(n-m)} \left(\bar{y} - R_m \bar{x} \right), \qquad (4.1.1)$$

and

$$T^*_m = R^*_m \bar{X} + \frac{(N-m)\,n}{N(n-m)} \left(\bar{y} - R^*_m \bar{x} \right), \qquad (4.1.2)$$

where

$$R^*_m = E_n(R_m) = \frac{1}{\binom{n}{m}} \sum_{Z_m}^{\binom{n}{m}} R_m .$$

Hartley-Ross (1954) unbiased ratio type estimator may be derived either from G^*_m (3.4.3) with $p = 1$ by taking $a(Z_m) = \bar{r}_m = \frac{1}{m} \sum_{j \in Z_m} r_j$ ($r_j = y_j / x_j$), for any m ($< n$); or from T^*_m (4.1.2) with $m = 1$. In our discussion we take T^*_1 to represent Hartley-Ross estimator:

$$T^*_1 = \bar{r} \bar{X} + \frac{(N-1)\,n}{N(n-1)} \left(\bar{y} - \bar{r} \bar{x} \right). \qquad (4.1.3)$$

We recall our discussion in section 2.4 on Mickey's unbiased ratio type estimators for the population ratio $R = \bar{Y} / \bar{X}$, and note that with the definitions of t_m(2.4.2) and t^*_m (2.4.5) we have $t_m \bar{X} = T_m$ (4.1.1) and $\bar{X} = T^*_m$ (4.1.2). The classes of unbiased ratio type estimators $\bar{t}_m$(2.4.3), t'_m(2.4.4) and t^*_m (2.4.5) mentioned there correspond to the classes of general unbiased estimators $\bar{G}_m$, $\bar{G}'_{n-m}$ and G^*_m defined in section 3.2. The available literature on Hartley-Ross and Mickey's unbiased ratio type estimators has already been reviewed in section 2.4, the main contributions being those of Robson (1957), Goodman and Hartley (1958), Rao (1967), ... and Beeagle (1967) and Rao (1969).

We now present our results beginning with a new class of unbiased ratio type estimators.

4.2. <u>A new class of unbiased ratio type estimators</u>:

For a single auxiliary variate x positively correlated with y, like the choice $a(Z_m) = \bar{y}_m / \bar{x}_m$, we may consider a choice $a(Z_m) = q\, \bar{y}_m / \bar{X}$, where q is a positive constant. With this choice we have the following theorem.

<u>Theorem 4.2.1</u>: Mickey's general unbiased estimator G_m^* (3.4.3), with a single auxiliary variate x positively correlated with y and the coefficient function $a(Z_m) = q\, \bar{y}_m / \bar{X}$, $q > 0$, results in an unbiased ratio type estimator

$$S^*_{(q)} = (q+1)\,\bar{y} - q \left[\frac{\bar{y}\,\bar{x}}{\bar{X}} - \left(\frac{1}{n} - \frac{1}{N} \right) \frac{s_{xy}}{\bar{X}} \right], \quad (4.2.1)$$

where $s_{xy} = \frac{1}{(n-1)} \sum_{j=1}^{n} y_j (x_j - \bar{x})$.

<u>Proof</u>: We note $a^*(Z_m) = E_n[a(Z_m)] = q\, E_n(\bar{y}_m)/\bar{X} = q\, \bar{y}/\bar{X}$, (4.2.2)

and $Cov_n[a(Z_m), \bar{x}_m] = q\, Cov_n[\bar{y}_m, \bar{x}_m]/\bar{X} = q \frac{(n-m)}{nm} \cdot \frac{s_{xy}}{\bar{X}}$

Using these in G^*_m (3.4.3) we obtain

$$S^*_{(q)} = \bar{y} - \frac{q\,\bar{y}}{\bar{X}} (\bar{x} - \bar{X}) + \frac{q\,(N-n)}{Nn} \cdot \frac{s_{xy}}{\bar{X}}$$

$$= (q+1)\,\bar{y} - q \left[\frac{\bar{y}\,\bar{x}}{\bar{X}} - \left(\frac{1}{n} - \frac{1}{N} \right) \frac{s_{xy}}{\bar{X}} \right].$$

Remark (a) : For different positive values of q, different estimators of the form $S^*_{(q)}$ are obtained. However, for a particular choice of q, for any m in the range 1 to n-1, $S^*_{(q)}$ is the same. This enables us to use the results on the exact variance of G^*_1 (3.4.4) given in section 3.4 a to obtain the exact variance of $S^*_{(q)}$, and determine the optimum choice for q from among all possible positive values from the point of view of variance minimisation. Apart from the optimum choice for q, it will be seen from next section that another choice of interest is q = 1.

Remark (b) : It is interesting to note that $S^*_{(q)}$ is a linear combination of $\bar{y}$ and Robson's (1957) unbiased product estimator for the population mean $\bar{Y}$:

$$\hat{\bar{Y}}_{RP} = \frac{\bar{y}\,\bar{x}}{\bar{X}} - \left(\frac{1}{n} - \frac{1}{N}\right) \frac{s_{xy}}{\bar{X}} \qquad (4.2.3)$$

itself $\hat{\bar{Y}}_{RP}$ is recommended as an unbiased estimator of $\bar{Y}$ when x is negatively correlated with y. Here, when x is positively correlated with y, appears with a negative coefficeint (since q > 0) in $S^*_{(q)}$. The necessity of taking q > 0, when x and y have positive correlation, will become evident from next section where we investigate the variance of $S^*_{(q)}$. It may, however, be noted that the proof of theorem 4.2.1 does not use the condition q > 0 and whatever be the choice of q (positive, zero, or negative) $S^*_{(q)}$ as defined in (4.2.1) is unbiased for $\bar{Y}$. In fact, for negative choices of q, the same method and the same form (4.2.1) of the estimator will be used in a subsequent chapter to obtain a class of unbiased product type estimators.

4.3. Exact variance of unbiased ratio type estimators:

In this section we derive and study the exact variance of those unbiased ratio type estimators which are obtainable as special cases of G_1^* with appropriate choices of the coefficient functions a_{ij}, ($i = 1, 2, \ldots, p$).

.3.(i) Hartley-Ross unbiased ratio type estimator:

We note that G_1^* (3.4.4) results in Hartley-Ross estimator T_1^* 4.1.3) when $p = 1$ and $a_j = y_j / x_j = r_j$. Hence from corollary 3.4.1b we educe

$$
\begin{aligned}
V(T_1^*) = \left(\frac{1}{n} - \frac{1}{N}\right) \Big[& \frac{1}{(N-1)} \sum_{(1)}^{N} (y_j - \bar{Y})^2 - \frac{2}{N(N-1)} \Big[\sum_{(2)}^{N} y_j x_j r_{j'} - \frac{1}{(N-2)} \sum_{(3)}^{N} y_j x_{j'} r_{j''} \Big] \\
& + \frac{1}{N^2(n-1)} \Big[\sum_{(2)}^{N} x_j^2 r_{j'}^2 + \frac{(N-n)}{(N-1)} \sum_{(2)}^{N} y_j y_{j'} + \frac{(Nn-2N+1)}{(N-1)(N-2)} \sum_{(3)}^{N} x_j^2 r_{j'} r_{j''} \\
& - \frac{2(N-n)}{(N-1)(N-2)} \sum_{(3)}^{N} y_j x_{j'} r_{j''} - \frac{1}{(N-2)} \sum_{(3)}^{N} x_j r_{j'}^2 x_{j''} \\
& - \frac{(Nn-3N+n+1)}{(N-1)(N-2)(N-3)} \sum_{(4)}^{N} x_j r_{j'} x_{j''} r_{j'''} \Big] \Big]. \qquad (4.3.1)
\end{aligned}
$$

Remark (a): The exact variance of T_1^* was first obtained by Robson(1957). It may be easily verified that (4.3.1) is identical with the result given by Robson.

When $N \gg n$, from corollary 3.4.1c we derive the simpler form the variance of T_1^*:

$$V(T_1^*) = \frac{1}{n}\left[V(y) - 2E(r)\,Cov(y,x) + E^2(r)\,V(x)\right] + \frac{1}{n(n-1)}\left[V(x)V(r) + Cov^2(x,r)\right]. \qquad (4.3.2)$$

Remark (b) :- The expression (4.3.2) for $V(T_1^*)$ was originally due to Goodman and Hartley (1958).

3 (ii) Rao-Pereira (1968) unbiased double ratio type estimator:

Rao and Pereira (1968) proposed two unbiased double ratio type estimators in sampling on two occasions for estimating the current population mean $\bar{Y}_1$, which are particular cases of Mickey's ratio type estimators with three auxiliary variates. If (y_1, x_1) and (y_0, x_0) denote variates y and x on the current and previous occasions respectively, then one of the double ratio type estimators is a multivariate unbiased estimator G_1^* (3.4.4) with the choice of coefficient functions:

$$a_{1j} = r_{1j} = \frac{y_{1j}}{x_{1j}}, \quad a_{2j} = r_{2j} = \frac{y_{1j}}{y_{oj}}, \text{ and } a_{3j} = -r_{3j} = -\frac{y_{1j}}{x_{oj}}.$$

It is, therefore, possible to obtain the exact variance of this double ratio type estimator from theorem 3.4.1, and when N >> n the simpler form of the variance from corollary 3.4.1a. These results are, however, not put down here explicitely. To derive the simpler form of the variance, Rao and Pereira (1968) have actually used (a) Goodman and Hartley's (1958) variance formula for Hartley-Ross estimator and (b) the formula for the covariance between two Hartley-Ross estimators as derived by Robson and Vithayasai (1961). But the variance may be obtained directly from the variance of G_1^* (3.4.4) by appropriate choice of a_{ij}.

4.3. (iii) Unbiased ratio type estimator $S^{*}_{(q)}$ (4.2.1) :

The exact variance of $S^{*}_{(q)}$ (4.2.1) may be obtained from corollary 3.4.1b by choosing $a_j = \frac{q\, y_j}{\bar{X}}$. The result is stated in the following theorem.

Theorem 4.3.1 : For the unbiased ratio type estimator $S^{*}_{(q)}$ (4.2.1) the exact variance is given by

$$V[S^{*}_{(q)}] = \left(\frac{1}{n} - \frac{1}{N}\right)\Bigg[\frac{1}{(N-1)}\sum_{(1)}^{N}(y_j - \bar{Y})^2 - \frac{2q}{N(N-1)\bar{X}}\Big[\sum_{(2)}^{N} y_j x_j y_{j'}$$

$$- \frac{1}{(N-2)}\sum_{(3)}^{N} y_j x_{j'} y_{j''}\Big] + \frac{q^2}{N^2(n-1)\bar{X}^2}\Big[\sum_{(2)}^{N} x_j^2 y_{j'}^2$$

$$+ \frac{(N-n)}{(N-1)}\sum_{(2)}^{N} x_j y_j x_{j'} y_{j'} + \frac{(Nn-2N+1)}{(N-1)(N-2)}\sum_{(3)}^{N} x_j^2 y_{j'} y_{j''}$$

$$- \frac{2(N-n)}{(N-1)(N-2)}\sum_{(3)}^{N} x_j y_j x_{j'} y_{j''} - \frac{1}{(N-2)}\sum_{(3)}^{N} x_j y_{j'}^2 x_{j''}$$

$$- \frac{(Nn-3N+n+1)}{(N-1)(N-2)(N-3)}\sum_{(4)}^{N} x_j y_{j'} x_{j''} y_{j'''}\Big]\Bigg] \qquad (4.3.3)$$

Corollary 4.3.1: When $N \gg n$, the variance of $S^{*}_{(q)}$ takes the simpler form:

$$V[S^{*}_{(q)}] = \frac{1}{n}\left[V(y) - 2qR\,\mathrm{Cov}(y,x) + q^2R^2V(x)\right] + \frac{q^2}{n(n-1)\bar{X}^2}\left[V(x)V(y) + \mathrm{Cov}^2(x,y)\right]$$

(4.3.4)

Proof: This follows from corollary 3.4.1c by writing $a = \frac{qy}{\bar{X}}$, $E(a) = qR$, etc.

Remark (c): From expression (4.3.4) which, infact, is valid for any choice of q, we note that, when x is positively correlated with y and $R > 0$, the simple unbiased estimator $\bar{y}$ is more efficient than $S^*_{(q)}$ for all choices of $q < 0$. For $q = 0$, $S^*_{(q)} = \bar{y}$. We, therefore, consider only positive values for q when x and y are positively associated. This explains the need to take $q > 0$ in the definition of the unbiased ratio type estimator $S^*_{(q)}$.

Corollary 4.3.2: In large samples, when the finite population correction is not ignored,

$$V[S^*_{(q)}] = \left(\frac{1}{n} - \frac{1}{N}\right)\left(S_y^2 - 2qR\,S_{xy} + q^2R^2\,S_x^2\right).$$

Proof: For $a(Z_1) = a = qy/\bar{X}$, it may be easily shown that the conditions (3.4.37) of theorem 3.4.2 are satisfied. Hence the result.

Theorem 4.3.4: The unbiased ratio type estimator:

$$S^*_{(1)} = 2\bar{y} - \left[\frac{\bar{y}\,\bar{x}}{\bar{X}} - \left(\frac{1}{n} - \frac{1}{N}\right)\frac{s_{xy}}{\bar{X}}\right] \qquad (4.3.5)$$

has the exact variance, for any N and n, given by (4.3.3) with $q = 1$; and when $N \to \infty$ it has the variance:

$$V[S^*_{(1)}] = \frac{1}{n}\left[V(y) - 2R\,\mathrm{Cov}(y,x) + R^2\,V(x)\right] + \frac{1}{n(n-1)\bar{X}^2}\left[V(x)V(y) + \mathrm{Cov}^2(x,y)\right].$$

(4.3.6)

Proof: Follows immediately from (4.2.1), (4.3.3) and (4.3.4) by taking $q = 1$.

Remark (1) : The unbiased ratio type estimator $S^*_{(1)}$ (4.3.5) has the same large sample variance (i.e. variance upto order n^{-1}) as the classical biased ratio estimator $\hat{\bar{Y}}_R = \left(\frac{\bar{y}}{\bar{x}}\right)\bar{X}$. It uses the same amount of information on population parameters, namely the population mean $\bar{X}$, as the biased ratio estimator does. Its variance formulae: (4.3.3) with q = 1, and (4.3.6), are exact and valid for any sample size n. As may be seen from a later section, applying the results of section 3.6, its variance can be unbiasedly estimated from the sample for any sample size. The advantages of $S^*_{(1)}$ over $\hat{\bar{Y}}_R$ are further emphasised by the following result.

Theorem 4.3.3 : If the population is large and (y, x) have a bivariate normal distribution, the unbiased ratio type estimator $S^*_{(1)}$ (4.3.5) is more efficient than the classical biased ratio estimator $\hat{\bar{Y}}_R$ provided

$$2\left(\frac{C_y}{C_x}\right)^2 + 5\left(\rho\frac{C_y}{C_x} - 3\right)\left(\rho\frac{C_y}{C_x} - 0.6\right) \geqslant 0, \qquad (4.3.7)$$

where C_y and C_x are the coefficients of variation of y and x, and ρ is the correlation coefficient between y and x. In particular, the condition (4.3.7) is satisfied when (i) $C_x = C_y$, $\rho \leqq 0.78$, or (ii) $3C_x = 2C_y$, $\rho \leqq 0.70$.

Proof: Assuming the sample size is large enough to make the condition $\left|\frac{\bar{x}-\bar{X}}{\bar{X}}\right| < 1$ valid, Sukhatme (1954, 1970) has given the relative mean square error of $\hat{\bar{Y}}_R$, upto order n^{-2}, in a large bivariate (y, x) normal population as :

$$\frac{MSE(\hat{\bar{Y}}_R)}{\bar{Y}^2} = \frac{1}{n}\left(C_y^2 + C_x^2 - 2\rho C_y C_x \right) + \frac{3C_x^2}{n^2}\left[(1+2\rho^2) C_y^2 - 6\rho C_y C_x + 3C_x^2 \right]. \tag{4.3.8}$$

From (4.3.6) we may write in a large population (without any assumption on he sample size and joint distribution of y and x),

$$\frac{[\bar{S}^*_{(1)}]}{\bar{Y}^2} = \frac{1}{n}\left(C_y^2 + C_x^2 - 2\rho C_y C_x \right) + \frac{C_x^2 C_y^2 (1+\rho^2)}{n(n-1)}. \tag{4.3.9}$$

omparing (4.3.8) and (4.3.9) we obtain

$$\frac{SE(\hat{\bar{Y}}_R) - V[\bar{S}^*_{(1)}]}{\bar{Y}^2} \doteq \frac{C_x^2}{n^2}\left[2C_y^2 + (5\rho^2 C_y^2 - 18\rho C_y C_x + 9C_x^2) \right]$$

$$= \frac{C_x^4}{n^2}\left[2\left(\frac{C_y}{C_x}\right)^2 + 5\left(\rho\frac{C_y}{C_x} - 3\right)\left(\rho\frac{C_y}{C_x} - 0.6\right) \right]. \tag{4.3.10}$$

may be noted that $V[\bar{S}^*_{(1)}]$ does not contain terms of order n^{-3} and above reas MSE ($\hat{\bar{Y}}_R$) has some contribution from such higher order terms.

e from (4.3.10), it follows that a <u>sufficient</u> condition for $S^*_{(1)}$ to be more lent than $\hat{\bar{Y}}_R$ is given by

$$2\left(\frac{C_y}{C_x}\right)^2 + 5\left(\rho\frac{C_y}{C_x} - 3\right)\left(\rho\frac{C_y}{C_x} - 0.6\right) \geqslant 0. \tag{4.3.11}$$

rticular, when $C_y = C_x$, this condition reduces to :

$$5(\rho^2 - 3.6\rho + 2.2) = 5(\rho - 0.78)(\rho - 2.82) \geqslant 0.$$

ence, when $C_y = C_x$ and $\rho \leq 0.78$, $MSE(\hat{\bar{Y}}_R) \geq V[\bar{S}^*_{(1)}]$.

Again when $2C_y = 3C_x$, (4.3.11) is equivalent to :

$$5\left(\frac{9}{4}\rho^2 - 5.4\rho + 2.7\right) = \frac{45}{4}(\rho - 0.7)(\rho - 1.7) \geq 0.$$

ence, when $2C_y = 3C_x$ and $\rho \leq 0.70$, again $MSE(\hat{\bar{Y}}_R) \geq V[\bar{S}^*_{(1)}]$.

heorem 4.3.4 : When $N \gg n$, for the unbiased ratio type estimator $^*_{(q)}$ (4.2.1), the optimum choice of q is given by

$$q_o = \rho\frac{C_y}{C_x}\left[1 + \frac{C_y^2(1+\rho^2)}{(n-1)}\right]^{-1}, \qquad (4.3.12)$$

$$V[\bar{S}^*_{(q_o)}] = \frac{\bar{Y}^2 C_y^2(1-\rho^2)}{n} + \frac{\bar{Y}^2 C_y^4 \rho^2(1+\rho^2)}{n[(n-1) + C_y^2(1+\rho^2)]}. \qquad (4.3.13)$$

oof : From (4.3.4) we may write

$$\bar{S}^*_{(q)}] = \frac{\bar{Y}^2}{n}\left[C_y^2 - 2q\rho C_y C_x + q^2 C_x^2\right] + \frac{\bar{Y}^2 q^2 C_x^2 C_y^2(1+\rho^2)}{n(n-1)}$$

$$= \frac{\bar{Y}^2 C_x^2}{n}\left[q^2\left(1 + \frac{C_y^2(1+\rho^2)}{(n-1)}\right) - 2q\rho\frac{C_y}{C_x} + \left(\frac{C_y}{C_x}\right)^2\right]. \qquad (4.3.14)$$

rentiating $V[\bar{S}^*_{(q)}]$ with respect to q and setting the derivative l to zero, we obtain

$$\frac{dV[S^*_{(q)}]}{dq} = \frac{2\bar{Y}^2 C_x^2}{n}\left[q\left(1 + \frac{C_y^2(1+\rho^2)}{(n-1)}\right) - \rho\frac{C_y}{C_x}\right] = 0,$$

hen $q = \rho\frac{C_y}{C_x}\left[1 + \frac{C_y^2(1+\rho^2)}{(n-1)}\right]^{-1} = q_o$.

ince $\frac{d^2V[S^*_{(q)}]}{dq^2} = \frac{2\bar{Y}^2 C_x^2}{n}\left[1 + \frac{C_y^2(1+\rho^2)}{(n-1)}\right] > 0$, the choice $q = q_o$

inimises $V[S^*_{(q)}]$ with respect to q. Substituting q_o for q in (4.3.14), we get

$$V[S^*_{(q_o)}] = \frac{\bar{Y}^2 C_x^2}{n}\left[\left(\frac{C_y}{C_x}\right)^2 - q_o\rho\frac{C_y}{C_x}\right]$$

$$= \frac{\bar{Y}^2}{n}\left[C_y^2 - \rho^2 C_y^2\left(1 + \frac{C_y^2(1+\rho^2)}{(n-1)}\right)^{-1}\right] \qquad (4.3.15)$$

$$= \frac{\bar{Y}^2}{n} C_y^2(1-\rho^2) + \frac{\bar{Y}^2 C_y^2 \rho^2}{n}\left[1 - \frac{(n-1)}{\{(n-1) + C_y^2(1+\rho^2)\}}\right]$$

= expression (4.3.13).

rk (e) : From (4.3.15) it follows that $S^*_{(q_o)}$ is more efficient than the ple unbiased estimator $\bar{y}$. Since q_o is the optimum choice for q, $S^*_{(q_o)}$ is rior to $S^*_{(1)}$ (4.3.5) which has the same large sample variance as the sical ratio estimator $\hat{\bar{Y}}_R$. But to compute $S^*_{(q_o)}$ we require information re population parameters.

Remark (f) : Since $\frac{C_y^2(1+\rho^2)}{(n-1)} \leqslant C_y^2$ for $n \geqslant 3$, a good approximation to the optimum choice q_o (4.3.12) is $q_o' = \frac{\rho C_y}{C_x}$. For the unbiased ratio type estimator $S^*_{(q_o')}$, when $N >> n$,

$$V\left[S^*_{(q_o')}\right] = \frac{\bar{Y}^2 C_y^2 (1-\rho^2)}{n} + \frac{\bar{Y}^2 C_y^4 \rho^2 (1+\rho^2)}{n(n-1)} \qquad (4.3.16)$$

The leading terms in the variances (4.3.13) and (4.3.16) are the same, and the difference occurs only in the second term of order n^{-2}

Remark (g) : Both $S^*_{(q_o)}$ and $S^*_{(q_o')}$ have the same large sample variance as the usual linear regression estimator $\hat{Y}_r = \bar{y} + b(\bar{X} - \bar{x})$, where b is the linear regression coefficient of y on x computed from the sample.

Remark (h) : Information on $q_o' = \rho \frac{C_y}{C_x}$ available from a pilot survey, previous census, or past experience may be used to construct the estimator $S^*_{(q_o')}$ without disturbing the unbiased nature of the estimator.

Remark (i) : Srivastava (1967) has considered a biased ratio estimator $\bar{y}_r = \bar{y}\left(\frac{\bar{X}}{\bar{x}}\right)^q$ where q is a constant and has shown that, in large samples, $q_o' = \rho \frac{C_y}{C_x}$ is the optimum choice for q. With q_o' the estimator has the same large sample variance as the linear regression estimator $\hat{Y}_r$.

Remark (j) : Singh (1969) has considered $\hat{Y}_q = \frac{\bar{y}\,\bar{X}}{\left[\bar{X} + q(\bar{x} - \bar{X})\right]}$ where q is constant such that $\left| \frac{q(\bar{x} - \bar{X})}{\bar{X}} \right| < 1$. He has shown that the optimum choice q is again $\rho \frac{C_y}{C_x}$. The estimator with the optimum choice has only second degree (n^{-2}) bias and its large sample variance is the same as that of linear regression estimator $\hat{\bar{Y}}_r$. In large samples the condition, $\left| \frac{\bar{x} - \bar{X}}{\bar{X}} \right| < 1$, is usually satisfied, but the condition: $\left| \frac{q(\bar{x} - \bar{X})}{\bar{X}} \right| < 1$, imposes a restriction on q and hence also on the optimum choice $\rho \frac{C_y}{C_x}$. Further, if an estimated value of $\rho \frac{C_y}{C_x}$ is used, to the extent the estimated value deviates from the true value the bias in the estimator $\hat{\bar{Y}}_q$ will have a component of order n^{-1}.

Remark (k): When an estimated value of $q_o' = \rho\, C_y / C_x$ is used, the estimator $S^*_{(q_o')}$ remains unbiased. Its variance formulae (both for any N, large N) and the theory of determination of optimum choice are valid for any sample size; whereas these are valid for the other biased or almost unbiased estimators (suggesting the use of $\rho\, C_y/C_x$) only in large samples. Using the results in section 3.6, it is possible to estimate unbiasedly the variance of $S^*_{(q_o')}$ from the sample for any sample size $n \geq 2$. For the other estimators variance estimation from the sample is possible only in large samples.

4.4. Large sample variance of Mickey's unbiased ratio type estimators:

In this section we investigate the variance of the unbiased ratio type estimator T^*_m (4.1.2) in large samples, using our theorems 3.4.2 and 3.4.3 on the large sample variance of the general unbiased estimator G^*_m (3.4.3).

Theorem 4.4.1: In large samples the unbiased ratio type estimator T^*_m (4.1.2) has the variance:

$$V(T^*_m) = \left(\frac{1}{n} - \frac{1}{N}\right) \left[S_y^2 - 2E(R_m)S_{xy} + E^2(R_m) S_x^2 \right] \qquad (4.4.1)$$

of: We consider two cases.

(i): n is large and m is <u>not small</u> compared to n

In this case, treating Z_m as a simple random sub-sample from the sample of size n, we have from the theory of ratio estimation

$$R^*_m = E_n(R_m) = \frac{\bar{y}}{\bar{x}} \left[1 + \left(\frac{1}{n} - \frac{1}{m}\right) \left[\left(\frac{s_x^2}{\bar{x}^2} - \frac{s_{xy}}{\bar{x}\,\bar{y}} \right) + \begin{matrix}\text{terms involving higher}\\ \text{order sample moments}\\ \text{with coefficients at most}\\ \text{of order } m^{-1}.\end{matrix} \right] \right] \qquad (4.4.2)$$

m (4.4.2) it follows that $\mu_2(R^*_m)$ and $\mu_4(R^*_m)$ are of same order in n $\mu_2(\bar{y}/\bar{x})$ and $\mu_4(\bar{y}/\bar{x})$ respectively. But in large samples we know $\mu_2(\bar{y}/\bar{x}) = O(n^{-1})$ and $\mu_4(\bar{y}/\bar{x}) = O(n^{-2})$.

Hence $\mu_2(R^*_m) = O(n^{-1})$ and $\mu_4(R^*_m) = O(n^{-2})$. (4.4.3)

since $\bar{x}_{n-m} = \dfrac{n\bar{x} - m\bar{x}_m}{n-m}$, using (4.4.2) we have

$$\text{Cov}_n(R_m, \bar{x}_{n-m}) = -\frac{m}{(n-m)} \text{Cov}_n(R_m, \bar{x}_m)$$

$$= -\frac{m}{(n-m)} \left[\bar{y} - E_n(R_m)\bar{x} \right]$$

$$= \frac{1}{n} \left[\frac{\bar{y}\,s_x^2}{\bar{x}^2} - \frac{s_{xy}}{\bar{x}} + \begin{matrix}\text{terms involving higher order}\\ \text{sample moments with coefficients}\\ \text{at most of order } m^{-1}\end{matrix} \right] \qquad (4.4.4)$$

rom (4.4.4) we get

$$\mu_2 \left[Cov_n (R_m , \bar{x}_{n-m}) \right] = O (n^{-3}). \qquad (4.4.5)$$

e note (4.4.3) and (4.4.5) imply that the conditions(3.4.37) of theorem 4.2 are satisfied for the coefficient function $a(Z_m) = R_m$. Hence, applying orem 3.4.2, we obtain the result (4.4.1) in this case.

se (ii): n is large and m is small compared to n.

Here we apply theorem 3.4.3 and obtain the result (4.4.1).

This completes the proof of the theorem.

llary 4.4.1 : In large samples for any large m

$$V (T^*_m) = (\frac{1}{n} - \frac{1}{N}) \left[S_y^2 - 2R\, S_{xy} + R^2 S_x^2 \right]. \qquad (4.4.6)$$

ich is same as the large sample variance of the classical ratio estimator $= (\bar{y} / \bar{x})\, \bar{X}$

f: When m is large, we may assume $\left| \frac{\bar{x}_m - \bar{X}}{\bar{X}} \right| < 1$ and obtain $R_m) = R + O (m^{-1})$. Substituting this in (4.4.1), we get (4.4.6).

ollary 4.4.2 : In large samples for any two different choices m_1 and m_2 n, $V (T^*_{m_2}) \leq V (T^*_{m_1})$ if and only if $| E (R_{m_2}) - \beta | \leq | E(R_{m_1}) - \beta |$, ere β is the population linear regression coefficient of y on x.

Follows immediately by re-writing (4.4.1) as

$$V(T^*_m) = (\frac{1}{n} - \frac{1}{N}) \left[S_y^2 (1 - \rho^2) + \left\{ E (R_m) - \beta \right\}^2 S_x^2 \right]. \qquad (4.4.7)$$

Remark (a) : It is difficult to determine the nature of the function $\left| E(R_m) - \beta \right|$ as m increases from 1 to $n-1$ for all types of finite populations. However, since ratio method of estimation is considered appropriate in populations where regression of y on x is linear, the following lemma on the behaviour of $\left| E(R_m) - \beta \right|$ in such populations is of interest.

Lemma 4.4.1 : *When regression of y on x is linear and passes through the origin, $\left| E(R_m) - \beta \right| = 0$ for all m. In populations where (i) regression of y on x is linear and does not pass through the origin, and (ii) the values of x on all units are of same sign (positive or negative), $\left| E(R_m) - \beta \right|$ is a strictly decreasing function of m.*

Proof : Since the regression of y on x is linear, we may write

$$y_j = \alpha + \beta x_j + e_j , \quad E(e_j / x_j) = 0 .$$

Then

$$E(R_m) = E\left(\frac{\alpha + \beta \bar{x}_m + \bar{e}_m}{\bar{x}_m} \right) = \alpha E\left(\frac{1}{\bar{x}_m} \right) + \beta, \text{ and}$$

$$\therefore \quad \left| E(R_m) - \beta \right| = |\alpha| \left| E\left(\frac{1}{\bar{x}_m} \right) \right| \qquad (4.4.8)$$

When the regression line passes through the origin, $\alpha = 0$ and from (4.4.8), $\left| E(R_m) - \beta \right| = 0$ for all m.

Suppose $|\alpha| \neq 0$. Then from (4.4.8) we note that the behaviour of $\left| E(R_m) - \beta \right|$ is same as that of $\left| E\left(\frac{1}{\bar{x}_m} \right) \right|$ as a function of m.

At this stage we observe that when x is a random variable assuming values of the same sign,

$$\left| E\left(\frac{1}{x} \right) \right| > \frac{1}{\left| E(x) \right|} \tag{4.4.9}$$

Now consider a set of sample elements Z_{m+1} and all the possible sub-samples Z_m that may be formed out of Z_{m+1}. For this set-up from (4.4.9) we have

$$\left| E\left(\frac{1}{\bar{x}_m} / Z_{m+1} \right) \right| > \frac{1}{\left| E\left(\bar{x}_m / Z_{m+1} \right) \right|} = \frac{1}{\left| \bar{x}_{m+1} \right|} \tag{4.4.10}$$

Taking expectations on both sides of (4.4.10) over all possible choices of Z_{m+1} from the population, we obtain

$$E\left[\left| E\left(\frac{1}{\bar{x}_m} / Z_{m+1} \right) \right| \right] \quad E\left[\frac{1}{\left| \bar{x}_{m+1} \right|} \right],$$

i.e., $$\left| E\left[E\left(\frac{1}{\bar{x}_m} / Z_{m+1} \right) \right] \right| > \left| E\left(\frac{1}{\bar{x}_{m+1}} \right) \right|,$$

i.e., $$\left| E\left(\frac{1}{\bar{x}_m} \right) \right| > \left| E\left(\frac{1}{\bar{x}_{m+1}} \right) \right|. \tag{4.4.11}$$

Hence from (4.4.8) and (4.4.11) it follows that $\left| E(R_m) - \beta \right|$ is a strictly decreasing function of m when (i) the regression of y on x is linear and does not pass through the origin, and (ii) values of x on all units in the population are of same sign.

Remark (b) : For any given sample of size n, one can construct n-1 unbiased ratio type estimators of the form T_m^* (4.1.2) corresponding to the n-1 possible choices of m in the range 1 to n-1. It is, therefore, useful if we can identify a choice of m, say m_o, such that $V(T^*_{m_o}) \leq V(T^*_m)$ for $1 \leq m \leq n-1$, >2. In large samples, by making use of the above results, we are now in position to give a partial solution to this problem of optimum choice of m or the unbiased ratio type estimator T_m^*.

Theorem 4.4.2 : When the regression of y on x is linear and passes through the origin, in large samples all choices of m in the range 1 to n-1 result in the same variance for the unbiased ratio type estimator T_m^* (4.1.2), which is equal to the variance of the classical ratio estimator $\hat{Y}_R$. When, however, the regression line does not pass through the origin, the optimum choice for m is n-1 if values of x on all units of the population are of same sign, and the estimator T^*_{n-1} has the same large sample variance as the classical estimator $\hat{Y}_R$.

Proof : When the regression of y on x is linear and passes through the origin, we know the classical ratio estimator $\hat{Y}_R = \left(\frac{\bar{y}}{\bar{x}}\right)\bar{X}$ is also biased and has the large sample variance

$$V(\hat{Y}_R) = \left(\frac{1}{n} - \frac{1}{N}\right) S_y^2 \left(1 - \rho^2\right). \tag{4.4.12}$$

Ing the first part of lemma 4.4.1 in the form (4.4.7) of $V(T_m^*)$, we note t in this case for all choices of m in the range 1 to n-1, $V(T_m^*)$ coincides (4.4.12). This establishes the first part of the theorem.

When regression of y on x is linear but does not pass through the origin, and values of x on all units in the population are of same sign; we use the second part of lemma 4.4.1 in the form (4.4.7) of $V(T^*_m)$ and obtain that $V(T^*_{n-1}) < V(T^*_m)$ for $1 \leq m \leq n-2$. Now from the corollary 4.4.1 we have $V(T^*_{n-1}) = V(\hat{\bar{Y}}_R)$ in large samples. This establishes the second part of the theorem.

4.5. Estimation of variance of some unbiased ratio type estimators:

In section 3.6 we have developed some procedures for estimating the variance of the averaged general unbiased estimator G^*_m (3.4.3). Here we apply those methods to the unbiased ratio type estimators T^*_m (4.1.2) and $S^*_{(q)}$ (4.2.1) with $q = 1$, and $q = q'_o = \rho \ C_y / C_x$.

We note that theorem 3.6.1 and corollary 3.6.1, theorem 3.6.2, and theorem 3.6.3 (in view of corollary 4.3.2 and theorem 4.4.1) may be directly applied to the unbiased ratio type estimators by making the following substitutions:

for T^*_m (4.1.2): $p = 1$, $a(Z_m) = R_m$; and

for $S^*_{(q)}$ (4.2.1): $p = 1$, $m = 1$, $a(Z_1) = \frac{q \ y_j}{\bar{X}}$

Thus from theorem 3.6.1 and corollary 3.6.1, for T^*_m we have

$$\hat{V}(T^*_m) = \frac{1}{\binom{n}{m}} \sum_{Z_m}^{\binom{n}{m}} \left[\hat{V}(T_m) - (T_m - T^*_m)^2 \right], \qquad (4.5.1)$$

where T_m is given by (4.1.1) and

$$\hat{V}(T_m) = \frac{(N-n)(N-m)}{N^2(n-m)} \cdot \frac{1}{(n-m-1)} \sum_{j \notin Z_m}^{n-m} \left[(y_j - \bar{y}_{n-m}) - R_m (x_j - \bar{x}_{n-m}) \right]^2$$

for $1 \leq m \leq n-2$ (4.5.2)

$$= \frac{(N-n+1)^2}{N^2} \left[U_{n-1}^2 (T) - y_j^2 \right] + \frac{(N-n+1)(N-n)}{N^2} s_y^2 \text{ for } m = 1, \quad (4.5.3)$$

where $U_{n-1}(T) = y_j - R_{n-1}(x_j - \bar{X}_{N-n+1})$, $j \notin Z_{n-1}$.

The other parameters appearing in the above formulae have already been defined in section 3.6.

Similarly for $S^*_{(q)}$, we have with $a(Z_1) = \frac{q\, y_j}{\bar{X}}$,

$$\hat{V}\left[S^*_{(q)} \right] = \frac{1}{n} \sum_{j=1}^{n} \left[\hat{V}(S^j_{(q)}) - (S^j_{(q)} - S^*_{(q)})^2 \right], \quad (4.5.4)$$

where we define using (3.4.1)

$$S^j_{(q)} = \left[\bar{y} - \frac{q\, y_j}{\bar{X}} (\bar{x} - \bar{X}) \right] - \frac{(N-n)}{N(n-1)} \left[(y_j - \bar{y}) - \frac{q\, y_j}{\bar{X}} (x_j - \bar{x}) \right], \quad (4.5.5)$$

$$\frac{1}{n} \sum_{j=1}^{n} S^j_{(q)} = S^*_{(q)},$$

using theorem (3.6.1)

$$\hat{V}^{-j}[S_{(q)}] = \frac{(N-n)(N-1)}{N^2(n-1)} \cdot \frac{1}{(n-2)} \sum_{j' \neq j}^{n-1} \left[(y_{j'} - \bar{y}_{n-1}) - \frac{q\, y_j}{\bar{X}}(x_{j'} - \bar{x}_{n-1})\right]^2$$

for $n > 2$, (4.5.6)

$$= \frac{(N-n+1)^2}{N^2}\left[U^2_{n-1}(S) - y^2_{j'}\right] + \frac{(N-n+1)(N-n)}{N^2} s_y^2 \text{ for } n=2, \quad (4.5.7)$$

where $U_{n-1}(S) = y_{j'} - \frac{q\, y_j}{\bar{X}}(x_{j'} - \bar{X}_{N-n+1})$, $(j' \neq j)$.

In large samples, in view of theorem 4.4.1, we may use theorem 3.6.3 and obtain a simpler estimator of variance of T^*_m as

$$\hat{V}(T^*_m) = \left(\frac{1}{n} - \frac{1}{N}\right)\left[s_y^2 - 2R^*_m s_{xy} + (R^*_m)^2 s_x^2\right], \quad (4.5.8)$$

where
$$R^*_m = \frac{1}{\binom{n}{m}} \sum_{Z_m}^{\binom{n}{m}} R_m$$

Similarly, using corollary (4.3.2) and theorem 3.6.3, we may have in large samples

$$\hat{V}[S^*_{(q)}] = \left(\frac{1}{n} - \frac{1}{N}\right)\left[s_y^2 - 2\frac{q\,\bar{y}}{\bar{X}} s_{xy} + \frac{q^2\,\bar{y}^2}{\bar{X}^2} s_x^2\right].$$

(4.5.9)

Alternately, for $S^*_{(1)}$ and $S^*_{(q'_o)}$, $q'_o = \rho\, C_y / C_x$, we may also use in large samples

$$\hat{V}'[S^*_{(1)}] = \left(\frac{1}{n} - \frac{1}{N}\right)\left[s_y^2 - 2\left(\frac{\bar{y}}{\bar{x}}\right)s_{xy} + \left(\frac{\bar{y}}{\bar{x}}\right)^2 s_x^2\right] \quad (4.5.10)$$

and $$\hat{v}'[S^*_{(q'_o)}] = \left(\frac{1}{n} - \frac{1}{N}\right)\left[s_y^2 - \frac{s_{xy}^2}{s_x^2}\right]. \qquad (4.5.11)$$

We do not propose to write down here the estimators of variance for T_1^* and $S^*_{(q)}$ which may be obtained using theorem 3.6.2, as no numerical illustration of these estimators of variance is attempted. For T_m^*, $S^*_{(1)}$ and $S^*_{(q'_o)}$ we now illustrate with numerical examples the estimators of variance given by (4.5.1), (4.5.4), (4.5.8), (4.5.10) and (4.5.11).

4.5a. <u>Illustration of variance estimation using (4.5.1) and (4.5.4)</u> :

We consider five live populations. The first population consists of data on number of workers (x) and number of absentees (y) for 43 factories (p. 398, Murthy, 1967). These 43 factories themselves constitute a simple random sample without replacement from a list of 345 factories situated in a district of India.

The other four populations are formed out of six characteristics recorded for each of the 22 cities and urban agglomerations with population size 100,000 and above, situated in the state of Uttar Pradesh of India. These are as follows:

Population 2 : x : 1961 census population(persons),
y : 1971 census population(persons),

Population 3 : x : 1971 census population (persons),
y : workers (persons) as per 1971 census;

Population 4 : x : 1971 census female population,
y : female workers as per 1971 census;

Population 5 : x : 1971 census female population,
y : female literates as per 1971 census.

As we make use of these populations in later chapters also, we present here the relevant data.

DATA FOR POPULATION 1

TABLE - 4.1

Number of workers and number of absentees for 43 factories

Factory Number (1)	Number of Workers (2)	Number of Absentees (3)	Factory Number (1)	Number of Workers (2)	Number of Absentees (3)
1	95	9	26	83	7
2	79	7	27	124	13
3	30	3	28	31	2
4	45	2	29	96	23
5	28	3	30	42	13
6	142	8	31	85	18
7	125	9	32	91	14
8	81	10	33	73	7
9	43	6	34	1[illegible]9	18
10	53	2	35	54	13
11	148	16	36	69	14
12	89	4	37	61	1
13	57	5	38	164	35
14	132	13	39	132	21
15	47	4	40	82	5
16	43	9	41	33	4
17	116	12	42	86	11
18	65	8	43	41	10
19	103	9			
20	52	8			
21	67	14			
22	64	6			
23	75	6			
24	69	8			
25	63	5			

DATA FOR POPULATIONS 2, 3, 4 AND 5

TABLE - 4.2

1961 Population and 1971 population (persons, females), workers(persons, females), and female literates of 22 cities and urban agglomerations with population 100,000 +

STATE: UTTAR PRADESH (Figures in 00's)

Sl. No.	Name of the city/urban agglomeration (100,000 +)	1961 Population Persons	1971 Population Persons	1971 Population Females
(1)	(2)	(3)	(4)	(5)
1	Kanpur	9710	12730	5504
2	Lucknow	6557	8262	3696
3	Agra	5087	6378	2909
4	Varanasi	4898	5829	2637
5	Allahabad	4307	5140	2260
6	Meerut	2840	3678	1653
7	Bareilly	2728	3261	1493
8	Moradabad	1918	2724	1245
9	Aligarh	1850	2540	1152
10	Gorakhpur	1802	2307	1024
11	Saharanpur	1852	2257	1026
12	Dehra Dun	1563	1994	873
13	Jhansi	1697	1981	933
14	Rampur	1354	1618	753
15	Shahjahanpur	1177	1441	668
16	Mathura	1253	1405	638
17	Firozabad	986	1339	610
18	Ghaziabad	704	1280	567
19	Muzaffarnagar	876	1149	525
20	Farrukhabad-cum-Fatehgarh	946	1114	507
21	Faizabad	883	1098	478
22	Mirzapur-cum-Vindhyachal	1001	1059	488

contd....

TABLE - 4.2 (contd.)

(Figures in OO's)

Sl. No.	Name of the city/urban agglomeration (100,000 +)	1971 Workers Persons	1971 Workers Females	1971 Literates Females
(1)	(2)	(6)	(7)	(8)
1	Kanpur	3866	204	2268
2	Lucknow	2390	144	1603
3	Agra	1695	88	945
4	Varanasi	1823	273	842
5	Allahabad	1499	130	934
6	Meerut	1058	47	622
7	Bareilly	888	43	499
8	Moradabad	771	46	418
9	Aligarh	668	34	377
10	Gorakhpur	617	39	424
11	Saharanpur	618	21	371
12	Dehra Dun	591	38	480
13	Jhansi	497	50	347
14	Rampur	453	14	178
15	Shahjahanpur	384	12	168
16	Mathura	377	20	233
17	Firozabad	361	14	185
18	Ghaziabad	368	15	222
19	Muzaffarnagar	279	11	201
20	Farrukhabad-cum-Fatehgarh	322	19	171
21	Faizabad	311	18	159
22	Mirzapur-cum-Vindhyachal	308	28	123

Source: Census of India 1971, Series-1, Paper-1 of 1971-Supplement Provisional Population Totals, Issued by Registrar General and Census Commissioner, India.

For the five populations under consideration, some parameters of interest are given in Table-4.3. These are population totals (Y, X), variances (σ_y^2 , σ_x^2), covariance (σ_{xy}), coefficients of variation (C_y, C_x), correlation coefficient (ρ), and linear regression coefficient of y on x (β).

TABLE - 4.3

Values of some population parameters

Variable/ Parameter	Population Number 1	2	3	4	5
y	Number of absentees	1971 Population (Persons)	1971 Workers (Persons)	1971 female workers	1971 female literates
x	Number of workers	1961 Population (Persons)	1971 Population (Persons)	1971 females	1971 females
Y	415	70584	20144	1308	11770
X	3417	55989	70584	31639	31639
σ_y^2	4[illegible], 13	8012202	731725	4534	264525
σ_x^2	1299, 32	4887102	8012202	1530114	1530114
σ_{xy}	154.60	6241901	2416160	68606	627658
C_y	0.67	0.88	0.93	1.13	0.96
C_x	0.45	0.87	0.88	0.86	0.86
ρ	0.6608	0.9976	0.9979	0.8237	0.9866
β	0.12	1.28	0.30	0.04	0.41

As estimators of the population total Y, we use the unbiased estimators: $N\bar{y}$, $N T^*_m$ ($1 \leq m \leq n-1$), and $N S^*_{(q)}$ ($q=1$, $q = q'_o = \rho\, C_y/C_x$).

The variances of these estimators are estimated by the following unbiased estimators:

$$\hat{V}(N\bar{y}) = N^2 \left(\frac{1}{n} - \frac{1}{N}\right) s_y^2 \ ,$$

$$\hat{V}(N T^*_m) = N^2 \hat{V}(T^*_m) \text{ where } \hat{V}(T^*_m) \text{ is given by (4.5.1),}$$

and $\hat{V}\left[N S^*_{(q)}\right] = N^2 \hat{V}\left[S^*_{(q)}\right]$ where $\hat{V}\left[S^*_{(q)}\right]$ is given by (4.5.4).

Four sample sizes n = 2, 3, 4 and 5 are tried. In a later chapter, we shall examine in small samples (of sizes 2, 3, 4 and 5) the relative efficiencies of (i) the unbiased ratio type estimators given here, (ii) some unbiased regression type estimators to be studied in a subsequent chapter, and (iii) the classical biased ratio and regression estimators, on the basis of a reasonable number of independent samples drawn from each of the five populations considered here and also some additional populations. As such the numerical examples on small samples presented here are mainly intended to illustrate the methods of estimation and variance estimation on the basis of single sample. No detailed efficiency comparisons can be attempted from these examples, although some general observations may be made. To serve the present purpose, from each of the five populations we take one sample of each of the four sizes 2, 3, 4 and 5. All samples are drawn with equal probability and without replacement. The serial numbers of the units (as in Tables 4.1 and 4.2) in the samples are given in Table - 4.4.

TABLE - 4.4

Simple random samples from the five populations

Sample size (n)	Population Number 1	2	3	4	5
2	11	13	14	4	4
	13	19	6	7	13
3	4	10	14	2	14
	31	8	17	15	17
	17	17	6	21	6
4	25	12	12	16	11
	8	7	7	7	12
	14	6	6	2	6
	29	14	14	19	5
5	24	14	19	10	19
	42	10	7	17	7
	27	13	6	1	6
	39	8	10	6	10
	18	22	5	19	5

To save space we omit the intermediate calculations on the sample and give directly the sample estimates of population total Y and [e]stimates of their variance separately for each population in Tables - 4.5 4.9.

POPULATION I

TABLE 4.5

Estimates of population total Y and variance estimates**

Estimator	Sample size 2 Estimate	Sample size 2 Estimate of variance	Sample size 3 Estimate	Sample size 3 Estimate of variance	Sample size 4 Estimate	Sample size 4 Estimate of variance	Sample size 5 Estimate	Sample size 5 Estimate of variance
$N\bar{y}$	451	5.33	459	3.75	548	2.41	525	0.94
$NS^*_{(1)}$	450	4.98 (*)	476	0.86	467	1.27	435	0.25
$NS^*_{(q'_o)}$	450	4.99 (*)	476	0.85	469	1.27	437	0.24
NT^*_1	373	-ve	462	1.62	472	1.55	441	0.23
NT^*_2			445	2.41 (*)	467	1.53	440	0.23
NT^*_3					465	-ve	440	0.23
NT^*_4							440	-ve

** Table values are to be multiplied by 10^4 to obtain actual values of estimates of variance.

(*) High value.

POPULATION 2

TABLE - 4.6

Estimates of population total Y and variance estimates**

Estimator	Sample size 2		Sample size 3		Sample size 4		Sample size 5	
	Estimate	Estimate of variance	Estimate	Estimate of variance	Estimate	Estimate of variance	Estimate	Estimate of variance
$N\bar{y}$	34430	76.14	46713	70.34	58031	96.52	42632	30.50
$NS^*_{(1)}$	52797	1612.82 (*)	65516	9.62	69028	9.51	59537	0.28
$NS^*_{(q'_o)}$	53038	1637.47 (*)	65763	8.70	69173	8.85	59759	-ve
NT^*_1	68154	3423.39 (*)	75781	4.53	69633	2.31	69457	0.98
NT^*_2			75805	3375.24 (*)	69648	2.76	69767	0.52
NT^*_3					69654	1371.30 (*)	69858	0.67
NT^*_4							69900	2667.84 (*)

** Table values are to be multiplied by 10^6 to obtain actual values of estimates of variance.

(*) Very high value.

POPULATION 3

TABLE - 4.7

Estimates of population total Y and variance estimates**

Estimator	Sample size 2: Estimate	Sample size 2: Estimate of variance	Sample size 3: Estimate	Sample size 3: Estimate of variance	Sample size 4: Estimate	Sample size 4: Estimate of variance	Sample size 5: Estimate	Sample size 5: Estimate of variance
$N\bar{y}$	16621	40.26	13728	19.98	16445	7.50	19100	15.79
$N\,S^*_{(1)}$	21466	213.85 (*)	18948	6.02	19749	4.25	20432	2.11
$N\,S^*_{(q'_o)}$	21741	225.09 (*)	19[illegible]43	4.26	19936	1.78	20507	2.09
$N\,T^*_1$	20199	163.03 (*)	19912	1.45	20000	1.10	19801	0.27
$N\,T^*_2$			19977	211.52 (*)	19994	1.17	19796	0.20
$N\,T^*_3$					19994	118.85 (*)	19794	0.18
$N\,T^*_4$							19794	32.74 (*)

** Table values are to be multiplied by 10^6 to obtain actual values of estimates of variance.

(*) Very high value.

POPULATION 4

TABLE 4.8

Estimates of population total Y and variance estimates**

Estimator	Sample size 2		Sample size 3		Sample size 4		Sample size 5	
	Estimate	Estimate of variance	Estimate	Estimate of variance	Estimate	Estimate of variance	Estimate	Estimate of variance
$N\bar{y}$	3476	58.19	1276	7.74	1199	3.71	1386	4.83
$N S^*_{(1)}$	2876	21.60 (*)	1710	5.81	1354	1.90	1371	1.93
$N S^*_{(q'_o)}$	2825	18.84 (*)	1747	5.10	1368	1.75	1370	1.76
$N T^*_1$	2991	27.80 (*)	1223	1.37	1141	0.56	1144	0.47
NT^*_2			1291	7.10 (*)	1151	0.56	1137	0.33
NT^*_3					1158	2.15 (*)	1137	0.33
NT^*_4							1139	-ve

** Table values are to be multiplied by 10^5 to obtain actual values of estimates of variance.

(*) Very high ve

POPULATION 5

TABLE - 4.9

Estimates of population total Y and variance estimates**

Estimator	Sample size 2		Sample size 3		Sample size 4		Sample size 5	
	Estimate	Estimate of variance	Estimate	Estimate of variance	Estimate	Estimate of variance	Estimate	Estimate of variance
$N \bar{y}$	13079	26.95	7223	9.01	13239	5.90	11797	5.46
$N S^*_{(1)}$	12857	18.62 (*)	10024	6.12	13562	1.41	12590	1.07
$N S^*_{(q'_o)}$	12834	17.77 (*)	10312	5.12	13595	1.56	12672	1.11
$N T^*_1$	9993	-ve	10334	2.30	13003	0.37	12197	0.29
$N T^*_2$			10566	62.88 (*)	13031	0.32	12208	0.28
$N T^*_3$					13046	-ve	12215	0.28
$N T^*_4$							12218	12.07*

** Table values are to be multiplied by 10^6 to obtain actual values of estimates of variance.

(*) Very high value.

marks on Tables 4.5 to 4.9 :

mark (a) : For the choice $m = n-1$, variance estimates of $N T^{*}_{m}$, $N S^{*}_{(1)}$ $N S^{*}_{(q_o')}$ are either negative or have very high positive values. But in st cases values of $N T^{*}_{n-2}$ and $N T^{*}_{n-1}$ as estimates of population total are rather close, suggesting $V(NT^{*}_{n-2})$ and $V(NT^{*}_{n-1})$ may not differ ch. This view will be further strengthened when in a later chpater we mine the relative efficiencies of several estimators in small samples, ing a reasonable number of independent samples of same size. It may be called that $\hat{V}(N T^{*}_{m})$, etc., are all derived from $\hat{V}(G^{*}_{m})$ given by .6.11), which itself is based on $\hat{V}(G_m)$ given by theorem 3.6.1. Now in orem 3.6.1, $\hat{V}(G_m)$ is not of the same form for all choices of m. For $< m \leq n-2$ it is given by (3.6.4) and for $m = n-1$ by (3.6.5). Actually estimator (3.6.4) breaks down for $m = n-1$ as there is only one obser-ion in the derived sample of size n-m and hence it becomes necessary to an alternate method (3.6.5) for $m = n-1$. From the results in Tables 4.5 4.9, it seems that the alternate method for $m = n-1$ is subjected to very h degree of variation and does not reflect the true situation. In the cumstances, it seems advisable to use the estimator of $V(N T^{*}_{n-2})$ as indicator of the precision of $N T^{*}_{n-1}$. However, when $n = 2$, the only ce is $m = n-1 = 1$. In this case, generally we may assume $N >> n$ and the estimator G^{*}_{1} (3.4.4), of which $N T^{*}_{1}$, $N S^{*}_{(1)}$ and $N S^{*}_{(q_o')}$ are ticular examples, we may use the variance form (3.4.32) in which one may stitute unbiased estimates of the expected values, variances, etc., to in an estimator of variance.

Remark (b) : In the light of Remark (a), it is likely that all the unbiased ratio type estimators are more efficient than the simple estimator $N\bar{y}$

Remark (c) : Among the ratio type estimators, $NS^*_{(1)}$ and $NS^*_{(q_o')}$ seem to have performed well in population 1, while in the other four populations the estimators NT^*_m appear to be more efficient. In large samples, we know from theorem 4.3.3 that $NS^*_{(1)}$ is more efficient than the classical ratio estimator when x and y have a joint normal distribution, $2C_y = 3C_x$ and $\rho \leq 0.70$. In this connection, it is interesting to note that in population 1, value of ρ is the least (0.66) and $2C_y \doteq 3C_x$, and in small samples from it $NS^*_{(1)}$ and $NS^*_{(q_o')}$ have done well.

Remark (d) : For the ratio type estimators NT^*_m, there is some visible trend of increasing efficiency as m increases from 1 to n-1. It may be recalled that, in large samples, theorem 4.4.2 indicates the possible superiority of the choice m = n-1 for the estimators NT^*_m. In addition to the empirical results in the above tables, more studies are needed to confirm such a trend in small samples.

4.5b. Illustration of variance estimation using (4.5.8), (4.5.10) and (4.5.11):

We now consider estimation of the approximate variances of NT^*_m, $NS^*_{(1)}$ and $NS^*_{(q_o')}$ upto order n^{-1}, using the variance estimators (4.5.8), (4.5.10) and (4.5.11) of T^*_m, $S^*_{(1)}$ and $S^*_{(q_o')}$ respectively. To compute $\hat{V}(T^*_m)$ given by (4.5.8), we need to compute $R^*_m = \frac{1}{\binom{n}{m}} \sum_{Z_m}^{\binom{n}{m}} R_m$. It may be noted that, even for a sample of size n = 10, the choice $m = \frac{n}{2} = 5$ involves the formation of $\binom{n}{m} = \binom{10}{5} = 252$ combinations of Z_m and computation of 252 values of R_m and their average. Consequently, even for a sample of size 10, it is rather difficult to estimate the variance of NT^*_m for all choices of m in the range 1 to n-1 in order to study the behaviour of the variance as m increases from 1 to n-1. Further, if efficiency comparisons are to be attempted on the basis of estimated variances, it is necessary to repeat the computations on more than one sample of same size.

In our illustration it is, therefore, imperative to choose the population size, sample size, values of m, and number of independent samples of same size, in such a way that on one hand reasonable care is taken to ensure the validity of the assumptions under which the approximate variance estimators are obtained, and on the other the computational load is also kept within feasible limits. Keeping in view all these considerations, we have selected the following computation plan in our numerical investigation.

Population size	Sample size (n)	Values of m for NT^*_m	Number of independent samples of same size	
	10	1, 3, 5, 7, 9	15	
N = 142	15	1, 2, 13, 14	10	(4.5.12)
	20	1, 2, 18, 19	8	
	25	1, 24	5	

Population data for this study are again taken from the census of India 1971 document referred below Table - 4.2. It consists of 1961 census population (x) and 1971 census population (y) for 142 cities/urban agglomerations of India with population size 100,000 and above. We make use of this population in later chapters also. The relevant data are presented in Table 4.10.

DATA FOR POPULATION 6

TABLE - 4.10

1961 population and 1971 population of 142 cities and urban agglomerations with population 100,000 and above.

INDIA (Figures in '00s)

Name of the city/ urban agglomeration (100,000 and above)	1961 population (persons)	1971 population (persons)
(2)	(3)	(4)
Hyderabad	12490	17989
Visakhapatnam	2112	3633
Vizayawada	2344	3437
Guntur	1871	2699
Warangal	1561	2071
Rajahmundry	1300	1888
Kakinada	1229	1643
Kurnool	1008	1367
Nellore	1068	1336
Eluru	1083	1270
Nizamabad	791	1149
Machilipatnam	1014	1126
Tenali	785	1029
Gauhati	1007	1230
Patna	3646	4903

contd...

Table 4.10 contd.

(1)	(2)	(3)	(4)
16	Jamshedpur	3280	4652
17	Dhanbad	2006	4331
18	Ranchi	1402	2560
19	Gaya	1511	1798
20	Bhagalpur	1438	1727
21	Darbhanga	1030	1321
22	Muzaffarpur	1091	1270
23	Bokaro Steel City	751*	1080
24	Monghyr	898	1025
25	Bihar	786	1001
26	Ahmedabad	11499	15884
27	Surat	2880	4718
28	Baroda	2984	4674
29	Rajkot	1941	3002
30	Bhavnagar	1765	2261
31	Jamnagar	1486	2149
32	Nadiad	790	1083
33	Rohtak	882	1248
34	Ambala Cantt.	1055	1025
35	Srinagar	2853	4036
36	Jammu	1027	1558
37	Cochin	2807	4384
38	Trivandrum	2398	4098
39	Calicut	1925	3340
40	Alleppey	1388	1601
41	Quilon	910	1241
42	Indore	3949	5726
43	Jabalpur	3670	5338
44	Gwalior	3006	4068
45	Bhopal	2229	3921
46	Durg-Bhilainagar	1332	2453
47	Ujjain	1442	2091
48	Raipur	1398	2059
49	Sagar	1047	1548
50	Bilaspur	867	1308

contd...

table 4.10 contd.

(1)	(2)	(3)	(4)
51	Ratlam	875	1186
52	Burhanpur	821	1053
53	Greater Bombay	41520	59685
54	Nagpur	6436	8661
55	Poona	5976	8532
56	Sholapur	3376	3981
57	Kolhapur	1874	2591
58	Amravati	1379	1936
59	Malegaon	1214	1918
60	Nasik	1311	1762
61	Thana	1011	1702
62	Akola	1158	1685
63	Ulhasnagar	1078	1681
64	Aurangabad	876	1505
65	Dhulia	989	1371
66	Nanded	811	1264
67	Ahmadnagar	970	1173
68	Sangli	738	1151
69	Jalgaon	804	1067
70	Bangalore	11526	16487
71	Hubli-Dharwar	2485	3796
72	Mysore	2539	3556
73	Mangalore	1745	2141
74	Belgaum	1468	2138
75	Gulbarga	971	1456
76	Bellary	857	1451
77	Devanagere	781	1210
78	Bijapur	789	1033
	Shimoga	638	1027
	Bhadravathi	658	1013
81	Cuttak	1463	1940
82	Rourkela	903	1725
83	Berhampur	769	1176
84	Bhubaneswar	382	1055
85	Amritsar	3763	4327

contd...

table 4.10 contd.

(1)	(2)	(3)	(4)
86	Ludhiana	2440	4011
87	Jullundur	2226	2961
88	Patiala	1252	1519
89	Jaipur	4034	6131
90	Jodhpur	2248	3189
91	Ajmer	2312	2625
92	Kota	1203	2130
93	Bikaner	1506	1886
94	Udaipur	1111	1629
95	Alwar	727	1008
96	Madras City	17292	24703
97	Madurai	4248	5483
98	Coimbatore	2863	3535
99	Salem	2492	3083
100	Tiruchirapalli	2499	3062
101	Tuticorin	1242	1548
102	Nagercoil	1062	1412
103	Thanjavur	1111	1405
104	Vellore	1137	1382
105	Dindigul	929	1274
106	Singanallur	247	1134
107	Tiruppur	798	1132
108	Kumbakonam	926	1130
109	Kanchipuram	927	1105
110	Tirunelveli	880	1085
111	Erode	738	1037
112	Cuddalore	792	1013
113	Kanpur	9710	12730
114	Lucknow	6557	8262
115	Agra	5087	6378
116	Varanasi	4898	5829
117	Allahabad	4307	5140
118	Meerut	2840	3678
119	Bareilly	2728	3261
120	Mordabad	1918	2744

contd...

table 4.10 contd.

(1)	(2)	(3)	(4)
121	Aligarh	1850	2540
122	Gorakhpur	1802	2307
123	Saharanpur	1852	2257
124	Dehra Dun	1563	1994
125	Jhansi	1697	1981
126	Rampur	1354	1618
127	Shahjahanpur	1177	1441
128	Mathura	1253	1405
129	Firozabad	986	1339
130	Ghaziabad	704	1280
131	Muzaffarnagar	876	1149
132	Farrukhabad-cum-Fatehgarh	946	1114
133	Faizabad	883	1098
134	Mirzapur-cum-Vindhyachal	1001	1059
135	Calcutta	57369	70054
136	Durgapur	417	2072
137	Kharagpur	1473	1619
138	Asansol	1034	1574
139	Burdwan	1082	1450
140	Chandigarh	993	2330
141	Delhi	23593	36298
142	Imphal	677	1006

* Bokaro Steel City came into existence only after 1961. To ensure uniformity, a hypothetical 1961 population is estimated based on growth rate in other cities of Bihar State in which Bokaro is situated.

For population 6 some parameters of interest are:

$$Y = 570161, \quad X = 411855, \quad \sigma_y^2 = 71896861, \quad \sigma_x^2 = 40607638, \quad \sigma_{xy} = 53750172$$

$$C_y = 2.11, \quad C_x = 2.60, \quad \rho = 0.9948, \quad \beta = 1.32 \qquad (4.5.13)$$

We now give in the following Tables 4.11 to 4.14, the contents (serial numbers of the units) of the simple random samples drawn without replacement from population 6 according to the plan (4.5.12).

TABLE - 4.11

Simple random samples of size 10 from population 6

Sample Number														
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
55	75	52	68	23	50	48	122	116	63	68	113	30	44	22
8	14	11	12	31	94	140	15	23	107	95	131	76	95	94
37	128	128	115	117	1	94	45	18	68	125	45	90	70	132
22	2	41	76	96	12	13	30	92	11	62	21	41	58	7
79	40	133	126	39	51	87	27	135	82	11	84	46	114	122
127	74	6	111	142	119	95	86	57	135	119	75	64	6	115
126	55	10	16	41	79	75	107	30	113	60	115	138	106	72
90	91	139	47	83	24	117	73	93	47	109	116	89	73	127
69	57	71	100	126	97	64	130	63	123	123	101	7	111	4
21	27	136	140	36	129	88	99	19	136	73	34	107	110	129

TABLE - 4.12

Simple random samples of size 15 from population 6

Sample Number									
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
45	36	114	55	26	44	70	133	123	4
83	41	13	76	137	45	88	118	128	33
142	100	40	[illegible]	32	77	[illegible]	99	53	131
93	3	20	23	73	91	77	59	50	125
25	115	102	19	45	25	9	107	136	71
87	129	76	81	90	32	84	68	102	23
46	19	134	113	99	54	58	6	139	93
39	59	41	58	89	77	86	9	61	87
66	35	89	24	15	69	68	114	8	116
96	140	125	46	114	63	97	53	84	92

contd....

table 4.12 contd.,

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
27	76	8	9	59	78	11	38	121	14
57	109	69	59	100	41	1	90	14	67
41	131	63	98	16	90	126	140	97	5
110	27	96	15	142	87	111	79	64	39
20	105	86	6	80	140	66	66	112	36

TABLE - 4.13

Simple random samples of size 20 from population 6

Sample Number							
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
141	40	44	115	36	27	42	128
82	13	87	87	114	2	110	116
127	95	96	113	40	29	94	103
72	36	97	9	18	44	105	44
113	72	37	22	128	100	128	15
102	90	132	39	2	112	16	52
87	19	98	133	1	139	25	6
57	[illegible]	[illegible]	105	[illegible]	102	2	89
92	68	70	109	130	[illegible]	[illegible]	32
117	79	71	27	41	53	78	33
75	[illegible]	[illegible]	[illegible]	3	[illegible]	[illegible]	[illegible]
4	78	86	51	115	25	127	101
23	137	138	119	35	9	23	115
38	71	3	29	74	40	66	70
129	21	78	88	62	95	112	72
[illegible]	68	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
70	26	10	79	10	13	33	76
50	109	111	86	82	58	130	47
112	27	68	121	48	21	116	96
121	38	5	74	101	117	29	80

TABLE - 4.14

Simple random samples of size 25 from population 6

Sample Number				
(1)	(2)	(3)	(4)	(5)
22	78	12	78	46
84	121	10	56	30
25	89	76	129	108
107	22	38	124	124
9	90	136	4	129
42	80	15	134	8
8	81	74	93	52
44	118	110	50	96
4	54	120	61	141
39	10	142	10	111
73	128	106	138	47
81	85	102	2	61
134	71	16	45	67
47	29	90	73	7
96	3	118	44	97
18	96	125	72	139
101	34	55	86	98
140	24	25	101	58
138	83	21	59	127
10	63	116	90	99
13	103	131	83	142
112	127	80	32	1
33	46	67	69	50
80	76	8	108	34
30	126	4	6	113

Omitting all intermediate calculations on the samples data, we now present in Tables 4.15 to 4.18 the estimates of approximate variances of the estimators NT^*_m, $NS^*_{(1)}$ and $NS^*_{(q^*_0)}$, computed in accordance with the formulae (4.5.8), (4.5.10) and (4.5.11) and the plan (4.5.12). For the purpose of comparison, we also include the usual estimators of variance of $N\bar{y}$ and $N\hat{\bar{Y}}_R$

POPULATION 6

TABLE - 4.15

Estimates** of approximate variances using samples of size 10

Estimator	Sample Number (1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	Average over samples
$N\bar{y}$	100.64	93.74	12.84	59.06	994.72	520.90	28.52	53.47	8603.97	8686.17	9.24	269.17	43.09	459.24	48.66	1332.23
$N\hat{Y}_R$	1.10	3.28	5.34	3.37	3.07	2.71	4.00	3.17	9.46	29.19	0.49	4.17	1.17	3.50	0.59	4.97
$NS^*_{(1)}$	1.10	3.28	5.34	3.37	3.07	2.71	4.00	3.17	9.46	29.19	0.49	4.17	1.17	3.50	0.59	4.97
$NS^*_{(q_o')}$	0.77	3.06	4.84	2.51	3.02	1.49	1.83	3.17	1.76	3.97	0.18	3.65	1.14	3.49	0.53	2.36
NT^*_1	1.58	3.53	6.25	4.37	3.08	3.54	5.33	3.17	223.25	2080.95	0.62	9.59	1.18	23.73	0.59	158.05
NT^*_3	1.22	3.34	5.42	3.68	3.04	3.77	4.55	3.18	95.94	441.88	0.52	4.74	1.18	3.67	0.60	38.45
NT^*_5	1.15	3.31	5.36	3.50	3.09	3.39	4.27	3.17	50.10	185.73	0.50	4.37	1.18	3.52	0.60	18.22
NT^*_7	1.12	3.29	5.35	3.43	3.10	3.06	4.12	3.17	26.66	85.75	0.49	4.24	1.17	3.50	0.55	9.94
NT_9	1.10	3.28	5.34	3.39	3.09	2.81	4.03	3.17	13.61	41.32	0.49	4.18	1.17	3.50	0.59	6.07

** Table values are to be multiplied by 10^8 to obtain actual values of the estimates.

POPULATION 6

TABLE 4.16

Estimates** of approximate variances using samples of size 15.

Estimator	Sample Number (1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	Average over samples
$N\bar{y}$	18.37	30.04	451.06	3703.54	178.08	48.07	368.91	2661.40	14.70	20.27	749.36
$N\hat{Y}_R$	1.31	2.06	1.88	9.83	1.94	1.69	0.98	2.68	2.99	1.65	2.70
$NS^*_{(1)}$	1.31	2.06	1.88	9.83	1.94	1.69	0.98	2.68	2.99	1.65	2.70
$NS^*_{(q'_o)}$	1.24	1.74	1.69	1.71	1.92	1.44	0.98	2.66	1.85	1.29	1.65
NT^*_1	1.34	2.23	2.25	64.81	1.98	1.86	2.16	7.01	5.35	1.73	9.07
NT^*_2	1.33	2.16	2.15	53.73	1.95	1.81	1.11	4.02	3.72	1.72	7.36
NT^*_{13}	1.31	2.06	1.89	12.21	1.94	1.69	0.98	2.68	3.01	1.65	2.94
NT^*_{14}	1.31	2.06	1.88	10.94	1.94	1.69	0.98	2.68	3.00	1.65	2.81

** Table values are to be multiplied by 10^8 to obtain actual values of the estimates.

POPULATION 6

TABLE - 4.17

Estimates** of approximate variances using samples of size 20.

Estimator	Sample Number (1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	Average over samples
$N\bar{y}$	597.23	96.04	298.21	64.81	133.17	1447.33	29.67	304.45	371.36
$N\hat{\bar{Y}}_R$	5.23	1.28	0.72	1.30	1.46	1.90	1.03	1.39	1.79
$NS^*_{(1)}$	5.23	1.28	0.72	1.30	1.46	1.90	1.03	1.39	1.79
$NS^*_{(q'_o)}$	3.83	1.26	0.65	1.10	1.46	1.06	0.91	1.14	1.43
NT^*_1	5.18	1.27	0.89	1.36	1.54	6.47	1.11	1.69	2.44
NT^*_2	5.71	1.27	0.83	1.38	1.50	5.06	1.09	1.63	2.31
NT^*_{18}	5.28	1.28	0.73	1.30	1.46	2.03	1.03	1.40	1.81
NT^*_{19}	5.26	1.28	0.73	1.30	1.46	1.96	1.03	1.40	1.80

** Table values are to be multiplied by 10^8 to obtain actual values of the estimates.

POPULATION 6

TABLE - 4.18

Estimates** of approximate variances using samples of size 25.

Estimator	Sample Number					Average over samples
	(1)	(2)	(3)	(4)	(5)	
$N\bar{y}$	146.32	152.44	23.79	8.07	504.49	167.02
$N\hat{Y}_R$	0.68	0.88	1.42	0.72	3.11	1.36
$NS^*_{(1)}$	0.68	0.88	1.42	0.72	3.11	1.36
$NS^*_{(q'_o)}$	0.68	0.72	1.25	0.72	1.88	1.05
NT^*_1	0.79	1.07	2.69	0.72	5.40	2.13
NT^*_{24}	0.68	0.88	1.42	0.72	3.13	1.37

** Table values are to be multiplied by 10^8 to obtain actual values of the estimates.

Remarks on Tables 4.15 to 4.18 :

Remark (a): When the variance is estimated upto order n^{-1}, all the ratio estimators have shown significant gains in precision over the simple unbiased estimator $N\bar{y}$.

Remark (b): The precision of the unbiased ratio type estimators NT^*_m is clearly on the increase as m increases from 1 to n-1.

Remark (c): The unbiased ratio type estimators $NS^*_{(1)}$ and NT^*_{n-1} are almost of equal precision as the classical ratio estimator $N\hat{Y}_R$.

Remark (d) : The unbiased ratio type estimator $N S^{*}_{(q_o')}$, using information on $\rho \; C_y / C_x$, is the most efficient estimator among the estimators compared.

Remark (e) : The above observations are consistent with our theoretical results in sections 4.3 and 4.4.

CHAPTER V

PRECISION OF UNBIASED RATIO TYPE ESTIMATORS UNDER SOME SUPER POPULATION MODELS.

5.0. Summary

In the previous chapter the behaviour of the variance of the unbiased ratio type estimator T_m^* as m varies in the range 1 to n-1 has been studied in large samples, and it has been shown that when the regression of y on x is linear and does not pass through the origin and all values of x in the population are of same sign, $V(T_m^*)$ is a strictly decreasing function of m and $V(T_m^*) \geq V(T_{n-1}^*) = V(\hat{\bar{Y}}_R)$. In this chapter we investigate the behaviour of the expected variance of T_m^* for different choices of m under two super-population models $\triangle$ and $\triangle'$. The model $\triangle$ assumes that (i) population is infinite, (ii) $y_j = \alpha + \beta x_j + u_j$, $E(u_j/x_j) = 0$; (iii) $V(u_j/x_j) = \delta$, a constant; and (iv) $Cov(u_j, u_{j'}/x_j, x_{j'}) = 0$ for $j \neq j'$. The model $\triangle'$ is a special case of $\triangle$. In addition to the four conditions of $\triangle$, the model $\triangle'$ assumes that x is a gamma variate. Under $\triangle$ the expected variance of T_m^* and the expected mean square error of $\hat{\bar{Y}}_R$ are evaluated upto order n^{-3}; while the derivations under $\triangle'$ are valid for any sample size. The expected variance of T_m^* is studied for different choices of m with a view to finding out the optimum choice of m, and it is also compared with the expected mean square error of $\hat{\bar{Y}}_R$ under the respective models.

When the sample is split at random into g groups each of size (n = mg) and a sub-set of n-m sample elements obtained by the exclusion of one such group from the sample is denoted by Z_{n-m}, we may compute unbiased ratio type estimators T_{n-m} corresponding to the g possible

sub-sets Z_{n-m}. The estimator T_{n-m} is obtained from Mickey's basic general unbiased estimator G_{n-m} by choosing for a single auxiliary variate x, $a(Z_{n-m}) = \bar{y}_{n-m} / \bar{x}_{n-m}$. Let the average of the g estimators, T_{n-m}, be represented by $\bar{T}'_{n-m}$. It may be easily seen that $\bar{T}'_{n-m}$ averaged over all possible splits of the sample into groups of size m results in T^*_{n-m}, and consequently, $V(\bar{T}'_{n-m}) \geqslant V(T^*_{n-m})$. Also $\bar{T}'_{n-1} = T^*_{n-1}$. Making use of these relations between $\bar{T}'_{n-m}$ and T^*_{n-m} and the results on the expected variance of T^*_m under the super-population model $\triangle$, we also investigate the behaviour of the expected variance of $\bar{T}'_{n-m}$ for different choices of m.

The chapter contains six sections. In the introductory section, the problems are formulated. The results reported in the other five sections are briefly presented in the following. Section 2 is fully devoted to the derivation of the expected variance of T^*_m under $\triangle$ for choices of m in the range $\frac{n}{2}$ to $n-1$. In the derivation, repeated use is made of Sukhatme's (1944) results on the moments and product moments of moment statistics for random samples from infinite populations. Without loss of generality, it is assumed that the units of measurement of the variate x are so chosen that the population mean $\bar{X} = 1$. The derivation involves very heavy algebra and the final result is presented in theorem 5.2.1. The expected variance of T^*_m under $\triangle$, upto order n^{-3}, turns out to be a function of (i) the parameters of the super population model, α and δ; (ii) the mean, the second and the third central moments of the variate x; (iii) the choice of m; and (iv) the sample size n.

In section 3, we determine the optimum choice of m for T_m^* and T_{n-m} under $\triangle$. When the expected variance is considered upto order n^{-3}, it is shown that n-1 is the optimum choice of m in the range $(\frac{n}{2}, n-1)$ for T_m^* under $\triangle$, provided the coefficient of variation of x is not less than the coefficient of skewness of x. It may be noted that the condition on the coefficient of variation of x is only a sufficient but not a necessary condition. For all negatively skewed or symmetric distributions of x, the condition holds and hence the result is true. In particular, under Durbin's (1959) super-population model 1, which is a special case of $\triangle$ with the further assumption that x is a normal variate, the result is true. If the sample size n is large enough to neglect the contribution of terms of order n^{-3} to the expected variance of T_m^* under $\triangle$, then the condition on the coefficient of variation of x may be dispensed with and the result is valid for all distributions of x.

It is interesting to note that the above results on the optimum choice of m for T_m^* in the range $(\frac{n}{2}, n-1)$ under $\triangle$ may be used to derive similar results for the class of unbiased ratio type estimators $\bar{T}'_{n-m}$ for values of m in the range $(1, \frac{n}{2})$. Following such an approach, in section 3 it is shown that, when the expected variance under $\triangle$ is considered upto order n^{-3}, the estimator $\bar{T}'_{n-1} = T^*_{n-1}$ is the optimum choice estimator in the class of estimators $\bar{T}'_{n-m}$ $(1 \leqslant m \leqslant \frac{n}{2})$, provided the coefficient of variation of x is not less than the coefficient of skewness of x. In particular, the result holds for all negatively skewed or symmetric distributions of x. This result, when x has a normal distribution, is originally due to Rao (1967). If the expected variance under $\triangle$ is considered only upto terms of order n^{-2}, then without any assumption on the coefficient of variation of x, $\bar{T}'_{n-1} = T^*_{n-1}$ is the optimum choice in the class of estimators $\bar{T}'_{n-m}$ ($1 \leq m \leq \frac{n}{2}$).

In section 4, we obtain a comparison of the optimum choice estimator T^*_{n-1} with the classical ratio estimator $\hat{\bar{Y}}_R$ under $\triangle$ and show that, when the expected variance or the mean square error is considered upto order n^{-3}, T^*_{n-1} is more efficient than $\hat{\bar{Y}}_R$ provided the coefficient of variation of x is not less than the coefficient of skewness of x. In particular, the results holds for all negatively skewed or symmetric distributions of x. This result, when x has a normal distribution, is originally due to Rao (1967). Summing up the results in sections 3 and 4, we observe that, when the expected variance or mean square error under $\triangle$ is considered upto order n^{-3}, the estimator T^*_{n-1} is the optimum choice estimator in the classes of estimators T^*_m ($\frac{n}{2} \leq m \leq n-1$) and $\bar{T}'_{n-m}$ ($1 \leq m \leq \frac{n}{2}$), and is also superior to the classical ratio estimator $\hat{\bar{Y}}_R$, provided the coefficient of variation of x is not less than the coefficient of skewness of x.

In section 5, we derive the expected variance of T^*_m under the super-population model $\triangle'$. The expected variance is applicable for any sample size n and any choice of m in the range (1, n - 1). In the derivation, we make use of the results due to Rao and Webster (1966) and Rao (1967) on the expected values of certain functions of mutually independent gamma variates. The expected variances of T_1 and T^*_{n-1}

under $\triangle'$, which are available from the results of Rao (1967), may be obtained as particular cases of our result. It may be noted that Rao (1967) has derived under the model $\triangle'$ the expected variance for the class of unbiased ratio type estimators $\bar{\bar{T}}'_{n-m}$ ($1 \leq m \leq \frac{n}{2}$) and numerically demonstrated that (i) $\bar{\bar{T}}'_{n-1} = T^*_{n-1}$ is the optimum choice for $\bar{\bar{T}}'_{n-m}$, (ii) $\bar{\bar{T}}'_{n-1}$ is superior to Hartley-Ross estimator T^*_1 and (iii) $\bar{\bar{T}}'_{n-1}$ is more efficient than the classical ratio estimator $\hat{\bar{Y}}_R$ for $n \geq 8$. From our result under the model $\triangle'$, taking x as a gamma variate with parameter h, one can obtain a numerical comparison of the expected variance of T^*_m and the expected mean square error of the classical ratio estimator $\hat{\bar{Y}}_R$ for specific choices of h, n and m. As an illustration, such a numerical comparison is made in section 6 for the choices:

$$h = 3, \quad n = 2, \quad n = 3 \quad \text{and} \quad n = 4$$
$$m = 1, \quad m = 1, 2 \qquad m = 1, 2, 3 ;$$

and it is shown that (i) the precision of T^*_m increases as m increases from 1 to $n-1$ for $n = 3$ and 4, (ii) T^*_{n-1} ($n = 3$, and 4) is more efficient than $\hat{\bar{Y}}_R$ and (iii) Hartley - Ross estimator T^*_1 is less efficient than $\hat{\bar{Y}}_R$.

5.1. Introduction

In chapter IV we have obtained the approximate variance (4.4.1) of Mickey's unbiased ratio type estimator T^*_m (4.1.2) upto order n^{-1} and observed that it is equal to the large sample variance of the classical ratio timator $\hat{\bar{Y}}_R$ for large values of m. Using the same approximate variance of T_m^*, we have also shown in theorem 4.4.2 that $m = n - 1$ is the optimum choice, when the regression of y on x is linear and does not pass through the origin and values of x on all units in the population are of same sign. We may also recall our discussion in section 2.4 (ch. II) wherein we have reviewed the available results due to Rao (1967), and Rao and Beegle (1967) on the precision of Mickey type unbiased ratio estimator $\bar{t}'_{n-m}$ (2.4.4) for the population ratio, using Durbin's (1959) super-population models 1 and 2 as stated in section 2.3 (ch. II). In terms of the general classes of unbiased estimators $\bar{G}'_{n-m}$ and G_m^* considered in section 3.2 (ch. III), it may be noted that $\bar{t}'_{n-m}$ corresponds to $\bar{G}'_{n-m}$ whereas T_m^* corresponds to G_m^*. However, $\bar{t}'_{n-1}\ \bar{X} = T_{n-1}$.

Against the above background, in this chapter we examine the relative efficiencies of the unbiased ratio type estimator T_m^* and the biased ratio estimator $\hat{Y}_R$ under some super-population models following two approaches. In one approach no assumption is made on the frequency distribution of the variate y or x, but we make some assumption on the sample size n and evaluate the variance/mean square error upto terms of order n^{-3}. In the second approach results valid for any sample size are derived under Durbin's super-population model 2 in which it is assumed that x is a gamma variate.

We consider the following super-population models $\triangle$ and $\triangle'$.

<u>Super-population mode $\triangle$</u> :

(i) Population is infinite; (ii) $y_j = \alpha + \beta x_j + u_j$, $E(u_j/x_j) = 0$;
(iii) $E(u_j^2/x_j) = \delta$ (a constant); and (iv) $E(u_j u_{j'}/x_j, x_{j'}) = 0, j \neq j'$ (5.1.1)

It may be noted that Durbin's (1959) super-population model 1 sists of the conditions (5.1.1) and the assumption x is a normal iate.

per-population model Δ'

In addition to the four conditions of the model Δ, x is a gamma riate. (5.1.2)

We may note that the model Δ' is Durbin's (1959) super lation model 2.

Under super-population model Δ, we derive the expected riance of T^*_m and the expected mean square error of $\hat{\bar{Y}}_R$ upto order and compare them. Also under the model Δ', we get an exact ression for the expected variance of T^*_m valid for any sample size n d obtain an empirical comparison between the expected variance of T^*_m the expected exact mean square error of $\hat{\bar{Y}}_R$. While the results ined are easily interpreted, the proofs are complicated and involve ry heavy algebra. We, therefore, prove the relevant theorems through ral lemmas.

2. Expected variance of unbiased ratio type estimator T^*_m (4.1.2) under super-population model Δ (5.1.1).

a 5.2.1: Under super-population model Δ (5.1.1), the expected ance of T^*_m (4.1.2) is given by

$$T^*_m) = E\left[\bar{u} + E_m\left\{\frac{\bar{u}_m}{\bar{x}_m}(\bar{X} - \bar{x}_{n-m})\right\}\right]^2 + a^2 E\left[E^2_n\left(\frac{\bar{X} - \bar{x}_{n-m}}{\bar{x}_m}\right)\right]. \quad (5.2.1)$$

Proof : We know the basic unbiased ratio type estimator T_m (4.1.1) results from the basic general unbiased estimator G_m (3.4.1) for p = 1 and $a(Z_m) = R_m$. Hence using the form (3.1.2) of G_m we have

$$T_m = \frac{m\,\bar{y}_m + (N-m)\,\bar{y}_{n-m}}{N} - \frac{(N-m)}{N} R_m (\bar{x}_{n-m} - \bar{X}_{N-m}). \quad (5.2.2)$$

Write $\bar{X}_{N-m} = \frac{N\bar{X} - m\bar{x}_m}{N-m} = \bar{X} + \frac{m}{(N-m)}(\bar{X} - \bar{x}_m)$ in (5.2.2) and

obtain $T_m = \frac{m\bar{y}_m + (N-m)\bar{y}_{n-m}}{N} - R_m(\bar{x}_{n-m} - \bar{X}) + \frac{m}{N} R_m(\bar{x}_{n-m} - \bar{x}_m)$.

Then $T_m^* = E_n(T_m) = \bar{y} - E_n[R_m(\bar{x}_{n-m} - \bar{X})] + \frac{m}{N} E_n[R_m(\bar{x}_{n-m} - \bar{x}_m)]$,

(5.2.3)

Where, as usual, $E_n(\,.\,)$ denotes expectation over all possible choices of Z_m from a given sample.

Under the model Δ (5.1.1) population size is infinite and so the third term in (5.2.3) may be neglected. Thus we have

$$T_m^* = \bar{y} + E_n[R_m(\bar{X} - \bar{x}_{n-m})]. \quad (5.2.4)$$

Using (5.1.1) we write in (5.2.4)

$\bar{y} = \alpha + \beta\bar{x} + \bar{u}$, and $R_m = \frac{\bar{y}_m}{\bar{x}_m} = \frac{\alpha + \bar{u}_m}{\bar{x}_m} + \beta$ and obtain

$$T_m^* = (\alpha + \beta\bar{x} + \bar{u}) + E_n\left[\frac{(\alpha + \bar{u}_m)}{\bar{x}_m}(\bar{X} - \bar{x}_{n-m})\right] + \beta E_n[\bar{X} - \bar{x}_{n-m}]$$

$$= (\alpha + \beta\bar{X}) + \left[\bar{u} + E_n\left\{\frac{(\alpha + \bar{u}_m)}{\bar{x}_m}(\bar{X} - \bar{x}_{n-m})\right\}\right]. \quad (5.2.5)$$

Note that under the model $\triangle$ (5.1.1), $E(T_m^*) = \bar{Y} = \alpha + \beta \bar{X}$.

Hence from (5.2.5)

$$V(T_m^*) = E(T_m^* - \bar{Y})^2 = E\left[\bar{u} + E_n \left\{ \frac{(\alpha + \bar{u}_m)}{\bar{x}_m} (\bar{X} - \bar{x}_{n-m}) \right\} \right]^2 \qquad (5.2.6)$$

Now we expand (5.2.6) and get the result (5.2.1) of the lemma, since the product term :

$$2E\left[\left\{ \bar{u} + E_n \left[\frac{\bar{u}_m}{\bar{x}_m} (\bar{X} - \bar{x}_{n-m}) \right] \right\} E_n \left\{ \frac{\alpha (\bar{X} - \bar{x}_{n-m})}{\bar{x}_m} \right\} \right] = 0$$

under the model $\triangle$ (5.1.1) in view of the assumption $E(u_j / x_j) = 0$.

<u>Remark (a)</u> : In what follows we assume, without loss of generality, that the units of measurement of the variate x are so chosen that $\bar{X} = 1$.

<u>Lemma 5.2.2</u> : For choices of m for which $\left| \frac{\bar{x}_m - \bar{x}}{\bar{x}} \right| < 1$,

$$\bar{u} + E_n \left[\frac{\bar{u}_m}{\bar{x}_m} (1 - \bar{x}_{n-m}) \right] \doteq \frac{\bar{u}}{\bar{x}} + \left(1 - \frac{n\bar{x}}{n-m}\right) \left[L_1 \left(\frac{m_{02}\bar{u}}{\bar{x}^3} - \frac{m_{11}}{\bar{x}^2} \right) + L_2 \left(\frac{m_{12}}{\bar{x}^3} - \frac{m_{03}\bar{u}}{\bar{x}^4} \right) + L_3 \left(\frac{m_{02}^2 \bar{u}}{\bar{x}^5} - \frac{m_{02} m_{11}}{\bar{x}^4} \right) + L_4 \left(\frac{m_{04}\bar{u}}{\bar{x}^5} - \frac{m_{13}}{\bar{x}^4} \right) \right], \qquad (5.2.7)$$

where

$$L_1 = \frac{(n - m)}{m(n-1)}, \quad L_2 = \frac{(n-m)(n-2m)}{m^2(n-1)(n-2)},$$

$$L_3 = \frac{3n(n-m)(n-m-1)(m-1)}{m^3(n-1)(n-2)(n-3)}, \quad L_4 = \frac{(n-m)(n^2 - 6mn + n + 6m^2)}{m^3(n-1)(n-2)(n-3)}, \qquad (5.2.8)$$

and $m_{p_1 p_2} = \frac{1}{n} \sum_{j=1}^{n} (u_j - \bar{u})^{p_1} (x_j - \bar{x})^{p_2}$. (5.2.9)

Proof : We have

$$\bar{u} + E_n \left[\frac{\bar{u}_m}{\bar{x}_m} (1 - \bar{x}_{n-m}) \right] = \bar{u} + E_n \left[\frac{\bar{u}_m}{\bar{x}_m} \left(1 - \frac{n\bar{x} - m\bar{x}_m}{n-m}\right) \right]$$

$$= \frac{n\bar{u}}{(n-m)} + \left(1 - \frac{n\bar{x}}{n-m}\right) E_n\left(\frac{\bar{u}_m}{\bar{x}_m}\right) \ldots \quad (5.2.10)$$

In (5.2.10) we use for $E_n\left(\frac{\bar{u}_m}{\bar{x}_m}\right)$ an approximation given by Sukhatme (p. 140, 1970), which is valid for values of m for which $\left|\frac{\bar{x}_m - \bar{x}}{\bar{x}}\right| < 1$. Then we obtain the expression (5.2.7) of the lemma. The approximation neglects the contribution to $E_n\left(\frac{\bar{u}_m}{\bar{x}_m}\right)$ from terms

$$E_n \left[(\bar{u}_m - \bar{u})^{p_1} (\bar{x}_m - \bar{x})^{p_2} \right] \text{ for } p_1 + p_2 \geq 5.$$

Remark (b) : To evaluate the first term of $V(T^*_m)$ given by lemma 5.2.1, we make use of lemma 5.2.2. Our approach is first to take the conditional expectation of the square of the expression (5.2.7) for given x under the model Δ (5.1.1). This would involve the conditional expectations for given x of :

$$\bar{u}^2,\ m_{11}^2,\ m_{11}\bar{u},\ m_{12}^2,\ m_{12}\bar{u},\ m_{13}^2,\ m_{13}\bar{u},\ m_{11}m_{12},\ m_{11}m_{13}, \text{ and } m_{12}m_{13}$$

Denoting the conditional expectation for given x by $E_x(\cdot)$, the following lemma contains the required conditional expectations.

Lemma 5.2.3 : Under the super population model Δ (5.1.1) we have

$$E_x(\bar{u}^2) = \frac{\delta}{n},\ E_x(m_{11}^2) = \frac{\delta}{n} m_2,\ E_x(m_{12}^2) = \frac{\delta}{n}(m_4 - m_2^2),\ E_x(m_{13}^2) = \frac{\delta}{n}(m_6 - m_3^2),$$

$$E_x(m_{11} m_{12}) = \frac{\delta}{n} m_3,\ E_x(m_{11} m_{13}) = \frac{\delta}{n} m_4,\ E_x(m_{12} m_{13}) = \frac{\delta}{n}(m_5 - m_2 m_3), \qquad (5.2.11)$$

$$E_x(m_{11}\ \bar{u}) = E_x(m_{12}\ \bar{u}) = E_x(m_{13}\ \bar{u}) = 0\ ;\ \text{where } m_{p_i} = m_{op_i}.$$

<u>**Proof :**</u> Under $\triangle$(5.1.1) we have

$$E_x(u_j) = 0,\ E_x(u_j^2) = \delta \text{ and } E_x(u_j\ u_{j'}) = 0 \text{ for } j \neq j'.$$

Hence $E_x(\bar{u}^2) = \frac{\delta}{n},\ E_x(u_j - \bar{u})^2 = \delta\left(1 - \frac{2}{n} + \frac{1}{n}\right) = \frac{n-1}{n}\delta.$

$$E_x\left[(u_j - \bar{u})(u_{j'} - \bar{u})\right] = \delta\left[-\frac{1}{n} - \frac{1}{n} + \frac{1}{n}\right] = -\frac{\delta}{n},\ \text{for } j \neq j'. \qquad \ldots (5.2.12)$$

Further

$$E_x(m_{1p_1} m_{1p_2}) = \frac{1}{n^2} E_x\left[\sum_{j=1}^{n}(u_j - \bar{u})^2 (x_j - \bar{x})^{p_1 + p_2} + \sum_{j \neq j'}^{n}\sum^{n}(u_j - \bar{u})(u_{j'} - \bar{u})(x_j - \bar{x})^{p_1}(x_{j'} - \bar{x})^{p_2}\right]$$

$$= \frac{1}{n^2}\left[\frac{\delta(n-1)}{n}\sum_{j=1}^{n}(x_j - \bar{x})^{p_1 + p_2} - \frac{\delta}{n}\sum_{j \neq j'}^{n}\sum^{n}(x_j - \bar{x})^{p_1}(x_{j'} - \bar{x})^{p_2}\right],$$

in view of (5.2.12)

$$= \frac{\delta}{n}\left[m_{p_1 + p_2} - m_{p_1}\ m_{p_2}\right]. \qquad \ldots (5.2.13)$$

From (5.2.13) we have $E_x(m_{11}^2) = \frac{\delta}{n} m_2,\ E_x(m_{12}^2) = \frac{\delta}{n}(m_4 - m_2^2),$

$$(m_{13}^2) = \frac{\delta}{n}(m_6 - m_3^2),\ E_x(m_{11} m_{12}) = \frac{\delta}{n} m_3,\ E_x(m_{11} m_{13}) = \frac{\delta}{n} m_4$$

$$E_x(m_{12}\ m_{13}) = \frac{\delta}{n}(m_5 - m_2 m_3). \qquad (5.2.14)$$

Also $E_x(m_{1p_1}\bar{u}) = E_x\left[\frac{}{n}\sum_{j=1}^{n}\bar{u}(u_j-\bar{u})(x_j-\bar{x})\right] = \left(\frac{\delta}{n}-\frac{}{n}\right)m_{p_1} = 0.$

Hence $E_x(m_{11}\bar{u}) = E_x(m_{12}\bar{u}) = E_x(m_{13}\bar{u}) = 0.$ ... (5.2.15)

Now the lemma follows from (5.2.12), (5.2.14) and (5.2.15).

Lemma 5.2.4 : Under super-population model Δ (5.1.1),

$$E_x\left[\bar{u}+E_n\left\{\frac{\bar{u}_m}{\bar{x}_m}(1-\bar{x}_{n-m})\right\}\right]^2 = \frac{\delta}{n}\left[\frac{1}{\bar{x}^2}+\left(1-\frac{n\bar{x}}{n-m}\right)^2 L_1^2\left(\frac{m_2^2}{\bar{x}^6}+\frac{m_2}{\bar{x}^4}\right)\right.$$
$$\left.+2\left(1-\frac{n\bar{x}}{n-m}\right)\left(\frac{L_1 m_2}{\bar{x}^4}-\frac{L_2 m_3}{\bar{x}^5}+\frac{L_3 m_2^2}{\bar{x}^6}\right)\right]+O(n^{-4}), \qquad (5.2.16)$$

provided choice of m is in the range $\left(\frac{n}{2}, n-1\right)$.

Proof : From lemma 5.2.2 we have, for choices of m for which $\left|\frac{\bar{x}_m-\bar{x}}{\bar{x}}\right| < 1$, $E_x\left[\bar{u}+E_n\left\{\frac{\bar{u}_m}{\bar{x}_m}(1-\bar{x}_{n-m})\right\}\right]^2 \doteq E_x\left[\text{Expression}(5.2.7)\right]^2$.

We now take the conditional expectation term by term in the square of expression (5.2.7) and use the results of lemma 5.2.3. Then we obtain

$$\left[\bar{u}+E_n\left\{\frac{\bar{u}_m}{\bar{x}_m}(1-\bar{x}_{n-m})\right\}\right]^2 \doteq \frac{\delta}{n}\left[\frac{1}{\bar{x}^2}+\left(1-\frac{n\bar{x}}{n-m}\right)^2\left\{L_1^2\left(\frac{m_2^2}{\bar{x}^6}+\frac{m_2}{\bar{x}^4}\right)\right.\right.$$
$$+L_2^2\left(\frac{m_4-m_2^2}{\bar{x}^6}+\frac{m_3^2}{\bar{x}^8}\right)+L_3^2\left(\frac{m_2^4}{\bar{x}^{10}}+\frac{m_2^3}{\bar{x}^8}\right)+L_4^2\left(\frac{m_4^2}{\bar{x}^{10}}+\frac{m_6-m_3^2}{\bar{x}^8}\right)$$

contd...

$$-2L_1L_2\left(\frac{m_3}{\bar{x}^5}+\frac{m_2m_3}{\bar{x}^7}\right)+2L_1L_3\left(\frac{m_2^3}{\bar{x}^8}+\frac{m_2^2}{\bar{x}^6}\right)$$

$$+2L_1L_4\left(\frac{m_2m_4}{\bar{x}^8}+\frac{m_4}{\bar{x}^6}\right)-2L_2L_3\left(\frac{m_2m_3}{\bar{x}^7}+\frac{m_2^2m_3}{\bar{x}^9}\right)$$

$$-2L_2L_4\left(\frac{m_5-m_2m_3}{\bar{x}^7}+\frac{m_3m_4}{\bar{x}^9}\right)+2L_3L_4\left(\frac{m_2^2m_4}{\bar{x}^{10}}+\frac{m_2m_4}{\bar{x}^8}\right)$$

$$+2\left(1-\frac{n\bar{x}}{n-m}\right)\left(\frac{L_1m_2}{\bar{x}^4}-\frac{L_2m_3}{\bar{x}^5}+\frac{L_3m_2^2}{\bar{x}^6}+\frac{L_4m_4}{\bar{x}^6}\right)\Big] \qquad (5.2.17)$$

We proposed to evaluate (5.2.17) upto $O(n^{-3})$ only. Since (5.2.17) is valid for values of m for which $\left|\frac{\bar{x}_m-\bar{x}}{\bar{x}}\right|<1$, we may consider the range $(\frac{n}{2}, n-1)$ as a possible range for m. Within this range of m, we first note the order of the coefficients L_1, L_2, L_3 and L_4. Thus from (5.2.8) we have :

Coefficient	$m=\frac{n}{2}$	$m=n-1$
L_1	$\frac{1}{n-1}=O(n^{-1})$	$\frac{1}{(n-1)^2}=O(n^{-2})$
L_2	0	$-\frac{1}{(n-1)^3}=O(n^{-3})$
L_3	$\frac{3(n-2)}{n(n-1)(n-3)}=O(n^{-2})$	0
L_4	$O(n^{-3})$	$O(n^{-4})$

(5.2.18)

From (5.2.18) it follows that in the range ($\frac{n}{2}$, n - 1) of m,

$$L_i^2 \ , \ L_i^2 \frac{n}{(n-m)} , \ \frac{L_i^2 n^2}{(n-m)^2} \text{ are atmost } O(n^{-4}), \ i = 2,3,4 ;$$

$$L_{i_1} L_{i_2} , \ \frac{L_{i_1} L_{i_2} n}{(n-m)} \ \frac{L_{i_1} L_{i_2} n^2}{(n-m)^2} \text{ are atmost } O(n^{-3}) \text{ for }$$

$(i_1 , i_2) = (1,2) , (1,3)$

and atmost $O(n^{-4})$ for $(i_1 , i_2) = (1,4), (2,3), (2,4)$ and $(3,4)$; and

$$L_4 , \ \frac{L_4 n}{(n-m)} \text{ are } O(n^{-3}). \qquad (5.2.19)$$

Now in view of (5.2.19) we obtain from (5.2.17) the result (5.2.16) of the lemma.

<u>Lemma 5.2.5</u> : Under super-population model Δ (5.1.1), when

$\left| \frac{\bar{x} - \bar{X}}{\bar{X}} \right| = \left| \bar{x} - 1 \right| < 1$, for choices of m in the range ($\frac{n}{2}$, n - 1)

$$E\left[\bar{u} + E_n \left\{ \frac{\bar{u}_m}{\bar{x}_m} (1 - \bar{x}_{n-m}) \right\} \right]^2 = \frac{6}{n} \left[1 + \frac{\mu_2}{(n-1)} + \frac{2n}{(n-2)(n-3)} \left(\frac{1}{n-1} + \frac{1}{m} \right) \mu_2^2 \right]$$

$$+ O(n^{-4}) , \qquad \dots (5.2.20)$$

where $\mu_i = E(x-1)^i$ is the i-th central moment of x.

<u>Proof :</u> Using lemma 5.2.4, we have

$$E\left[\bar{u} + E_n \left\{ \frac{\bar{u}_m}{\bar{x}_m} (1 - \bar{x}_{n-m}) \right\} \right]^2 = \frac{6}{n} E\left[\frac{1}{\bar{x}^2} + L_1^2 m_2^2 \left(\frac{1}{\bar{x}^6} - \frac{2n}{n-m} \cdot \frac{1}{\bar{x}^5} + \frac{n^2}{(n-m)^2 \bar{x}^4} \right) \right.$$

$$\left. + L_1^2 m_2 \left(\frac{1}{\bar{x}^4} - \frac{2n}{(n-m)\bar{x}^3} + \frac{n^2}{(n-m)^2 \bar{x}^2} \right) + 2L_1 m_2 \left(\frac{1}{\bar{x}^4} - \frac{n}{(n-m)\bar{x}^3} \right) \right.$$

contd...

$$-2L_2 m_3 \left(\frac{1}{\bar{x}^5} - \frac{n}{n-m} \cdot \frac{1}{\bar{x}^4}\right) + 2L_3 m_2^2 \left(\frac{1}{\bar{x}^6} - \frac{n}{n-m} \cdot \frac{1}{\bar{x}^5}\right) \Big] + O(n^{-4}).$$

(5.2.21)

From (5.2.18) it may be noted that

$$L_1 = O(n^{-1}) = \frac{L_1 n}{(n-m)},\ L_1^2 = O(n^{-2}) = \frac{L_1^2\ n}{(n-m)},\ \frac{L_1^2 n^2}{(n-m)^2} = O(n^{-2}),$$

$$L_2 = O(n^{-3}),\ \frac{L_2 n}{(n-m)} = O(n^{-2}),\ L_3 = O(n^{-2}) = \frac{L_3 n}{(n-m)}.$$

(5.2.22)

Assuming that $\left|\frac{\bar{x} - \bar{X}}{\bar{X}}\right| = \left|\bar{x} - 1\right| < 1$, we have

$$\frac{1}{\bar{x}^i} = 1 - \binom{i}{1} m_1 + \binom{i+1}{2} m_1^2 - \binom{i+2}{3} m_1^3 + \binom{i+3}{4} m_1^4 - \binom{i+4}{5} m_1^5$$

$$+ \binom{i+5}{6} m_1^6 + \ldots \qquad (5.2.23)$$

where $m_1 = (\bar{x} - 1)$.

Also from Sukhatme's (1944) results we note that

$$E(m_1^4) = O(n^{-2}),\ E(m_1^5) = O(n^{-3}),\ E(m_2^2 m_1) = O(n^{-1}),$$

$$E(m_2\ m_1) = O(n^{-1}),\ E(m_2 m_1^2) = O(n^{-1}),\ E(m_2 m_1^3) = O(n^{-2})$$

and $E(m_3\ m_1) = O(n^{-1})$. (5.2.24)

Now substituting the expansion (5.2.23) for $\frac{1}{\bar{x}^i}$, ($i = 1, 2, \ldots, 6$), in (5.2.21), and using (5.2.22) and (5.2.24) so as to neglect the contribution of any term which is atmost of order n^{-4}, we obtain

$$E\left[\bar{u} + E_n\left\{\frac{\bar{u}_m}{\bar{x}_m}(1-\bar{x}_{n-m})\right\}\right]^2 = \frac{6}{n} E\left[(1-2m_1+3m_1^2-4m_1^3+5m_1^4)\right.$$

$$+\frac{L_1^2 m^2}{(n-m)^2}(m_2^2+m_2) + 2L_1 m_2\left\{(1-4m_1+10m_1^2) - \frac{n}{(n-m)}(1-3m_1+6m_1^2)\right\}$$

$$\left. + \frac{2L_2 m}{(n-m)} \cdot m_3 - \frac{2L_3 m}{(n-m)} m_2^2\right] + O(n^{-4}),$$

$$= \frac{6}{n} E\left[(1-2m_1+3m_1^2-4m_1^3+5m_1^4) + \left\{\frac{L_1^2 m^2}{(n-m)^2} - \frac{2L_3 m}{(n-m)}\right\} m_2^2\right.$$

$$+\left\{\frac{L_1^2 m^2}{(n-m)^2} - \frac{2L_1 m}{(n-m)}\right\} m_2 - \frac{2(n-4m)}{(n-m)} L_1 m_1 m_2$$

$$\left. + \frac{2(4n-10m)}{(n-m)} L_1 m_2 m_1^2 + \frac{2L_2 m}{(n-m)} \cdot m_3\right] + O(n^{-4}), \quad (5.2.25)$$

To evaluate the expectation in (5.2.25) we again make use of Sukhatme's (1944) results and get

$$E\left[\bar{u} + E_n\left\{\frac{\bar{u}_m}{\bar{x}_m}(1-\bar{x}_{n-m})\right\}\right]^2 = \frac{6}{n}\left[\left(1+\frac{3\mu_2}{n} - \frac{4\mu_3}{n^2} + \frac{15\mu_2^2}{n^2}\right)\right.$$

$$+\left\{\frac{L_1^2 m^2}{(n-m)^2} - \frac{2L_3 m}{(n-m)}\right\}\frac{(n-1)}{n}\mu_2^2 + \left\{\frac{L_1^2 m^2}{(n-m)^2} - \frac{2L_1 m}{(n-m)}\right\}\frac{(n-1)}{n}\mu_2$$

$$-2\frac{(n-4m)}{(n-m)} L_1 \cdot \frac{(n-1)}{n^2}\mu_3 + 2\frac{(4n-10m)}{(n-m)} L_1 \cdot \frac{(n-1)}{n^2}\mu_2^2$$

$$\left. + \frac{2L_2 m}{(n-m)} \frac{(n-1)(n-2)}{n^2} \mu_3\right] + O(n^{-4}), \quad (5.2.26)$$

where $\mu_i = E(x-1)^i$.

Now using the definitions of L_1, L_2 and L_3 given in (5.2.8), we have within the bracket $\frac{\delta}{n}$ [.] of (5.2.26)

$$\text{Coefficient of } \mu_2 = \frac{3}{n} + \left[\frac{L_1^2 m^2}{(n-m)^2} - \frac{2L_1 m}{(n-m)} \right] \frac{(n-1)}{n} = \frac{3}{n} + \left[\frac{1}{(n-1)^2} - \frac{2}{(n-1)} \right] \frac{n-1}{n}$$

$$= \frac{3}{n} + \frac{(3-2n)}{n(n-1)} = \frac{1}{(n-1)} \qquad (5.2.27)$$

$$\text{Coefficient of } \mu_3 = -\frac{4}{n^2} - 2\frac{(n-4m)}{(n-m)} \cdot \frac{L_1(n-1)}{n^2} + \frac{2L_2 m(n-1)(n-2)}{(n-m)n^2}$$

$$= -\frac{4}{n^2} - \frac{2(n-4m)}{mn^2} + \frac{2(n-2m)}{mn^2} = O, \ldots (5.2.28)$$

$$\text{Coefficient of } \mu_2^2 = \frac{15}{n^2} + \left[\frac{L_1^2 m^2}{(n-m)^2} - \frac{2L_3 m}{(n-m)} \right] \frac{(n-1)}{n} + \frac{2(4n-10m)L_1}{(n-m)} \cdot \frac{(n-1)}{n^2}$$

$$= \frac{15}{n^2} + \frac{1}{(n-1)n} - \frac{6(n-m-1)(m-1)}{(n-2)(n-3)m^2} + \frac{2(4n-10m)}{mn^2}$$

$$= \frac{2n}{(n-2)(n-3)} \left(\frac{1}{n-1} + \frac{1}{m} \right) + O(n^{-3}) \ldots (5.2.29)$$

Now using (5.2.27) to (5.2.29) in (5.2.26), we finally obtain the result (5.2.20).

emark (c) : It may be noted that with the above lemma the evaluation of e first term in the expected variance of T_m^* (5.2.1) is completed. We proceed to evaluate the second term in (5.2.1).

Lemma 5.2.6 : When the choice of m is such that $\left|\frac{\bar{x}_m - \bar{x}}{\bar{x}}\right| < 1$,

$$E_n\left(\frac{1-\bar{x}_{n-m}}{\bar{x}_m}\right) \doteq \left[\left(\frac{1}{\bar{x}} - 1\right) + \left(1 - \frac{n\bar{x}}{n-m}\right)\left(\frac{L_1 m_2}{\bar{x}^3} - \frac{L_2 m_3}{\bar{x}^4} + \frac{L_3 m_2^2}{\bar{x}^5} + \frac{L_4 m_4}{\bar{x}^5}\right)\right]$$

(5.2.30)

Proof: We have $E_n\left(\frac{1-\bar{x}_{n-m}}{\bar{x}_m}\right) = E_n\left(\frac{1}{\bar{x}_m} - \frac{n\bar{x} - m\bar{x}_m}{(n-m)\bar{x}_m}\right)$

$$= -\frac{m}{(n-m)} + \left(1 - \frac{n\bar{x}}{n-m}\right) E_n\left(\frac{1}{\bar{x}_m}\right).$$

(5.2.31)

For values of m for which $\left|\frac{\bar{x}_m - \bar{x}}{\bar{x}}\right| < 1$, we now use in (5.2.31) an approximation for $E_n\left(\frac{1}{\bar{x}_m}\right)$ similar to the approximation used for $E_n\left(\frac{u_m}{\bar{x}_m}\right)$ in lemma 5.2.2. This gives the result (5.2.30).

Lemma 5.2.7 : When $|\bar{x} - 1| < 1$, for choices of m in the range $\left(\frac{n}{2}, n-1\right)$

$$a^2 E\left[E_n^2\left(\frac{1-\bar{x}_{n-m}}{\bar{x}_m}\right)\right] = a^2\left[\frac{\mu_2}{n} + \frac{2\mu_2^2}{n}\left(\frac{1}{n-1} + \frac{1}{m}\right) - \frac{2\mu_2\mu_3}{m(n-2)}\left(\frac{1}{m} + \frac{6(n-2m)}{(n-1)(n-3)}\right)\right.$$
$$\left. + \frac{\mu_2^3}{m(n-3)}\left(\frac{7}{m} + \frac{6(m+2n)}{n(n-1)}\right)\right] + O(n^{-4}). \quad \ldots \ (5.2.32)$$

Proof: Assuming that lemma 5.2.6 holds when m is in the range $\left(\frac{n}{2}, n-1\right)$, we have from (5.2.30)

$$^2E\left[E_n^2\left(\frac{1-\bar{x}_{n-m}}{\bar{x}_m}\right)\right] \doteq a^2 E\left[\left(\frac{1}{\bar{x}}-1\right)+\left(1-\frac{n\bar{x}}{n-m}\right)\left(\frac{L_1 m_2}{\bar{x}^3}-\frac{L_2 m_3}{\bar{x}^4}+\frac{L_3 m_2^2}{\bar{x}^5}+\frac{L_4 m_4}{\bar{x}^5}\right)\right]^2. \quad (5.2.33)$$

In (5.2.33) we now use the orders of magnitude given in (5.2.19) and obtain

$$^2E\left[E_n^2\left(\frac{1-\bar{x}_{n-m}}{\bar{x}_m}\right)\right] = a^2 E\left[\left(1-\frac{2}{\bar{x}}+\frac{1}{\bar{x}^2}\right)+\left(1-\frac{2n\bar{x}}{(n-m)}+\frac{n^2\bar{x}^2}{(n-m)^2}\right)\left(\frac{L_1^2 m_2^2}{\bar{x}^6}-\frac{2L_1L_2 m_2 m_3}{\bar{x}^7}+\frac{2L_1L_3 m_2^3}{\bar{x}^8}\right)+2\left(\frac{1}{\bar{x}}+\frac{n\bar{x}}{(n-m)}-\frac{(2n-m)}{(n-m)}\right)\left(\frac{L_1 m_2}{\bar{x}^3}-\frac{L_2 m_3}{\bar{x}^4}+\frac{L_3 m_2^2}{\bar{x}^5}+\frac{L_4 m_4}{\bar{x}^5}\right)\right]+O(n^{-4}).$$

(5.2.34)

In addition to the orders of magnitude noted in (5.2.19) and (5.2.22), we further observe that in the range $(\frac{n}{2}, n-1)$ of m ,

$$\frac{(2n-m)}{(n-m)} = O(n^{-1}),\quad \frac{L_2(2n-m)}{(n-m)} = O(n^{-2}) = \frac{L_3(2n-m)}{(n-m)},\ \text{and}\ \frac{L_4(2n-m)}{(n-m)} = O(n^{-3}).$$

(5.2.35)

Also from Sukhatme (1944) we have

$$(m_1^7) = O(n^{-4}),\ E(m_2^2 m_1^3) = O(n^{-2}),\ E(m_2 m_3 m_1) = O(n^{-1}),\ E(m_2^3 m_1) = O(n^{-1}),$$

$$(m_2 m_1^5) = O(n^{-3}),\ E(m_3 m_1^3) = O(n^{-2}),\ \text{and}\ E(m_4 m_1) = O(n^{-1}).$$

(5.2.36)

Now we substitute the series expansion (5.2.23) for $\frac{1}{\bar{x}_i}$ in (5.2.34) and use the orders of magnitude (5.2.19), (5.2.22), (5.2.35) and (5.2.36) to neglect the contribution of any term which is atmost of order n^{-4}. Thus we obtain

$$a^2 E\left[E_n^2 \left(\frac{1-\bar{x}_{n-m}}{\bar{x}_m} \right) \right] = a^2 E\Big[1 - 2(1 - m_1 + m_1^2 - m_1^3 + m_1^4 - m_1^5 + m_1^6)$$

$$+ (1 - 2m_1 + 3m_1^2 - 4m_1^3 + 5m_1^4 - 6m_1^5 + 7m_1^6) + L_1^2 m_2^2 \Big\{ (1 - 6m_1 + 21m_1^2)$$

$$- \frac{2n}{(n-m)} (1 - 5m_1 + 15m_1^2) + \frac{n^2}{(n-m)^2} (1 - 4m_1 + 10m_1^2) \Big\}$$

$$- \frac{2L_1 L_2 m^2}{(n-m)^2} \cdot m_2 m_3 + \frac{2L_1 L_3 m^2}{(n-m)^2} m_2^3 + 2L_1 m_2 \Big\{ (1 - 4m_1 + 10m_1^2$$

$$- 20m_1^3 + 35m_1^4) + \frac{n}{(n-m)} (1 - 2m_1 + 3m_1^2 - 4m_1^3 + 5m_1^4)$$

$$- \frac{(2n-m)}{(n-m)} (1 - 3m_1 + 6m_1^2 - 10m_1^3 + 15m_1^4) \Big\} - 2L_2 m_3 \Big\{ (1 - 5m_1 + 15m_1^2)$$

$$+ \frac{n}{(n-m)} (1 - 3m_1 + 6m_1^2) - \frac{(2n-m)}{(n-m)} (1 - 4m_1 + 10m_1^2) \Big\}$$

$$+ 2L_3 m_2^2 \Big\{ (1 - 6m_1 + 21m_1^2) + \frac{n}{(n-m)} (1 - 4m_1 + 10m_1^2)$$

$$+ \frac{(2n-m)}{(n-m)} (1 - 5m_1 + 15m_1^2) \Big\} + 2L_4 m_4 \left(1 + \frac{n}{(n-m)} - \frac{(2n-m)}{(n-m)} \right) \Big].$$

$$+ O(n^{-4}). \qquad (5.2.37)$$

On simplifying (5.2.37) we have

$$a^2 E\left[E_n^2 \left(\frac{1-\bar{x}_{n-m}}{\bar{x}_m} \right) \right] = a^2 E\Big[(m_1^2 - 2m_1^3 + 3m_1^4 - 4m_1^5 + 5m_1^6)$$

$$+ L_1^2 \Big\{ \frac{m^2}{(n-m)^2} \cdot m_2^2 + \frac{2m(n-3m)}{(n-m)^2} \cdot m_1 m_2^2 + \left(\frac{10m^2}{(n-m)^2} + \frac{(n-11m)}{(n-m)} \right) m_1^2 m_2^2 \Big\}$$

contd...

$$- \frac{2L_1L_2\, m^2}{(n-m)^2} \cdot m_2 m_3 + \frac{2L_1L_3\, m^2}{(n-m)^2} \cdot m_2^3 + 2L_1 \left\{ \frac{m}{(n-m)} \cdot m_1 m_2 \right.$$

$$\left. + \frac{(n-4m)}{(n-m)} \cdot m_1^2 m_2 - \frac{(4n-10m)}{(n-m)} \cdot m_1^3 m_2 + \frac{(10n-20m)}{(n-m)} \cdot m_1^4 m_2 \right\}$$

$$- 2L_2 \left\{ \frac{m}{(n-m)} \cdot m_1 m_3 + \frac{(n-5m)}{(n-m)} \cdot m_1^2 m_3 \right\} + 2L_3 \left\{ \frac{m}{(n-m)} \cdot m_1 m_2^2 \right.$$

$$\left. + \frac{(n-6m)}{(n-m)} \cdot m_1^2 m_2^2 \right\} \Big] + O(n^{-4}). \quad \ldots \quad (5.2.38)$$

Now applying Sukhatme's (1944) results we obtain

$$a^2 E\left[\bar{E}_n^2 \left(\frac{1-\bar{x}_{n-m}}{\bar{x}_m} \right) \right] = a^2 \Big[\left\{ \frac{\mu_2}{n} - \frac{2\mu_3}{n^2} + \frac{3\mu_4}{n^3} + \frac{9(n-1)\mu_2^2}{n^3} - \frac{40(n-1)\mu_2\mu_3}{n^4} \right.$$

$$\left. + \frac{75(n-1)(n-2)\mu_2^3}{n^5} \right\} + L_1^2 \left\{ \frac{m^2}{(n-m)^2} \left(\frac{(n-1)^2}{n^3} \mu_4 + \frac{(n-1)(n^2-2n+3)}{n^3} \mu_2^2 \right) \right.$$

$$\left. + \frac{2m(n-3m)}{(n-m)^2} \cdot \frac{2(n-1)}{n^2} \mu_2 \mu_3 + \left(\frac{10m^2}{(n-m)^2} + \frac{(n-11m)}{(n-m)} \right) \frac{(n-1)(n-2)}{n^3} \mu_2^3 \right\}$$

$$- \frac{2L_1L_2\, m^2}{(n-m)^2} \cdot \frac{(n-1)}{n} \mu_2 \mu_3 + \frac{2L_1L_3\, m^2}{(n-m)^2} \cdot \frac{(n-1)(n-2)}{n^2} \mu_2^3 + 2L_1 \left\{ \frac{m(n-1)}{(n-m)n^2} \mu_3 \right.$$

$$+ \frac{(n-4m)}{(n-m)} \left(\frac{(n-1)}{n^3} \mu_4 + \frac{(n-1)(n-3)}{n^3} \mu_2^2 \right) - \frac{(4n-10m)}{(n-m)} \cdot \frac{4(n-1)}{n^3} \mu_2 \mu_3$$

$$\left. + \frac{(10n-20m)}{(n-m)} \cdot \frac{3(n-1)(n-2)}{n^4} \mu_2^3 \right\} - 2L_2 \left\{ \frac{m(n-1)(n-2)}{(n-m)\,n^3} (\mu_4 - 3\mu_2^2) \right.$$

$$\left. + \frac{(n-5m)(n-1)}{(n-m)\,n^2} \mu_2 \mu_3 \right\} + 2L_3 \left\{ \frac{m}{(n-m)} \cdot \frac{2(n-1)}{n^2} \mu_2 \mu_3 + \frac{(n-6m)(n-1)(n-2)}{(n-m)\,n^3} \mu_2^3 \right\} \Big]$$

$$+ O(n^{-4}). \quad (5.2.39)$$

Using the definitions (5.2.8) of L_1 in the expression (5.2.39), we have within the bracket $n^2[\cdot]$

$$\text{Coefficient of } \mu_3 = -\frac{2}{n^2} + \frac{2L_1 m(n-1)}{(n-m)n^2} - -\frac{2}{n^2} + \frac{2}{n^2} = 0,$$

$$\text{Coefficient of } \mu_2^2 = \frac{9(n-1)}{n^3} + \frac{L_1^2 m^2 (n-1)(n^2-2n+3)}{(n-m)^2 n^3} + \frac{2L_1(n-4m)(n-1)(n-3)}{(n-m)n^3}$$

$$+ \frac{6L_2 m(n-1)(n-2)}{(n-m)n^3}$$

$$= \frac{9(n-1)}{n^3} + \frac{(n^2-2n+3)}{(n-1)n^3} + \frac{2(n-4m)(n-3)}{mn^3} + \frac{6(n-2m)}{mn^3}$$

$$\frac{2}{n}\left(\frac{1}{n-1} + \frac{1}{m}\right), \qquad (5.2.41)$$

$$\text{Coefficient of } \mu_4 = \frac{3}{n^3} + \frac{L_1^2 m^2}{(n-m)^2} \cdot \frac{(n-1)^2}{n^3} + \frac{2L_1(n-4m)(n-1)}{(n-m)n^3} - \frac{2L_2 m(n-1)(n-2)}{(n-m)n^3}$$

$$= \frac{3}{n^3} + \frac{1}{n^3} + \frac{2(n-4m)}{mn^3} - \frac{2(n-2m)}{mn^3}$$

$$= 0, \qquad (5.2.42)$$

$$\text{Coefficient of } \mu_2\mu_3 = -\frac{40(n-1)}{n^4} - \frac{4L_1^2 m(n-3m)(n-1)}{(n-m)^2 n^2} - \frac{2L_1L_2 m^2(n-1)}{(n-m)^2 n}$$

$$-\frac{8L_1(4n-10m)(n-1)}{(n-m)n^3} - \frac{2L_2(n-5m)(n-1)}{(n-m)n^2} + \frac{4L_3(n-1)m}{(n-m)n^2}$$

contd..

$$- \frac{40(n-1)}{n^4} - \frac{4(n-3m)}{m(n-1)n^2} - \frac{2(n-2m)}{mn(n-1)(n-2)} - \frac{8(4n-10m)}{mn^3}$$

$$- \frac{2(n-5m)(n-2m)}{m^2n^2(n-2)} + \frac{12(n-m-1)(m-1)}{m^2 n(n-2)(n-3)}$$

$$= \frac{2}{m(n-2)} \left[\frac{1}{m} + \frac{6(n-2m)}{(n-1)(n-3)} \right] + O(n^{-4}), \qquad (5.2.43)$$

and finally Coefficient of μ_2^3

$$= \frac{75(n-1)(n-2)}{n^5} + \frac{L_1^2(n^2-12mn+21m^2)}{(n-m)^2} \cdot \frac{(n-1)(n-2)}{n^3}$$

$$+ \frac{2L_1L_3\, m^2(n-1)(n-2)}{(n-m)^2 n^2} + \frac{60\,L_1(n-2m)(n-1)(n-2)}{(n-m)\; n^4} + \frac{2L_3(n-6m)(n-1)(n-2)}{(n-m)\,n^3}$$

$$= \frac{75(n-1)(n-2)}{n^5} + \frac{(n^2-12mn+21m^2)(n-2)}{m^2\,n^3(n-1)} + \frac{6(n-m-1)(m-1)}{m^2n(n-1)(n-3)}$$

$$+ \frac{60(n-2m)(n-2)}{m\,n^4} + \frac{6(n-6m)(n-m-1)(m-1)}{m^3n^2(n-3)}$$

$$= \frac{1}{m(n-3)} \left[\frac{7}{m} + \frac{6(m+2n)}{n(n-1)} \right] + O(n^{-4}). \qquad (5.2.44)$$

Hence using (5.2.40) to (5.2.44) in (5.2.39) we get

$$a^2E\left[\Xi_n^2 \left(\frac{1-\bar{x}_{n-m}}{\bar{x}_m} \right)^4 \right] = a^2\left[\frac{\mu_2}{n} + \frac{2}{n}\left(\frac{1}{n-1} + \frac{1}{m} \right)\mu_2^2 - \frac{2\mu_2\mu_3}{m(n-2)}\left(\frac{1}{m} + \frac{6(n-2m)}{(n-1)(n-3)} \right) \right.$$

$$\left. + \frac{\mu_2^3}{m(n-3)}\left(\frac{7}{m} + \frac{6(m+2n)}{n(n-1)} \right) \right] + O(n^{-4}).$$

is completes the proof of the lemma.

Theorem 5.2.1 : Let the units of measurement of the variate x be so chosen that $\bar{X} = 1$. Let $|\bar{x} - 1| < 1$ and m be in the range $(\frac{n}{2}, n-1)$. Then under super-population model Δ (5.1.1) the expected variance of the unbiased ratio type estimator T^*_m (4.1.2') is given upto order n^{-3} by.

$$V(T^*_m) = \frac{\delta}{n}\left[1 + \frac{\mu_2}{(n-1)} + \frac{2n\,\mu_2^2}{(n-2)(n-3)}\left(\frac{1}{n-1} + \frac{1}{m}\right)\right] + a^2\left[\frac{\mu_2}{n} + \frac{2\mu_2^2}{n}\left(\frac{1}{n-1} + \frac{1}{m}\right)\right.$$
$$\left. - \frac{2\mu_2\,\mu_3}{m(n-2)}\left(\frac{1}{m} + \frac{6(n-2m)}{(n-1)(n-3)}\right) + \frac{\mu_2^3}{m(n-3)}\left(\frac{7}{m} + \frac{6(m+2n)}{n(n-1)}\right)\right], \tag{5.2.45}$$

where $\mu_i = E(x-1)^i$ is the i-th central moment of x.

Proof : Follows from lemmas 5.2.1, 5.2.5, and 5.2.7

5.3. Optimum choice of m for T^*_m and $\bar{T}'_{n-m}$ (5.3.5) under super-population model Δ (5.1.1).

In the previous section we have obtained the expected variance of T^*_m for choices of m in the range $(\frac{n}{2}, n-1)$ under the conditions stated in theorem 5.2.1. We now study the variance expression (5.2.45) with a view to determining the optimum choice of m in the range $(\frac{n}{2}, n-1)$.

Theorem 5.3.1 : Under the conditions stated in theorem 5.2.1, when the expected variance of the unbiased ratio type estimator T^*_m is considered only upto order n^{-2}, the optimum choice of m in the range $(\frac{n}{2}, n-1)$ is n - 1 and

$$V(T^*_{n-1}) = \frac{\delta}{n}\left[1 + \frac{\mu_2}{(n-1)}\right] + a^2\left[\frac{\mu_2}{n} + \frac{4\mu_2^2}{n(n-1)}\right]. \tag{5.3.1}$$

Proof : From theorem 5.2.1, upto order n^{-2},

$$V(T_m^*) = \frac{\delta}{n}\left[1 + \frac{\mu_2}{(n-1)}\right] + a^2\left[\frac{\mu_2}{n} + \frac{2\mu_2^2}{n}\left(\frac{1}{n-1} + \frac{1}{m}\right)\right], \quad (5.3.2)$$

which is clearly a decreasing function of m. Hence the result follows.

Theorem 5.3.2 : Under the conditions stated in theorem 5.2.1, when the expected variance of the unbiased ratio type estimator T_m^* is considered upto order n^{-3}, the optimum choice of m in the range $(\frac{n}{2}, n-1)$ is $n-1$, provided the coefficient of variation of $x \geq$ coefficient of skewness of x. In perticular, the result holds for all negatively skewed or symmetric distributions of x. For the optimum estimator T_{n-1}^*

$$V(T_{n-1}^*) = \frac{\delta}{n}\left[1 + \frac{\mu_2}{(n-1)} + \frac{4n\,\mu_2^2}{(n-1)(n-2)(n-3)}\right] + a^2\left[\frac{\mu_2}{n} + \frac{4\mu_2^2}{n(n-1)} + \frac{2(5n-9)\mu_2\mu_3}{(n-1)^2(n-2)(n-3)} + \frac{(25n-6)\mu_2^3}{n(n-1)^2(n-3)}\right]. \quad (5.3.3)$$

Proof: Let us re-write (5.2.45) as

$$V(T_m^*) = \frac{\delta}{n}\left[1 + \frac{\mu_2}{(n-1)} + \frac{2n\,\mu_2^2}{(n-2)(n-3)}\left(\frac{1}{n-1} + \frac{1}{m}\right)\right] + a^2\left[\frac{\mu_2}{n} + \frac{2\mu_2^2}{n}\left(\frac{1}{n-1} + \frac{1}{m}\right) + \frac{24\,\mu_2\mu_3}{(n-1)(n-2)(n-3)} + \frac{6\,\mu_2^3}{n(n-1)(n-3)} + \frac{2}{m(n-2)}\left(\frac{1}{m} + \frac{6n}{(n-1)(n-3)}\right)(\mu_2^3 - \mu_2\mu_3) + \left\{\frac{1}{m^2}\left(\frac{7}{n-3} - \frac{2}{n-2}\right) - \frac{24}{m(n-1)(n-2)(n-3)}\right\}\mu_2^3\right]. \quad (5.3.4)$$

In (5.3.4) we note that the coefficients of μ_2^2 and $(\mu_2^3 - \mu_2\mu_3)$ are positive decreasing functions of m. Further, the coefficient of μ_2^3 involving m, vi

$$f(m) = \frac{1}{m^2}\left(\frac{7}{n-3} - \frac{2}{n-2}\right) - \frac{24}{m(n-1)(n-2)(n-3)} = \frac{5n^2-13n-24m+8}{m^2(n-1)(n-2)(n-3)} > 0,$$

for $n \geq 8$ when $\frac{n}{2} \leq m \leq n-1$.

Also f (m) is then decreasing function of m, since

$$\frac{d\,f(m)}{dm} = -\frac{2(5n^2-13n-12m+8)}{m^3(n-1)(n-2)(n-3)} < 0 .$$

We may take $n \geq 8$ as we have already assumed $| \bar{x} - 1 | < 1$, which is likely to be valid only in fairly large samples. Hence from (5.3.4) it follows that $V(T^*_{n-1}) \leq V(T^*_m)$ provided $\mu_2^3 - \mu_2\mu_3 \geq 0$.

It may be recalled that the units of measurement are so chosen that $\bar{X} = E(x) = 1$, which is equivalent to working with the variate $\frac{x}{\bar{X}}$ instead of x. Hence when $\bar{X} = 1$, $\mu_2^3 \geq \mu_2\mu_3$ is equivalent to the condition

$$\frac{\sqrt{\mu_2}}{\bar{X}} \geq \frac{\mu_3}{\mu_2^{3/2}}$$

Thus, when the coefficient of variation of $x \geq$ coefficient of skewness of x, we have $V(T^*_{n-1}) \leq V(T^*_m)$. In particular, the condition on the coefficient of variation is trivially satisfied for all negatively skewed or symmetric distributions of x. Alternately, in this case since $\mu_3 \leq 0$, one may observe from (5.2.45) itself that $V(T^*_m)$ is a decreasing function of m. Finally, we obtain (5.3.3) from (5.2.45) for $m = n-1$. This completes the proof of the theorem.

Remark (a) : The condition: coefficient of variation of $x \geq$ coefficient of skewness of x, is only a sufficient condition. Even when x has a positively skewed distribution with the coefficient of variation less than the skewness coefficient, it may be possible that T^*_{n-1} is the optimum choice.

Remark (b): Since normal distribution is symmetric, theorem 5.3.2. holds under Durbin's (1959) population model 1, which is equivalent to (5.1.1) with the further assumption that x is a normal variate.

Remark (c): It is interesting to note that the results in theorems 5.3.1 nd 5.3.2 may be used to derive similar results for the class of unbiased ratio-type estimators $\overline{T}'_{n-m} = \overline{t}'_{n-m}\ \bar{X}$, where $\overline{t}'_{n-m}$ is defined by 2.4.4). These results are contained in the following theorems.

mma.5.3.1: In the class of estimators $\overline{T}'_{n-m}$, $V(\overline{T}'_{n-m}) \geqslant V(\overline{T}'_{n-1})$ $V(T^*_{n-m}) \geqslant V(T^*_{n-1})$ for $\frac{n}{2} \leqslant (n-m) \leqslant n-1$.

oof: To construct $\overline{T}'_{n-m}$ as stated in section 2.4, we split the sample size n at random into g groups each of size m ($n = mg$). Correspond-g to a group $S_{n-m}(j)$, obtained by exclusion of the j-th ($j = 1, 2, \ldots, g$) oup of size m from the total sample, we construct an estimator $T_{n-m}(j)$ the form (4.1.1). Then

$$\overline{T}'_{n-m} = \frac{1}{g} \sum_{j=1}^{g} T_{n-m}(j) \qquad (5.3.5)$$

Evidently $2 \leqslant g \leqslant n$ and hence $1 \leqslant m \leqslant n/2$. Thus for $\overline{T}'_{n-m}$ have $\frac{n}{2} \leqslant (n-m) \leqslant n-1$. As observed in section 2.4, average of $_{-m}$ over all possible splits of the sample into groups of size m is ual to T^*_{n-m} and hence $V(\bar{T}'_{n-m}) \geqslant V(T^*_{n-m})$. Also $\bar{T}'_{n-1} = T^*_{n-1}$.

Now, if $V(T^*_{n-m}) \geqslant V(T^*_{n-1})$ for $\frac{n}{2} \leqslant (n-m) \leqslant n-1$, it ollows that $V(\bar{T}'_{n-m}) \geqslant V(T^*_{n-m}) \geqslant V(T^*_{n-1}) = V(\bar{T}'_{n-1})$.

Hence the lemma is proved.

Theorem 5.3.3: Under super-population model $\triangle$(5.1.1), if $|\bar{x}-1| < 1$, $'_{n-1}$ is the optimum choice in the class of unbiased ratio type estimators $'_{n-m}$ $(1 \leqslant m \leqslant \frac{n}{2})$ when the variance is considered upto order n^{-2}.

Proof: We apply theorem 5.3.1 and lemma 5.3.1.

Theorem 5.3.4: Under super-population model $\triangle$ (5.1.1), if $|\bar{x}-1| < 1$ nd coefficient of variation of $x \geqslant$ coefficient of skewness of x, $\bar{T}'_{n-1}$ is the optimum choice in the class of unbiased ratio type estimators $\bar{T}'_{n-m}$ $(1 \leqslant m \leqslant \frac{n}{2})$ when the variance is considered upto order n^{-3} In particular, the result holds for negatively skewed or symmetric istributions of x.

Proof: We apply theorem 5.3.2 and lemma 5.3.1.

Remark (d): The result of theorem 5.3.4 holds under Durbin's population del 1, since normal distribution is symmetric. This result under rbin's model 1 is due to Rao (1967). In deriving the result, Rao has ssumed $\left| \frac{\bar{x}_{n-m} - \bar{X}}{\bar{X}} \right| = |\bar{x}_{n-m} - 1| < 1$ for $\frac{n}{2} \leqslant (n-m) \leqslant n-1$. In r derivation the assumptions involved are: $\left| \frac{\bar{x}_{n-m} - \bar{x}}{\bar{x}} \right| < 1$ for $\leqslant (n-m) \leqslant n-1$, and $|\bar{x}-1| < 1$. These are weaker than Rao's sumption.

Remark (e) : From (5.3.3), under Durbin's model 1, we have

$$V(T^*_{n-1}) = \frac{\delta}{n}\left[1 + \frac{\mu_2}{n-1} + \frac{4n\ \mu_2^2}{(n-1)(n-2)(n-3)}\right] + a^2\left[\frac{\mu_2}{n} + \frac{4\ \mu_2^2}{n(n-1)} + \frac{(25n-6)\mu_2^3}{n(n-1)^2\ (n-3)}\right]. \qquad (5.3.6)$$

Rao's (1967) result also gives $V(T^*_{n-1})$ under Durbin's model 1. His notation has, however, been slightly different. He has assumed $E(u_j^2 / x_j) = n\delta$ and $V(\bar{x}) = h$ where both δ and h are constants of order n^{-1} His result, in terms of our notation, may be given as

$$V(T^*_{n-1}) = \frac{\delta}{n}\left[1 + \frac{\mu_2}{n-1} + \frac{(4n-5)}{(n-1)^3}\ \mu_2^2\right] + a^2\left[\frac{\mu_2}{n} + \frac{4\mu_2^2}{n(n-1)} + \frac{(25n-29)}{n(n-1)^3}\ ^3\mu_2\right]. \qquad (5.3.7)$$

It may be noted that (5.3.6) and (5.3.7) are same except for some negligible differences in the terms of order n^{-3}

5.4. Comparison of the optimum choice estimator T^*_{n-1} with the classical ratio estimator $\hat{\bar{Y}}_R$ under super population model $\triangle$ (5.1.1):

In the previous section we have shown in theorems 5.3.1 and 5.3.2 that T^*_{n-1} is more efficient than any other choice of T^*_m for $\frac{n}{2} \leq m \leq n-1$, under super-population model $\triangle$ (5.1.1) and some other conditions. In this section we first obtain the mean square error of the classical biased ratio estimator $\hat{\bar{Y}}_R = \frac{\bar{y}}{\bar{x}}\ \bar{X}$ under the model $\triangle$ (5.1.1) and then compare it with the variance of T^*_{n-1}.

Lemma 5.4.1 : Under super-population model $\triangle$ (5.1.1), when $\bar{X} = 1$ and

$|\bar{x} - 1| < 1$, upto order n^{-3} the mean square error of the classical ratio estimator $\hat{\bar{Y}}_R$ is given by

$$MSE(\hat{\bar{Y}}_R) = \frac{\delta}{n}\left[1 + \frac{3\mu_2}{n} - \frac{4\mu_3}{n^2} + \frac{15\mu_2^2}{n^2}\right] + \alpha^2\left[\frac{\mu_2}{n} - \frac{2\mu_3}{n^2} + \frac{3\mu_4}{n^3} + \frac{9(n-1)}{n^3}\mu_2^2 - \frac{40\mu_2\mu_3}{n^3} + \frac{75\mu_2^3}{n^3}\right]. \quad (5.4.1)$$

Proof : Under the model Δ (5.1.1) we have

$$\hat{\bar{Y}}_R = \frac{\bar{y}}{\bar{x}}\bar{X} = \frac{\bar{y}}{\bar{x}} = \frac{\alpha + \bar{u}}{\bar{x}} + \beta$$

$$(\hat{\bar{Y}}_R - \bar{Y})^2 = \left[\frac{\alpha + \bar{u}}{\bar{x}} + \beta - (\alpha + \beta)\right]^2 = \left[\alpha\left(\frac{1}{\bar{x}} - 1\right) + \frac{\bar{u}}{\bar{x}}\right]^2.$$

and hence

$$MSE(\hat{\bar{Y}}_R) = E(\hat{\bar{Y}}_R - \bar{Y})^2 = \alpha^2 E\left(\frac{1}{\bar{x}^2} - \frac{2}{\bar{x}} + 1\right) + \frac{\delta}{n} E\left(\frac{1}{\bar{x}^2}\right). \quad (5.4.2)$$

Now we assume $|\bar{x} - 1| < 1$ and use the series expansion (5.2.23) for $\frac{1}{\bar{x}^r}$ in (5.4.2). On the basis of Sukhatme's (1944) results, we neglect the terms whose contributions is atmost of order n^{-4}. Then we obtain

$$MSE(\hat{\bar{Y}}_R) = \frac{\delta}{n} E(1 - 2m_1 + 3m_1^2 - 4m_1^3 + 5m_1^4) + \alpha^2 E(m_1^2 - 2m_1^3 + 3m_1^4 - 4m_1^5 + 5m_1^6)$$

$$= \text{expression } (5.4.1).$$

Theorem 5.4.1 : Under super-population model Δ (5.1.1), when $|\bar{x} - 1| < 1$ and the variance or mean square error is considered upto order n^{-2}, the unbiased ratio type estimator T^*_{n-1} is more efficient than the classical biased ratio estimator $\hat{\bar{Y}}_R$ if coefficient of variation of $x \geq \frac{2(n-1)}{(5n-9)}$, coefficien of skewness of x. In particular, the result holds for all negatively skewed or symmetric distributions of x.

Proof : From (5.3.1) and (5.4.1), we have upto order n^{-2}

$$MSE(\hat{Y}_R) - V(T^*_{n-1}) = \frac{8}{n}\left(\frac{3}{n} - \frac{1}{n-1}\right)\mu_2 + \frac{a^2}{n}\left[\left(\frac{9}{n} - \frac{4}{n-1}\right)\mu_2^2 - \frac{2\mu_3}{n}\right]$$

$$= \frac{8(2n-3)\mu_2}{n^2(n-1)} + \frac{a^2}{n^2(n-1)}\left[(5n-9)\mu_2^2 - 2(n-1)\mu_3\right]. \qquad (5.4.3)$$

From (5.4.3) it follows that a sufficient condition for the unbiased estimator T^*_{n-1} to be more efficient than the biased estimator $\hat{Y}_R$ is given by $\mu_2^2 \geqslant \frac{2(n-1)\mu_3}{(5n-9)}$. Since $\bar{X} = 1$, this condition is equivalent to the condition: coefficient of variation of x $\geqslant \frac{2(n-1)}{(5n-9)}$. coefficient of skewness of x. Hence the theorem is proved.

Theorem 5.4.2 : Under super-population model $\triangle$ (5.1.1), when $|\bar{x}-1| < 1$ and the variance or mean square error is considered upto order n^{-3}, the unbiased ratio type estimator T^*_{n-1} is more efficient than the classical biased ratio estimator $\hat{Y}_R$ if coefficient of variation of x $\geqslant$ coefficient of skewness of x. In particular, the result holds for all negatively skewed or symmetric distributions of x.

Proof : From (5.3.3) and (5.4.1), we have

$$MSE(\hat{Y}_R) - V(T^*_{n-1}) = \frac{8}{n}\left[\left(\frac{3}{n} - \frac{1}{n-1}\right)\mu_2 + \frac{4(\mu_2^2 - \mu_3)}{n^2} + \left(\frac{11}{n^2} - \frac{4n}{(n-1)(n-2)(n-3)}\right)\mu_2^2\right]$$

contd...

$$+\alpha^2 \left[\frac{-2(\mu_2^2 - \mu_3)}{n^2} + \left(\frac{7n-9}{n^3} - \frac{4}{n(n-1)} \right) \mu_2^2 + \frac{3\mu_4}{n^3} \right.$$

$$+ \left(-\frac{40}{n^3} + \frac{2(5n-9)}{(n-1)^2(n-2)(n-3)} \right) (\mu_2^3 - \mu_2\mu_3)$$

$$\left. + \left(\frac{35}{n^3} - \frac{2(5n-9)}{(n-1)^2(n-2)(n-3)} - \frac{(25n-6)}{n(n-1)^2(n-3)} \right) \mu_2^3 \right]$$

$$= \frac{\beta}{n} \left[-\frac{(2n-3)}{n(n-1)} \mu_2 + \frac{4(\mu_2^2 - \mu_3)}{n^2} + \frac{7n\,\mu_2^2}{(n-1)(n-2)(n-3)} \right]$$

$$+\alpha^2 \left[-\frac{2(\mu_2^2 - \mu_3)}{n^2} + \frac{(3n-16)\mu_2^2}{n^2(n-1)} + \frac{3\mu_4}{n^3} + \frac{50n(\mu_2^3 - \mu_2\mu_3)}{(n-1)^2(n-2)(n-3)} \right] + O(n^{-4}),$$

(5.4.4)

From (5.4.4) it follows that, when terms of order n^{-4} and less are neglected, T^*_{n-1} is more efficient than $\hat{\bar{Y}}_R$ if $\mu_2^2 \geq \mu_3$. Since $\bar{X} = 1$, this sufficient condition is equivalent to the condition:

coefficient of variation of x $\geq$ coefficient of skewness of x.

Remark (a) : It is interesting to note that, when the variance or mean square error is considered upto order n^{-3}, under the same sufficient conditions coefficient of variation of $x \geq$ coefficient of skewness of x, T^*_{n-1} is the optimum choice for T^*_m [for $\frac{n}{2} \leq m \leq n-1$ (theorem 5.3.2)] and also for $\hat{\bar{T}}'_{n-m}$ [for $1 \leq m \leq n/2$ (theorem 5.3.4)] and is further superior to $\hat{\bar{Y}}_R$ (theorem 5.4.2).

Remark (b) : Since normal distribution is symmetric, theorem 5.4.2 holds under Durbin's population model I. The result under Durbin's model I is due to Rao (1967).

5.5. Expected variance of T_m^* under super-population model $\triangle'$ (5.1.2):

As stated in section 5.1, super-population model $\triangle'$ (5.1.2) involves the conditions of the model $\triangle$ (5.1.1) and the additional condition that x is a gamma variate, say, with parameter h. Under this model, we obtain the expected variance of the unbiased ratio type estimator T_m^* which would be applicable for any sample size n and any choice of m in the range (1, n - 1). In deriving the result we use lemma 5.2.1 and the following lemmas.

Lemma 5.5.1 : Under super-population model $\triangle$ (5.1.1)

$$E\left[\bar{u} + E_n\left\{\frac{\bar{u}_m}{\bar{x}_m}(\bar{X} - \bar{x}_{n-m})\right\}\right]^2 = \frac{\delta}{n} + E\left[E_n^2\left\{\frac{\bar{u}_m}{\bar{x}_m}(\bar{X} - \bar{x}_{n-m})\right\}\right] \tag{5.5.1}$$

Proof : We note, under the model $\triangle$ (5.1.1), $E(\bar{u}^2) = \frac{\delta}{n}$ and

$E_x(\bar{u}_m \bar{u}) = E_x(\bar{u}^2) = \frac{\delta}{n}$ Hence

$$E\left[\bar{u} + E_n\left\{\frac{\bar{u}_m}{\bar{x}_m}(\bar{X} - \bar{x}_{n-m})\right\}\right]^2 = E(\bar{u}^2) + E\left[E_n^2\left\{\frac{\bar{u}_m}{\bar{x}_m}(\bar{X} - \bar{x}_{n-m})\right\}\right]$$

$$+ 2E\left[\bar{u}\, E_n\left\{\frac{\bar{u}_m}{\bar{x}_m}(\bar{X} - \bar{x}_{n-m})\right\}\right]$$

$$= \frac{\delta}{n} + E\left[E_n^2\left\{\frac{\bar{u}_m}{\bar{x}_m}(\bar{X} - \bar{x}_{n-m})\right\}\right] + \frac{2\delta}{n}E\left[E_n\left(\frac{\bar{X} - \bar{x}_{n-m}}{\bar{x}_m}\right)\right] \tag{5.5.2}$$

As we observe $E\left[E_n\left(\frac{\bar{X} - \bar{x}_{n-m}}{\bar{x}_m}\right)\right] = E\left(\frac{\bar{X} - \bar{x}_{n-m}}{\bar{x}_m}\right) = 0,$

since $\bar{x}_m$ and $\bar{x}_{n-m}$ are independently distributed in sampling from an infinite population. Hence from (5.5.2) the lemma follows.

Remark (a) : In what follows, we denote $\bar{u}_m = \bar{u}(Z_m)$, $\bar{x}_m = \bar{x}(Z_m)$ and $\bar{x}_{n-m} = \bar{x}'(Z_m)$.

Lemma 5.5.2 : Under super-population model $\triangle$ (5.1.1)

$$E\left[E_n^2 \left\{ \frac{\bar{u}_m}{\bar{x}_m} (\bar{X} - \bar{x}_{n-m}) \right\} \right] = \frac{\delta}{\binom{n}{m}} \left[\frac{V(x)}{m(n-m)} E\left(\frac{1}{\bar{x}^2(Z_m)} \right) + E\left\{ \sum_{s=0}^{m-1} \frac{s}{m^2} \binom{m}{s}\binom{n-m}{m-s} \frac{\left[\bar{X} - \bar{x}'(Z_m)\right]\left[\bar{X} - \bar{x}'(Z_m^s)\right]}{\bar{x}(Z_m)\,\bar{x}(Z_m^s)} \right\} \right],$$

(5.5.3)

where, for a given sample, Z_m and Z_m^s denote two distinct sub-sets having s ($0 \leq s \leq m-1$) sample elements in common and each containing m sample elements.

Proof : We have $E\left[E_n^2 \left\{ \frac{\bar{u}_m}{\bar{x}_m} (\bar{X} - \bar{x}_{n-m}) \right\} \right] = E\left[\frac{1}{\binom{n}{m}} \sum_{Z_m} \frac{\bar{u}(Z_m)\left\{\bar{X} - \bar{x}'(Z_m)\right\}}{\bar{x}(Z_m)} \right]^2$

$$= \frac{1}{\binom{n}{m}^2} E\left[\sum_{Z_m} \frac{\bar{u}^2(Z_m)\left\{\bar{X} - \bar{x}'(Z_m)\right\}^2}{\bar{x}^2(Z_m)} \right] + \frac{1}{\binom{n}{m}^2} E\left[\sum_{Z_m \neq \tilde{Z}_m} \sum \frac{\bar{u}(Z_m)\left\{\bar{X} - \bar{x}'(Z_m)\right\}}{\bar{x}(Z_m)} \cdot \frac{\bar{u}(\tilde{Z}_m)\left\{\bar{X} - \bar{x}'(\tilde{Z}_m)\right\}}{\bar{x}(\tilde{Z}_m)} \right].$$

(5.5.4)

e first term in (5.5.4) may be re-written as :

$$\frac{1}{\binom{n}{m}} E\left[E_n \left\{ \frac{\bar{u}(Z_m)\left[\bar{X} - \bar{x}'(Z_m)\right]^2}{\bar{x}^2(Z_m)} \right\} \right] = \frac{1}{\binom{n}{m}} E\left[\frac{\bar{u}^2(Z_m)\left\{\bar{X} - \bar{x}'(Z_m)\right\}^2}{\bar{x}^2(Z_m)} \right]$$

$$= \frac{\delta}{\binom{n}{m} m} E\left[\frac{\left\{\bar{X} - \bar{x}'(Z_m)\right\}^2}{\bar{x}^2(Z_m)} \right]$$

$$= \frac{\delta\, V(x)}{\binom{n}{m} m(n-m)} E\left(\frac{1}{\bar{x}^2(Z_m)} \right), \qquad (5.5.5)$$

since $\bar{x}(Z_m)$ and $\bar{x}'(Z_m)$ are independently distributed in sampling from an infinite population.

We now consider the second term in (5.5.4). We note that, when $Z_m \neq \breve{Z}_m$, they may have s ($0 \leq s \leq m-1$) common sample elements. Also, the number of sub-sets $\breve{Z}_m$ having s common elements with any specified Z_m is given by $\binom{m}{s}\binom{n-m}{m-s}$, provided $(n-m) \geq (m-s)$. If $(n-m) < (m-s)$, there is no sub-set $\breve{Z}_m$ having s common elements with Z_m. Let us, therefore, denote by Z_m^s a sub-set $\breve{Z}_m$ having s common elements with a specified Z_m. Given a Z_m, all the $\binom{n}{m} - 1$ sub-sets $\breve{Z}_m$ distinct from Z_m are accounted by Z_m^s, $s = 0, 1, \ldots, m-1$, because of the identity : $\binom{n}{m} - 1 = \sum_{s=0}^{m-1} \binom{m}{s}\binom{n-m}{m-s}$, with the understanding $\binom{n-m}{m-s} = 0$ whenever $(n-m) < (m-s)$.

With the above notation, we may now re-write the second term in (5.5.4) as :

$$\frac{1}{\binom{n}{m}^2} E\left[\sum_{Z_m}^{\binom{n}{m}} \left\{ \frac{\bar{u}(Z_m)\left[\bar{X} - \bar{x}'(Z_m)\right]}{\bar{x}(Z_m)} \cdot \sum_{s=0}^{m-1} \sum_{Z_m^s}^{\binom{m}{s}\binom{n-m}{m-s}} \frac{\bar{u}(Z_m^s)\left\{\bar{X} - \bar{x}'(Z_m^s)\right\}}{\bar{x}(Z_m^s)} \right\} \right]$$

$$= \frac{1}{\binom{n}{m}} E\left[\sum_{s=0}^{m-1} \binom{m}{s}\binom{n-m}{m-s} \cdot \frac{\bar{u}(Z_m)\left\{\bar{X} - \bar{x}'(Z_m)\right\}}{\bar{x}(Z_m)} \quad \frac{\bar{u}(Z_m^s)\left\{\bar{X} - \bar{x}'(Z_m^s)\right\}}{\bar{x}(Z_m^s)} \right]$$

contd...

$$= \frac{s\,\delta}{m^2\binom{n}{m}}\, E\Big[\sum_{s=0}^{m-1}\binom{m}{s}\binom{n-m}{m-s}\frac{\{\bar{X}-\bar{x}'(Z_m)\}\{\bar{X}-\bar{x}'(Z_m^s)\}}{\bar{x}(Z_m)\,\bar{x}(Z_m^s)}\Big],$$

since $E_x\big[\bar{u}(Z_m)\,\bar{u}(Z_m^s)\big] = \frac{s\,\delta}{m^2}$

Using (5.5.5) and (5.5.6) in (5.5.4), we obtain (5.5.3). Hence the lemma is proved.

Lemma 5.5.3: Under super-population model $\triangle$ (5.1.1)

$$E\Big[E_n^2\Big(\frac{\bar{X}-\bar{x}_{n-m}}{\bar{x}_m}\Big)\Big] = \frac{1}{\binom{n}{m}}\Big[\frac{V(x)}{(n-m)}\,E\Big(\frac{1}{\bar{x}^2(Z_m)}\Big)$$

$$+ E\Big\{\sum_{s=0}^{m-1}\binom{m}{s}\binom{n-m}{m-s}\frac{\{\bar{X}-\bar{x}'(Z_m)\}}{\bar{x}(Z_m)}\,\frac{\{\bar{X}-\bar{x}'(Z_m^s)\}}{\bar{x}(Z_m^s)}\Big\}\Big]. \qquad (5.5.7)$$

Proof : It is similar to the proof of lemma 5.5.2..

Lemma 5.5.4 : Under super-population model $\triangle$(5.1.1), the expected variance of the unbiased ratio type estimator T_m^* may be expressed as:

$$V(T_m^*) = \frac{\delta}{n} + \Big(a^2 + \frac{\delta}{m}\Big)\frac{V(x)}{(n-m)\binom{n}{m}}\,E\Big(\frac{1}{\bar{x}^2(Z_m)}\Big)$$

$$+ \frac{1}{\binom{n}{m}}\,E\Big[\sum_{s=0}^{m-1}\Big(a^2+\frac{s\,\delta}{m^2}\Big)\binom{m}{s}\binom{n-m}{m-s}\frac{\{\bar{X}-\bar{x}'(Z_m)\}\{\bar{X}-\bar{x}'(Z_m^s)\}}{\bar{x}(Z_m)\,\bar{x}(Z_m^s)}\Big]. \qquad (5.5.8)$$

Proof : We use lemmas 5.2.1, 5.5.1, 5.5.2 and 5.5.3.

Lemma 5.5.5 When sampling is from an infinite population and x is a gamma variate with parameter h,

$$\eta_s = E\left[\frac{\{\bar{X} - \bar{x}'(Z_m)\}\{\bar{X} - \bar{x}'(Z_m^s)\}}{\bar{x}(Z_m)\ \bar{x}(Z_m^s)}\right] = \left(\frac{m}{n-m}\right)^2 \left[\{(m-s)^2 h^2\right.$$

$$+ (n-2m+s)h\} E\left(\frac{1}{(\xi_1+\xi_2)(\xi_1+\xi_3)}\right) - 2(m-s)h\, E\left(\frac{\xi_2}{(\xi_1+\xi_2)(\xi_1+\xi_3)}\right)$$

$$\left. + E\left(\frac{\xi_2\ \xi_3}{(\xi_1+\xi_2)(\xi_1+\xi_3)}\right)\right], \qquad (5.5.9)$$

where ξ_1, ξ_2 and ξ_3 are mutually independent gamma variates with parameters sh, $(m-s)h$, and $(m-s)h$ respectively. For $s = 0$, however, $\xi_1 \equiv 0$.

Proof : We have from the definition

$$\eta_s = m^2 \bar{X}^2 E\left[\frac{1}{x(Z_m)\, x(Z_m^s)}\right] - \frac{m^2 \bar{X}}{(n-m)} E\left[\frac{x'(Z_m) + x'(Z_m^s)}{x(Z_m)\, x(Z_m^s)}\right]$$

$$+ \frac{m^2}{(n-m)^2} E\left[\frac{x'(Z_m)\, x'(Z_m^s)}{x(Z_m)\, x(Z_m^s)}\right], \qquad (5.5.10)$$

where $m\bar{x}(Z_m) = x(Z_m)$, $(n-m)x'(Z_m) = x'(Z_m)$, etc.

Let us now write

$$x(Z_m) = \xi_1 + \xi_2,\ x(Z_m^s) = \xi_1 + \xi_3,\ x'(Z_m) = \xi_3 + \xi_4 \text{ and } x'(Z_m^s) = \xi_2 + \xi_4, \qquad (5.5.11)$$

where

ξ_1 = Sum of x values of the s common elements belonging to Z_m and Z_m^s,

ξ_2 = Sum of x values of the m-s elements of Z_m not contained in Z_m^s.

ξ_3 = Sum of x values of the m-s elements of Z_m^s not contained in Z_m and

ξ_4 = Sum of x values of the (n-2m+s) elements of the sample not contained in either Z_m or Z_m^s.

Since sampling is from an infinite population and x is a gamma variate with parameter h, we note that ξ_1, ξ_2, ξ_3 and ξ_4 are all mutually independent gamma variates with parameters sh, (m-s)h, (m-s)h, and (n-2m+s)h respectively. For s = 0, however, $\xi_1 \equiv 0$.

Now using (5.5.11) we may re-write (5.5.10) as :

$$m^2h^2 E\left[\frac{1}{(\xi_1+\xi_2)(\xi_1+\xi_3)}\right] - \frac{m^2h}{(n-m)} E\left[\frac{\xi_2+\xi_3+2\xi_4}{(\xi_1+\xi_2)(\xi_1+\xi_3)}\right] + \frac{m^2}{(n-m)^2} E\left[\frac{(\xi_2+\xi_4)(\xi_3+\xi_4)}{(\xi_1+\xi_2)(\xi_1+\xi_3)}\right].$$

$$= E\left[\frac{1}{(\xi_1+\xi_2)(\xi_1+\xi_3)}\right]\left[m^2h^2 - \frac{2(n-2m+s)m^2h^2}{(n-m)} + \frac{m^2(n-2m+s)h}{(n-m)^2}\{(n-2m+s)h+1\}\right]$$

$$+ E\left[\frac{(\xi_2+\xi_3)}{(\xi_1+\xi_2)(\xi_1+\xi_3)}\right]\left[\frac{m^2(n-2m+s)h}{(n-m)^2} - \frac{m^2h}{(n-m)}\right] + \frac{m^2}{(n-m)^2} E\left[\frac{\xi_2\xi_3}{(\xi_1+\xi_2)(\xi_1+\xi_3)}\right].$$

= expression (5.5.9), since $E\left[(\xi_1+\xi_2)^{-1}(\xi_1+\xi_3)^{-1}\xi_2\right] = E\left(\frac{\xi_3}{(\xi_1+\xi_2)(\xi_1+\xi_3)}\right)$.

Hence the lemma is proved.

Remark (b) : The expectations of the functions of mutually independent gamma variates ξ_i, appearing in expression (5.5.9), may be evaluated for integer h, using the results of Rao and Webster (1966), and Rao (1967). These results are given in the following lemma.

Lemma 5.5.6 : [Rao and Webster (1966) and Rao (1967)] : When ξ_1, ξ_2 and ξ_3 are mutually independent gamma (Γ) variates with parameters sh, (m - s)h, and (m - s) h respectively, and h is an integer:

$$E[(\xi_1+\xi_2)^{-1}(\xi_1+\xi_3)^{-1}] = \frac{\Gamma(2mh-sh-2)}{\Gamma^2(mh-sh)\Gamma(sh)} C(mh-sh, sh),$$

$$E[\xi_2(\xi_1+\xi_2)^{-1}(\xi_1+\xi_3)^{-1}] = \frac{\Gamma(2mh-sh-1)}{\Gamma^2(mh-sh)\Gamma(sh)} C(mh-sh, sh) - \frac{1}{(mh-1)},$$

$$E[\xi_2\xi_3(\xi_1+\xi_2)^{-1}(\xi_1+\xi_3)^{-1}] = \frac{\Gamma(2mh-sh)}{\Gamma^2(mh-sh)\Gamma(sh)} C(mh-sh+1, sh), \text{ where}$$

(5.5.12)

$$C(mh-sh, sh) = \left[\sum_{k=0}^{(mh-sh-2)} (-1)^k \frac{\Gamma^2(mh-sh-k-1)\Gamma(sh+k)}{\Gamma(2mh-sh-k-2)}\right]$$

$$+(-1)^{mh-sh-1}\left[2\sum_{k=1}^{mh-2}(-1)^{k+1}\cdot\frac{1}{(mh-k-1)^2}+(-1)^{mh}\frac{\pi^2}{6}\right],$$

for $mh - sh \geq 2$, ,

$$= 2\sum_{k=1}^{sh-1}(-1)^{k+1}\frac{1}{(sh-k)^2}+(-1)^{sh+1}\frac{\pi^2}{6} \text{ for } mh-sh=1\text{; and}$$

$$C(mh-sh+1, sh) = \left[\sum_{k=0}^{mh-sh-1}(-1)^k\frac{\Gamma^2(mh-sh-k)\Gamma(sh+k)}{\Gamma(2mh-sh-k)}\right]$$

$$+(-1)^{mh-sh}\left[2\sum_{k=1}^{mh-1}(-1)^{k+1}\frac{1}{(mh-k)^2}+(-1)^{mh+1}\frac{\pi^2}{6}\right] \text{ for } (mh-sh)\geq 1.$$

Theorem 5.5.1: Under super-population model $\triangle'$ (5.1.2), the expected variance of the unbiased ratio type estimator T_m^* valid for any sample size n and any choice of m in the range (1, n-1) is given by

$$V(T_m^*) = \frac{\delta}{n} + \left(a^2 + \frac{\delta}{m}\right)\frac{m^2 h}{(n-m)\binom{n}{m}} \cdot \frac{1}{(mh-1)(mh-2)}$$

$$+ \frac{1}{\binom{n}{m}} \sum_{s=0}^{m-1} \left(a^2 + \frac{s\,\delta}{m^2}\right)\binom{m}{s}\binom{n-m}{m-s}\ \eta_s \qquad (5.5.13)$$

where η_s may be evaluated explicitely using lemmas 5.5.5 and 5.5.6.

Proof: We use lemma 5.5.4, and note $V(x) = h$ and

$$E\left[\frac{1}{\bar{x}^2(Z_m)}\right] = \frac{m^2}{(mh-1)(mh-2)}$$

Corollary 5.5.1: Under super-population model $\triangle'$ (5.1.2), the expected variance of the Hartley-Ross unbiased ratio type estimator T_1^* is given by

$$V(T_1^*) = \frac{\delta}{n} + \frac{(a^2+\delta)h}{n(n-1)(h-1)(h-2)} + \frac{a^2\left[1+(n-2)h\right]}{n(n-1)(h-1)^2}, \text{ for } h > 2.$$

$$(5.5.14)$$

Proof: Putting, m = 1 in (5.5.13), we obtain

$$V(T_1^*) = \frac{\delta}{n} + \frac{(a^2+\delta)h}{n(n-1)(h-1)(h-2)} + \frac{(n-1)}{n}a^2\eta_0\text{, where from}$$

(5.5.9)

$$\eta_0 = \frac{1}{(n-1)^2}\left[\left\{h^2+(n-2)h\right\}E\left(\frac{1}{\xi_2\xi_3}\right) - 2hE\left(\frac{1}{\xi_3}\right)+1\right],\ \text{since } \xi_1 = 0$$

for m = 1 and s = 0;

contd...

$$= \frac{1}{(n-1)^2} \left[\frac{h^2 + (n-2)h}{(h-1)^2} - \frac{2h}{(h-1)} + 1 \right] = \frac{1 + (n-2)h}{(n-1)^2 (h-1)^2} .$$

Hence the result.

Corollary 5.5.2 : Under super-population model $\triangle'$ (5.1.2), the expected variance of the unbiased ratio type estimator T^*_{n-1} is given by

$$V(T^*_{n-1}) = \frac{\delta}{n} + (\alpha^2 + \frac{\delta}{n-1}) \frac{h(n-1)^2}{n[(n-1)h-1][(n-1)h-2]} + [\alpha^2 + \frac{(n-2)\delta}{(n-1)^2}] \frac{(n-1)}{n} \eta_{n-2} , \quad (5.5.15)$$

where $\eta_{n-2} = (n-1)^2 [h^2 E(\frac{1}{(\xi_1+\xi_2)(\xi_1+\xi_3)}) - 2h E(\frac{\xi_2}{(\xi_1+\xi_2)(\xi_1+\xi_3)}) + E(\frac{\xi_2 \xi_3}{(\xi_1+\xi_2)(\xi_1+\xi_3)})]$

ξ_1, ξ_2 and ξ_3 are mutually independent gamma variates with parameters $(n-2)h$, h, and h respectively; and the expectations in η_{n-2} are given by lemma 5.5.6 with $m = n-1$ and $s = n-2$.

Proof : Follows directly from theorem 5.5.1.

Remark (c) : The results of corollary 5.5.1 and corollary 5.5.2 are originally due to Rao (1967). It may be noted that in Rao's (1967) notation $V(u/x) = n\delta$ where δ is a constant of order n^{-1} and $\frac{x}{n}$ is a gamma variate with parameter h. His results are also for the unbiased ratio estimators of the population ratio. Infact, he has derived the expected variance for the class of unbiased ratio type estimators t'_{n-m} (2.4.4) and has numerically demonstrated that (i) $\bar{t}'_{n-1} = T^*_{n-1} / \bar{X}$ is the optimum choice for $\bar{t}'_{n-m}$, (ii) $\bar{t}'_{n-1}$ is superior to Hartley-Ross estimator $T^*_1 / \bar{X}$, and (iii) $\bar{t}'_{n-1}$ is more efficient than the classical ratio estimator $\hat{\bar{Y}}_R/\bar{X}$ for $n \geq 8$.

5.6. Comparison of T_m^* with the classical ratio estimator $\hat{Y}_R$ under super-population model $\triangle'$ (5.1.2):

Under super-population model $\triangle'$ (5.1.2), it can be easily shown that the expected mean square error of the classical ratio estimator $\hat{Y}_R$ is given by

$$MSE(\hat{Y}_R) = \frac{a^2(nh+2) + \delta nh^2}{(nh-1)(nh-2)} \qquad (5.5.16)$$

Now for specific choices for h, n and m, one can compute the coefficients of a^2 and δ in $V(T_m^*)$ given by (5.5.13) and MSE ($\hat{Y}_R$) given by (5.5.16) and obtain a comparison of the unbiased ratio type estimator T_m^* with the classical ratio estimator $\hat{Y}_R$ under super-population model $\triangle'$ (5.1.2). As an illustration, such a numerical comparison is attempted here for the choices: $h = 3$; $n = 2$, $m = 1$; $n = 3$, $m = 1, 2$ and $n = 4$, $m = 1, 2, 3$.

The computed coefficients of a^2 and δ, say $C(a^2)$ and $C(\delta)$, are presented in Table - 5.1 .

TABLE - 5.1

efficients of α^2 and δ in $V(T_m^*)$ and $MSE(\hat{\bar{Y}}_R)$ under $\triangle'$ for h = 3

timator	n = 2		n = 3		n = 4	
	$C(\alpha^2)$	$C(\delta)$	$C(\alpha^2)$	$C(\delta)$	$C(\alpha^2)$	$C(\delta)$
T_1^*	0.875	1.250	0.417	0.583	0.271	0.375
T_2^*	-	-	0.191	0.431	0.143	0.296
T_3^*	-	-	-	-	- 0.022	0.258
$\hat{\bar{Y}}_R$	0.404	0.900	0.196	0.482	0.127	0.327

emark (d) : Under super-population model $\triangle'$ (5.1.2) for h = 3, the sults in Table - 5.1 show that:

i) Hartley-Ross unbiased ratio type estimator T_1^* is less efficient than the classical ratio estimator $\hat{Y}_R$.

ii) For n = 3 as well as 4, the precision of the unbiased ratio type estimator T_m^* increases as m increases from 1 to n-1.

iii) The unbiased ratio type estimator T^*_{n-1} is more efficient than the classical ratio estimator $\hat{Y}_R$ for n = 3 and 4.

CHAPTER VI

BIASED PRODUCT AND RATIO-CUM-PRODUCT TYPE ESTIMATORS

0. Summary :

In chapters IV and V, our investigations have been related to tle method of estimation in which information on an auxiliary variate itively associated with the variate under study is used to improve the ecision of the estimator of the population mean of the study variate. In this pter we concentrate our attention on product method of estimation which es information on an auxiliary variate negatively associated with the study riate. When several auxiliary variates of which some having positive and rs negative correlation with the study variate are available, we construct biased ratio-cum-product type estimators. Our approach is to derive the lased product and ratio-cum-product type estimators from Mickey's timator G^*_m for specific choices of the coefficient functions $a_i(Z_m)$. The riance of these estimators is also investigated by making use of the results the variance and estimation of variance of G^*_m. Section-wise details of the sults are briefly given in the following.

The chapter contains six sections. Section 1 provides a brief roduction. In section 2 we show that, for a single auxiliary variate x atively correlated with the study variate y and the choice $a(Z_m) = \frac{q\,\bar{y}_m}{\bar{X}}$, $q < 0$, the estimator G^*_m yields an unbiased product type estimator $S^*_{(q<0)}$. In particular, for $q = -1$, the estimator $S^*_{(q<0)}$ reduces to

Robson's (1957) unbiased product type estimator $\hat{\bar{Y}}_{RP}$. Thus Mickey's general class of estimators includes unbiased product type estimators of which Robson's estimator $\hat{\bar{Y}}_{RP}$ is a particular case. It is interesting to note that the unbiased product type estimator $S^*_{(q<o)}$ given by (6.2.1) and the unbiased ratio type estimator $S^*_{(q)}$ (4.2.1) have the same 'form', except that we take $q < 0$ for the product type estimator and $q > 0$ for the ratio type estimator. Thus for the choice $a(Z_m) = \frac{q\,\bar{y}_m}{\bar{X}}$, the estimator G^*_m results in the unbiased product type estimator $S^*_{(q<o)}$ or the simple unbiased estimator $\bar{y}$ or the unbiased ratio type estimator $S^*_{(q)}$ according as the choice of $q \lesseqgtr 0$.

A second class of unbiased product type estimators H^*_m (6.2.3) is derived from G^*_m in section 2 for the choice $a(Z_m) = -\frac{P_m}{\bar{X}}$ where

$P_m = \frac{1}{m\,\bar{x}_m}\,(\sum_{j\in Z_m} y_j\,x_j)$. Robson's estimator $\hat{\bar{Y}}_{RP}$ is a particular case of H^*_m for $m = 1$. A third class of unbiased product type estimators L^*_m (6.2.5) is shown to result from G^*_m for the choice $a(Z_m) = -\frac{\bar{r}_m\,\bar{x}_m}{\bar{X}}$ where $r_j = y_j / x_j$ and $\bar{r}_m = \frac{1}{m}(\sum_{j\in Z_m} r_j)$. This class also contains Robson's estimator $\hat{\bar{Y}}_{RP}$ since $L^*_1 = \hat{\bar{Y}}_{RP}$. Further, given any two specific choices of m, say m' and m'' in the range 1 to n-1, we show that the entire class of estimators L^*_m for $1 \leq m \leq n-1$ may be generated as weighted linear combinations of $L^*_{m'}$ and $L^*_{m''}$. In particular, when the finite population correction is ignored and n is even, the estimator $L^*_{(n/2)}$ (6.2.16) is of simple form and, consequently, L_m^* for any m in the range 1 to n-1 may be generated as a weighted linear combination of L_1^* and $L_{(n/2)}^*$. Finally, in section 2 we observe an interesting property of the estimator G_m^* which says that for $m = 1$, whatever be the choice of the coefficient function $a(Z_1) = a_j$, the estimator involves Robson's product type estimator $\hat{\bar{A}}_{RP}$ for the population mean $\bar{A}$ of a_j.

In section 3, we consider the formation of certain classes of unbiased ratio-cum-product type estimators. Assuming that p_1 (say, $x_1, x_2, \ldots, x_{p_1}$) of the p auxiliary variates ($x_1, x_2, \ldots, x_{p_1}, x_{p_1+1}, \ldots, x_p$) have positive correlation and the remaining ($p - p_1$) variates have negative correlation with the study variate y, we show that the multi-variate unbiased estimator G_m^* (3.4.3) yields the unbiased ratio-cum-product type estimators:

(i) $S^*_{(\pm 1)}$ (6.3.2) for the choice: $a_i(Z_m) = \dfrac{W_i \bar{y}_m}{\bar{X}_i}$, $(i = 1, 2, \ldots, p_1)$,

$$= \frac{-W_i \bar{y}_m}{\bar{X}_i}, (i = p_1+1, \ldots, p),$$

$$\sum_{i=1}^{p} W_i = 1;$$

(ii) TH_m^* (6.3.4) for the choice: $a_i(Z_m) = W_i R_{im}$, $(i = 1, 2, \ldots, p_1)$,

$$= \frac{-W_i P_{im}}{\bar{X}_i}, (i = p_1+1, \ldots, p).$$

...re R_{im} and P_{im} are the functions R_m and P_m defined on y and x_i;

and (iii) TL^*_m (6.3.6) for the choice: $a_i(Z_m) = W_i R_{im}$, ($i=1, 2, \ldots, p_1$),

$$= \frac{-W_i \bar{r}_{im} \bar{x}_{im}}{\bar{X}_i}, (i=p_1+1, \ldots, p),$$

where $\bar{r}_{im}$ is the function $\bar{r}_m$ defined on y and x_i.

In particular, for m = 1, both TH^*_m and TL^*_m reduce to the unbiased ratio-cum-product type estimator (6.3.7) which is a weighted combination of Hartley-Ross unbiased ratio type estimators and Robson's unbiased product estimators. Srivastava (1965), Rao and Mudholkar (1967) etc., have earlier mentioned this unbiased ratio-cum-product type estimator without, however, determining the optimum choice of the weight vector $W=(W_1, W_2, \ldots, W_p)$.

In section 4, we derive the exact variances of the unbiased product type estimator $S^*_{(q < o)}$ and the unbiased ratio-cum-product type estimators (±1) and TH^*_1. We show that Robson's (1957) result on the exact variance of his unbiased product type estimator $\hat{\bar{Y}}_{RP}$ is a particular case of our result on the variance of $S^*_{(q<o)}$ for the choice q = -1. When N >> n, for the estimator $S^*_{(q<o)}$ we give the optimum choice q_o of q and the corresponding minimum variance. The estimator $S^*_{(q_o<o)}$ is more efficient than Robson's unbiased product type estimator $\hat{\bar{Y}}_{RP}$. A good approximation to q_o is $q_o' = \rho C_y/C_x$. In large samples, both $S^*_{(q_o<o)}$ and $S^*_{(q_o'<o)}$ have the same variance as the classical linear regression estimator $\hat{\bar{Y}}_r$ and are superior to the conventional product estimator $\hat{\bar{Y}}_p = \frac{\bar{y}\,\bar{x}}{\bar{X}}$. Finally, in this

ection, when $N >> n$, using the exact variance formulae of $S^*_{(\pm 1)}$ and TH^*_1, we determine the optimum choices for the weight vector W which is used in the construction of these estimators.

In section 5, we obtain the large sample variance of the unbiased oduct type estimators: H^*_m and L^*_m, and of the unbiased ratio-cum-product ype estimators: TH^*_m and TL^*_m, by establishing that for these estimators e conditions, under which the large sample variance of G^*_m is derived in ction 3.4 b, are satisfied. In large samples, we show that the unbiased oduct type estimator H^*_{n-1} is more or equally or less efficient compared o Robson's unbiased product type estimator $\hat{Y}_{RP}$ or the conventional product stimator $\hat{Y}_P$ according as $\left| \rho \; C_y/C_x \right| \gtreqless (1 + \frac{1}{2} \; C_x^2)^{-1}$. When the pulation ratio $R = \bar{Y}/\bar{X}$ and the population mean of ratios $\bar{r}_N = \frac{1}{N} \sum_{j=1}^{N} r_j$ re equal, it is seen that the large sample variance of L^*_m is same for all hoices of m in the range 1 to n-1. Denoting the population linear regression oefficient of y on x by β and assuming R and $\bar{r}_N$ are positive, we also ow that the large sample variance of L^*_m decreases as m increases when er $R > \bar{r}_N > |\beta|$ or $R < \bar{r}_N < |\beta|$; so that in these cases m=n-1 is e optimum choice for L^*_m. Finally, for the unbiased ratio-cum-product pe estimators TH^*_m and TL^*_m, we determine the optimum choices of the ight vector W by making use of the large sample variance of these timators.

In section 6, estimators of variance are developed for the unbiased product type estimators: $S^*_{(q_0' < o)}$, H^*_m and L^*_m, using the results in section 3.6 on G^*_m. These include unbiased varinace estimators valid for samples of size $n \geq 2$ and biased but simpler estimators of variance applicable in large samples. By a similar approach, one can also formulate estimators of variance for the unbiased ratio-cum-product type estimators: $S^*_{(\pm 1)}$, TH^*_m and TL^*_m. The computation of the estimates of variance is illustrated for samples of sizes 2, 3, 4, 5, 9 and 12 from a population of size 45. The population used in these examples is a special selection of cities/urban agglomerations of size 100,000 and above (as per 1971 census of India). The 45 cities/urban agglomerations are situated in different states of India. In these cities/urban agglomerations, high or low female literacy rates are associated with low or high female work participation rates. Taking female literacy rate as auxiliary variate (x), product estimates of the average female work participation rate ($\overline{Y}$) and their variance estimates are computed. For each of the sizes 2, 3, 4 and 5, five independent samples are chosen and on each sample unbiased variance estimates are computed. For the sample sizes 9 and 12, we select 5 and 4 independent samples respectively and compute on each sample biased estimates of variance.

The small sample calculations presented in section 6 are primarily meant for illustration and not to make any firm efficiency comparisons. In a later chapter, we will study the relative efficiencies of (i) the unbiased product type estimators given here, (ii) some unbiased regression type estimators presented in the next chapter, and (iii) the conventional biased product and linear regression estimators, in small samples of sizes 2, 3, 4 and 5 drawn from the same population on the basis of a reasonable number of independent samples of each size. The tentative observations from the computations in this section are as follows. Among the product type estimators, $5^{*}_{(q_0' < o)}$ has performed well in most of the samples of different sizes. In samples of sizes 9 and 12, different choices of m resulted in almost equal precision for either H^{*}_{m} or L^{*}_{m}; and L^{*}_{m} is subjected to a slightly smaller sampling error than H^{*}_{m}.

6.1. Introduction :

In the last two chapters we have been concerned with the use of information on an auxiliary variate, positively associated with the variate under study, in improving the precision of the estimator of the population mean of the study variate through ratio method of estimation. We now turn our attention to unbiased techniques of estimation which are capable of using the knowledge of an auxiliary variate negatively associated with the study variate. In line with the common usage, these may be included under product method of estimation. A brief review of the literature on product estimators has already been given in section 2.5 and 2.6.

When several auxiliary variates of which some having positive and others negative correlation with the study variate are available, ratio-cum-product method of estimation has been suggested by Srivastava (1965), Rao and Mudholkar (1967), Singh (1967) etc. Earlier to these attempts, Olkin (1958) has utilised information on several supplementary variates positively correlated with the study variate to develop a composite multivariate biased ratio estimator.

In this chapter we derive some classes of unbiased product and ratio-cum-product type estimators from Mickey's general class of unbiased estimators G_m^* (3.4.3) with appropriate choices of the coefficient functions $a_i(Z_m)$. In particular, it is shown that Robson's (1957) unbiased product estimator for the population mean $\bar{Y}$ belongs to Mickey's general class of unbiased estimators. Using the fundamental results on the variance and estimation of variance of G_m^*, obtained in chapter III, we investigate the precision of the unbiased product and ratio-cum-product type estimators.

6.2. Unbiased product type estimators:

Suppose information is available on a single auxiliary variate x having negative correlation with y. For the general unbiased estimator G_m^* (3.4.3), we consider three possible choices for the coefficient function $a(Z_m)$ leading to unbiased product type estimators. These are contained in the following results.

Theorem 6.2.1: With a single auxiliary variate x negatively associated with y, and the choice $a(Z_m) = q\,\bar{y}_m/\bar{X}$, $q < 0$, the unbiased estimator G_m^* (3.4.3) yields an unbiased product type estimator:

$$S^*_{(q<0)} = (q+1)\,\bar{y} - q\left[\frac{\bar{y}\,\bar{x}}{\bar{X}} - \left(\frac{1}{n} - \frac{1}{N}\right)\frac{s_{xy}}{\bar{X}}\right]. \qquad (6.2.1)$$

<u>Proof</u> : Same as the proof of theorem 4.2.1 in chapter IV.

<u>Corollary 6.2.1</u> : For $q = -1$, the estimator $S^*_{(q<o)}$ reduces to Robson's (1957) unbiased product type estimator:

$$\hat{\bar{Y}}_{RP} = S^*_{(-1)} = \frac{\bar{y}\ \bar{x}}{\bar{X}} - \left(\frac{1}{n} - \frac{1}{N}\right)\frac{s_{xy}}{\bar{X}} . \qquad (6.2.2)$$

<u>Remark</u> (a) : The form of the unbiased ratio type estimator $S^*_{(q)}$ (4.2.1) and that of the unbiased product type estimator $S^*_{(q<o)}$ (6.2.1) are the same, excepting that we take $q > 0$ for ratio type estimator and $q < 0$ for product type estimator.

<u>Theorem 6.2.2</u> : For the choice $a(Z_m) = -\frac{P_m}{\bar{X}}$ where $P_m = \frac{1}{m\bar{x}_m}(\sum_{j \in Z_m} y_j)$ the unbiased estimator G^*_m (3.4.3) results in an unbiased product type estimator given by

$$H^*_m = (\bar{y} - P^*_m) + \left[\frac{\bar{x}\,P^*_m}{\bar{X}} - \frac{m(N-n)}{(n-m)N\bar{X}}\left(\frac{1}{n}\sum_{j=1}^{n} y_j\, x_j - P^*_m\,\bar{x}\right)\right],$$

$$\text{where} \quad P^*_m = \frac{1}{\binom{n}{m}} \sum_{Z_m}^{\binom{n}{m}} P_m . \qquad (6.2.3)$$

<u>Proof</u> : We have

$$a^*(Z_m) = E_n\left[a(Z_m)\right] = -\frac{E_n(P_m)}{\bar{X}} = -\frac{P^*_m}{\bar{X}}, \qquad (6.2.4)$$

and

$$\mathrm{Cov}_n\left[a(Z_m), \bar{x}_m\right] = -\frac{\mathrm{Cov}_n(P_m, \bar{x}_m)}{\bar{X}} = -\frac{1}{\bar{X}}\left[E_n(P_m\bar{x}_m) - P^*_m\,\bar{x}\right]$$
$$= -\frac{1}{\bar{X}}\left[\frac{1}{n}\sum_{j=1}^{n} y_j x_j - P^*_m\,\bar{x}\right].$$

Using (6.2.4) in G^*_m (3.4.3), we get the product type estimator H^*_m (6.2.3).

Corollary 6.2.2: For $m = 1$, H^*_m (6.2.3) reduces to Robson's unbiased product type estimator $\hat{\hat{Y}}_{RP}$ (6.2.2).

Proof: This follows from the observation $P^*_m = \bar{y}$ for $m = 1$.

Theorem 6.2.3: For the choice $a(Z_m) = -\dfrac{\bar{r}_m \bar{x}_m}{\bar{X}}$ where

$\bar{r}_m = \frac{1}{m}\left(\sum_{j \in Z_m} r_j\right)$ and $r_j = y_j / x_j$, the unbiased estimator G^*_m (3.4.3) provides an unbiased product type estimator:

$$L^*_m = \left[\frac{\bar{y}\,\bar{x}}{\bar{X}} - \left(\frac{1}{n} - \frac{1}{N}\right)\frac{s_{xy}}{\bar{X}}\right] + \frac{n(m-1)}{(n-1)m\bar{X}}\left[\frac{(N-n)}{N(n-2)}\, m^{rx}_{12} - (\bar{x} - \bar{X})\, m^{rx}_{11}\right], \tag{6.2.5}$$

where $m^{rx}_{ik} = \frac{1}{n}\sum_{j=1}^{n} (r_j - \bar{r})^i (x_j - \bar{x})^k$

Proof: Here $a^*(Z_m) = E_n[a(Z_m)] = -\frac{1}{\bar{X}} E_n(\bar{r}_m \bar{x}_m)$

$$= -\frac{1}{\bar{X}}\left[\mathrm{Cov}_n(\bar{r}_m, \bar{x}_m) + \bar{r}\,\bar{x}\right],$$

$$= -\frac{1}{\bar{X}}\left[\frac{(n-m)}{(n-1)m}\, m^{rx}_{11} + \bar{r}\,\bar{x}\right]. \tag{6.2.6}$$

om Sukhatme (1970, p 191) we have

$$\mathrm{Cov}_n\left[a(Z_m), \bar{x}_m\right] = -\frac{1}{\bar{X}}\quad E_n\left[\bar{r}_m \bar{x}_m (\bar{x}_m - \bar{x})\right].$$

$$= -\frac{1}{\bar{X}} E_n \left[(\bar{r}_m - \bar{r})(\bar{x}_m - \bar{x})^2 + \bar{x}\,\bar{r}_m(\bar{x}_m - \bar{x}) + \bar{r}\,(\bar{x}_m - \bar{x})^2 \right]$$

$$= -\frac{1}{\bar{X}} \left[-\frac{(n-m)(n-2m)}{m^2(n-1)(n-2)} m_{12}^{rx} + \frac{(n-m)}{m(n-1)} (\bar{x}\, m_{11}^{rx} + \bar{r}\, m_{02}^{rx}) \right]. \qquad (6.2.7)$$

Substituting (6.2.6) and (6.2.7) in G_m^* (3.4.3) we obtain

$$L_m^* = \bar{y} + \left[-\frac{(n-m)}{m(n-1)} m_{11}^{rx} + \bar{r}\,\bar{x} \right] \frac{(\bar{x} - \bar{X})}{\bar{X}} - \frac{(N-n)}{N(n-1)\bar{X}} \left[-\frac{(n-2m)}{m(n-2)} m_{12}^{rx} + \bar{x}\, m_{11}^{rx} + \bar{r}\, m_{02}^{rx} \right]. \qquad (6.2.8)$$

Hence
$$L_1^* = \bar{y} + (m_{11}^{rx} + \bar{r}\,\bar{x}) \frac{(\bar{x} - \bar{X})}{\bar{X}} - \frac{(N-n)}{N(n-1)\bar{X}} \left[m_{12}^{rx} + \bar{x}\, m_{11}^{rx} + \bar{r}\, m_{02}^{rx} \right]. \qquad (6.2.9)$$

Using (6.2.8) and (6.2.9) we may write

$$L_m^* = L_1^* + \left[-\frac{(n-m)}{m(n-1)} - 1 \right] \frac{(\bar{x} - \bar{X})}{\bar{X}} m_{11}^{rx} - \frac{(N-n)}{N(n-1)\bar{X}} \left[-\frac{(n-2m)}{m(n-2)} - 1 \right] m_{12}^{rx},$$

$$= L_1^* + \frac{n(m-1)}{(n-1)m\bar{X}} \left[-\frac{(N-n)}{N(n-2)} m_{12}^{rx} - (\bar{x} - \bar{X})\, m_{11}^{rx} \right]. \qquad (6.2.10)$$

Now in (6.2.9) we note $m_{11}^{rx} + \bar{r}\,\bar{x} = \frac{1}{n} \sum_{j=1}^{n} r_j\, x_j = \bar{y}$, and

$$m_{12}^{rx} + \bar{x}\, m_{11}^{rx} + \bar{r}\, m_{02}^{rx} = \frac{1}{n} \left[\sum_{j=1}^{n} (r_j - \bar{r})(x_j - \bar{x})^2 + \bar{x} \sum_{j=1}^{n} r_j (x_j - \bar{x}) + \bar{r} \sum_{j=1}^{n} (x_j - \bar{x})^2 \right]$$

$$= \frac{1}{n} \sum_{j=1}^{n} r_j\, x_j\, (x_j - \bar{x}) = \frac{n-1}{n} s_{xy}.$$

Hence (6.2.9) reduces to Robson's unbiased product type estimator $\hat{\bar{Y}}_{RP}$ (6.2.2). Thus we have

$$L_1^* = \frac{\bar{y}\,\bar{x}}{\bar{X}} - \left(\frac{1}{n} - \frac{1}{N}\right)\frac{s_{xy}}{\bar{X}} .$$

Writing the above expression for L_1^* in (6.2.10), we obtain the expression (6.2.5) for L_m^*. This completes the proof.

<u>Corollary 6.2.3</u>: For $m = 1$, the unbiased product type estimator L_m^* (6.2.5) reduces to Robson's unbiased product type estimator $\hat{\bar{Y}}_{RP}$ (6.2.2).

<u>Remark</u> (b): The unbiased product type estimator L_m^* depends on m only through the coefficient $\frac{(m-1)}{m}$ of the second term in expression (6.2.5). In view of this we have the following interesting property possessed by the estimators L_m^* (6.2.5) for $1 \leq m \leq n-1$.

<u>Theorem 6.2.4</u>: Given any two specific choices of m, say m' and m'', in the range 1 to $n-1$, the unbiased product type estimator L_m^* (6.2.5) may be expressed as

$$L_m^* = W' L_{m'}^* + W'' L_{m''}^* , \qquad (6.2.11)$$

where $W' = \frac{m'(m - m'')}{m(m' - m'')}$, $W'' = \frac{m''(m' - m)}{m(m' - m'')}$, $W' + W'' = 1$.

<u>Proof</u>: Writing $\psi = \frac{n}{(n-1)\bar{X}} \left[-\frac{(N-n)}{N(n-2)} m_{12}^{rx} - (\bar{x} - \bar{X})\, m_{11}^{rx} \right]$, we have from (6.2.5), $L_m^* = L_1^* + \left(1 - \frac{1}{m}\right)\psi$.

$$(6.2.12)$$

Hence $$L^*_{m'} - L^*_{m''} = \frac{(m' - m'')}{m' m''}\ \psi ,$$

and $$\psi = \frac{m' m''}{(m' - m'')}\ (L^*_{m'} - L^*_{m''}). \qquad (6.2.13)$$

Using (6.2.12) and (6.2.13) we have

$$L^*_1 = L^*_{m'} - \frac{m''(m' - 1)}{(m' - m'')}\ (L^*_{m'} - L^*_{m''}). \qquad (6.2.14)$$

Now substituting for ψ and L^*_1 from (6.2.13) and (6.2.14) in (6.2.12) we obtain the result (6.2.11).

Remark (c) : It may be noted that ψ, as defined above, is a zero function in the sense that its expected value is zero.

Theorem 6.2.5 : When the finite population correction is ignored and the sample size n is even, the unbiased product type estimator L^*_m (6.2.5) may be expressed in a simple form as

$$L^*_m = \frac{(n - 2m)}{m(n-2)}\ L^*_1 + \frac{n(m-1)}{m(n-2)}\ L^*_{(n/2)} , \qquad (6.2.15)$$

where $$L^*_1 = \frac{\bar{y}\,\bar{x}}{\bar{X}} - \frac{1}{n}\ \frac{s_{xy}}{\bar{X}} ,$$

$$L^*_{(n/2)} = \left(\frac{n-2}{n-1}\right)(\bar{y} - \bar{r}\,\bar{x}) + \frac{\bar{r}}{\bar{X}} \cdot \left(\bar{x}^2 - \frac{s_x^2}{n}\right). \qquad (6.2.16)$$

Proof: We use theorem 6.2.4 with $m' = 1$ and $m'' = \frac{n}{2}$ and obtain

$$L_m^* = \frac{(n-2m)}{m(n-2)} L_1^* + \frac{n(m-1)}{m(n-2)} L_{(n/2)}^*$$

When the finite population correction is ignored, $L_1^* = \hat{Y}_{RP}$(6.2.2) is given by

$$L_1^* = \frac{\bar{y}\,\bar{x}}{\bar{X}} - \frac{1}{n}\,\frac{s_{xy}}{\bar{X}}$$

For $m = \frac{n}{2}$, ignoring the finite population correction, we have from the form (6.2.8) of L_m^*:

$$L_{(n/2)}^* = \bar{y} + \left[-\frac{m_{11}^{rx}}{(n-1)} + \bar{r}\,\bar{x} \right] \frac{(\bar{x}-\bar{X})}{\bar{X}} - \frac{1}{(n-1)\bar{X}} \left[\bar{x}\, m_{11}^{rx} + \bar{r}\, m_{02}^{rx} \right]$$

$$= \bar{y} - \left[\frac{m_{11}^{rx}}{(n-1)} + \bar{r}\,\bar{x} \right] + \frac{\bar{r}}{\bar{X}} \left[\bar{x}^2 - \frac{m_{02}^{rx}}{(n-1)} \right]. \quad (6.2.17)$$

We note $m_{11}^{rx} = (\bar{y} - \bar{r}\,\bar{x})$, and $\frac{m_{02}^{rx}}{(n-1)} = \frac{s_x^2}{n}$

Hence (6.2.17) reduces to (6.2.16). This completes the proof of the theorem.

Remark (d): The following theorem brings out an interesting feature of Mickey's general unbiased estimator G_m^* (3.4.3). It says that with a single auxiliary variate x, when m = 1, whatever may be the choice of the coefficient function $a(Z_1) = a_j$, the estimator G_m^* involves product method of estimation.

Theorem 6.2.6: When m = 1, with a single auxiliary variate x, the general unbiased estimator G_m^* (3.4.3) may be expressed in the form

$$G_1^* = \bar{y} + \bar{X}(\bar{a} - \hat{\bar{A}}_{RP}), \tag{6.2.18}$$

where $\bar{a} = \frac{1}{n}\sum_{j=1}^{n} a_j$, $\bar{A} = \frac{1}{N}\sum_{j=1}^{N} a_j$;

and $\hat{\bar{A}}_{RP} = \frac{\bar{a}\,\bar{x}}{\bar{X}} - \left(\frac{1}{n} - \frac{1}{N}\right)\frac{s_{ax}}{\bar{X}}$ is an unbiased product type estimator of $\bar{A}$ similar to $\hat{\bar{Y}}_{RP}$ (6.2.2).

Proof : With a single auxiliary variate x, we have from (3.4.4)

$$G_1^* = \bar{y} - \bar{a}(\bar{x} - \bar{X}) + \left(\frac{1}{n} - \frac{1}{N}\right)s_{ax},$$

$$= \bar{y} + \bar{X}(\bar{a} - \hat{\bar{A}}_{RP}).$$

Remark (e) : The function $(\bar{a} - \hat{\bar{A}}_{RP})$, appearing as the coefficient of $\bar{X}$ in G_1^*, is a zero function since $E(\bar{a}) = \bar{A} = E(\hat{\bar{A}}_{RP})$.

6.3. Unbiased ratio-cum-product type estimators:

When several auxiliary variates of which some having positive and others negative correlation with the study variate are available, we may adopt ratio-cum-product estimation in which ratio method is applied to positively correlated auxiliary variates and product method is employed for negatively associated auxiliary variates. In this connection we recall our discussion in section 3.5 on the effect of introducing an additional auxiliary variate in the general multi-variate unbiased estimator G_m^* (3.4.3). In the light of that discussion we now frame the following unbiased ratio-cum-product type estimators.

Let p_1 $(x_1, x_2, \ldots, x_{p_1})$ of the p $(x_1, x_2, \ldots, x_{p_1}, x_{p_1+1}, \ldots, x_p)$ auxiliary variates have positive correlation and the remaining $(p - p_1)$ variates have negative correlation with the study variate y.

Theorem 6.3.1: For the choice of the coefficient functions:

$$\hat{a}_i(Z_m) = \frac{W_i \bar{y}_m}{\bar{X}_i}, \quad (i = 1, 2, \ldots, p_1),$$

$$= \frac{-W_i \bar{y}_m}{\bar{X}_i}, \quad (i = p_1 + 1, \ldots, p), \tag{6.3.1}$$

$$\sum_{i=1}^{p} W_i = 1 ;$$

the multivariate unbiased estimator G_m^* (3.4.3) yields an unbiased ratio-cum-product type estimator:

$$1) = \sum_{i=1}^{p_1} W_i \left[2\bar{y} - \left\{ \frac{\bar{y}\,\bar{x}_i}{\bar{X}_i} - \left(\frac{1}{n} - \frac{1}{N}\right) \frac{s_{x_i y}}{\bar{X}_i} \right\} \right]$$

$$+ \sum_{i=p_1+1}^{p} W_i \left[\frac{\bar{y}\,\bar{x}_i}{\bar{X}_i} - \left(\frac{1}{n} - \frac{1}{N}\right) \frac{s_{x_i y}}{\bar{X}_i} \right]. \tag{6.3.2}$$

of: We use the unbiased ratio type estimator $S^*_{(1)}$ (4.3.5) and the biased product type estimator $S^*_{(-1)} = \hat{\bar{Y}}_{RP}$ (6.2.2) in the multivariate based estimator G_m^* (3.5.5), which itself has been obtained from

G_m^* (3.4.3) in section 3.5 and considered as the appropriate multivariate estimator when each of the coefficient functions is defined on a single auxiliary variate. Here the vector $W = (W_1, W_2, \ldots, W_p)$ is taken as a non-random weight vector.

Theorem 6.3.2 : For the choice of the coefficient functions:

$$\tilde{a}_i(Z_m) = W_i R_{im}, \quad R_{im} = \frac{\bar{y}_m}{\bar{x}_{im}}, \quad (i = 1, 2, \ldots, p_1),$$

$$= \frac{-W_i P_{im}}{\bar{X}_i}, \quad P_{im} = \frac{\sum_{j \in Z_m} y_j x_{ij}}{m \bar{x}_{im}}, (i = p_1+1, \ldots, p)$$

$$\sum_{i=1}^{p} W_i = 1 ; \qquad (6.3.3)$$

the multivariate unbiased estimator G_m^* (3.4.3) provides an unbiased ratio-cum-product type estimator:

$$TH_m^* = \sum_{i=1}^{p_1} W_i \left[R_{im}^* \bar{X}_i + \frac{(N-m)n}{N(n-m)} (\bar{y} - R_{im}^* \bar{x}_i) \right] + \sum_{i=p_1+1}^{p} W_i \left[(\bar{y} - P_{im}^*) \right.$$

$$\left. + \frac{\bar{x}_i P_{im}^*}{\bar{X}_i} - \frac{m(N-n)}{(n-m)N \bar{X}_i} \left(\frac{1}{n} \sum_{j=1}^{n} y_j x_{ij} - \bar{x}_i P_{im}^* \right) \right] \quad (6.3.4)$$

where $R_{im}^* = \frac{1}{\binom{n}{m}} \sum_{Z_m}^{\binom{n}{m}} R_{im}$ and $P_{im}^* = \frac{1}{\binom{n}{m}} \sum_{Z_m}^{\binom{n}{m}} P_{im}$.

Proof : We use the unbiased ratio type estimator (4.1.1) and the unbiased product type estimator (6.2.3) in the multivariate unbiased estimator (3.5.5).

<u>Theorem 6.3.3</u> : *For the choice of the coefficient functions*:

$$\tilde{a}_i(z_m) = W_i R_{im}, \quad (i = 1, 2, \ldots, p_1),$$

$$= \frac{-W_i \bar{r}_{im} \bar{x}_{im}}{\bar{X}_i}, \quad (i = p_1 + 1, \ldots, p), \qquad (6.3.5)$$

the multivariate unbiased estimator G^*_m (3.4.3) yields an unbiased ratio-cum-product type estimator:

$$TL^*_m = \sum_{i=1}^{p_1} W_i T^*_{im} + \sum_{i=p_1+1}^{p} W_i L^*_{im}, \qquad (6.3.6)$$

where, for an auxiliary variate x_i, T^*_{im} represents an unbiased ratio type estimator (4.1.1) and L^*_{im} denotes an unbiased product type estimator (6.2.5).

<u>Proof</u> : Follows from the use of (4.1.1) and (6.2.5) in (3.5.5).

<u>Corollary 6.3.1</u> : For $m = 1$, both TH^*_m (6.3.4) and TL^*_m (6.3.6) reduce to an unbiased ratio-cum-product type estimator:

$$TH^*_1 = TL^*_1 = \sum_{i=1}^{p_1} W_i \left[\bar{r}_i \bar{X}_i + \frac{(N-1)\,n}{N(n-1)} \left(\bar{y} - \bar{r}_i \bar{x}_i \right) \right]$$

$$+ \sum_{i=p_1+1}^{p} W_i \left[\frac{\bar{y}\,\bar{x}_i}{\bar{X}_i} - \left(\frac{1}{n} - \frac{1}{N} \right) \frac{s_{x_i y}}{\bar{X}_i} \right], \qquad (6.3.7)$$

which is a weighted combination of Hartley-Ross (1954) unbia estimators and Robson's (1957) unbiased product type estimators.

Remark (a) : Srivastava (1965), Rao and Mudholkar (1967), etc., have earlier mentioned TH_1^* (6.3.7) in their articles without determining the optimum choice for the weight vector W.

Remark (b) : For the unbiased ratio-cum-product type estimators given in theorems 6.3.1, 6.3.2 and 6.3.3, we consider in section 5 of this chapter, the question of determining the optimum choice for the weight vector W.

6.4. Exact variance of some unbiased product and ratio-cum-product type estimators:

We have given in section 3.4a some results on the exact variance of the unbiased estimator G_m^* (3.4.3) for the particular choice $m = 1$. In the previous two sections of this chapter unbiased product, and ratio-cum-product type estimators have been derived from G_m^*. In particular, for $m = 1$, we have the unbiased product type estimator $S^*_{(q<o)}$ (6.2.1) which includes Robson's estimator $S^*_{(-1)} = \hat{Y}_{RP}$ (6.2.2), and the unbiased ratio-cum-product type estimators $S^*_{(\pm 1)}$ (6.3.2) and $TH_1^* = TL_1^*$(6.3.7). It is, therefore, possible to apply the results on the exact variance of G_1^* (given in section 3.4a) to the estimators: $S^*_{(q<o)}$ and hence $\hat{Y}_{RP}$, $S^*_{(\pm 1)}$, and TH_1^*

Theorem 6.4.1 : The exact variance of the unbiased product type estimator $S^*_{(q<o)}$ (6.2.1) is given by expression (4.3.3). When $N >> n$, the exact variance may be expressed in the form (4.3.4).

Proof: We note that $S^*_{(q)}$ (4.2.1) and $S^*_{(q<o)}$ (6.2.1) have the same expression. Hence the results follow.

Corollary 6.4.1: The exact variance of Robson's unbiased product type estimator $S^*_{(-1)} = \hat{Y}_{RP}$ (6.2.2) is given by

$$V(\hat{Y}_{RP}) = \left(\frac{1}{n}-\frac{1}{N}\right)\Big[\frac{1}{(N-1)}\sum_{(1)}^{N}(y_j-\bar{Y})^2 + \frac{2}{N(N-1)\bar{X}}\Big\{\sum_{(2)}^{N} y_j x_j y_{j'} - \frac{1}{(N-2)}\sum_{(3)}^{N} y_j x_{j'} y_{j''}\Big\}$$

$$+ \frac{1}{N^2(n-1)\bar{X}^2}\Big\{\sum_{(2)}^{N} x_j^2 y_{j'}^2 + \frac{(N-n)}{(N-1)}\sum_{(2)}^{N} x_j y_j x_{j'} y_{j'} + \frac{(Nn-2N+1)}{(N-1)(N-2)}\sum_{(3)}^{N} x_j^2 y_{j'} y_{j''}$$

$$- \frac{2(N-n)}{(N-1)(N-2)}\sum_{(3)}^{N} x_j y_j x_{j'} y_{j''} - \frac{1}{(N-2)}\sum_{(3)}^{N} x_j y_{j'}^2 x_{j''}$$

$$- \frac{(Nn-3N+n+1)}{(N-1)(N-2)(N-3)}\sum_{(4)}^{N} x_j y_{j'} x_{j''} y_{j'''}\Big\}\Big]. \qquad (6.4.1)$$

and when $N >> n$

$$V(\hat{Y}_{RP}) = \frac{1}{n}\Big[V(y) + 2R\,\mathrm{Cov}(y,x) + R^2V(x)\Big] + \frac{1}{n(n-1)\bar{X}^2}\Big[V(x)\,V(y) + \mathrm{Cov}^2(x,y)\Big]. \qquad (6.4.2)$$

Proof: We substitute $q = -1$ in the expressions (4.3.3) and (4.3.4) and obtain respectively the expressions (6.4.1) and (6.4.2).

Remark (a): With a little algebra it may be shown that (6.4.1) is identical with the exact variance formula given by Robson (1957). Robson has also given the form (6.4.2). It may be noted that the unbiased estimator $\hat{Y}_{RP}$ and the biased product estimator $\hat{Y}_P = \frac{\bar{y}\,\bar{x}}{\bar{X}}$ have the same large sample variance.

Theorem 6.4.2 : When the finite population correction is ignored, for the unbiased product type estimator $S^*_{(q<o)}$ (6.2.1), the optimum choice q_o of q is given by (4.3.12) and the optimum variance by (4.3.13). The estimator $S^*_{(q<o)}$ is more efficient than Robson's unbiased product type estimator $\hat{Y}_{RP}$. A good approximation to q_o is $q_o' = \rho \frac{C_y}{C_x}$ and the variance of $S^*_{(q_o'<o)}$, is given by (4.3.16). In large samples, the unbiased product type estimators $S^*_{(q_o<o)}$ and $S^*_{(q_o'<o)}$ have the same variance as the classical linear regression estimator $\hat{Y}_r$ and are superior to the conventional product estimator $\hat{Y}_P$.

Proof : These results are similar to the results given in theorem 4.3.4, and remarks (f) and (g) of section 4.3.

Remark (b) : Srivastava (1967) has studied a biased product estimator: $\hat{Y}_P' = \bar{y}\left(\frac{\bar{X}}{\bar{x}}\right)^q$ with $q < 0$, and observed that $q_o' = \rho \frac{C_y}{C_x}$ is the optimum choice of q in large samples. With q_o' the estimator has the same large sample variance as the classical regression estimator $\hat{Y}_r$.

Remark (c) : Nagi Reddy (1972) has considered the use of the product estimator: $\hat{Y}_{(q<o)} = \frac{\bar{y}\,\bar{X}}{[\bar{X} + q(\bar{x} - \bar{X})]}$ with $q < 0$, assuming that $\left|\frac{q(\bar{x} - \bar{X})}{\bar{X}}\right| < 1$. Again in large samples the optimum choice of q is $\rho \frac{C_y}{C_x}$.

The estimator with the optimum choice of q has only second degree (n^{-2}) bias and its large sample variance is the same as that of the classical regression estimator $\hat{\bar{Y}}_r$

Remark (d) : Our observations under remark (k) of section 4.3 (in respec of the unbiased and biased ratio estimators using information on $\rho \frac{C_Y}{C_X}$) are equally applicable to the above unbiased and biased product estimators using $\rho \frac{C_Y}{C_X}$.

We now apply to the unbiased ratio-cum-product type estimators $S^*_{(\pm 1)}$ (6.3.2) and TH^*_1 (6.3.7) the exact variance formula (3.4.32) of the estimator G^*_1 (3.4.4) and determine the optimum choice for the weight vector W. In this connection the following theorem is useful.

Theorem 6.4.3 : Let $a_i (Z_1) = a_{ij}$ be a function of y and x_i, and define $\tilde{a}_i (Z_1) = W_i a_{ij}$ with $\sum_{i=1}^{p} W_i = 1$. Then G^*_1 (3.4.4) using the coefficient functions $\tilde{a}_i (Z_1)$ may be expressed in the form:

$$G^*_1 = \sum_{i=1}^{p} W_i \; G^*_{i1} \; , \tag{6.4.3}$$

$$\text{where } G^*_{i1} = \bar{y} - \bar{a}_i (\bar{x}_i - \bar{X}_i) + (\frac{1}{n} - \frac{1}{N}) s_{a_i x_i} . \tag{6.4.4}$$

Further, when $N >> n$,

$$V (G^*_1) = W F W' , \tag{6.4.5}$$

where $W = (W_1, W_2, \ldots, W_p)$ and $F = (f_{ii'})$ is a (p x p) matrix with

$$f_{ii'} = \frac{1}{n} \left[V(y) - E(a_i)\, Cov(y, x_i) - E(a_{i'})\, Cov(y, x_{i'}) + E(a_i) E(a_{i'})\, Cov(x_i, x_{i'}) \right]$$

$$+ \frac{1}{n(n-1)} \left[Cov(x_i, x_{i'})\, Cov(a_i, a_{i'}) + Cov(x_i, a_{i'})\, Cov(x_{i'}, a_i) \right]. \qquad (6.4.6)$$

When F^{-1} exists, the optimum choice for W which minimises $V(G_1^*)$, subject to $\sum_{i=1}^{p} W_i = 1$, is given by

$$W_o = \frac{e\, F^{-1}}{e\, F^{-1} e'}, \qquad (6.4.7)$$

and with the optimum choice W_o the minimum variance is

$$V_o(G_1^*) = \frac{1}{e\, F^{-1} e'}, \qquad (6.4.8)$$

where e is a unit row vector with p elements.

Proof : The result (6.4.3) follows from (3.5.5) for m = 1. The result (6.4.5) is obtained by writing $W_i a_i$ for a_i in the variance formula (3.4.32 and expressing the result in matrix notation. Finally the results (6.4.7) and (6.4.8) are obtained following the well-known procedure due to Olkin (1958).

Remark (c) : While determining the optimum choice for W, if it is decided to consider the variance only upto order n^{-1}, we may define a (p x p) matrix $F_1 = (f_{ii'}^{(1)})$, where

$$f_{ii'}^{(1)} = \frac{1}{n} \left[V(y) - E(a_i)\, Cov(y, x_i) - E(a_{i'})\, Cov(y, x_{i'}) + E(a_i) E(a_{i'})\, Cov(x_i, x_{i'}) \right]; \qquad (6.4.9)$$

and minimise $V^{(1)}(G_1^*) = W\, F_1\, W'$. In that case the optimum choice is

$W_o^{(1)} = \frac{e F_1^{-1}}{e F_1^{-1} e'}$, and the minimum variance upto order n^{-1} is

$V_o^{(1)}(G_1^*) = \frac{1}{e F_1^{-1} e'}$. In the above theorem, terms of order n^{-2} are also retained in finding the optimum W with a view to accomodating even small sample sizes.

<u>Theorem 6.4.4</u>: When $N >> n$, the exact variance of the unbiased ratio-cum-product type estimator $S^*_{(\pm 1)}$ (6.3.2) is given by

$$V[S^*_{(\pm 1)}] = W F_s W', \tag{6.4.10}$$

where $F_s = (f^s_{ii'})$ is a $(p \times p)$ matrix,

$$f^s_{ii'} = f_{ii'}\ (6.4.6) \text{ with } a_{ij} = \frac{y_j}{\bar{X}_i} \text{ for } (i=1,2,\ldots,p_1),$$
$$= -\frac{y_j}{\bar{X}_i} \text{ for } (i=p_1+1,\ldots,p). \tag{6.4.11}$$

Assuming that F_s^{-1} exists, the optimum choice for W which minimises $V[S^*_{(\pm 1)}]$, subject to $\sum_{i=1}^{p} W_i = 1$, is given by

$$W_o^s = \frac{e F_s^{-1}}{e F_s^{-1} e'}, \tag{6.4.12}$$

and with W_o^s the minimum variance is

$$V_o[S^*_{(\pm 1)}] = \frac{1}{e F_s^{-1} e'} \tag{6.4.13}$$

<u>Proof</u>: Follows from the definition of $S^*_{(\pm 1)}$ (6.3.2) and theorem 6.4.3.

Remark (f) : To compute the optimum weight vector W_o^* (6.4.12), it is necessary to compute $f_{ii'}^*$ (6.4.11) which involves the population parameters

$\overline{X}_i$, $R_i = \frac{\overline{Y}}{\overline{X}_i}$, $V(y)$, $Cov(y, x_i)$, and $Cov(x_i, x_{i'})$ for$(i, i')=1, 2, \ldots, p$.

Information on $\overline{X}_i$ is assumed to be available. For other parameters information obtained from past experience may be used. The unbiased nature of the estimator $S^*_{(\pm 1)}$ is not disturbed due to errors in the estimates of the weights W_i.

Theorem 6.4.5: When $N >> n$, the exact variance of the unbiased ratio-cum-product type estimator TH_1^* (6.3.7) is given by

$$V(TH_1^*) = W F_{th} W', \tag{6.4.14}$$

where $F_{th} = (f_{ii'}^{th})$ is a $(p \times p)$ matrix,

$$\begin{aligned} f_{ii'}^{th} &= f_{ii'} \text{ (6.4.6) with } a_{ij} = r_{ij} \text{ for } (i=1, 2, \ldots, p_1), \\ &= -\frac{y_j}{X_i} \text{ for } (i=p_1+1, \ldots, p). \end{aligned} \tag{6.4.15}$$

Assuming that F_{th}^{-1} exists, the optimum choice for W which minimises $V(TH_1^*)$, subject to $\sum_{i=1}^{p} W_i = 1$, is given by

$$W_o^{th} = \frac{e F_{th}^{-1}}{e F_{th}^{-1} e'}, \tag{6.4.16}$$

and with W_o^{th} the minimum variance is $V_o(TH_1^*) = \frac{1}{e F_{th}^{-1} e'}$. (6.4.17)

<u>Proof</u>: Follows from the definition of TH_1^* (6.3.7) and theorem (6.4.3).

<u>Remark</u> (g): To compute $f_{ii'}^{th}$ we need information on the following population parameters:

$$V(y),\ Cov(y, x_i),\ Cov(x_i, x_{i'})\ for (i, i') = 1, 2, \ldots, p;$$

$$E(r_i),\ Cov(y, r_i),\ Cov(r_i, r_{i'})\ for\ (i, i') = 1, 2, \ldots, p_1;$$

$$\bar{X}_i\ ,\quad R_i\quad for\ i = p_1+1, \ldots, p;$$

$$Cov(r_i, x_{i'})\quad for\ i = 1, 2, \ldots, p_1,$$
$$i' = 1, 2, \ldots, p_1, p_1+1, \ldots, p.$$

We may use information obtained from past experience without disturbing the unbiased nature of the estimator TH_1^*. Comparing the above requirement of population parameters with that in remark (f), we note that information is needed on <u>less</u> number of population parameters to compute the optimum weight vector for $S^*_{(\pm 1)}$.

6.5. <u>Large sample variance of unbiased product and ratio-cum-product type estimators:</u>

In this section we investigate, upto order n^{-1}, the variance of the unbiased product and ratio-cum-product type estimators: H_m^* (6.2.3), L_m^* (6.2.5), TH_m^* (6.3.4), and TL_m^* (6.3.6), making use of our results in section 3.4b on the large sample variance of the general unbiased estimator G_m^* (3.4.3).

<u>Theorem 6.5.1</u> : In large samples the variance of the unbiased product type estimator H_m^* (6.2.3) is given by

$$V(H_m^*) = \left(\frac{1}{n} - \frac{1}{N}\right)\left[S_y^2 + 2\,\frac{E(P_m)}{\bar{X}}\,S_{xy} + \frac{E^2(P_m)}{\bar{X}^2}\,S_x^2\right], \quad (6.5.1)$$

where $P_m = \left(\sum_{j \in Z_m} y_j x_j\right) / m\,\bar{x}_m$.

<u>Proof</u> : Let us consider two cases.

<u>Case</u> (1): n is large and m is <u>not small</u> compared to n.

We recall that H_m^* (6.2.3) results from G_m^* (3.4.3) with a single auxiliary variate x and $a(Z_m) = -\frac{P_m}{\bar{X}}$ Writing

$$-\bar{X}\,a\,(Z_m) = P_m = \frac{\sum_{j \in Z_m} y_j (x_j - \bar{x}_m)}{m\,\bar{x}_m} + \bar{y}_m$$

$$= \left(\frac{m-1}{m}\right)\frac{s_{xy}(Z_m)}{\bar{x}_m} + \bar{y}_m , \qquad (6.5.2)$$

We have

$$-\bar{X}\,a^*(Z_m) = -\bar{X}\,E_n\left[a(Z_m)\right] = \left(\frac{m-1}{m}\right)E_n\left[\frac{s_{xy}(Z_m)}{\bar{x}_m}\right] + \bar{y}. \qquad (6.5.3)$$

Since m is not small compared to n, we may assume $\left|\frac{\bar{x}_m - \bar{x}}{\bar{x}}\right| < 1$.

Then it may be seen that

$$E_n\left[\frac{s_{xy}(Z_m)}{\bar{x}_m}\right] \doteq \frac{s_{xy}}{\bar{x}}\left[1 + \frac{V_n(\bar{x}_m)}{\bar{x}^2} - \frac{Cov_n\{\bar{x}_m, s_{xy}(Z_m)\}}{\bar{x}\;s_{xy}}\right],$$

$$\doteq \frac{s_{xy}}{\bar{x}} \left[1 + \left(\frac{1}{n} - \frac{1}{m} \right) \left(\frac{s_x^2}{\bar{x}^2} - \frac{m_{12}^{yx}}{\bar{x}\, s_{xy}} \right) \right], \qquad (6.5.4)$$

where $m_{12}^{yx} = \frac{1}{n} \sum_{j=1}^{n} (y_j - \bar{y}) (x_j - \bar{x})^2$

From (6.5.3) and (6.5.4) we have

$$a^*(Z_m) \doteq - \frac{1}{\bar{X}} \left[\bar{y} + \left(\frac{m-1}{m} \right) \frac{s_{xy}}{\bar{x}} \left\{ 1 + \left(\frac{1}{n} - \frac{1}{m} \right) \left(\frac{s_x^2}{\bar{x}^2} - \frac{m_{12}^{yx}}{\bar{x}\, s_{xy}} \right) \right\} \right]$$

(6.5.5)

From (6.5.5), it now follows that, in large samples, $\mu_2 \left[a^*(Z_m) \right]$ and $\mu_4 \left[a^*(Z_m) \right]$ are of same order in n as $\mu_2 (\bar{y})$ and $\mu_4 (\bar{y})$ respectiv
Hence we obtain

$$\mu_2 \left[a^*(Z_m) \right] = O (n^{-1}) , \text{ and } \mu_4 \left[a^*(Z_m) \right] = O (n^{-2}). \qquad (6.5.6)$$

Now consider

$$\mu_2 \left[Cov_n \left\{ a(Z_m),\ \bar{x}_{n-m} \right\} \right] = \mu_2 \left[- \frac{m}{(n-m)} \, Cov_n \left\{ a(Z_m),\ \bar{x}_m \right\} \right] \qquad (6.5.7)$$

In view of (6.5.2) and (6.5.4) we have

$$- \bar{X}\, Cov_n \left[a(Z_m), \bar{x}_m \right] = Cov_n \left[\left(\frac{m-1}{n} \right) \frac{s_{xy}(Z_m)}{\bar{x}_m} + \bar{y}_m ,\ \bar{x}_m \right],$$

$$= \left(\frac{m-1}{m} \right) \left[E_n \left\{ s_{xy}(Z_m) \right\} - \bar{x}\, E_n \left\{ \frac{s_{xy}(Z_m)}{\bar{x}_m} \right\} \right] + Cov_n (\bar{y}_m ,$$

$$\doteq \left(\frac{1}{n} - \frac{1}{m} \right) s_{xy} \left[\left(\frac{m-1}{m} \right) \left(\frac{m_{12}^{yx}}{\bar{x}\, s_{xy}} - \frac{s_x^2}{\bar{x}^2} \right) + 1 \right].$$

$$\text{Hence } -\frac{m}{(n-m)} \; Cov_n [\, a(Z_m), \bar{x}_m \,] \doteq \left(\frac{s_{xy}}{n\bar{X}} \right) [\, 1 + \left(\frac{m-1}{m} \right) \left(\frac{m_{12}^{yx}}{\bar{x}\, s_{xy}} - \frac{s_x^2}{\bar{x}^2} \right)]. \tag{6.5.8}$$

From (6.5.7) and (6.5.8) it follows that, in large samples,

$$\mu_2 [\, Cov_n \{ a(Z_m), \; \bar{x}_{n-m} \} \,] = O(n^{-3}) . \tag{6.5.9}$$

Now (6.5.6) and (6.5.9) imply that the conditions (3.4.37) of theorem 3.4.2 are satisfied for the coefficient function $a(Z_m) = -\frac{P_m}{\bar{X}}$. Consequently, in this case, from theorem 3.4.2 we obtain the result (6.5.1).

<u>Case</u> (ii) : n is large and m is <u>small</u> compared to n .

Here we apply theorem 3.4.3. and obtain the result (6.5.1).

This completes the proof of the theorem.

<u>Corollary 6.5.1</u> : In large samples Robson's unbiased product type estimator $\hat{\bar{Y}}_{RP}$ (6.2.2) has the variance:

$$V(\hat{\bar{Y}}_{RP}) = \left(\frac{1}{n} - \frac{1}{N} \right) [\, S_y^2 + 2R\, S_{xy} + R^2 S_x^2 \,]. \tag{6.5.10}$$

<u>Proof</u> : From corollary 6.2.2 we note $H_1^* = \hat{\bar{Y}}_{RP}$ (6.2.2). When m=1, $P_m = y_j$ and $\frac{E(P_m)}{\bar{X}} = R$. Hence from theorem 6.5.1 we obtain $V(H_1^*) = V(\hat{\bar{Y}}_{RP}) = (6.5.10)$.

<u>Corollary 6.5.2</u> : In large samples, for any large m,

$$V(H_m^*) = \left(\frac{1}{n} - \frac{1}{N} \right) [\, S_y^2 + 2\left(R + \frac{S_{xy}}{\bar{X}^2} \right) S_{xy} + \left(R + \frac{S_{xy}}{\bar{X}^2} \right)^2 S_x^2 \,]. \tag{6.5.11}$$

Proof : When m is large, we may assume $\left|\frac{\bar{x}_m - \bar{X}}{\bar{X}}\right| < 1$. Then from (6.5.2) we have

$$\frac{E(P_m)}{\bar{X}} = \frac{1}{\bar{X}} E\left[\left(\frac{m-1}{m}\right) \frac{s_{xy}(Z_m)}{\bar{x}_m} + \bar{y}_m \right],$$

$$\doteq \left(\frac{S_{xy}}{\bar{X}^2} + R \right) + O(m^{-1}). \qquad (6.5.12)$$

Substituting (6.5.12) in (6.5.1) we obtain, upto order n^{-1}, $V(H_m^*) = (6.5.11)$.

Theorem 6.5.2 : For the unbiased product type estimator H_m^* (6.2.3) in large samples, from the point of view of precision the choice $m = n-1$ is superior or equal or inferior to the choice $m = 1$, according as

$$\left|\rho \frac{C_y}{C_x}\right| \lesseqgtr \frac{2}{2 + C_x^2}. \qquad (6.5.13)$$

Proof : Since $H_1^* = \hat{\bar{Y}}_{RP}$ (6.2.2), to obtain a comparison of the large sample variances of H_1^* and H_{n-1}^* the variances (6.5.10) and (6.5.11) may be used. Thus we have

$$V(H_{n-1}^*) - V(H_1^*) = \left(\frac{1}{n} - \frac{1}{N}\right)\left[\frac{2 S_{xy}^2}{\bar{X}^2} + \frac{S_{xy}^2 S_x^2}{\bar{X}^4} + \frac{2R\, S_{xy} S_x^2}{\bar{X}^2} \right],$$

$$= 2\left(\frac{1}{n} - \frac{1}{N}\right) \frac{S_{xy}^2}{\bar{X}^2} \left[1 + \frac{C_x^2}{2} + \frac{C_x}{\rho C_y} \right].$$

(6.5.14)

$\rho < 0$, we have

$$\left[1 + \frac{C_x^2}{2} + \frac{C_x}{\rho C_y} \right] \lesseqgtr 0, \text{ according as } \left|\frac{C_x}{\rho C_y}\right| \gtreqless \frac{2 + C_x^2}{2}$$

Hence from (6.5.14) it follows that

$$V(H^*_{n-1}) \lesseqgtr V(H^*_1), \text{ according as } \left|\frac{\rho C_y}{C_x}\right| \lesseqgtr \frac{2}{2+C_x^2}.$$

Remark (a) : Since the conventional product estimator $\hat{\bar{Y}}_P = \frac{\bar{y}\,\bar{x}}{\bar{X}}$ and the unbiased product type estimator $\hat{\bar{Y}}_{RP}$ (6.2.2) have the same large sample variance, and $\hat{\bar{Y}}_{RP} = H^*_1$, theorem 6.5.2 incidentally provides an efficiency comparison of H^*_{n-1} with $\hat{\bar{Y}}_P$ and $\hat{\bar{Y}}_{RP}$.

We now consider the derivation of the large sample variance of the unbiased product type estimator L^*_m (6.2.5).

Theorem 6.5.3 : In large samples the variance of the unbiased product type estimator L^*_m (6.2.5) is given by

$$V(L^*_m) = \left(\frac{1}{n} - \frac{1}{N}\right)\left[S_y^2 + 2\left\{\frac{R + (m-1)\bar{r}_N}{m}\right\} S_{xy} + \left\{\frac{R + (m-1)\bar{r}_N}{m}\right\}^2 S_x^2\right], \tag{6.5.15}$$

where $\bar{r}_N = E(r)$.

Proof : We note that L^*_m is obtained from G^*_m (3.4.3) with a single auxiliary variate x and the choice $a(Z_m) = -\frac{\bar{r}_m \bar{x}_m}{\bar{X}}$. For this choice of the coefficient function $a(Z_m)$, from (6.2.6) it may be seen that $\mu_2[a^*(Z_m)] = O(n^{-1})$ and $\mu_4[a^*(Z_m)] = O(n^{-2})$. Using (6.2.7) we have

$$\text{Cov}_n[a(Z_m), \bar{x}_{n-m}] = -\frac{m}{(n-m)} \text{Cov}_n[a(Z_m), \bar{x}_m]$$

$$= \frac{1}{\bar{X}} \left[\frac{(n-2m)\, m_{12}^{rx}}{m(n-1)(n-2)} + \frac{1}{(n-1)} \left(\bar{x}\, m_{11}^{rx} + \bar{r}\, m_{02}^{rx} \right) \right].$$

Hence it follows that

$$\mu_2 \left[\mathrm{Cov}_n \left\{ a(Z_m),\ \bar{x}_{n-m} \right\} \right] = O(n^{-3}) .$$

Thus all the three conditions given in (3.4.37) hold for the coefficient function $a(Z_m) = -\frac{\bar{r}_m\, \bar{x}_m}{\bar{X}}$ for any choice of m in the range 1 to $n-1$. Consequently, applying theorem 3.4.2, in large samples we obtain

$$V(L_m^*) = \left(\frac{1}{n} - \frac{1}{N}\right) \left[S_y^2 + \frac{2E(\bar{r}_m\, \bar{x}_m)}{\bar{X}} S_{xy} + \frac{E^2(\bar{r}_m\, \bar{x}_m)}{\bar{X}^2} S_x^2 \right] \quad (6.5.16)$$

Now we note

$$\frac{E(\bar{r}_m\, \bar{x}_m)}{\bar{X}} = \frac{1}{\bar{X}} \left[\mathrm{Cov}(\bar{r}_m, \bar{x}_m) + \bar{r}_N\, \bar{X} \right]$$

$$= \frac{1}{\bar{X}} \left[\frac{(N-m)}{Nm} \cdot \frac{1}{(N-1)} \left(\sum_{j=1}^{N} r_j x_j - N\, \bar{r}_N\, \bar{X} \right) + \bar{r}_N\, \bar{X} \right]$$

$$= \frac{(N-m)}{m(N-1)}\, R + \frac{N(m-1)}{m(N-1)}\, \bar{r}_N$$

$$= \frac{R + (m-1)\, \bar{r}_N}{m} + O(N^{-1}). \quad (6.5.17)$$

Using (6.5.17) in (6.5.16) we get the result (6.5.15).

Corollary 6.5.3 : In large samples the variance of the unbiased product type estimator L_1^*, which is identical with Robson's unbiased product type estimator $\hat{Y}_{RP}$ (6.2.2), is given by (6.5.10).

Proof : Writing $m = 1$ in (6.5.15) of theorem 6.5.3, we obtain (6.5.10).

Theorem 6.5.4 : The large sample variance (6.5.15) of the unbiased product type estimator L_m^* (6.2.5) is same for any choice of m in the range 1 to n-1 if $R = \bar{r}_N$. When R and $\bar{r}_N$ are positive, and if either $R > \bar{r}_N > |\beta|$ or $R < \bar{r}_N < |\beta|$, the variance decreases as m increases; so that in these cases $m = n-1$ is the optimum choice for L_m^*.

Proof : When $R = \bar{r}_N$, (6.5.15) reduces to

$$V(L_m^*) = \left(\frac{1}{n} - \frac{1}{N}\right)\left[S_y^2 + 2\bar{r}_N S_{xy} + \bar{r}_N^2 S_x^2\right], \qquad (6.5.18)$$

which is the same for $1 \leqslant m \leqslant n-1$.

Suppose $R \neq \bar{r}_N$. We may write (6.5.15) in the form:

$$V(L_m^*) = \left(\frac{1}{n} - \frac{1}{N}\right)\left[\left(S_y^2 - \frac{S_{xy}^2}{S_x^2}\right) + \left(\frac{R + (m-1)\bar{r}_N}{m} + \beta\right)^2 S_x^2\right], (6.5.19)$$

and observe that $V(L_m^*)$ decreases as m increases, provided

$$\left|\frac{R + (m-1)\bar{r}_N}{m} + \beta\right| = \left|\frac{(R - \bar{r}_N)}{m} + (\bar{r}_N + \beta)\right| \text{ decreases.}$$

Now, if $(R - \bar{r}_N)$ and $(\bar{r}_N + \beta)$ are of same sign,

$$\left|\frac{(R - \bar{r}_N)}{m} + (\bar{r}_N + \beta)\right| = \frac{|R - \bar{r}_N|}{m} + |\bar{r}_N + \beta|$$

decreases as m increases. Consequently, when $(R - \bar{r}_N)$ and $(\bar{r}_N + \beta)$ of same sign, $V(L_m^*)$ given by (6.5.19) decreases as m increases, and $m = n-1$ is the optimum choice for L_m^*.

Now assume $R > 0$ and $\bar{r}_N > 0$. Since x and y are negatively associated, $\beta < 0$. Then both $(R - \bar{r}_N)$ and $(\bar{r}_N + \beta)$ are positive if $R > \bar{r}_N > |\beta|$. Similarly both $(R - \bar{r}_N)$ and $(\bar{r}_N + \beta)$ are negative if $R < \bar{r}_N < |\beta|$. This completes the proof.

We are now in a position to study the large sample variance of e unbiased ratio-cum-product type estimators TH_m^* (6.3.4) and TL_m^* (6.3.6).

heorem 6.5.5: In large samples the variance of the unbiased ratio-cum-roduct type estimator TH_m^* (6.3.4) is given by

$$V(TH_m^*) = W F_{thm} W', \tag{6.5.20}$$

ere $F_{thm} = (f_{ii'}^{thm})$ is a $(p \times p)$ matrix,

$$f_{ii'}^{thm} = \left(\frac{1}{n} - \frac{1}{N}\right)\Big[S_y^2 - E\{a_i(Z_m)\} S_{x_i y} - E\{a_{i'}(Z_m)\} S_{x_{i'} y} + E\{a_i(Z_m)\} E\{a_{i'}(Z_m)\} S_{x_i x_{i'}} \Big], \tag{6.5.21}$$

$$\begin{aligned} a_i(Z_m) &= R_{im} \quad \text{for } (i = 1, 2, \ldots, p_1), \\ &= -\frac{P_{im}}{\bar{X}_i} \quad \text{for } (i = p_1+1, \ldots, p). \end{aligned} \tag{6.5.22}$$

Assuming that F_{thm}^{-1} exists, the optimum choice for W which imises $V(TH_m^*)$, subject to $\sum_{i=1}^{p} W_i = 1$, is given by

$$W_o^{thm} = \frac{e\, F_{thm}^{-1}}{e\, F_{thm}^{-1}\, e'}, \tag{6.5.23}$$

and with W_o^{thm} the minimum variance is

$$V_o (TH_m^*) = \frac{1}{e F_{thm}^{-1} e'} , \qquad (6.5.24)$$

where e is a unit row vector of p elements.

Proof: In theorems 4.4.1 and 6.5.1 we have shown that, for the coefficient functions $a_i (Z_m)$ given in (6.5.22), the conditions (3.4.37) are satisfied when n is large and m is not small compared to n. Hence in that case, by applying theorem 3.4.2 to the estimator TH_m^* (6.3.4), we get the result (6.5.20). When n is large and m is small compared to n, we apply theorem 3.4.3 and obtain (6.5.20). The results (6.5.23) and (6.5.24) are obtained from (6.5.20) following the method due to Olkin (1958).

Remark (b) : $E [a_i (Z_m)]$ appearing in (6.5.21) may be explicitely written for the choices m = 1 and m = n-1.

For m = 1, $E [a_i (Z_m)] = E (r_{1j}) = \bar{r}_{iN}$ for $(i = 1, 2, \ldots, p_1)$,

$$= \frac{E (- y_j)}{\bar{X}_i} = - R_i \text{ for } (i = p_1 + 1, \ldots , p). \qquad (6.5.25)$$

For m = n-1, from (6.5.12),

$$E [a_i (Z_m)] = E (R_{i\,n-1}) \doteq R_i \text{ for } (i = 1, 2, \ldots , p_1),$$

$$= - \frac{E (P_{i\,n-1})}{\bar{X}_i} = -[R_i + \frac{S_{x_i y}}{\bar{X}_i^2}]$$

$$\text{for } (i = p_1 + 1, \ldots , p). \qquad (6.5.26)$$

Theorem 6.5.6 : In large samples the variance of the unbiased ratio-cum-product type estimator TL_m^* (6.3.6) is given by

$$V(TL_m^*) = W \quad F_{tlm} W' , \tag{6.5.27}$$

where $F_{tlm} = (f_{ii'}^{tlm})$ is a $(p \times p)$ matrix, $f_{ii'}^{tlm}$ = expression (6.5.21) with $a_i(Z_m) = R_{im}$ for $(i = 1, 2, \ldots, p_1)$,

$$= - \frac{\bar{r}_{im} \bar{x}_{im}}{\bar{X}_i} \text{ for } (i = p_1+1, \ldots, p). \tag{6.5.28}$$

Assuming that F_{tlm}^{-1} exists, the optimum choice for W which minimises $V(TL_m^*)$, subject to $\sum_{i=1}^{p} W_i = 1$, is given by

$$W_o^{tlm} = \frac{e \; F_{tlm}^{-1}}{e \; F_{tlm}^{-1} \; e'} , \tag{6.5.29}$$

and with W_o^{tlm} the minimum variance is

$$V_o [TL_m^*] = \frac{1}{e \; F_{tlm}^{-1} \; e'} . \tag{6.5.30}$$

Proof : It is similar to the proof of theorem 6.5.5. Here we use theorems 4.4.1, 6.5.3, 3.4.2 and 3.4.3.

Remark (c) : In the above theorem, for $m = 1$,

$$E[a_i(Z_m)] = \bar{r}_{iN} \text{ for } (i = 1, 2, \ldots, p_1),$$

$$= - R_i \text{ for } (i = p_1+1, \ldots, p).$$

For $m = n-1$, $E[\, a_i(Z_m)\,] \doteq R_i$ for $(i = 1, 2, \ldots, p_1)$,

$$\doteq -\left[\frac{R_i + (n-2)\,\bar{r}_{iN}}{(n-1)} \right] \text{ for } (i = p_1+1, \ldots, p). \tag{6.5.31}$$

6.6. <u>Estimation of variance of unbiased product type estimators:</u>

In this section we develop estimators of variance for the unbiased product type estimators: $S^*_{(q_o' < o)}$, H^*_m and L^*_m, using the results on G^*_m (3.4.3) in section 3.6.

Using theorem 3.6.1 and corollary 3.6.1, it may be noted that an unbiased estimator of variance of $S^*_{(q_o' < o)}$ is given by (4.5.4) with $q = q_o' = \rho \dfrac{C_y}{C_x}$, and that of H^*_m is obtained as

$$\hat{V}(H^*_m) = \frac{1}{\binom{n}{m}} \sum_{Z_m}^{\binom{n}{m}} \left[\hat{V}(H_m) - (H_m - H^*_m)^2 \right], \tag{6.6.1}$$

where $\hat{V}(H_m) = \dfrac{(N-n)(N-m)}{N^2(n-m)} \cdot \dfrac{1}{(n-m-1)} \sum_{j \notin Z_m}^{n-m} \left[(y_j - \bar{y}_{n-m}) + \dfrac{P_m}{\bar{X}} (x_j - \bar{x}_{n-m}) \right]^2$

for $1 \leq m \leq n-2$,

$$= \frac{(N-n+1)^2}{N^2} \left[U^2_{n-1}(H) - y_j^2 \right] + \frac{(N-n+1)(N-n)}{N^2} s_y^2 \text{ for } m = n-1,$$

$U_{n-1}(H) = y_j + \dfrac{P_{n-1}}{\bar{X}} (x_j - \bar{X}_{N-n+1})$, $j \notin Z_{n-1}$, and

$$H_m = \left[\bar{y} + \frac{P_m}{\bar{X}} (\bar{x} - \bar{X}) \right] - \frac{m(N-n)}{(n-m)N} \left[(\bar{y}_m - \bar{y}) + \frac{P_m}{\bar{X}} (\bar{x}_m - \bar{x}) \right].$$

Similarly, an unbiased estimator of variance of L_m^* is :

$$\hat{V}(L_m^*) = \frac{1}{\binom{n}{m}} \sum_{Z_m}^{\binom{n}{m}} \left[\hat{V}(L_m) - (L_m - L_m^*)^2 \right], \qquad (6.6.2)$$

where $\hat{V}(L_m) = \frac{(N-n)(N-m)}{N^2(n-m)} \cdot \frac{1}{(n-m-1)} \sum_{j \notin Z_m}^{n-m} \left[(y_j - \bar{y}_{n-m}) + \frac{\bar{r}_m \bar{x}_m}{\bar{X}} (x_j - \bar{x}_{n-m}) \right]^2$

for $1 \leq m \leq n-2$,

$$= \frac{(N-n+1)^2}{N^2} \left[U_{n-1}^2(L) - y_j^2 \right] + \frac{(N-n+1)(N-n)}{N^2} s_y^2 \quad \text{for } m = n-1,$$

$$U_{n-1}(L) = y_j + \frac{\bar{r}_{n-1} \bar{x}_{n-1}}{\bar{X}} (x_j - \bar{X}_{N-n+1}), \quad j \notin Z_{n-1}, \text{ and}$$

$$L_m = \left[\bar{y} + \frac{\bar{r}_m \bar{x}_m}{\bar{X}} (\bar{x} - \bar{X}) \right] - \frac{m(N-n)}{(n-m)N} \left[(\bar{y}_m - \bar{y}) + \frac{\bar{r}_m \bar{x}_m}{\bar{X}} (\bar{x}_m - \bar{x}) \right].$$

In particular, for Robson's unbiased product type estimator $\hat{\bar{Y}}_{RP}$ (6.2.2), one may use $\hat{V}(H_1^*)$ or $\hat{V}(L_1^*)$ as unbiased estimators of variance.

In large samples, in view of corollary 4.3.2 and corollary 6.5.1, we may use as estimators of variance of $S^*_{(q_o' < o)}$ and $\hat{\bar{Y}}_{RP}$ the following:

$$\hat{V}' \left[S^*_{(q_o' < o)} \right] = \left(\frac{1}{n} - \frac{1}{N} \right) \left[s_y^2 - \frac{s_{xy}^2}{s_x^2} \right], \qquad (6.6.3)$$

and $$\hat{V}'(\hat{\bar{Y}}_{RP}) = \left(\frac{1}{n} - \frac{1}{N} \right) \left[s_y^2 + 2 \left(\frac{\bar{y}}{\bar{x}} \right) s_{xy} + \left(\frac{\bar{y}}{\bar{x}} \right)^2 s_x^2 \right]. \qquad (6.6.4)$$

Incidentally (6.6.4) also estimates the large sample variance of the conventional product estimator $\hat{Y}_P = \frac{\bar{y}\,\bar{x}}{\bar{X}}$

Using the results (6.5.1), (6.5.16) and (3.6.13), in large samples, for the estimators H_m^* and L_m^* we may have the variance estimators:

$$\hat{V}(H_m^*) = \left(\frac{1}{n}-\frac{1}{N}\right)\left[s_y^2 + 2\,\frac{P_m^*}{\bar{X}}\, s_{xy} + \left(\frac{P_m^*}{\bar{X}}\right)^2 s_x^2 \right], \qquad (6.6.5)$$

$$\hat{V}(L_m^*) = \left(\frac{1}{n}-\frac{1}{N}\right)\left[s_y^2 + \frac{2\,(\bar{r}_m\,\bar{x}_m)^*}{\bar{X}}\, s_{xy} + \left\{(\bar{r}_m\,\bar{x}_m)^*\right\}^2 \frac{s_x^2}{\bar{X}^2} \right], \qquad (6.6.6)$$

where $(\bar{r}_m\,\bar{x}_m)^* = \frac{1}{\binom{n}{m}} \sum_{Z_m}^{\binom{n}{m}} \bar{r}_m\,\bar{x}_m$.

It may be noted that for the unbiased ratio-cum-product type estimators: $S^*_{(\pm 1)}$ (6.3.2), TH_m^* (6.3.4), and TL_m^* (6.3.6), for any given weight vector W used in their formulation, one can construct unbiased estimators of variance valid for any sample size n (≥ 2) and simpler estimators of variance in large samples by direct application of theorem 3.6.1, corollary 3.6.1 and theorem 3.6.3. We do not, however, present the relevant formulae as no numerical illustration of the estimation of variance of the ratio-cum-product type estimators is attempted here.

We now illustrate with numerical examples the computation of the proposed estimators of variance for the unbiased product type estimators. For the sake of comparison, we also include in small samples the unbiased estimator $\bar{y}$, and in large samples the estimators $\bar{y}$ and $\hat{Y}_p$.

As our population, we have chosen 45 cities/urban agglomeration (u.as) with population size 100, 000 and above from the Census of India (1971) document referred below the Table 4.2. In these 45 cities/u.as, situated all over India, high/low female literacy rates are associated with low/high female work participation rates. Taking female literacy rate as the auxiliary varinte (x), we propose to estimate the average female work participation rate ($\bar{Y}$), applicable to this set of 45 cities/u.as. The relevant population data are presented in Table - 6.1.

DATA FOR POPULATION 7

TABLE - 6.1

Female literacy rate and work participation rate (1971) in 45 selected cities and urban agglomerations of India with population 100, 000 and above

Sl. No.	Name of the city/urban agglomeration (100, 000 and above)	State/ Union territory	Female literacy rate	Female work participation rate
(1)	(2)	(3)	(4)	(5)
1	Vizayawnda	Andhra Pradesh	45.79	6 07
2	Guntur	" "	37.42	15.81
3	Kurnool	" "	36.49	14.75
4	Nizamabad	" "	26.40	16.12
5	Gauhati	Assam	41.87	8 26
6	Ranchi	Bihar	49.39	4 83
7	Bokaro Steel City	"	16.25	12.23
8	Bihar	"	24.82	9 14
9	Ahmedabad	Gujarat	49.76	4 96
0	Rajkot	"	51.41	4 27

contd...

(1)	(2)	(3)	(4)	(5)
11	Bhavanagar	Gujarat	45.54	5.51
12	Jamnagar	"	43.91	5.39
13	Nadiad	"	52.63	5.42
14	Rohtak	Haryana	47.15	3.19
15	Ambala Cantt.	"	46.60	3.88
16	Jammu	Jammu and Kashmir	49.53	4.39
17	Calicut	Kerala	58.38	6.84
18	Indore	Madhya Pradesh	46.60	5.59
19	Ratlam	" "	44.23	4.88
20	Sholapur	Maharashtra	33.24	11.93
21	Kolhapur	" "	48.46	5.40
22	Malegaon	" "	31.73	9.83
23	Ulhasnagar	" "	48.07	4.26
24	Belgaum	Mysore	50.35	5.10
25	Bellary	"	35.77	8.55
26	Cuttak	Orissa	46.90	4.64
27	Berhampur	"	37.94	7.48
28	Bhubaneswar	"	50.46	5.08
29	Amritsar	Punjab	51.50	3.71
30	Ludhiana	"	50.66	2.88
31	Jullundur	"	50.98	2.78
32	Patiala	"	51.92	5.12
33	Jodhpur	Rajasthan	34.69	7.38
34	Ajmer	"	46.95	5.42
35	Dindigul	Tamil Nadu	47.44	5.99
36	Kumbakonam	" "	48.36	4.52
37	Kanpur	Uttar Pradesh	41.21	3.71
38	Lucknow	" "	43.38	3.90
39	Varanasi	" "	31.92	10.36
40	Gorakhpur	" "	41.42	3.81
41	Dehra Dun	" "	54.97	4.40
42	Calcutta	West Bengal	50.48	4.75
43	Asansol	" "	48.95	4.38
44	Delhi	Delhi	50.88	5.13
45	Imphal	Manipur	43.53	18.97

For population 7, some parameters of interest are:

$\bar{Y} = 6.69$, $\bar{X} = 44.14$, $\sigma_y^2 = 14.14$, $\sigma_x^2 = 72.00$, $\sigma_{xy} = -20.74$

$C_y = 0.56$, $C_x = 0.19$, $\rho = -0.6501$, $\beta = -0.29$.

Five independent simple random samples are drawn without replacement from population 7 for each of the sample sizes 2, 3, 4 and 5. The contents (i.e., serial numbers of the units as in Table 6.1) of the samples are given in Table 6.2 .

TABLE - 6.2

Simple random samples of sizes 2, 3, 4 and 5 from population 7

Sample Number	Sample Size			
	2	3	4	5
1	43, 7	2, 28, 9	43, 22, 19, 12	10, 11, 39, 21, 32
2	27, 4	3, 5, 13	20, 43, 40, 18	44, 36, 22, 1, 12
3	33, 45	2, 24, 29	28, 37, 33, 3	31, 37, 12, 33, 5
4	3, 16	18, 4, 24	8, 1, 21, 31	9, 19, 25, 39, 35
5	10, 4	44, 4, 40	14, 19, 3, 23	38, 14, 34, 37, 27

In a later chapter, we study the relative efficiencies of (i) the unbiased product type estimators given in this chapter, (ii) some unbiased regression type estimators of next chapter, and (iii) the conventional biased product and regression estimators, in small samples (of sizes 2, 3, 4 and 5) from population 7 using a reasonable number of independent samples of each size. The small sample calculations presented here are mainly to illustrate the estimators of variance of the unbiased product type estimators and to make some general observations.

For the unbiased product type estimators $S^*_{(q'_n < o)}$, H^*_m and L^*_m we use respectively the unbiased estimators of variance given by (4.5.4), (6.6.1) and (6.6.2). For $\bar{y}$, the estimator of variance is $\left(\frac{1}{n} - \frac{1}{N}\right) s^2_y$. The final calculations are presented in Tables 6.3 to 6.6, separately for each sample size.

POPULATION 7

TABLE - 6.3

Estimates of population mean $\bar{Y}$ and variance estimates based on samples of size 2

Sample Number	Estimator					
	$\bar{y}$		$\hat{Y}_{RP}$		$S^*_{(q'_n < o)}$	
	Estimate	Estimate of variance	Estimate	Estimate of variance	Estimate	Estimate of variance
1	8.30	14.72	7.52	2.30	6.82	-ve
2	11.80	17.83	9.14	-ve	6.74	-ve
3	13.17	32.09	11.12	-ve	9.26	-ve
4	9.57	25.64	10.06	34.69(*)	10.49	43.25(*)
5	10.39	33.55	10.59	41.03(*)	10.95	48.04(*)

(*) High value.

POPULATION 7

TABLE - 6.4

Estimates of population mean $\overline{Y}$ and variance estimates based on samples of size 3

Estimator	Sample Number 1: Estimate	1: Estimate of variance	2: Estimate	2: Estimate of variance	3: Estimate	3: Estimate of variance	4: Estimate	4: Estimate of variance	5: Estimate	5: Estimate of variance
$\bar{y}$	8.62	12.07	9.48	7.12	8.21	13.64	8.94	12.06	8.35	14.21
$\hat{Y}_{RP}$	9.28	11.34	9.63	4.21	9.00	13.37	8.89	8.22	8.01	12.28
$S^*_{(q_o^* < o)}$	9.87	9.60	9.76	3.85	9.71	11.63	8.84	5.09	7.69	12.22
L_2^*	9.37	24.73 (*)	9.61	9.36 (*)	9.11	28.29 (*)	9.04	13.44 (*)	8.03	8.94
H_2^*	9.21	22.00 (*)	9.64	9.85 (*)	8.91	24.96 (*)	8.80	9.42	7.97	8.07

(*) Very High Value

POPULATION 7

TABLE - 6.5

Estimates of population mean $\bar{Y}$ and variance estimates based on samples of size 4

Estimator	Sample Number 1		2		3		4		5	
	Estimate	Estimate of variance	Estimate	Estimate of variance	Estimate	Estimate of variance	Estimate	Estimate of variance	Estimate	Estimate of variance
$\bar{y}$	6.12	1.43	6.43	3.19	7.73	5.51	5.85	1.56	6.77	6.56
$\hat{Y}_{RP}$	5.95	0.71	6.31	2.53	7.22	4.24	5.78	0.41	6.89	5.94
$S^{*}_{(q'_o < o)}$	5.79	0.33	6.20	2.19	6.77	4.18	5.72	-ve	6.99	5.36
L^{*}_{2}	5.95	0.72	6.32	2.59	7.19	3.92	5.84	0.39	6.90	6.02
H^{*}_{2}	5.94	0.70	6.30	2.48	7.25	4.30	5.76	0.40	6.87	5.91
L^{*}_{3}	5.95	-ve	6.32	1.82	7.18	-ve	5.86	1.58 (*)	6.91	8.25 (*)
H^{*}_{3}	5.94	-ve	6.30	1.59	7.25	-ve	5.75	0.41	6.87	7.72 (*)

(*) High Value.

POPULATION 7

TABLE - 6.6

Estimates of population mean $\bar{\bar{Y}}$ and Variance estimates based on samples of size 5

Estimator	Sample Number 1		2		3		4		5	
	Estimate	Estimate of variance	Estimate	Estimate of variance	Estimate	Estimate of variance	Estimate	Estimate of variance	Estimate	Estimate of variance
$\bar{y}$	6.13	1.04	6.19	0.79	5.50	0.97	6.95	1.04	4.74	0.54
$\hat{Y}_{RP}$	6.45	0.49	6.25	0.34	5.34	0.58	6.65	0.35	4.67	0.43
$S^*_{(q_o' < 0)}$	6.73	0.07	6.30	0.13	5.19	0.45	6.39	0.21	4.60	0.39
L_2^*	6.49	0.55	6.26	0.36	5.33	0.57	6.64	0.29	4.67	0.44
H_2^*	6.42	0.50	6.24	0.34	5.35	0.58	6.66	0.35	4.67	0.43
L_3^*	6.50	0.58	6.27	0.37	5.33	0.57	6.64	0.27	4.67	0.44
H_3^*	6.42	0.50	6.23	0.34	5.35	0.58	6.66	0.35	4.67	0.43
L_4^*	6.51	5.30(*)	6.27	1.70(*)	5.33	-ve	6.64	-ve	4.67	-ve
H_4^*	6.41	4.19(*)	6.23	1.24(*)	5.35	-ve	6.66	-ve	4.67	-ve

Remarks on Tables 6.3 to 6.6 :

Remark (a) : For the choice $m = n-1$, the performance of the unbiased estimator of variance of each of the product type estimators: $S^*_{(q_0' < 0)}$, L^*_m and H^*_m, is erratic; the variance estimator taking either negative or high/very high positive value. In this connection remark (a) of section 4.5 may be recalled, wherein a similar situation for the unbiased ratio type estimators has been discussed in detail. The points made there are equally valid for the unbiased product type estimators also. The important observation is that, for the choices $m = n-1$ and $n-2$, there is not much difference in the values of the estimates of population mean $\overline{Y}$ either in the case of L^*_m or H^*_m. As such there is no reason to expect a significant difference between the variances of L^*_{n-1} and L^*_{n-2} and similarly between H^*_{n-1} and H^*_{n-2}.

Remark (b) : In the light of remark (a), it is likely that all the product type estimators are more efficient than the simple estimator $\overline{y}$.

Remark (c) : Among the product type estimators, $S^*_{(q_0' < 0)}$ has the least value for the estimate of variance in more number of cases.

Remark (d) : The results do not indicate clear differences among different choices of m for L^*_m or H^*_m. More number of samples of same size may be required to examine this issue.

Remark (e) : For any particular choice of m (> 1), it is not clear whether L^*_m and H^*_m significantly differ in precision. On this issue also, more number of samples have to be examined.

We now illustrate the computation of the estimators of variance (6.6.3), (6.6.4), (6.6.5) and (6.6.6), using 5 independent samples of size 9 and 4 independent samples of size 12 selected from the 45 cities/u.as of population 7. The serial numbers of the cities/u.as selected in these samples are given in Table -6.7. The estimates of variance are presented in Tables -6.8 and 6.9. The formulae (6.6.5) and (6.6.6) have been computed for the choices m = 1, 3, 5, 7 and 8 with the sample size n = 9, and for m = 1 and 11 with n = 12.

TABLE - 6.7

Simple random samples of sizes 9 and 12 from population 7

Sample size	9					12			
Sample No.	1	2	3	4	5	1	2	3	4
Serial numbers of selected cities/u.as	7	9	14	4	38	37	9	14	2
	41	38	42	22	6	6	33	13	7
	33	28	21	35	40	5	41	29	6
	13	3	30	9	17	33	1	33	36
	40	39	12	8	10	14	24	32	37
	18	20	44	16	15	27	28	38	17
	23	[illegible]8	15	44	2	39	12	30	44
	25	25	45	38	32	20	2	20	29
	29	1	24	13	42	38	13	11	34
						43	18	41	8
						16	29	10	9
						17	14	37	28

POPULATION 7

TABLE - 6.8

Estimates of approximate variances using samples of size 9

Estimator	Sample Number 1	2	3	4	5	Average over samples
$\bar{y}$	0.70	1.17	2.16	1.37	1.30	1.34
$\hat{Y}_P$	0.18	0.67	1.98	0.59	1.13	0.91
$\hat{Y}_{RP}$	0.18	0.67	1.98	0.59	1.13	0.91
$S^*_{(q_0 < 0)}$	0.13	0.64	1.42	0.44	1.07	0.74
L_3^*	0.14	0.68	1.96	0.55	1.12	0.89
H_3^*	0.22	0.71	1.97	0.66	1.13	0.94
L_5^*	0.14	0.68	1.96	0.54	1.12	0.89
H_5^*	0.[illegible]2	0.71	1.97	0.67	1.13	0.94
L_7^*	0.14	0.68	1.96	0.53	1.12	0.88
H_7^*	0.22	0.71	1.97	0.67	1.13	0.94
L_8^*	0.14	0.68	1.96	0.53	1.12	0.88
H_8^*	0.22	0.71	1.97	0.67	1.13	0.94

POPULATION 7

TABLE - 6.9

Estimates of approximate variances using samples of size 12

Estimator	Sample Number 1	2	3	4	Average over samples
$\bar{y}$	0.48	0.65	0.37	0.86	0.59
$\hat{Y}_P$	0.32	0.48	0.25	0.52	0.39
$\hat{Y}_{RP}$	0.32	0.48	0.25	0.52	0.39
$S^*_{(q'_o < o)}$	0.30	0.36	0.20	0.50	0.34
L^*_{11}	0.31	0.47	0.24	0.50	0.38
H^*_{11}	0.32	0.48	0.25	0.53	0.40

Remarks on Tables 6.8 and 6.9 :

Remark (f) : All product estimators have shown gain in precision compared to the simple unbiased estimator $\bar{y}$.

Remark (g) : The unbiased product type estimator $S^*_{(q'_o < o)}$ is the most fficient among the product estimators.

Remark (h) : All the choices of m tried have resulted in almost equal precision both for L^*_m and H^*_m.

Remark (i) : For any particular choice of m (>1), L^*_m has gained slightly in precision over H^*_m.

CHAPTER VII

UNBIASED REGRESSION TYPE ESTIMATORS

7.0 Summary :

Auxiliary information available on one or more characters which are positively or negatively associated with the character under study may be used in regression method of estimation. Mickey (1959) has shown that his general unbiased estimators G_m and G_m^* result in the unbiased regression type estimators D_m (7.1.1) and D_m^* (7.1.2) for a single auxiliary variate x and the choice $a(Z_m) = b_m$, the linear regression coefficient of y on x based on Z_m. When p (> 1) auxiliary variates are available we may consider for G_m^* the choice $a_i(Z_m) = b_{im}$, where b_{im} is the partial regression coefficient of y on x_i based on Z_m, and denote the resulting multivariate unbiased regression type estimator by D_{pm}^*. It may be noted that D_m^* and D_{pm}^* are defined for $2 \leq m \leq n-1$. By assuming the knowledge of population variances and covariances of the auxiliary variates, we also formulate unbiased regression type estimators D_1^* and D_{p1}^* as special cases of G_m^* for the choice $m = 1$, mainly to facilitate theoretical studies on the exact variance of the unbiased regression type estimators. Broadly, two important problems are studied in this chapter. One is to investigate in detail the variance and the efficiency of the unbiased regression type estimators D_1^*, D_{p1}^*, D_m^* and D_{pm}^*. Second problem is to determine the optimum forms of the coefficient functions $a_i(Z_m)$ for the general unbiased estimator G_m^* in large samples and relate the optimum forms to the forms

so far used such as those yielding unbiased ratio, product, ratio-cum-product and regression type estimators.

The chapter is divided into five sections. In the introductory section we formulate the estimators D_1^*, D_{p1}^*, D_m^* and D_{pm}^* With a single auxiliary variate x, D_1^* is obtained from G_m^* for the choice m=1 and the coefficient function $a(Z_1) = a_j = \frac{y_j(x_j - \overline{X})}{\sigma_x^2}$ where σ_x^2 is the population variance of x. Denoting by S_{XX} the matrix ($S_{x_i x_{i'}}$) of the population mean squares and products among the p auxiliary variates x_i, the multivariate estimator D_{p1}^* is derived from G_m^* for m=1 and the column vector of coefficient functions:

$$A_j = (a_{1j}, a_{2j}, \ldots, a_{pj})' = \frac{N}{(N-1)} S_{XX}^{-1} y_j (X_j - \underset{\sim}{\overline{X}}) \text{ where}$$

$$X_j = (x_{1j}, x_{2j}, \ldots, x_{pj})' \text{ and } \underset{\sim}{\overline{X}} = (\overline{X}_1, \overline{X}_2, \ldots, \overline{X}_p)'.$$

In section 2, we obtain the exact variance of D_1^* using the result on the exact variance of G_1^* given in section 3.4a. The exact variance has terms upto order n^{-2}. It is observed that D_1^* and the classical linear regression estimator $\hat{\overline{Y}}_r$ have the same large sample variance. Under two super-population models, we also compare the expected variance of D_1^* and the expected mean square error of $\hat{\overline{Y}}_r$ by retaining terms upto order n^{-2} The first super-population model is characterised by the conditions: (i) population is infinite, (ii) $y_j = \alpha_1 + \alpha_2(x_j - \overline{X}) + \alpha_3(x_j - \overline{X})^2 + u_j$,

$E(u_j/x_j) = 0$, $V(u_j/x_j) = \delta$ (a constant), and $Cov(u_j, u_{j'}/x_j, x_{j'}) = 0$ for $j \neq j'$, and (iii) x has a normal distribution. The second super-population model assumes conditions (i) and (ii) of the first model and that x is a gamma variate. Under both the super-population models it is shown that the classical biased regression estimator is more efficient than the unbiased regression type estimator D_1^*.

In section 3, we show that the conditions, under which the large sample variance has been derived for G_m^* in section 3.4b, hold good for the estimators D_m^*, D_{pm}^* and D_{p1}^*; and thereby obtain the large sample variance of these unbiased regression type estimators. It is seen that in large samples for any large value of m the unbiased regression type estimator D_m^* is of same precision as the classical regression estimator $\hat{Y}_r$. Also in large samples the choice $m = n-1$ is an optimum choice for D_m^* and the estimators D_{n-1}^* and $\hat{\bar{Y}}_r$ are equally efficient. Similar results are obtained in respect of the multivariate unbiased regression type estimator D_{pm}^* and the multivariate biased regression estimator

$$\hat{\bar{Y}}_{pr} = \bar{y} + \sum_{i=1}^{p} b_{in}(\bar{X}_i - \bar{x}_i),$$ where b_{in} is the partial regression

coefficient of y on x_i based on the sample of size n. It is also noted that the estimators D_{p1}^* and $\hat{\bar{Y}}_{pr}$ are of equal precision in large samples.

In section 4, we consider in large samples the problem of determining optimum forms of the coefficient functions $a_i(Z_m)$ for the general unbiased estimator G_m^*. It may be recalled that in section 3.4b of Chapter - III an expression has been obtained for the large sample variance of G_m^* which is valid for all coefficient functions $a_i(Z_m)$ when m is small in relation to n, and for such $a_i(Z_m)$ for which the moment conditions (3.4.37) are satisfied. Using this result, in Chpaters - IV, VI and VII we have shown that for the estimators G_m^* involving the coefficient functions $a_i(Z_m)$ such as:

$$\frac{q\,\bar{y}_m}{\bar{X}_i}\ ,\ R_{im}\ ,\ -\frac{P_{im}}{\bar{X}_i}\ ,\ -\frac{\bar{r}_{im}\,\bar{x}_{im}}{\bar{X}_i}\ ,$$

linear regression coefficient of y on x_i based on Z_m, and

partial regression coefficient of y on x_i based on Z_m,

the large sample variance can be expressed in the form (7.4.1). Similarly, it may be possible to identify some more choices of $a_i(Z_m)$ in respect of which the large sample variance of the corresponding estimator G_m^* has the form (7.4.1). It is, therefore, natural to seek optimum choices for $a_i(Z_m)$ in the class of all coefficient functions for which the large sample variance of the corresponding estimator G_m^* can be expressed in the form (7.4.1). It is easily shown that the possible minimum of the large sample variance (7.4.1) is given by (7.4.3). With a single auxiliary variate x, the minimum possible large sample variance, which is equal to the large sample variance of the classical linear regression estimator $\hat{\bar{Y}}_r$, is attained for the choices:

$a(Z_m) = \dfrac{q_o'\,\bar{y}_m}{\bar{X}}$ where $q_o' = \rho\, C_y/C_x$, $a(Z_m) = b_m$ for large m, and

$$a(Z_m) = \frac{y_j(x_j - \bar{X})}{\sigma_x^2} \quad \text{for } m = 1 .$$

The corresponding estimators are the unbiased ratio type estimator $S^*_{(q_o')}$, the unbiased product type estimator $S^*_{(q_o' < o)}$, and the unbiased regression type estimators D^*_m and D^*_1. When several auxiliary variates are used, the minimum possible variance (7.4.3), which is equal to the large sample variance of the multivariate biased regression estimator $\hat{\bar{Y}}_{pr}$, is attained for the choices:

$$A(Z_m) = [\, a_1(Z_m),\ a_2(Z_m),\ \ldots,\ a_p(Z_m) \,]' = (\, b_{1m},\ b_{2m},\ \ldots,\ b_{pm})'$$

for large m

$$\text{and } A(Z_1) = A_j = (a_{1j}, a_{2j}, \ldots, a_{pj})' = \frac{N}{(N-1)} \; S_{XX}^{-1} \; y_j (X_j - \underset{\sim}{\bar{X}}).$$

The corresponding estimators are the multivariate unbiased regression type estimators D^*_{pm} and D^*_{p1}. Thus Mickey's general class of unbiased estimators G^*_m contains no estimator which is more efficient than the classical regression estimator in large samples.

In section 5, we develop variance estimators for the unbiased regression type estimators D^*_1 and D^*_m, making use of the results given in section 3.6 on the estimation of variance of G^*_m. For the multivariate unbiased regression type estimators D^*_{p1} and D^*_{pm}, variance estimation presents no new problems and may be attempted on similar lines. Unbiased estimates of variance are computed for D^*_1 and D^*_m by utilising (a) samples of sizes 2, 3, 4 and 5 drawn from the populations 1 to 5 and used in Chapter - IV for illustrating variance estimation of the unbiased ratio type estimators, and (b) samples of sizes 2, 3, 4 and 5 selected from population 7 and given in Chapter - VI to illustrate the estimation of variance of the unbiased product type estimators. The tentative conclusions are as follows. In small samples, the performance of the unbiased regression type estimators is poor in relation to the unbiased ratio type estimators studied in Chapter - IV. Rao (1969) and Samford (1969) have made a similar observation in small samples on the relative efficiency of the classical biased regression and ratio estimators. Similarly in the case of a negatively correlated auxiliary variate, in small samples, the unbiased regression type estimators are inferior to the unbiased product type estimators studied in Chapter - VI. Following a different approach, comparisons based on more number of small samples are given in a later chapter. In this section, we also compute biased estimates of variance of D_1^* and D_m^* applicable for large samples. For this purpose, we use (a) amples of sizes 10, 15, 20 and 25 from population 6 given in Chapter-IV and (b) samples of sizes 9 and 12 from population 7 given in Chapter - VI. ased on these computations, the conclusions are as follows. In large amples, the precision of the unbiased regression type estimator D_m^* creases as m increases from 2 to n-1. The estimators D_{n-1}^* and D_1^* is as efficient as the classical linear regression estimator $\hat{\bar{Y}}_r$. With a tively correlated auxiliary variate, in large samples, the unbiased ression type estimators D_{n-1}^* and D_1^* are more efficient than the unbiased ratio type esti[redacted]s T_m^* but are of same precision as the unbiased ratio type estimator $S^*_{(q_o')}$. With a negatively correlated auxiliary variate, in large samples, D^*_{n-1} and D^*_1 are of greater precision than the unbiased product type estimators H^*_m and L^*_m but have the same efficiency as the unbiased product type estimator $S^*_{(q_o' < o)}$.

7.1 <u>Introduction</u> :

So far we have been concerned with such choices of the coefficient functions $a_i(Z_m)$ for which the general unbiased estimator G^*_m (3.4.3) yields either a ratio or product or ratio-cum-product type estimator. In this chapter we turn our attention to some choices of $a_i(Z_m)$ for which G^*_m results in an unbiased regression type estimator.

With a single auxiliary variate x and the choice $a(Z_m) = b_m$, the linear regression coefficient of y on x based on Z_m, the estimators G_m (3.4.1) and G^*_m (3.4.3) yield the following unbiased regression type estimators for the population mean $\bar{Y}$:

$$D_m = \bar{y} + b_m(\bar{X} - \bar{x}) + \frac{m(N-n)}{N(n-m)} \left[(\bar{y} - \bar{y}_m) - b_m(\bar{x} - \bar{x}_m) \right], \qquad (7.1.1)$$

$$D^*_m = \bar{y} + b^*_m(\bar{X} - \bar{x}) + \frac{m(N-n)}{N(n-m)} \cdot \frac{1}{\binom{n}{m}} \sum_{Z_m}^{\binom{n}{m}} b_m(\bar{x}_m - \bar{x}), \qquad (7.1.2)$$

where $$b^*_m = \frac{1}{\binom{n}{m}} \sum_{Z_m}^{\binom{n}{m}} b_m .$$

In section 2.9, we have already reviewed the literature on the unbiased regression type estimators: $N D_m = d_m$ (2.9.1), $N \bar{D}_m = \bar{d}_m$ (2.9.2), $N \bar{D}'_{n-m} = \bar{d}'_{n-m}$ (2.9.3), and $N D^*_m = d^*_m$ (2.9.4).

It may be noted that D^*_m is defined for $2 \leq m \leq n-1$. It is also instructive to construct an unbiased regression type estimator for the choice $m = 1$ as a particular case of the estimator G^*_1 (3.4.4), even by assuming the knowledge of some population parameters in addition to $\bar{X}$, so that we may use the results on the exact variance of G^*_1 to study in more detail the precision of an unbiased regression type estimator. With this objective, we may assume that both σ^2_x and $\bar{X}$ are known and consider an unbiased regression type estimator:

$$D^*_1 = \bar{y} - \frac{(\bar{x} - \bar{X})}{\sigma^2_x} \cdot \frac{1}{n} \sum_{j=1}^{n} y_j (x_j - \bar{X}) + \left(\frac{1}{n} - \frac{1}{N}\right) \frac{1}{(n-1)\sigma^2_x} \sum_{j=1}^{n} y_j (x_j - \bar{X})(x_j - \bar{x}) \tag{7.1.3}$$

which results from G^*_1 (3.4.4) with a single auxiliary variate and the choice $a(Z_1) = a_j = \frac{y_j (x_j - \bar{X})}{\sigma^2_x}$

When several auxiliary variates are available, we may consider for the unbiased estimator G^*_m (3.4.3) the choice $a_i(Z_m) = b_{im}$ for $2 \leq m \leq n-1$, where $\underset{\sim}{b}_m = (b_{1m}, b_{2m}, \ldots, b_{pm})'$ is the column vector of partial regression coefficients of y on $x_1, x_2, \ldots, x_p$, based Z_m. We denote the resulting multivariate unbiased regression type estimator by D^*_{pm}. Similarly we may have a multivariate extension D^*_{p1}

for D_1^* (7.1.3), by assuming that in addition to the vector $\underset{\sim}{\overline{X}} = (\overline{X}_1, \overline{X}_2, \ldots, \overline{X}_p)'$ the (p x p) matrix $(S_{x_i x_{i'}}) = S_{XX}$ is also known, and choosing the coefficient functions:

$$A_j = (a_{1j} , a_{2j} , \ldots , a_{pj})' = \frac{N}{(N-1)} S_{XX}^{-1} y_j (X_j - \underset{\sim}{\overline{X}}), \quad (7.1.4)$$

where $X_j = (x_{1j} , x_{2j} , \ldots , x_{pj})'$.

7.2 Exact variance of unbiased regression type estimators:

In this section we investigate the exact variance of the unbiased regression type estimator D_1^* (7.1.3) using the results on the exact variance of G_1^* (3.4.4) given in section 3.4a. We also compare the exact variance of D_1^* with the mean square error (upto order n^{-2}) of the classical linear regression estimator $\overset{\triangle}{Y}_r$ under some super-population models.

Theorem 7.2.1 : When $N >> n$, the exact variance of the unbiased regression type estimator D_1^* (7.1.3) is given by

$$V(D_1^*) = \frac{1}{n} (\mu_{20} - \frac{\mu_{11}^2}{\mu_{02}}) + \frac{1}{n(n-1)} \left[- \frac{(\mu_{22} - \mu_{11}^2)}{\mu_{02}} + 2 (\frac{\mu_{12}}{\mu_{02}} + \overline{Y})^2 - \frac{\mu_{12}^2}{\mu_{02}^2} \right], \quad (7.2.1)$$

where $\mu_{ij} = E \left[(y - \overline{Y})^i (x - \overline{X})^j \right]$.

Proof : We note that D_1^* (7.1.3) results from C_1^* (3.4.4) with a single auxiliary variate for the choice $a_j = \frac{y_j (x_j - \overline{X})}{\mu_{02}}$.

Hence to obtain the exact variance of D_1^* for $N >> n$, we may use corollary 3.4.1c. For this we observe

$$E(a_j) = \frac{1}{\mu_{02}} \cdot E[\, y_j (x_j - \bar{X}) \,] = \frac{\mu_{11}}{\mu_{02}}, \qquad (7.2.2)$$

$$V(a_j) = \frac{1}{\mu_{02}^2} \cdot V[\, (y_j - \bar{Y})(x_j - \bar{X}) + \bar{Y}(x_j - \bar{X}) \,]$$

$$= \frac{1}{\mu_{02}^2} [\, (\mu_{22} - \mu_{11}^2) + \bar{Y}^2 \mu_{02} + 2\bar{Y}\mu_{12} \,], \qquad (7.2.3)$$

and $$\mathrm{Cov}(a_j, x_j) = E[\, a_j (x_j - \bar{X}) \,]$$

$$= \frac{1}{\mu_{02}} \cdot E[\, (y_j - \bar{Y})(x_j - \bar{X})^2 + \bar{Y}(x_j - \bar{X})^2 \,]$$

$$= \frac{1}{\mu_{02}} (\mu_{12} + \bar{Y}\mu_{02}) . \qquad (7.2.4)$$

Now using (7.2.2) to (7.2.4) in (3.4.35) we obtain the result (7.2.1).

Remark (a) : The exact variance of D_1^* (7.1.3), valid for any N and n, may be written down from corollary 3.4.1b with $a_j = \frac{y_j (x_j - \bar{X})}{\sigma_x^2}$.

Remark (b) : The unbiased regression type estimator D_1^* (7.1.3) has the same large sample variance as the classical linear regression estimator $\hat{Y}_r$.

Remark (c) : From the theory of linear regression it is known that the classical regression estimator $\hat{Y}_r$ possesses certain properties like unbiasedness and minimum variance in the class of linear unbiased estimates when the population is infinite, regression of y on x is linear, and the residual variance of y about the regression line is constant. In populations where regression of y on x departs from linearity and, say, is better described by a quadratic curve, one may still like to use the linear regression estimator $\hat{Y}_r$ despite some bias in the estimator. In such situations it is worth examining how the biased estimator $\hat{Y}_r$ compares with the unbiased regression type estimator D_1^* (7.1.3). In this context the following theorem on the mean square error of $\hat{Y}_r$ due to Cochran (1942) is of use.

<u>Theorem 7.2.2</u> (Cochran, 1942) : When the population is infinite and $V(y/x)$ is a constant, upto terms of order n^{-2} the mean square error of the classical linear regression estimator $\hat{Y}_r$ is given by

$$MSE(\hat{Y}_r) = \frac{1}{n}\left(\mu_{20} - \frac{\mu_{11}^2}{\mu_{02}}\right) + \frac{1}{n^2}\left[\left(\mu_{20} - \frac{\mu_{11}^2}{\mu_{02}}\right) + \frac{1}{\mu_{02}^2}\left(\mu_{12} - \frac{\mu_{11}\mu_{03}}{\mu_{02}}\right)^2\right]. \tag{7.2.5}$$

<u>Proof</u> : Cochran in his book (2nd ed. 1963) has given the mean square error of $\hat{Y}_r$ under the stated conditions in the form:

$$MSE(\hat{Y}_r) = \frac{S_y^2(1-\rho^2)}{n}\left[1 + \frac{1+\rho_2^2(2+\gamma_2)}{n}\right], \tag{7.2.6}$$

where γ_2 is Fisher's measure of relative kurtosis for the distribution of x, ρ_2 is the linear correlation coefficient between $(x-\bar{X})^2$ and $(y-\bar{Y})-\beta(x-\bar{X})$, and β is the linear regression coefficient of y on x.

definition $$\rho_2^2 = \frac{Cov^2[(x-\overline{X})^2, (y-\overline{Y})-\beta(x-\overline{X})]}{V[(x-\overline{X})^2]\, V[(y-\overline{Y})-\beta(x-\overline{X})]} \qquad (7.2.7)$$

our notation $Cov[(x-\overline{X})^2, (y-\overline{Y})-\beta(x-\overline{X})] = \mu_{12} - \frac{\mu_{11}\mu_{03}}{\mu_{02}}$,

$$V[(x-\overline{X})^2] = \mu_{02}^2\,(2+\gamma_2), \text{ and}$$

$$[(y-\overline{Y})-\beta(x-\overline{X})] = S_y^2(1-\rho^2) = (\mu_{20} - \frac{\mu_{11}^2}{\mu_{02}}) . \qquad (7.2.8)$$

ence from (7.2.7) we note

$$S_y^2(1-\rho^2)\,\rho_2^2(2+\gamma_2) = \frac{1}{\mu_{02}^2}(\mu_{12} - \frac{\mu_{11}\mu_{03}}{\mu_{02}})^2 . \qquad (7.2.9)$$

Now using (7.2.8) and (7.2.9) in (7.2.6) we obtain the form (7.2.5).

Remark (d): When the population is infinite and the variance of y_j in arrays in which the x_j are fixed is a constant, a comparison of the mean square error (considered upto terms of order n^{-2}) of the classical regression estimator $\hat{Y}_r$ with the variance of the unbiased regression type estimator D_1^* (7.1.3) reduces to a comparison of the second terms in (7.2.5) and (7.2.1). It is possible to make an empirical comparison. With no assumptions on the nature of regression of y and x and the distribution of x, a theoretical comparison seems to be difficult. As stated in remark(c) the comparison is of interest when the regression of y on x departs from linearity. We, therefore, undertake a theoretical comparison under two super-population models formulated as follows.

Definition 7.2.1 : Super-population model $\triangle_1$: This is characterised by the conditions

(i) population is infinite, (7.2.10)

(ii) $y_j = a_1 + a_2(x_j - \bar{X}) + a_3(x_j - \bar{X})^2 + u_j$, with $E(u_j/x_j) = 0$,

$V(u_j/x_j) = \delta$, a constant, and $Cov(u_j, u_{j'}/x_j, x_{j'}) = 0$ for $j \neq j'$;

and (iii) x has a normal distribution.

Definition 7.2.2 : Super-population model $\triangle_2$: Conditions (i) and (ii) are same as under $\triangle_1$ and (iii) x has a gamma distribution with rameter h. (7.2.11)

ma 7.2.1 : Under condition (ii) of $\triangle_1$ (7.2.10) or $\triangle_2$ (7.2.11),

$$\bar{Y} = a_1 + a_3 \mu_{02}, \quad \mu_{11} = a_2 \mu_{02} + a_3 \mu_{03},$$

$$\mu_{20} = a_2^2 \mu_{02} + a_3^2 (\mu_{04} - \mu_{02}^2) + 2 a_2 a_3 \mu_{03} + \delta,$$

$$\mu_{12} = a_2 \mu_{03} + a_3 (\mu_{04} - \mu_{02}^2), \tag{7.2.12}$$

$$\mu_{22} = a_2^2 \mu_{04} + a_3^2 (\mu_{06} + \mu_{02}^3 - 2\mu_{02}\mu_{04}) + 2 a_2 a_3 (\mu_{05} - \mu_{02}\mu_{03}) + \delta \mu_{02}$$

of : The lemma is easily obtained from the definition of the moments and condition (ii) of (7.2.10).

<u>Theorem 7.2.3</u>: Under super-population model $\triangle_1$ (7.2.10), the classical linear regression estimator $\hat{\bar{Y}}_r$ is more efficient than the unbiased regression type estimator D_1^* (7.1.3) when the expected mean square error of $\hat{\bar{Y}}_r$ is considered upto terms of order n^{-2}.

Proof: When x has a normal distribution, we note that

$$\mu_{03} = 0 = \mu_{05}, \ \mu_{04} = 3\mu_{02}^2, \text{ and } \mu_{06} = 15\mu_{02}^3 .$$

From lemma 7.2.1, using the above, we obtain under super-population model $\triangle_1$ (7.2.10):

$$\bar{Y} = a_1 + a_3\mu_{02}, \ \mu_{11} = a_2\mu_{02}, \ \mu_{20} = a_2^2\mu_{02} + 2a_3^2\mu_{02}^2 + \delta,$$

$$\mu_{12} = 2a_3\mu_{02}^2, \text{ and } \mu_{22} = 3a_2^2\mu_{02}^2 + 10a_3^2\mu_{02}^3 + \delta\mu_{02}. \quad (7.2.13)$$

From (7.2.13) we have $\left(\mu_{20} - \dfrac{\mu_{11}^2}{\mu_{02}}\right) = \delta + 2a_3^2\mu_{02}^2, \quad (7.2.14)$

$$\left(\mu_{20} - \frac{\mu_{11}^2}{\mu_{02}}\right) + \frac{1}{\mu_{02}^2}\left(\mu_{12} - \frac{\mu_{11}\mu_{03}}{\mu_{02}}\right)^2 = \delta + 6a_3^2\mu_{02}^2, \quad (7.2.15)$$

$$\frac{(\mu_{22} - \mu_{11}^2)}{\mu_{02}} + 2\left(\frac{\mu_{12}}{\mu_{02}} + \bar{Y}\right)^2 - \frac{\mu_{12}^2}{\mu_{02}^2} = 2a_2^2\mu_{02} + 10a_3^2\mu_{02}^2 + \delta$$
$$+ 2(a_1 + 3a_3\mu_{02})^2 - 4a_3^2\mu_{02}^2$$
$$= (\delta + 6a_3^2\mu_{02}^2) + 2a_2^2\mu_{02} + 2(a_1 + 3a_3\mu_{02})^2 .$$
(7.2.16)

Using (7.2.14) to (7.2.16) in (7.2.1) and (7.2.5), we obtain under super-population model $\triangle_1$ (7.2.10),

$$V(D_1^*) = \frac{1}{n}\left(\delta + 2a_3^2\ \mu_{02}^2\right) + \frac{1}{n(n-1)}\left[\left(\delta + 6\ a_3^2\ \mu_{02}^2\right) + 2a_2^2\mu_{02} + 2(a_1 + 3a_3\mu_{02})^2\right], \quad (7.2.17)$$

$$MSE(\hat{\bar{Y}}_r) = \frac{1}{n}\left(\delta + 2a_3^2\ \mu_{02}^2\right) + \frac{1}{n^2}\left(\delta + 6\ a_3^2\ \mu_{02}^2\right) + O(n^{-3}). \quad (7.2.18)$$

Comparing (7.2.17) and (7.2.18) we have the result of the theorem.

<u>Theorem 7.2.4</u>: Under super-population model $\triangle_2$ (7.2.11), the classical linear regression estimator $\hat{\bar{Y}}_r$ is more efficient than the unbiased regression type estimator D_1^* (7.1.3) when the expected mean square error of $\hat{\bar{Y}}_r$ is considered upto terms of order n^{-2}

<u>Proof</u>: When x has a gamma distribution with parameter h,

$$\mu_{02} = h,\ \mu_{03} = 2h,\ \mu_{04} = 3h(h+2),\ \mu_{05} = 4h(5h+6), \text{ and } \mu_{06} = 5h(h+2)(3h+20). \quad (7.2.19)$$

From lemma 7.2.1 and (7.2.19) we obtain under super-population model $\triangle_2$ (7.2.11):

$$\bar{Y} = a_1 + a_3 h,\ \ \mu_{11} = (a_2 + 2a_3)h,\ \ \mu_{20} = a_2^2 h + 2a_3^2 h(h+3) + 4a_2 a_3 h + \delta,$$

$$\mu_{12} = 2a_2 h + 2a_3 h(h+3),\ \ \mu_{22} = 3a_2^2 h(h+2) + a_3^2 h(10h^2 + 118h + 200) + 12a_2 a_3 h(3h+4) + \delta h. \quad (7.2.20)$$

Using (7.2.19) and (7.2.20) we have

$$\left(\mu_{20} - \frac{\mu_{11}^2}{\mu_{02}}\right) = a_2^2 h + 2a_3^2 h(h+3) + 4a_2 a_3 h + \delta - (a_2 + 2a_3)^2 h$$

$$= \delta + 2a_3^2 h(h+1), \qquad (7.2.21)$$

$$\left(\mu_{20} - \frac{\mu_{11}^2}{\mu_{02}}\right) + \frac{1}{\mu_{02}^2}\left(\mu_{12} - \frac{\mu_{11}\mu_{03}}{\mu_{02}}\right)^2 = \delta + 2a_3^2 h(h+1) + \left[2a_2 + 2a_3(h+3) - 2(a_2+2a_3)\right]^2$$

$$= \delta + 2a_3^2(h+1)(3h+2), \qquad (7.2.22)$$

$$\frac{(\mu_{22} - \mu_{11}^2)}{\mu_{02}} + 2\left(\frac{\mu_{12}}{\mu_{02}} + \bar{Y}\right)^2 - \frac{\mu_{12}^2}{\mu_{02}^2} = 3a_2^2(h+2) + a_3^2(10h^2 + 118h + 200) + 12a_2 a_3(3h+4)$$

$$+ \delta - (a_2 + 2a_3)^2 h + 2\left[2a_2 + 2a_3(h+3) + a_1 + a_3 h\right]^2$$

$$- 4a_2^2 - 4a_3^2(h^2 + 6h + 9) - 8a_2 a_3(h+3),$$

$$= \left[\delta + 2a_3^2(h+1)(3h+2)\right] + 2(h+1)(a_2 + 6a_3)^2 + 8a_3^2(h+11)$$

$$+ 2\left[2a_2 + 3a_3(h+2) + a_1\right]^2 \qquad (7.2.23)$$

Using (7.2.21) to (7.2.23) in (7.2.1) and (7.2.5), we obtain under super-population model Δ_2 (7.2.11),

$$V(D_1^*) = \frac{1}{n}\left[\delta + 2a_3^2 h(h+1)\right] + \frac{1}{n(n-1)}\left[\left\{\delta + 2a_3^2(h+1)(3h+2)\right\} + 2(h+1)(a_2+6a_3)^2\right.$$

$$\left. + 8a_3^2(h+11) + 2\left\{2a_2 + 3a_3(h+2) + a_1\right\}^2\right], \qquad (7.2.24)$$

$$MSE(\hat{\bar{Y}}_r) = \frac{1}{n}\left[\, 6 + 2a_3^2\, h(h+1)\right] + \frac{1}{n^2}\left[\, 6 + 2a_3^2\,(h+1)(3h+2)\right] + O\,(n^{-3}). \tag{7.2.25}$$

Now comparing (7.2.24) and (7.2.25) we have the result of the theorem.

7.3 Large sample variance of unbiased regression type estimators:

In this section we make use of the results of section 3.4b on the large sample variance of the general unbiased estimator G_m^* (3.4.3) to obtain the large sample variances of the unbiased regression type estimator D_m^* (7.1.2) and the multivariate unbiased regression type estimators D_{pm}^* and D_{p1}^* defined in section 7.1.

Theorem 7.3.1 : In large samples the variance of the unbiased regression type estimator D_m^* (7.1.2) is given by

$$V(D_m^*) = \left(\frac{1}{n} - \frac{1}{N}\right)\left[\, S_y^2 - 2\,E(b_m)\, S_{xy} + E^2(b_m)\, S_x^2 \,\right]. \tag{7.3.1}$$

Proof : We may consider two cases.

Case (i) : n is large and m is not small compared to n.

By definition $b_m = \dfrac{s_{xy}(Z_m)}{s_x^2(Z_m)}$. In this case we may assume $\dfrac{s_x^2(Z_m) - s_x^2}{s_x^2} < 1$. For D_m^* (7.1.2), we have $a(Z_m) = b_m$.

Then it may be seen that

$$a^*(Z_m) = E_n(b_m) \doteq b \left[1 + \frac{V_n\{s_x^2(Z_m)\}}{s_x^4} - \frac{Cov_n\{s_{xy}(Z_m), s_x^2(Z_m)\}}{s_{xy}\, s_x^2} \right],$$

$$\text{where } b = \frac{s_{xy}}{s_x^2} \qquad (7.3.2)$$

In random samples of size m from n elements, we note that $V_n[s_x^2(Z_m)]$ and $Cov_n[s_{xy}(Z_m), s_x^2(Z_m)]$ are both of order m^{-1}. Hence from (7.3.2) it follows that $\mu_2[a^*(Z_m)]$ and $\mu_4[a^*(Z_m)]$ are of same order in n as $\mu_2(b)$ and $\mu_4(b)$. In large samples of size n, it is known that $\mu_2(b) = O(n^{-1})$ and $\mu_4(b) = O(n^{-2})$. Hence, for $a(Z_m) = b_m$, we obtain in this case

$$\mu_2[a^*(Z_m)] = O(n^{-1}) \text{ and } \mu_4[a^*(Z_m)] = O(n^{-2}). \qquad (7.3.3)$$

Again under the assumption $\left| \frac{s_x^2(Z_m) - s_x^2}{s_x^2} \right| < 1$, *it may be shown that*

$$n(b_m, \bar{x}_{n-m}) = -\frac{m}{(n-m)} Cov_n(b_m, \bar{x}_m)$$

$$\doteq -\frac{m}{(n-m)} \cdot \frac{(n-m)}{nm} b \left(\frac{m_{12}^{yx}}{s_{xy}} - \frac{m_{03}^{yx}}{s_x^2} \right)$$

$$= \frac{b}{n} \left(\frac{m_{03}^{yx}}{s_x^2} - \frac{m_{12}^{yx}}{s_{xy}} \right), \qquad (7.3.4)$$

where $m_{ik}^{yx} = \frac{1}{n} \sum_{j=1}^{n} (y_j - \bar{y})^i (x_j - \bar{x})^k$

From (7.3.4) it now follows that, in large samples when m is not small in comparison with n,

$$\mu_2 \left[\mathrm{Cov}_n \left\{ a(Z_m),\ \bar{x}_{n-m} \right\} \right] = \mu_2 \left[\mathrm{Cov}_n (b_m,\ \bar{x}_{n-m}) \right] = O(n^{-3}). \tag{7.3.5}$$

In view of (7.3.3) and (7.3.5), the conditions (3.4.37) of theorem 3.4.2 are satisfied for the coefficient function $a(Z_m) = b_m$. Hence in this case we obtain the result (7.3.1).

<u>Case</u> (ii) : n is large and m is small compared to n.

Here we apply theorem 3.4.3 and obtain the result (7.3.1).

This completes the proof of the theorem.

<u>Corollary 7.3.1</u> : In large samples, when m is also large,

$$V (D_m^*) = \left(\frac{1}{n} - \frac{1}{N} \right) \left(S_y^2 - \frac{S_{xy}^2}{S_x^2} \right), \tag{7.3.6}$$

which is same as the large sample variance of the classical linear regression estimator $\hat{\bar{Y}}_r$.

<u>Proof</u> : When m is large, we may assume $\left| \frac{s_x^2 (Z_m) - S_x^2}{S_x^2} \right| < 1$ and obtain $E(b_m) = \beta + O(m^{-1})$, where $\beta = S_{xy} / S_x^2$. Substituting $\beta + O(m^{-1})$ for $E(b_m)$ in (7.3.1), we get (7.3.6).

<u>Corollary 7.3.2</u> : In large samples the choice $m = n - 1$ is an optimum choice for the unbiased regression type estimator D_m^* (7.1.2).

Proof : We may write (7.3.1) in the form :

$$V(D_m^*) = \left(\frac{1}{n} - \frac{1}{N}\right)\left[\left(S_y^2 - \frac{S_{xy}^2}{S_x^2}\right) + (E(b_m) - \beta)^2 S_x^2\right] \qquad (7.3.7)$$

From (7.3.7) we note that the large sample variance of the unbiased regression type estimator D_m^* can never be less than $\left(\frac{1}{n} - \frac{1}{N}\right)\left(S_y^2 - \frac{S_{xy}^2}{S_x^2}\right)$, which is the large sample variance of the classical linear regression estimator $\hat{Y}_r$.

Now using corollary 7.3.1, it follows that in large samples the choice $m = n-1$ is an optimum choice for D_m^*.

Theorem 7.3.2 : In large samples the multivariate unbiased regression type estimator D_{pm}^*, which results from G_m^* (3.4.3) for the choice

$$A(Z_m) = [a_1(Z_m), a_2(Z_m), \ldots, a_p(Z_m)]' = (b_{1m}, b_{2m}, \ldots, b_{pm})' = \underset{\sim}{b}_m \qquad (7.3.8)$$

where $\underset{\sim}{b}_m$ is the column vector of partial regression coefficients of y on $x_1, x_2, \ldots, x_p$ based on Z_m, has variance given by

$$V(D_{pm}^*) = \left(\frac{1}{n} - \frac{1}{N}\right)\left[S_y^2 - 2\sum_{i=1}^{p} E(b_{im}) S_{x_i y} + \sum_{i=1}^{p}\sum_{i'=1}^{p} E(b_{im})E(b_{i'm}) S_{x_i x_{i'}}\right] \qquad (7.3.9)$$

Proof : We consider two cases.

Case (i) : n is large and m is not small compared to n.

On Z_m we define $s_{Xy}(Z_m) = [s_{x_1y}(Z_m), s_{x_2y}(Z_m), \ldots, s_{x_py}(Z_m)]'$

and the (p x p) matrix, $s_{XX}(Z_m)$, with $s_{x_ix_{i'}}(Z_m)$ as its (i, i')th element. Similarly on the sample of size n define s_{Xy} and s_{XX}; and on the population define S_{Xy} and S_{XX}. Similar to $\underset{\sim}{b}_m$ based on Z_m, we define respectively on the sample and the population the column vectors:

$$\underset{\sim}{b}_n = (b_{1n} , b_{2n} , \ldots , b_{pn})'$$

$$\text{and } \underset{\sim}{\beta} = (\beta_1 , \beta_2 , \ldots , \beta_p)'.$$

Then $\underset{\sim}{b}_m = s^{-1}_{XX}(Z_m)\, s_{Xy}(Z_m)$, $\underset{\sim}{b}_n = s^{-1}_{XX}\, s_{Xy}$, and $\underset{\sim}{\beta} = S^{-1}_{XX}\, S_{Xy}$.

(7.3.10)

Let $e_j = (y_j - \bar{y}) - \sum_{i=1}^{p} b_{in} (x_{ij} - \bar{x}_i)$.

$$\text{Then } s_{x_{i'}y}(Z_m) = \frac{1}{(m-1)} \sum_{j \in Z_m} (x_{i'j} - \bar{x}_{i'm}) y_j$$

$$= \frac{1}{(m-1)} \sum_{j \in Z_m} (x_{i'j} - \bar{x}_{i'm}) [e_j + \bar{y} + \sum_{i=1}^{p} b_{in}(x_{ij} - \bar{x}_{im} + \bar{x}_{im} - \bar{x}_i)]$$

$$= \frac{1}{(m-1)} \sum_{j \in Z_m} (x_{i'j} - \bar{x}_{i'm}) e_j + \sum_{i=1}^{p} b_{in} [\frac{1}{(m-1)} \sum_{j \in Z_m} (x_{i'j} - \bar{x}_{i'm})(x_{ij} - \bar{x}_{im})]$$

$$= s_{x_{i'}e}(Z_m) + \sum_{i=1}^{p} b_{in}\, s_{x_ix_{i'}}(Z_m).$$

Hence, denoting $s_{Xe}(Z_m) = [s_{x_1e}(Z_m), s_{x_2e}(Z_m), \ldots, s_{x_pe}(Z_m)]'$, we have

$$s_{Xy}(Z_m) = s_{Xe}(Z_m) + s_{XX}(Z_m)\, b_n \qquad (7.3.11)$$

It now follows that $b_m = s_{XX}^{-1}(Z_m)\, s_{Xy}(Z_m) = s_{XX}^{-1}(Z_m)\, s_{Xe}(Z_m) + b_n$,

$$E_n(b_m) = b_n + E_n\left[s_{XX}^{-1}(Z_m)\, s_{Xe}(Z_m) \right]. \qquad (7.3.12)$$

We also note from (7.3.11)

$$E_n\left[s_{Xe}(Z_m) \right] = E_n\left[s_{Xy}(Z_m) - s_{XX}(Z_m)\, b_n \right]$$

$$= s_{Xy} - s_{XX}\, b_n$$

$$= s_{Xy} - s_{Xy} = 0 \text{ (null column vector).} \qquad (7.3.13)$$

Writing $s_{XX}^{-1}(Z_m) = C_{XX}(Z_m)$, from (7.3.12) and (7.3.13) we have

$$b_{im}^{*} = E_n(b_{im}) = b_{in} + E_n\left[\sum_{i'=1}^{p} C_{x_i x_{i'}}(Z_m)\, s_{x_{i'} e}(Z_m) \right]$$

$$= b_{in} + \sum_{i'=1}^{p} \mathrm{Cov}_n\left[C_{x_i x_{i'}}(Z_m),\ s_{x_{i'} e}(Z_m) \right]. \qquad (7.3.14)$$

We note $$C_{x_i x_{i'}}(Z_m) = \frac{Q_{ii'}(Z_m)}{\left| s_{XX}(Z_m) \right|}, \qquad (7.3.15)$$

where $Q_{ii'}(Z_m)$ is the co-factor of the element $s_{x_i x_{i'}}(Z_m)$ in the matrix $s_{XX}(Z_m)$, and $\left| s_{XX}(Z_m) \right|$ is the determinant of $s_{XX}(Z_m)$. Further, in random samples of size m from n elements, $V_n\left[s_{x_i x_i}(Z_m) \right]$, is of order $K_{m,n} = \frac{(n - m)}{nm}$. Hence it may be shown that

$V_n [Q_{ii'}(Z_m)]$, $V_n [\{ s_{XX}(Z_m) \}]$ and $Cov_n [Q_{ii'}(Z_m), \{ s_{XX}(Z_m) \}]$

are all of order $K_{m,n}$.

Consequently, when m is not small in comparison with n, from (7.3.15) it follows that

$$V_n [C_{x_i x_{i'}}(Z_m)] \doteq V_n [Q_{ii'}(Z_m)] + V_n [\{ s_{XX}(Z_m) \}]$$
$$-2 \; Cov_n [Q_{ii'}(Z_m), \{ s_{XX}(Z_m) \}]$$

is also of order $K_{m,n}$. We also note that $V_n [s_{x_{i'}e}(Z_m)]$ is of order $K_{m,n}$. Hence we obtain

$$Cov_n [C_{x_i x_{i'}}(Z_m) , s_{x_{i'} e} (Z_m)] \text{ is of order } K_{m,n}. \quad (7.3.16)$$

Now using (7.3.16) in (7.3.14) we infer that $\mu_2 (b^*_{im})$ and $\mu_4 (b^*_{im})$ are of same order in n as $\mu_2 (b_{in})$ and $\mu_4 (b_{in})$ respectively. By a similar argument as above, in large samples of size n it can be shown that $\mu_2 (b_{in}) = O (n^{-1})$ and $\mu_4 (b_{in}) = O (n^{-2})$. Hence in the present case we obtain

$$\mu_2 (b^*_{im}) = O (n^{-1}) \text{ and } \mu_4 (b_{im}) = O (n^{-2}). \qquad (7.3.17)$$

Now consider $Cov_n (b_{im} , \bar{x}_{i\,n-m}) = - \frac{m}{(n-m)} Cov_n(b_{im}, \bar{x}_{im})$.

When m is not small in comparison with n, we note that $Cov_n(b_{im}, \bar{x}_{im})$ is of order $K_{m,n}$. Hence

$$\mathrm{Cov}_n\,(\,b_{im},\,\bar{x}_{i\,n-m}) = -\frac{m}{(n-m)}\cdot K_{m,n}\ \text{(terms involving sample moments of order two and above)}$$

$$= -\frac{1}{n}\ \text{(terms involving sample moments of order two and above)}.$$

In this case it, therefore, follows that

$$\mu_2\,[\,\mathrm{Cov}_n\,(\,b_{im}\,,\,\bar{x}_{i\,n-m}\,)\,] = O(\,n^{-3}\,). \qquad (7.3.18)$$

From (7.3.17) and (7.3.18) we observe that the conditions (3.4.37) of theorem 3.4.2 are satisfied for the coefficient functions $a_i(Z_m) = b_{im}$ Consequently, in this case we obtain the result (7.3.9).

Case (ii) : n is large and m is small compared to n.

We apply theorem 3.4.3 and obtain the result (7.3.9).

This completes the proof of the theorem.

Corollary 7.3.3 : In large samples for any large m

$$V(D^*_{pm}) = (\,\frac{1}{n} - \frac{1}{N})\,[\,S_y^2 - 2\sum_{i=1}^{p}\beta_i S_{x_i y} + \sum_{i=1}^{p}\sum_{i'=1}^{p}\beta_i\beta_{i'} S_{x_i x_{i'}}\,] \qquad (7.3.19)$$

$$= (\,\frac{1}{n} - \frac{1}{N})\,[\,S_y^2 - S'_{Xy}\ S^{-1}_{XX}\ S_{Xy}\,]. \qquad (7.3.20)$$

which is also the large sample variance of the multivariate biased regression estimator $\hat{\bar{Y}}_{pr} = \bar{y} + \sum_{i=1}^{p} b_{in}\,(\,\bar{X}_i - \bar{x}_i\,)$.

Proof: When m is large, similar to the results (7.3.14) and (7.3.16), it can be shown that $E(b_{im}) = \beta_i + O(m^{-1})$. Substituting this in (7.3.9) we obtain (7.3.19).

Md. Anwar Hossain Talukdar (1968) has considered $\hat{Y}_{pr}$ and shown that its large sample variance is given by (7.3.20), which is an alternate form of (7.3.19).

Corollary 7.3.4 : In large samples for the multivariate unbiased regression type estimator D^*_{pm}, the choice $m = n-1$ is an optimum choice.

Proof : Writing (7.3.9) in the alternate form :

$$V(D^*_{pm}) = \left(\frac{1}{n} - \frac{1}{N}\right)\left[(S^2_y - S'_{Xy} S^{-1}_{XX} S_{Xy}) + (E(b_m) - \beta)' S_{XX}(E(b_m) - \beta)\right], \qquad (7.3.21)$$

It may be noted that the large sample variance of D^*_{pm} can never be less than (7.3.20). Now using corollary 7.3.3, it follows that $m = n-1$ is an optimum choice for D^*_{pm} in large samples.

Theorem 7.3.3 : The multivariate unbiased regression type estimator D^*_{p1}, which results from G^*_m (3.4.3) for the choice $m = 1$ and the coefficient functions A_j (7.1.4), has in large samples the variance given by

$$V(D^*_{p1}) = \left(\frac{1}{n} - \frac{1}{N}\right)(S^2_y - S'_{Xy} S^{-1}_{XX} S_{Xy}). \qquad (7.3.22)$$

Proof : From the definition (7.1.4) of A_j, we note

$$E(A_j) = S^{-1}_{XX} S_{Xy} = \beta . \qquad (7.3.23)$$

Since $m = 1$, we apply theorem 3.4.3 and obtain the large sample variance of D^*_{p1}, which in view of (7.3.23) reduces to (7.3.22).

Remark (a) : The multivariate unbiased regression type estimator D^*_{p1} and the multivariate biased regression estimator $\hat{\bar{Y}}_{pr}$ have the same precision in large samples.

7.4 Optimum forms of the coefficient functions $a_i(Z_m)$ for the general unbiased estimator G^*_m (3.4.3) in large samples:

In section 3.4b we have shown in theorems 3.4.3 and 3.4.2 that in large samples the variance of the general unbiased estimator G^*_m (3.4.3) may be expressed in the form :

$$V(G^*_m) = \left(\frac{1}{n} - \frac{1}{N}\right)\left[S_y^2 - 2\sum_{i=1}^{p} E(a_i(Z_m))S_{x_iy} + \sum_{i=1}^{p}\sum_{i'=1}^{p} E(a_i(Z_m))E(a_{i'}(Z_m))S_{x_ix_{i'}} \right] \tag{7.4.1}$$

for all coefficient functions $a_i(Z_m)$ when m is small in comparison with n, and for such coefficient functions $a_i(Z_m)$ for which the moment conditions (3.4.37) are satisfied. Using this result, so far we have shown that for the estimators G^*_m, involving the coefficient functions $a_i(Z_m)$ such as:

$$\frac{q\,\bar{y}_m}{\bar{\bar{X}}_1}, \; R_{im}, \; -\frac{p_{im}}{\bar{\bar{X}}_i}, \; -\frac{\bar{r}_{im}\,\bar{x}_{im}}{\bar{\bar{X}}_1},$$

linear regression coefficient of y on x_i based on Z_m, and partial regression coefficient b_{im}, the large sample variance can be expressed in the form (7.4.1). Similarly it may be possible to identify some more choices of the coefficient functions $a_i(Z_m)$ in respect of which the large sample variance of the corresponding estimator G^*_m is provided by (7.4.1).

Rewriting (7.4.1) in the form :

$$V(G^*_m) = (\frac{1}{n} - \frac{1}{N}) \left[(S^2_y - S'_{Xy} S^{-1}_{XX} S_{Xy}) + (E \{ A(Z_m)\} - \beta)' S_{XX} (E \{A(Z_m)\} - \beta) \right] \quad (7.4.2)$$

it is clear that the possible minimum of (7.4.2) is given by

$$V_o (G^*_m) = (\frac{1}{n} - \frac{1}{N}) (S^2_y - S'_{Xy} S^{-1}_{XX} S_{Xy}) . \quad (7.4.3)$$

When we work with a single auxiliary variable, in large samples, from theorem 4.2.1 and corollary 4.3.2, and corollary 7.3.1; we know that the minimum (7.4.3) is attained for the choices:

$$a (Z_m) = \rho \frac{C_y}{C_x} \frac{\bar{y}_m}{\bar{X}} , \quad (7.4.4)$$

$$= b_m \text{ (for large } m \text{), and}$$

$$= \frac{y_j (x_j - \bar{X})}{\sigma^2_x} \text{ for } m = 1 .$$

Consequently, in large samples, the choices (7.4.4) may be termed optimum choices for the coefficient function $a(Z_m)$ in the class of all coefficient functions $a(Z_m)$ in respect of which the variance of the corresponding estimator G^*_m is given by

$$V(G_m^*) = (\frac{1}{n} - \frac{1}{N}) \left[S_y^2 - 2E (a (Z_m)) S_{xy} + E^2(a (Z_m)) S_x^2 \right].$$

(7.4.5)

Similarly when several auxiliary variables are available, from corollary 7.3.3 and theorem 7.3.3 we infer that in large samples the choices:

b_m (for large m), and

$$\frac{N}{(N-1)} S_{XX}^{-1} y_j (X_j - \bar{X}) \text{ given by (7.1.4) for } m = 1, \quad (7.4.6)$$

are <u>optimum choices</u> for $A(Z_m)$ in the class of all coefficient functions $A(Z_m)$ for which the corresponding estimator G_m^* has the variance (7.4.2).

In general, from (7.4.2) it may be noted that in large samples any choice of $A(Z_m)$ for which $E\left[A(Z_m) \right] \doteq \beta$ is an optimum choice in the class of coefficient functions $A(Z_m)$ for which G_m^* has the variance (7.4.2).

7.5 <u>Estimation of variance of unbiased regression type estimators</u>:

We now apply the methods developed in section 3.6 for estimating the variance of the unbiased estimator G_m^*(3.4.3) to estimate the variance of the unbiased regression type estimators D_1^* (7.1.3) and D_m^* (7.1.2). It may be stated that the estimation of variance of the multivariate unbiased regression type estimators D_{p1}^* and D_{pm}^* presents no new problems and may be attempted on similar lines. We, therefore, confine ourselves to the case of single auxiliary variate in illustrating the variance estimators.

Applying theorem 3.6.1 and corollary 3.6.1, an unbiased estimator of variance of

(i) D_1^* (7.1.3) is given by

$$\hat{V}(D_1^*) = \frac{1}{n} \sum_{j=1}^{n} \left[\hat{V}(D_1^j) - (D_1^j - D_1^*)^2 \right], \qquad (7.5.1)$$

where

$$\hat{V}(D_1^j) = \frac{(N-n)(N-1)}{N^2(n-1)} \cdot \frac{1}{(n-2)} \sum_{j' \neq j}^{n-1} \left[(y_{j'} - \bar{y}_{n-1}) - \frac{y_j(x_j - \bar{X})}{\sigma_x^2} (x_{j'} - \bar{x}_{n-1}) \right]^2 \text{ for } n > 2,$$

$$= \frac{(N-n+1)^2}{N^2} \left[U_{n-1}^2(D_1) - y_{j'}^2 \right] + \frac{(N-n+1)(N-n)}{N^2} s_y^2 \quad \text{for } n = 2,$$

$$U_{n-1}(D_1) = y_{j'} - \frac{y_j(x_j - \bar{X})}{\sigma_x^2} (x_{j'} - \bar{X}_{N-n+1}), \; j' \neq j, \text{ and}$$

$$D_1^j = \left[\bar{y} - \frac{y_j(x_j - \bar{X})}{\sigma_x^2} (\bar{x} - \bar{X}) \right] - \frac{(N-n)}{N(n-1)} \left[(y_j - \bar{y}) - \frac{y_j(x_j - \bar{X})}{\sigma_x^2} (x_j - \bar{x}) \right];$$

and (ii) D_m^* (7.1.2) is obtained as

$$\hat{V}(D_m^*) = \frac{1}{\binom{n}{m}} \sum_{Z_m}^{\binom{n}{m}} \left[\hat{V}(D_m) - (D_m - D_m^*)^2 \right], \qquad (7.5.2)$$

where

$$\hat{V}(D_m) = \frac{(N-n)(N-m)}{N^2(n-m)} \cdot \frac{1}{(n-m-1)} \sum_{j \notin Z_m}^{n-m} \left[(y_j - \bar{y}_{n-m}) - b_m(x_j - \bar{x}_{n-m}) \right]^2$$

$$\text{for } 2 \leq m \leq n-2,$$

$$= \frac{(N-n+1)^2}{N^2} \left[U_{n-1}^2(D_m) - y_j^2 \right] + \frac{(N-n+1)(N-n)}{N^2} s_y^2 \quad \text{for } m = n-1,$$

$U_{n-1}(D_m) = y_j - b_{n-1}(x_j - \bar{X}_{N-n+1})$, $j \notin Z_{n-1}$, and D_m is given by (7.1.1).

In large samples, from remark (b) of section 7.2 it may be noted that as an estimator of variance of D_1^* as well as of the classical regression estimator $\hat{\bar{Y}}_r$, we may use

$$\check{V}'(D_1^*) = \check{V}'(\hat{\bar{Y}}_r) = \left(\frac{1}{n} - \frac{1}{N}\right)\left(s_y^2 - \frac{s_{xy}^2}{s_x^2}\right). \tag{7.5.3}$$

From theorems 7.3.1 and 3.6.3, in large samples, for D_m^* we may have the variance estimator:

$$\check{V}(D_m^*) = \left(\frac{1}{n} - \frac{1}{N}\right)\left[s_y^2 - 2 b_m^* s_{xy} + (b_m^*)^2 s_x^2 \right], \tag{7.5.4}$$

where $b_m^* = \frac{1}{\binom{n}{m}} \sum_{Z_m}^{\binom{n}{m}} b_m$.

We now illustrate the computation of

(i) the variance estimators (7.5.1) and (7.5.2) using

a) populations 1 to 5 (Tables 4.1 and 4.2) and the samples of sizes 2, 3, 4 and 5 (Table 4.4) drawn from those populations, and

b) population 7 (Table 6.1) and the samples of sizes 2, 3, 4 and 5 (Table 6.2) drawn from it;

and (ii) the variance estimators (7.5.3) and (7.5.4) using

a) population 6 (Table 4.10) and from it the samples of sizes 10, 15, 20 and 25 (Tables 4.11 to 4.14), and

b) population 7 (Table 6.1) and from it the samples of sizes 9 and 12 (Table 6.7).

The final computations are now presented in Tables 7.1 to 7.15.

POPULATION 1

TABLE - 7.1

Regression estimates of population total Y and variance estimates**

Estimator	Sample Size 2		Sample Size 3		Sample Size 4		Sample Size 5	
	Estimate	Estimate of variance	Estimate	Estimate of variance	Estimate	Estimate of variance	Estimate	Estimate of variance
ND_1^*	944	72.79 (*)	537	-ve	512	-ve	454	1.58
ND_2^*			164	-ve	254	16.70	460	0.43
ND_3^*					244	-ve	430	0.38
ND_4^*							432	-ve

** Table values are to be multiplied by 10^4 to obtain actual values of estimates of variance.

(*) Very high value.

POPULATION 2

TABLE - 7.2

Regression estimates of population total Y and variance estimates**

Estimator	Sample Size 2		Sample Size 3		Sample Size 4		Sample Size 5	
	Estimate	Estimate of variance	Estimate	Estimate of variance	Estimate	Estimate of variance	Estimate	Estimate of variance
ND_1^*	24440	-ve	38785	117.07	58120	143.21	34906	44.32
ND_2^*			100028	7389.91 (*)	78193	-ve	89813	-ve
ND_3^*					69841	1394.38 (*)	83475	-ve
ND_4^*							79794	3984.82 (*)

** Table values are to be multiplied by 10^6 to obtain actual values of estimates of variance.

(*) Very high value.

POPULATION 3

TABLE - 7.3

Regression estimates of population total Y and variance estimates**

Estimator	Sample Size 2		Sample Size 3		Sample Size 4		Sample Size 5	
	Estimate	Estimate of variance	Estimate	Estimate of variance	Estimate	Estimate of variance	Estimate	Estimate of variance
ND_1^*	18014	86.01	13598	45.91	16421	11.72	20051	17.03
ND_2^*			20144	217.61 (*)	20416	-ve	19909	0.50
ND_3^*					19933	116.78 (*)	19871	0.02
ND_4^*							19863	34.46 (*)

** Table values are to be multiplied by 10^6 to obtain actual values of estimates of variance

(*) Very high value.

POPULATION 4

TABLE - 7.4

Regression estimates of population total Y and variance estimates**

Estimator	Sample Size 2		Sample Size 3		Sample Size 4		Sample Size 5	
	Estimate	Estimate of variance	Estimate	Estimate of variance	Estimate	Estimate of variance	Estimate	Estimate of variance
ND_1^*	32C5	41.48	2484	7.74	1734	5.82	2105	9.37
ND_2^*			2175	37.0[illegible] (*)	986	-ve	1014	0.62
ND_3^*					1178	2.46	1022	-ve
ND_4^*							1106	-ve

** Table values are to be multiplied by 10^5 to obtain actual values of estimates of variance.

(*) Very high value.

POPULATION 5

TABLE - 7.5

Regression estimates of population total Y and variance estimates**

Estimator	Sample Size 2		Sample Size 3		Sample Size 4		Sample Size 5	
	Estimate	Estimate of variance	Estimate	Estimate of variance	Estimate	Estimate of variance	Estimate	Estimate of variance
ND_1^*	17596	159.30(*)	7298	17.37	14065	9.40	12385	6.46
ND_2^*			12069	94.76(*)	15159	-ve	12539	1.82
ND_3^*					13700	15.49(*)	12516	0.48
ND_4^*							12470	17.18(*)

** Table values are to be multiplied by 10^6 to obtain actual values of estimates of variance.

(*) Very high value.

POPULATION 7

TABLE - 7.6

Regression estimates of population mean $\bar{Y}$ and variance estimates based on samples of size 2

Estimator	Sample 1		Sample 2		Sample 3		Sample 4		Sample 5	
	Estimate	Estimate of variance	Estimate	Estimate of variance	Estimate	Estimate of variance	Estimate	Estimate of variance	Estimate	Estimate of variance
D_1^*	21.95	421.46 (*)	-ve	-ve	12.04	4.22	14.77	149.51 (*)	27.23	660.89 (*)

(*) Very high value.

POPULATION 7

TABLE - 7.7

Regression estimates of population mean $\bar{Y}$ and variance estimates based on samples of size 3

Estimator	Sample 1		Sample 2		Sample 3		Sample 4		Sample 5	
	Estimate	Estimate of variance	Estimate	Estimate of variance	Estimate	Estimate of variance	Estimate	Estimate of variance	Estimate	Estimate of variance
D_y^*	11.48	37.01	12.00	21.38	11.34	39.70	15.46	47.43 (*)	11.75	143.95 (*)
D_2^*	12.20	83.55 (*)	11.60	49.48 (*)	9.38	32.02	9.50	21.55	10.35	50.14 (*)

POPULATION 7

TABLE - 7.8

Regression estimates of population mean $\bar{\bar{Y}}$ and variance estimates based on samples of size 4

Estimator	Sample 1		Sample 2		Sample 3		Sample 4		Sample 5	
	Estimate	Estimate of variance	Estimate	Estimate of variance	Estimate	Estimate of variance	Estimate	Estimate of variance	Estimate	Estimate of variance
D_1^*	6.93	4.02	7.34	9.75	7.08	5.42	8.79	0.51	7.72	11.02
D_2^*	5.10	14.95 (*)	6.67	5.71	5.51	77.01 (*)	3.77	11.87	8.43	11.33
D_3^*	6.02	0.25	7.22	13.44 (*)	7.21	-ve	3.56	-ve	7.98	22.99 (*)

(*) Very high value.

POPULATION 7

TABLE - 7.9

Regression estimates of population mean $\bar{Y}$ and variance estimates based on samples of size 5

Estimator	Sample 1		Sample 2		Sample 3		Sample 4		Sample 5	
	Estimate	Estimate of variance	Estimate	Estimate of variance	Estimate	Estimate of variance	Estimate	Estimate of variance	Estimate	Estimate of variance
D_1^*	7.63	-ve	7.33	2.14	5.59	0.62	7.23	5.98 (*)	4.88	0.88
D_2^*	6.99	8.42 (*)	6.44	2.04	5.62	95.24 (*)	6.69	0.07	1.63	-ve
D_3^*	6.82	0.77	6.46	0.54	4.96	0.52	6.66	-ve	4.99	1.41
D_4^*	6.98	11.06 (*)	6.59	5.39 (*)	4.87	-ve	6.40	-ve	4.93	2.27

(*) Very high value.

POPULATION 6

TABLE - 7.10

Estimates** of approximate variances of regression estimators using samples of size 10

Estimator	Sample Number															Average over samples
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
$N\hat{\bar{Y}}_r$	0.77	3.06	4.84	2.51	3.02	1.49	1.83	3.17	1.76	3.97	0.18	3.65	1.14	3.49	0.53	2.36
ND_1^*	0.77	3.06	4.84	2.51	3.02	1.49	1.83	3.17	1.76	3.97	0.18	3.65	1.14	3.49	0.53	2.36
ND_3^*	0.77	3.39	6.03	2.54	3.51	22.73	2.44	3.55	305.62	339.75	0.18	5.47	1.61	10.19	1.05	47.26
ND_5^*	1.02	3.32	5.73	2.51	3.74	5.34	1.89	3.20	44.26	111.65	0.18	3.68	1.17	5.39	0.81	12.93
ND_7^*	0.89	3.16	5.25	2.51	3.91	2.45	1.84	3.17	11.10	10.32	0.18	3.68	1.17	3.92	0.63	3.61
ND_9^*	0.78	3.07	4.89	2.51	3.29	1.57	1.83	3.17	2.68	4.08	0.18	3.66	1.15	3.53	0.54	2.46

** Table values are to be multiplied by 10^8 to obtain actual values.

POPULATION 6

TABLE - 7.11

Estimates** of approximate variances of regression estimators using samples of size 15

Estimator	Sample Number										Average over samples
	1	2	3	4	5	6	7	8	9	10	
$N\hat{\bar{Y}}_r$	1.24	1.74	1.69	1.71	1.92	1.44	0.98	2.66	1.85	1.29	1.65
ND_1^*	1.24	1.74	1.69	1.71	1.92	1.44	0.98	2.66	1.85	1.29	1.65
ND_2^*	56.70	124.67	50.95	53.49	28.65	227.95	2.75	39.03	264.07	7.78	85.60
ND_{13}^*	1.24	1.74	1.71	2.21	1.93	1.44	0.98	3.59	1.87	1.30	1.80
ND_{14}^*	1.24	1.74	1.70	1.83	1.92	1.44	0.98	2.91	1.85	1.29	1.69

** Table values are to be multiplied by 10^8 to obtain actual values.

POPULATION 6

TABLE - 7.12

Estimates** of approximate variances of regression estimators using samples of size 20

ati- ator	Sample Number 1	2	3	4	5	6	7	8	Average over samples
$\hat{Y}_r$	3.83	1.26	0.65	1.10	1.46	1.06	0.91	1.14	1.43
D_1^*	3.83	1.26	0.65	1.10	1.46	1.06	0.91	1.14	1.43
D_2^*	4.07	7.71	10.52	4.98	7.16	185.56	3.44	8.86	29.04
D_{18}^*	3.86	1.26	0.65	1.10	1.47	1.12	0.91	1.14	1.44
D_{19}^*	3.84	1.26	0.65	1.10	1.46	1.08	0.91	1.14	1.43

** Table values are to be multiplied by 10^8 to obtain actual values.

POPULATION 6

TABLE - 7.13

Estimates** of approximate variances of regression estimators using samples of size 25

imator	Sample Number 1	2	3	4	5	Average over samples
r	0.68	0.72	1.25	0.72	1.88	1.05
	0.68	0.72	1.25	0.72	1.88	1.05
4	0.68	0.72	1.25	0.72	1.87	1.05

** Table values are to be multiplied by 10^8 to obtain actual values.

POPULATION 7

TABLE - 7.14

Estimates of approximate variances of regression estimators using samples of size 9

Estimator	Sample Number					Average over samples
	1	2	3	4	5	
$\hat{Y}_r$	0.13	0.64	1.42	0.44	1.07	0.74
D_1^*	0.13	0.64	1.42	0.44	1.07	0.74
D_3^*	0.15	0.64	1.46	0.50	1.10	0.77
D_5^*	0.14	0.64	1.44	0.44	1.08	0.75
D_7^*	0.13	0.64	1.42	0.44	1.07	0.74
D_8^*	0.13	0.64	1.42	0.44	1.07	0.74

POPULATION 7

TABLE - 7.15

Estimates of approximate variances of regression estimators using samples of size 12

Estimator	Sample Number				Average over samples
	1	2	3	4	
$\hat{Y}_r$	0.30	0.36	0.20	0.50	0.34
D_1^*	0.30	0.36	0.20	0.50	0.34
D_{11}^*	0.30	0.36	0.20	0.50	0.34

Remarks on Tables 7.1 to 7.5 (Sample sizes 2, 3, 4 and 5):

Remark (a) : For m = n-1, the estimators of variance of ND_1^* and ND_m^* have assumed either negative or very high positive values. This is similar to the situation observed in respect of unbiased ratio type estimators (see remark (a) of section 4.5) and consequently in this case the estimates of variance do not reflect the true picture.

Remark (b) : In the light of (a), in samples of size 5, a trend of increasing precision may be noted with increasing m.

Remark (c) : Comparing the results of :

Table 7.1 with Table 4.5,

Table 7.2 with Table 4.6,

Table 7.3 with Table 4.7,

Table 7.4 with Table 4.8,

and Table 7.5 with Table 4.9 ;

it is clear that the performance of the unbiased regression type estimators is poor relative to the unbiased ratio type estimators and in most cases even the simple estimator $N\bar{y}$. In small samples Rao (1969) and Sampford (1969) have also made a similar observation on the relative performance of classical regression and ratio estimators.

Remarks on Tables 7.6 to 7.9 (Sample sizes 2, 3, 4, and 5):

Remark (d): For m = n-1, here again estimates of variance of D_1^* and D_m^*

are unsatisfactory. In the same population it has been so for the unbiased product type estimators also (see remark (a) of section 6.6).

Remark (e) : Making a comparison of the results of

Table 7.6 with Table 6.3,

Table 7.7 with Table 6.4,

Table 7.8 with Table 6.5,

and Table 7.9 with Table 6.6;

it may be seen that in small samples the unbiased regression type estimators are inferior to the unbiased product type estimators and in most cases even the simple unbiased estimator $\bar{y}$.

Remarks on Tables 7.10 to 7.13 (Sample sizes 10, 15, 20 and 25):

Remark (f) : When the estimator of variance is based on the large sample variance formula, the precision of the unbiased regression type estimator ND_m^* has increased as m increases from 2 to n-1. The estimator ND_{n-1} has the same efficiency as the classical regression estimator $N\hat{\bar{Y}}_r$ or the unbiased regression type estimator ND_1^*.

Remark (g) : Comparing the results of

Table 7.10 with Table 4.15,

Table 7.11 with Table 4.16,

Table 7.12 with Table 4.17,

and Table 7.13 with Table 4.18;

it may be seen that, unlike in small samples, the unbiased regression type estimator ND^*_{n-1} is superior to the unbiased ratio type estimator NT^*_{n-1} (which itself is superior to NT^*_m for $1 \leq m \leq n-2$), and has the same precision as the unbiased ratio type estimator $NS^*_{(q_o')}$.

Remarks on Tables 7.14 and 7.15 (Sample sizes 9 and 12) :

Remark (h): Here again the estimator of variance is based on the large sample variance formula. The results show that the choice $m = n-1$ is the best for D^*_m ($2 \leq m \leq n-1$). The estimator D^*_{n-1} has the same precision as D^*_1 and the classical regression estimator $\hat{Y}_r$.

Remark (i): Comparing the results of

Table 7.14 with Table 6.8,

and Table 7.15 with Table 6.9,

it is seen that, unlike in small samples, the unbiased regression type estimator D^*_{n-1} is more efficient than the unbiased product type estimators L^*_m and H^*_m, and has the same precision as the unbiased product type estimator $S^*_{(q_o' < o)}$.

CHAPTER VIII

AN EMPIRICAL STUDY IN SMALL SAMPLES

8.0 Summary :

The results of an empirical study conducted in small samples of sizes 2, 3, 4 and 5 on the relative performance of (a) biased and unbiased ratio and regression estimators and (b) biased and unbiased product and regression estimators, are reported in this chpater. It may be recalled that in chapters IV, VI and VII unbiased estimates of variance have been computed for the unbiased ratio, product and regression type estimators in small samples of sizes 2, 3, 4 and 5 for purposes of illustration and some general remarks have been made on the relative efficiency of the unbiased estimators. In those illustrations only a few samples of each size have been used and the classical biased ratio, product and regression estimators have not been included because their mean square errors cannot be estimated on the basis of single sample of small size. To enable comparison of biased and unbiased estimators in small samples, we follow a different approach in this chapter. For each sample size, a reasonable number (i.e., a multiple of $\frac{N}{n}$ where N and n are the population and sample sizes) of independent samples are selected from a population using the same sampling scheme. To estimate a population parameter Λ, if we use an estimator λ taking values λ_i ($i = 1, 2, \ldots, k$) on the k independent samples of same size; then the mean square among the k independent estimates λ_i, say $\hat{V}(\lambda)$, is taken as an unbiased estimator

of the variance of λ. When λ is a biased estimator and the 'true' value of Λ is known (as is the case in our empirical study), the mean square error of λ is estimated by $\hat{V}(\lambda) + (\bar{\lambda} - \Lambda)^2$ where $\bar{\lambda}$ is the mean of λ_i. For the study on ratio and regression estimators, we make use of data of ten populations (namely, the populations 1 to 6 of chapter IV and populations 8 to 11 given in Appendix A-1 of this chapter). Population 7 of chapter VI is used for the study on product and regression estimators.

The present chapter contains three sections. After a brief introductory section, in section 2 we present the study on ratio and regression estimators. The estimators included in the comparison are the simple unbiased estimator $N\bar{y}$, classical biased ratio estimator $N\hat{\bar{Y}}_R$, unbiased ratio type estimators $N T^*_m$ ($1 \leq m \leq n-1$), $N S^*_{(1)}$ and $N S^*_{(q'_o)}$, classical biased regression estimator $N\hat{\bar{Y}}_r$ and unbiased regression type estimators $N D^*_m$ ($2 \leq m \leq n-1$). To compute estimates of variance or mean square error of these estimators, independent samples are drawn using simple random sampling without replacement. In addition, for each sample size, using Midzuno-Sen (1952) sampling scheme, independent samples are also selected to compute an unbiased estimate of variance of Lahiri's (1951) unbiased ratio estimator. Care is taken to see that same number of samples are selected for both simple random sampling and Midzuno-Sen schemes. The actual number of independent samples of each sample size drawn from each of the ten populations is given in Table -8.2.

The sets of independent samples selected according to the two sampling schemes are presented in Appendix A - 2 and Appendix A - 3. The computed estimates of variance or mean square error are shown in Tables - 8.3 to 8.12, separately for each of the ten populations. The tentative conclusions are as follows.

The unbiased ratio type estimator NT^*_{n-1} is, in general, more efficient than other unbiased ratio type estimators based on simple random sampling design. The estimator NT^*_{n-1} is comparable in precision with the classical biased ratio estimator $N\hat{\bar{Y}}_R$ in only 12 out of 38 cases, thereby indicating that the unbiased estimator is not likely to be superior to the biased estimator in small samples. The unbiased regression type estimators are less efficient than the classical biased regression estimator $N\hat{\bar{Y}}_r$. The classical biased ratio estimator $N\hat{\bar{Y}}_R$ is superior to the classical biased regression estimator $N\hat{\bar{Y}}_r$ in small samples. The empirical results of Rao (1969) and Sampford (1969) also support this observation. The performance of the classical biased ratio estimator $N\hat{\bar{Y}}_R$ is remarkable even in small samples. Lahiri's (1951) unbiased ratio estimator based on Midzuno-Sen sampling scheme is either superior or equal in efficiency with the classical biased ratio estimator $N\hat{\bar{Y}}_R$ based on simple random sampling scheme in 25 out of 38 cases. Thus, in small samples, Lahiri's unbiased ratio estimator seems to be preferable to all other unbiased or biased ratio or regression estimators.

In section 3, we compare the performance of the simple unbiased estimator $\bar{y}$, classical biased product estimator $\hat{\bar{Y}}_P$, unbiased product type estimators $\hat{\bar{Y}}_{RP}$, $S^*_{(q_0' < 0)}$, L^*_m and H^*_m ($2 \leq m \leq n-1$), classical biased regression estimator $\hat{\bar{Y}}_r$ and unbiased regression type estimators D^*_m ($2 \leq m \leq n-1$). From the population 7 given in chapter VI, 45,30,25 and 20 independent simple random samples are drawn for the sample sizes 2, 3, 4 and 5 respectively. The estimates of variance or mean square error are given in Table - 8.13. The tentative conclusions are as follows. Among the unbiased product type estimators, the estimator $S^*_{(q_0' < 0)}$ is the best. The classical biased product estimator $\hat{\bar{Y}}_P$ is as efficient as the unbiased estimator $S^*_{(q_0' < 0)}$. The classical biased regression estimator $\hat{\bar{Y}}_r$ is more efficient than the unbiased regression type estimators $D^*_m(2 \leq m \leq n-1)$. The regression estimators are inferior to the product estimators in small samples.

8.1 <u>Introduction</u> :

By way of illustrating the unbiased estimators of variance of unbiased ratio, product and regression type estimators, we have used in chapters IV, VI and VII small samples of sizes 2, 3, 4 and 5 selected with equal probability and without replacement from some <u>known</u> populations; and made some general remarks on the relative performance of ratio and regression, and product and regression type unbiased estimators in small samples. In these illustrations only a few samples of each size have been used, and the classical biased ratio, product and regression estimators have not been included since it is not possible to estimate their mean square errors from a small sample. In this chapter we conduct an empirical study on the relative performance of the biased and unbiased ratio and regression, and product and regression estimators in small samples on the basis of a reasonable number of independent samples of each of the sizes 2, 3, 4 and 5 selected from the populations 1 to 7 (given in previous chapters) and some additional populations numbered 8 to 11.

To estimate a population parameter, say Λ, if we use a statistic λ taking values λ_i ($i = 1, 2, \ldots, k$) on k independent samples of same size and selected from the population according to the same sampling design, then an unbiased estimator of the variance of λ is

$$\hat{V}(\lambda) = \frac{1}{(k-1)} \sum_{i=1}^{k} (\lambda_i - \bar{\lambda})^2, \qquad (8.1.1)$$

where $\bar{\lambda} = \frac{1}{k} \sum_{i=1}^{k} \lambda_i$.

When λ is a biased estimator and Λ is known, as is the case in our empirical study, we may use $(\bar{\lambda} - \Lambda)$ to estimate the bias in λ. Then as an estimator of the mean square error of λ we take

$$\widehat{MSE}(\lambda) = \hat{V}(\lambda) + (\bar{\lambda} - \Lambda)^2. \qquad (8.1.2)$$

It may be noted that while $\hat{V}(\lambda)$ is unbiased for $V(\lambda)$, $\widehat{MSE}(\lambda)$ is not an unbiased estimator of $MSE(\lambda)$.

In our study, to obtain a comparison of the estimators, we use for biased estimators MSE (λ) and for unbiased estimators $\hat{V}(\lambda)$. To estimate MSE (λ), it is possible to use an unbiased estimator:

$$mse(\lambda) = \frac{1}{k}\sum_{i=1}^{k}(\lambda_i - \Lambda)^2 = \frac{1}{k}\sum_{i=1}^{k}(\lambda_i - \bar{\lambda})^2 + (\bar{\lambda} - \Lambda)^2$$

$$= \left(\frac{k-1}{k}\right)\hat{V}(\lambda) + (\bar{\lambda} - \Lambda)^2. \qquad (8.1.3)$$

When the number of independent samples k is fairly large, the relative performance of the estimators on the basis of (8.1.2) and (8.1.3) has not been much different and the conclusions drawn have not altered.

Empirical studies on ratio and regression estimators conducted by Rao and Beegle (1967), Rao (1969), Sampford (1969), Hutchison (1971), and Rao and Kuzik (1973) have already been reviewed in chapter II. We now proceed with our study first on the ratio and regression estimators and then on product and regression estimators.

8.2 Comparison of biased and unbiased ratio and regression estimators in small samples:

We consider estimation of the population total Y of the character under study (y). The estimators used are:

the simple unbiased estimator : $N\bar{y}$,

classical biased ratio estimator: $N\hat{Y}_R$, Lahiri's unbiased ratio estimator in Midzuno-Sen sampling scheme : $N \hat{\bar{Y}}_{MR}$,

unbiased ratio type estimators : $N\, T_m^*(1 \leq m \leq n-1)$, $NS^*_{(1)}$ and $N\, S^*_{(q'_0)}$,

classical biased regression estimator: $N \hat{\bar{Y}}_r$, and

unbiased regression type estimators : $ND_m^*\,(2 \leq m \leq n-1)$.

The study has been conducted on 10 populations. They are populations 1 to 6 given in chapter IV and four additional populations described here. For the 142 cities/urban agglomerations given in Table - 4.10, we define

Population 8 : x : 1971 census population (persons), y : 1971 workers (persons) as per census;

Population 9 : x : 1971 census female population, y : 1971 female workers as per census;

Population 10 : x : 1971 census female population, y : 1971 female literates as per census.

For 20 districts of the State of Andhra Pradesh (India), we define

Population 11 : x : 1951 census population (persons), y : 1961 census population (persons).

The relevant data of populations 8 to 11 are presented in Appendix A - 1.

We have given some parameters of interest for populations 1 to 5 in Table - 4.3, and for population 6 in (4.5.13). Similar information in respect of the populations 8 to 11 is provided in Table-8.1.

TABLE - 8.1

Values of some population parameters

rite/ ameter	Population Number 8	9	10	11
y	1971 workers (persons)	1971 female workers	1971 female literates	1961 population (persons)
x	1971 population (persons)	1971 females	1971 females	1951 population (persons)
Y	170683	16901	120029	35984
X	570161	257560	257560	31116
σ_y^2	8004652	55115	3470724	270994
σ_x^2	71896861	12920363	12920363	223167
xy	23879190	800762	6672646	244440
	2.35	1.97	2.20	0.29
C_x	2.11	1.98	1.98	0.30
	0.9954	0.9489	0.9964	0.9940
	0.33	0.06	0.52	1.10

We consider four sample sizes: 2, 3, 4 and 5. From a population of size N we select a multiple of $\frac{N}{n}$ independent samples of size n, llowing two sampling schemes:

i) simple random sampling without replacement,

nd ii) Midzuno-Sen (1952) sampling scheme.

Table - 8.2 shows the number of samples selected for each sample size from each of the 10 populations.

TABLE - 8.2

Number of independent samples of different sizes selected from various populations

Population Number	Population Size (N)	Sample Size (n)			
		2	3	4	5
1	43	50	35	25	20
2 to 5	22	40	25	20	15
6, 8 to 10	142	70	50	35	30
11	20	30	20	-	-

The sets of independent samples selected according to simple random sampling scheme and Midzuno-Sen scheme are presented respectively in Appendix A-2 and Appendix A-3.

As stated in the introduction, we have computed for unbiased estimators the estimator of variance $\hat{V}(\lambda)$ given by (8.1.1), and for biased estimators the estimator of mean square error $\widehat{MSE}(\lambda)$ given by (8.1.2). The final computations on the samples are now presented in Tables - 8.3 to 8.12.

POPULATION 1

TABLE - 8.3

Estimates** of Variance/MSE of some estimators of population total (Y)

Estimator	Sample Size 2	3	4	5
$N\bar{y}$	3.46	3.01	1.07	1.22
$N\hat{Y}_R$	1.78	0.88	1.02	0.68
$N\hat{Y}_{MR}$	2.42	1.38	1.02	0.85
$NS^*_{(1)}$	2.44	0.89	0.95	0.79
$NS^*_{(q'_o)}$	2.43	0.89	0.94	0.79
NT^*_1	2.21	0.83	1.03	0.73
NT^*_2	-	0.82	1.03	0.74
NT^*_3	-	-	1.02	0.74
NT^*_4	-	-		0.74
$N\hat{Y}_r$	16.63(*)	4.54	1.21	0.77
ND^*_2		108.11(*)	7.39	2.35
ND^*_3	-	-	1.52	2.05
ND^*_4		-	-	1.13

** Table values are to be multiplied by 10^4 to obtain actual values.

(*) Very high value.

POPULATION 2

TABLE - 8.4

Estimates** of Variance/MSE of some estimators of population total (Y)

Estimator	Sample Size 2	3	4	5
$N\bar{y}$	247.96 (*)	158.98 (*)	63.59 (*)	82.50 (*)
$N\hat{Y}_R$	1.67	1.69	0.53	0.69
$N\hat{Y}_{MR}$	0.82	0.80	0.24	0.24
$NS^*_{(1)}$	401.36 (*)	37.55 (*)	11.84 (*)	10.42 (*)
$NS^*_{(q_0^1)}$	412.52 (*)	37.88 (*)	11.86 (*)	10.78 (*)
NT^*_1	3.12	2.05	0.65	0.81
NT^*_2	-	1.74	0.47	0.69
NT^*_3	-	-	0.43	0.69
NT^*_4	-	-	-	0.70
$N\hat{Y}_r$	288.67 (*)	11.46	1.31	0.90
ND^*_2	-	293.27 (*)	4.52	80.00 (*)
ND^*_3	-		5.66	2.05
ND^*_4	-	-		1.50

** Table values are to be multiplied by 10^7 to obtain actual values.

(*) Very high value.

POPULATION 3

TABLE - 8.5

Estimates** of Variance/MSE of some estimators of population total (Y)

Estimator	Sample Size			
	2	3	4	5
$N\bar{y}$	227.65 (*)	92.66 (*)	57.09 (*)	71.12 (*)
$N\hat{Y}_R$	1.15	0.99	0.67	0.28
$N\hat{Y}_{MR}$	1.06	0.80	0.40	0.53
$NS^*_{(1)}$	372.52 (*)	35.25 (*)	9.70	8.88
$NS^*_{(q'_0)}$	418.76 (*)	36.12 (*)	9.58	10.14
NT^*_1	2.90	1.47	0.73	0.48
NT^*_2		1.30	0.69	0.46
NT^*_3	-	-	0.68	0.45
NT^*_4	-	-	-	0.46
$N\hat{Y}_r$	60.99 (*)	4.80	0.83	0.56
ND^*_2	-	18.15 (*)	240.37 (*)	0.68
ND^*_3	-		0.84	0.49
ND^*_4	-	-	-	0.51

** Table values are to be multiplied by 10^6 to obtain actual values.

(*) Very high value.

POPULATION 4

TABLE 8.6

Estimates** of Variance/MSE of some estimators of population total (Y)

Estimator	Sample Size			
	2	3	4	5
$N\bar{y}$	1.19	0.70	0.23	0.29
$N\hat{\bar{Y}}_R$	0.32	0.23	0.12	0.17
$N\hat{\bar{Y}}_{MR}$	0.16	0.06	0.20	0.15
$NS_{(1)}$	1.92	0.52	0.15	0.24
$NS^*_{(q'_o)}$	2.16	0.57	0.16	0.27
NT^*_1	0.58	0.26	0.10	0.14
NT^*_2		0.31	0.11	0.17
NT^*_3	-	-	0.12	0.19
NT^*_4	-	-	-	0.21
$N\hat{\bar{Y}}_r$	1.94	0.55	0.14	0.38
ND^*_2	-	113.29 (*)	0.31	0.35
ND^*_3		-	0.25	0.40
ND^*_4	-	-		0.57

** Table values are to be multiplied by 10^6 to obtain actual values.

(*) Very high value.

POPULATION - 5

TABLE - 8.7

Estimates** of Variance/MSE of some estimators of population total (Y)

Estimator	Sample Size			
	2	3	4	5
$N\bar{y}$	83.20 (*)	32.17 (*)	23.59 (*)	26.21 (*)
$N\hat{\bar{Y}}_R$	2.32	1.48	1.30	0.78
$N\hat{\bar{Y}}_{MR}$	2.48	1.37	1.08	0.88
$NS^*_{(1)}$	139.20 (*)	12.25 (*)	5.20	3.45
$NS^*_{(q_o')}$	171.60 (*)	12.75 (*)	5.00	4.16
NT^*_1	4.86	1.34	1.82	0.93
NT^*_2	-	1.43	1.62	0.92
NT^*_3	-		1.59	0.92
NT^*_4	-	-	-	0.92
$N\hat{\bar{Y}}_r$	92.95 (*)	4.07	1.45	0.85
ND^*_2	-	1131.69 (*)	16.12 (*)	2.88
ND^*_3	-	-	2.90	1.00
ND^*_4			-	1.26

** Table values are to be multiplied by 10^6 to obtain actual values.

(*) Very high value.

POPULATION 6

TABLE 8.8

Estimates** of Variance/MSE of some estimators of population total (Y)

Estimator	Sample Size			
	2	3	4	5
$N\bar{y}$	57.14 (*)	38.74 (*)	41.99 (*)	17.13 (*)
$N\hat{\bar{Y}}_R$	0.37	0.39	0.13	0.22
$N\hat{\bar{Y}}_{MR}$	0.30	0.23	0.15	0.12
NT_1^*	0.98	0.62	0.61	0.61
NT_2^*	-	1.02	0.44	0.33
NT_3^*	-		0.38	0.29
NT_4^*	-	-	-	0.29
$N\hat{\bar{Y}}_r$	49.90 (*)	3.87	0.40	0.45
ND_2^*	-	23.29 (*)	213.58 (*)	18.94 (*)
ND_3^*	-		3.61	0.94
ND_4^*	-	-	-	1.16

** Table values are to be multiplied by 10^{10} to obtain actual values.

(*) Very high value

POPULATION 8

TABLE - 8.9

Estimates** of Variance/MSE of some estimators of population total (Y)

Estimator	Sample Size 2	3	4	5
$N\bar{y}$	72.83 (*)	44.02 (*)	47.81 (*)	17.48 (*)
$N\hat{Y}_R$	0.22	0.16	0.17	0.15
$N\hat{Y}_{MR}$	0.33	0.29	0.15	0.27
NT_1^*	5.59	2.42	1.23	0.54
NT_2^*	-	2.33	1.19	0.43
NT_3^*	-	-	1.19	0.40
NT_4^*	-	-	-	0.40
$N\hat{Y}_r$	6.42	0.54	0.17	0.18
ND_2^*	-	20.02 (*)	36.98 (*)	91.29 (*)
ND_3^*	-	-	1.44	1.22
ND_4^*	-	-	-	0.87

** Table values are to be multiplied by 10^9 to obtain actual values.

(*) Very high value.

POPULATION 9

TABLE - 8.10

stimates** of Variance/MSE of some estimators of population total (Y)

Estimator	Sample Size			
	2	3	4	5
$N\bar{y}$	7.00	2.98	2.94	0.83
$N\hat{\bar{Y}}_R$	0.64	0.20	0.18	0.13
$N\hat{\bar{Y}}_{MR}$	0.21	0.32	0.17	0.15
NT_1^*	1.26	0.68	2.36	0.15
NT_2^*	-	0.62	2.62	0.16
NT_3^*	-	-	2.85	0.16
NT_4^*	-	-	-	0.16
$N\hat{\bar{Y}}_r$	42.45 (*)	1.09	0.36	0.29
ND_2^*	-	57.91 (*)	***	22.14 (*)
ND_3^*	-	-	12.54	5.05
ND_4^*	-	-		1.71

*** For 7 out of 35 samples, the estimator ND_2^* has taken negative values. Hence no estimate of variance is computed.

** Table values are to be multiplied by 10^8 to obtain actual values.

(*) Very high value.

POPULATION 10

TABLE - 8.11

Estimates** of Variance/MSE of some estimators of population total (Y)

Estimator	Sample Size			
	2	3	4	5
$N\bar{y}$	29.68*	18.94*	19.27(*)	7.58
$N\hat{Y}_R$	0.42	0.35	0.21	0.16
$N\hat{\bar{Y}}_{MR}$	0.25	0.14	0.30	0.14
NT_1^*	2.37	0.61	0.96	0.34
NT_2^*	-	0.68	1.05	0.33
NT_3^*	-	-	1.18	0.34
NT_4^*	-	-	-	0.36
$N\hat{Y}_T$	9.63	2.50	0.57	0.45
ND_2^*	-	24.42(*)	147.64(*)	10.99(*)
ND_3^*	-	-	3.52	2.32
ND_4^*	-	-	-	2.21

** Table values are to be multiplied by 10^9 to obtain actual values.

(*) Very high value.

POPULATION 11

TABLE - 8.12

Estimates** of Variance/MSE of some estimators of population total (Y)

Estimator	Sample Size	
	2	3
$N\bar{y}$	44.05(*)	29.07(*)
$N\hat{\bar{Y}}_R$	0.74	0.43
$N\hat{\bar{Y}}_{MR}$	0.79	0.65
NT_1^*	0.69	0.41
NT_2^*	-	0.39
$N\hat{Y}_r$	5.63	0.69
ND_2^*	-	4.35

** Table values are to be multiplied by 10^6 to obtain actual values.

(*) Very high value.

Remarks on Tables 8.3 to 8.12 :

Remark (a) : Some of the estimates of variance/mean square error recorded are very high values. This is due to the fact that the corresponding estimators of population total Y have assumed for some samples either too low or too high values. This has been observed when the sample-wise estimates of population total Y, based on different estimators, have been examined in detail. In view of limitations of space, we have not given in this thesis sample wise estimates of population total Y. They are, however, available in the computer output.

Remark (b) : It may be noted that our remarks on Tables 4.5 to 4.9 in chapter IV on the performance of the unbiased ratio type estimators $NS^*_{(1)}$, $NS^*_{(q_o')}$ and NT^*_m and the remarks on Tables 7.1 to 7.5 in chapter VII on the performance of the unbiased regression type estimators ND^*_m ($m \geq 2$), which have been made on the basis of unbiased estimators of variance computed from single sample, are generally consistent with the evidence obtained in Tables 8.3 to 8.12 of this chapter by following a different approach. In addition, the new information thrown up by Tables 8.3 to 8.12 relates to the following interesting observations.

Remark (c) : The classical biased ratio estimator $N\hat{\bar{Y}}_R$ performed remarkably well even in small samples of sizes 2, 3, 4 and 5.

Remark (d) : The classical biased regression estimator $N\hat{\bar{Y}}_r$ is definitely inferior to the classical ratio estimator $N\hat{\bar{Y}}_R$ in small samples.

Remark (e) : The unbiased ratio type estimator NT^*_{n-1} ($n > 2$), which is, in general, better than NT^*_m ($1 \leq m \leq n-2$), is comparable in precision with the classical biased ratio estimator $N\hat{\bar{Y}}_R$ in only 12 out of 38 cases, thereby indicating that the unbiased estimator is slightly inferior to the biased estimator in small samples.

Remark (f) : All the unbiased regression estimators ND^*_m ($m \geq 2$) are less efficient than the biased regression estimator $N\hat{Y}_r$ in small samples.

Remark (g) : Lahiri's unbiased ratio estimator $N\hat{Y}_{MR}$, based on a different sampling scheme (Midzuno-Sen), is either superior to or equal in efficiency with the biased ratio estimator $N\hat{Y}_R$ in 25 out of 38 cases, thereby indicating that in small samples Lahiri's unbiased ratio estimator is likely to be more efficient than the classical biased ratio estimator $N\hat{Y}_R$ as well as Mickey's unbiased ratio estimators NT^*_m $(1 \leq m \leq n-1)$.

8.3. Comparison of biased and unbiased product and regression estimators in small samples:

Here we estimate the population mean $\bar{Y}$ of the character under study, using information on a negatively associated character x. The estimators used are:

the simple unbiased estimator: $\bar{y}$,

conventional biased product estimator : $\hat{Y}_P$,

unbiased product type estimators : $\hat{Y}_{RP}$, $S^*_{(q'_o < o)}$, L^*_m and $H^*_m (2 \leq m \leq n-1)$;

classical biased regression estimator : $\hat{Y}_r$ and

unbiased regression type estimators : D^*_m $(2 \leq m \leq n-1)$.

The empirical investigation has been carried out using population 7 of Table 6.1 (chapter VI). Again four sample sizes 2, 3, 4 and 5 are tried. For the sample sizes 2, 3, 4 and 5, the numbers of independent simple random samples drawn without replacement from population 7 are respectively 45, 30, 25 and 20.

The sets of independent samples are given in Table A-2.5 of Appendix A-2. The estimates of variance/mean square error, computed for the different estimators using formulae (8.1.1) and (8.1.2), are now presented in Table 8.13.

POPULATION 7

TABLE - 8.13

Estimates of Variance/MSE of some estimators of population mean ($\bar{Y}$)

Estimator	Sample Size 2	3	4	5
$\bar{y}$	5.35	5.10	3.94	2.00
$\hat{\bar{Y}}_P$	3.68	3.96	1.84	1.15
$\hat{\bar{Y}}_{RP}$	4.16	4.32	1.97	1.08
$S^*_{(q'_0 < 0)}$	3.99	4.07	1.51	1.08
L^*_2	-	4.27	1.84	1.08
H^*_2	-	4.38	2.03	1.09
L^*_3	-	-	1.81	1.08
H^*_3	-	-	2.05	1.09
L^*_4	-	-	-	1.08
H^*_4	-	-	-	1.09
$\hat{\bar{Y}}_r$	32.69 (*)	9.04	3.61	1.35
D^*_2	-	612.11 (*)	21.82 (*)	16.64 (*)
D^*_3	-		3.88	2.23
D^*_4	-	-	-	1.96

(*) Very high value.

Remarks on Table 8.13 :

Remark (a) : The results in Table 8.13 are, in general, consistent with our remarks on Tables 6.3 to 6.6 in chapter VI on the performance of the unbiased product type estimators and our remarks on Tables 7.6 to 7.9 on the performance of the unbiased regression type estimators in chapter VII. The additional information from this study relates to the following observations.

Remark (b) : The biased product estimator $\hat{\bar{Y}}_p$ is almost as efficient as the unbiased product type estimator $S^*_{(q_o', \angle, o)}$, which is the best among the unbiased estimators in this study.

Remark (c) : The classical biased regression estimator $\hat{\bar{Y}}_r$ is superior to the unbiased regression type estimators D^*_m ($m \geq 2$) but is inferior to product estimators in the small samples.

APPENDIX A-1

DATA FOR POPULATIONS 8*, 9 AND 10

TABLE - A-1.1

INDIA (1971)

Workers, Females, Female workers and Female literates of 142 cities and urban agglomerations with population 100,000 +

(Figures in '00s)

Sl. No.	Name of the city/ urban agglomeration (100,000 +)	Workers (persons)	Females	Female workers	Female literates
(1)	(2)	(3)	(4)	(5)	(6)
1	Hyderabad	4369	8653	708	3608
2	Visakhapatnam	1000	1737	99	681
3	Vizayawada	1018	1662	101	761
4	Guntur	888	1331	210	498
5	Warangal	566	997	90	306
6	Rajahmundry	575	929	91	411
7	Kakinada	458	816	63	340
8	Kurnool	433	668	99	244
9	Nellore	397	653	50	305
10	Eluru	391	639	66	279
11	Nizamabad	397	555	90	147
12	Machilipatnam	311	550	43	269
13	Tenali	309	508	47	216
14	Gauhati	441	481	40	201
15	Patna	1373	2164	128	873
16	Jamshedpur	1332	2069	119	876
17	Dhanbad	1597	1728	131	516
18	Ranchi	688	1141	55	564
19	Gaya	463	823	42	272
20	Bhagalpur	447	772	31	283
21	Darbhanga	341	605	29	169
22	Muzaffarpur	403	542	50	212
23	Bokaro Steel City	471	425	52	69
24	Monghyr	241	469	22	153
25	Bihar	293	467	43	116

contd..

* 1971 census population (persons), used as auxiliary character for population 8, may be obtained from Table 4.10 in chapter IV.

Table - A-1.1 (contd.)

(1)	(2)	(3)	(4)	(5)	(6)
26	Ahmedabad	4525	7221	358	3594
27	Surat	1492	2221	170	1084
28	Baroda	1305	2150	115	1173
29	Rajkot	768	1440	61	741
30	Bhavnagar	578	1080	60	492
31	Jamnagar	572	1027	55	451
32	Nadiad	291	509	28	268
33	Rohtak	301	578	18	273
34	Ambala Cantt.	264	492	19	229
35	Srinagar	1083	1856	55	420
36	Jammu	438	707	31	350
37	Cochin	1195	2144	200	1373
38	Trivandrum	1154	2037	244	1307
9	Calicut	823	1659	113	968
0	Alleppey	410	798	73	516
41	Quilon	330	616	63	396
42	Indore	1479	2650	148	1235
43	Jabalpur	1541	2400	163	1072
44	Gwalior	1069	1860	86	660
45	Bhopal	1127	1772	91	753
46	Durg-Bhilainagar	751	1112	75	431
47	Ujjain	540	992	59	400
48	Raipur	607	970	99	396
49	Sagar	464	709	74	286
50	Bilaspur	391	619	64	264
51	Ratlam	298	562	27	249
52	Burhanpur	287	508	37	169
53	Greater Bombay	21862	24915	2087	13486
54	Nagpur	2350	4104	319	1946
55	Poona	2495	3992	356	2131
56	Sholapur	1159	1898	227	631
57	Kolhapur	684	1217	66	590
58	Amravati	511	907	62	426
59	Malegaon	530	823	[illegible]	293
60	Nasik	498	833	80	425

contd...

Table - A-1.1 (contd.)

(1)	(2)	(3)	(4)	(5)	(6)
61	Thana	575	745	67	410
62	Akola	456	787	50	359
63	Ulhasnagar	473	788	34	379
64	Aurangabad	401	682	65	281
65	Dhulia	347	644	47	297
66	Nanded	327	590	34	174
67	Ahmadnagar	327	557	65	293
68	Sangli	319	536	35	229
69	Jalgaon	278	503	38	237
70	Bangalore	4884	7690	632	3929
71	Hubli-Dharwar	1055	1784	137	761
72	Mysore	937	1689	115	826
73	Mangalore	771	1068	253	612
74	Belgaum	585	1004	51	508
75	Gulbarga	383	690	46	228
76	Bellary	361	596	51	213
77	Devanagere	366	566	54	229
78	Bijapur	251	489	22	186
79	Shimoga	299	483	45	236
80	Bhadravathi	271	478	23	188
81	Cuttack	579	848	39	397
82	Rourkela	584	737	47	302
83	Berhampur	322	567	42	215
84	Bhubaneswar	339	444	23	224
85	Amritsar	1297	1963	73	1011
86	Ludhiana	1208	1791	52	908
87	Jullundur	804	1368	38	697
88	Patiala	425	691	35	359
89	Jaipur	1655	2829	124	1025
90	Jodhpur	854	1467	108	509
91	Ajmer	661	1235	67	580
92	Kota	618	956	45	336
93	Bikaner	461	878	39	292
94	Udaipur	494	742	73	301
5	Alwar	263	457	22	157

contd...

Table - A-1.1 (contd.)

(1)	(2)	(3)	(4)	(5)	(6)
96	Madras City	7155	11715	784	6174
97	Madurai	1559	2670	220	1424
98	Coimbatore	1102	1671	140	942
99	Salem	987	1501	153	658
100	Tiruchirapalli	887	1490	108	830
101	Tuticorin	441	767	63	411
102	Nagercoil	385	704	57	448
103	Thanjavur	368	693	44	370
104	Vellore	412	676	51	352
105	Dindigul	353	627	38	297
106	Singanallur	385	546	81	227
107	Tiruppur	410	544	72	226
108	Kumbakonam	305	562	25	272
109	Kanchipuram	343	543	41	231
110	Tirunelveli	306	539	36	266
111	Erode	344	498	48	223
112	Cuddalore	271	501	33	212
113	Kanpur	3866	5504	204	2268
114	Lucknow	2390	3696	144	1603
115	Agra	1695	2909	88	945
116	Varanasi	1823	2637	273	842
117	Allahabad	1499	2260	130	934
118	Meerut	1058	1653	47	622
119	Bareilly	888	1493	43	499
120	Moradabad	771	1245	46	418
121	Aligarh	668	1152	34	377
122	Gorakhpur	617	1024	39	424
123	Soharanpur	618	1026	21	371
124	Dehra Dun	591	873	38	480
125	Jhansi	497	933	50	347
126	Rampur	453	753	14	178
127	Shahjahanpur	384	668	12	168
128	Mathura	377	638	20	233
129	Firozabad	361	610	14	185
130	Ghaziabad	368	567	15	222

contd...

TABLE A-1.1 (contd.)

(1)	(2)	(3)	(4)	(5)	(6)
131	Muzaffarnagar	279	525	11	201
132	Farrukhabad-cum-Fatehgarh	322	507	19	171
133	Faizabad	311	478	18	159
134	Mirzapur-cum-Vindhyachal	308	488	28	123
135	Calcutta	22826	28876	1370	14575
136	Durgapur	660	906	47	418
137	Kharagpur	401	755	67	363
138	Asansol	428	673	29	329
139	Burdwan	387	649	33	283
140	Chandigarh	774	1002	67	579
141	Delhi	11151	16110	827	8197
142	Imphal	286	499	95	217

APPENDIX A-1 (contd.)

DATA FOR POPULATION II

TABLE - A-1.2

STATE: ANDHRA PRADESH

1951 and 1961 census population of 20 districts (Figures in '000s)

Serial Number	Name of the district	1951 population	1961 population
1	Srikakulam	2123	2341
2	Visakhapatnam	2073	2291
3	East Godavari	2302	2608
4	West Godavari	1698	1978
5	Krishna	1736	2077
6	Guntur	2560	3010
7	Nellore	1795	2034
8	Chittor	1666	1915
9	Cuddapah	1163	1342
10	Anantapur	1484	1767
11	Kurnool	1617	1909
12	Mahbubnagar	1447	1591
13	Hyderabad	1822	2063
14	Medak	1110	1227
15	Nizamabad	835	1022
16	Adilabad	832	1009
17	Karimnagar	1428	1622
18	Warangal	1330	1545
19	Khammam	808	1058
20	Nalgonda	1287	1575

APPENDIX A - 2

POPULATION 1

TABLE - A-2.1

Independent simple random samples without replacement (Sample Units)

Sample Size	Sample Number 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2	8	11	25	43	21	14	11	9	18	21	39	18	37	21	19	26	42	4	8	42
	3	23	16	16	39	21	13	42	40	23	14	30	41	10	21	12	38	35	16	14
3	31	35	13	12	35	34	4	25	41	31	27	33	14	24	25	10	6	21	15	9
	6	7	20	5	8	36	31	4	31	39	23	22	30	20	22	4	30	26	42	37
	39	18	9	36	32	38	17	16	34	42	29	32	11	43	15	33	5	13	3	31
4	13	7	20	5	11	34	2	17	5	17	22	24	3	25	31	26	15	26	9	36
	28	38	18	3	13	5	18	9	25	1	6	42	25	8	9	3	25	2	8	33
	16	4	19	25	6	17	29	33	1	10	40	21	10	14	19	37	30	14	33	18
	43	1	14	17	12	25	20	41	11	18	8	34	13	20	26	24	32	24	1	6
5	34	24	41	7	12	10	12	42	27	16	38	42	29	10	13	11	18	29	33	33
	17	42	31	20	43	7	24	39	22	19	1	15	2	6	9	28	33	13	6	23
	40	27	40	11	17	43	22	43	20	23	4	41	15	23	5	23	28	4	42	16
	16	39	15	34	38	11	19	17	41	8	27	27	20	40	7	20	16	10	20	34
	37	18	8	8	19	18	37	8	15	9	15	12	33	39	41	7	35	38	27	30

TABLE - A-2.1 (contd.)

Sample Size	Sample Number														
	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
2	40 34	32 17	36 15	36 6	8 23	38 33	41 2	39 5	9 14	34 25	28 2	21 13	11 3	18 25	21 39
3	27 13 18	10 35 37	42 36 19	24 13 27	38 3 6	35 21 37	19 2 21	11 30 35	29 25 11	6 5 29	12 36 34	31 3 37	32 26 40	37 6 1	42 40 10
4	20 32 31 28	6 28 7 21	16 40 26 7	9 29 16 23	42 28 16 27										

contd. . .

Sample Size	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
2	33 26	41 37	40 35	29 4	2 36	42 17	17 20	36 23	20 16	24 8	29 18	18 38	28 11	24 21	41 17

APPENDIX A-2 (contd.)

POPULATIONS 2 to 5

TABLE - A-2.2

Independent simple random samples without replacement (Sample Units)

Sample Size	Sample Number 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2	2	17	18	4	20	8	6	4	18	19	17	2	3	13	14
	17	5	7	10	16	3	1	13	15	14	19	1	22	19	6
3	6	1	11	14	7	1	5	8	9	3	8	2	21	22	4
	21	14	17	19	19	11	7	6	16	17	3	15	8	15	1
	20	12	10	13	9	12	15	11	7	21	9	21	12	3	13
4	17	17	3	11	12	6	22	13	16	16	9	8	7	16	11
	2	16	10	19	8	17	11	3	15	7	18	19	9	15	12
	14	3	9	4	16	16	2	10	22	2	14	1	3	7	6
	3	9	18	13	13	9	1	9	10	19	15	7	22	9	5
5	7	18	6	19	16	9	10	6	3	3	21	10	18	21	14
	10	16	18	7	13	12	3	4	4	7	11	17	17	4	10
	9	13	10	6	6	18	21	10	17	13	9	1	19	12	13
	5	9	5	10	10	19	8	2	16	9	5	6	4	15	8
	1	4	1	5	18	8	5	3	20	1	12	19	15	16	22

Table - A-2.2 (contd.)

Sample Size	Sample Number														
	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
2	1 17	16 17	15 2	22 12	19 21	16 15	21 3	6 22	17 4	13 10	19 18	7 8	14 3	21 17	1 20
3	10 8 17	16 22 17	14 17 6	16 18 17	11 15 3	18 10 21	10 19 21	13 4 16	4 22 14	3 5 10					
4	12 7 6 14	2 18 8 15	5 21 17 18	12 20 2 3	6 12 2 13										

contd....

Sample Size	31	32	33	34	35	36	37	38	39	40
2	15 21	6 22	1 9	13 2	18 16	18 7	19 12	14 19	18 2	4 7

APPENDIX A-2 (contd.)

POPULATIONS 6, 8, 9 and 10

TABLE - A-2.3

Independent simple random samples without replacement (Sample Units)

Sample Size	Sample Number 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
2	126	19	139	121	114	61	118	9	30	78	97	93	7	28	40	41	37	12
	21	39	45	120	17	78	29	45	19	124	83	53	62	25	31	47	75	94
3	33	109	70	88	100	48	85	94	103	67	42	119	138	98	121	117	79	133
	20	134	133	71	49	6	71	60	96	99	90	140	126	56	33	13	132	111
	113	55	81	45	36	27	124	79	51	97	35	15	32	78	67	35	86	126
4	115	4	61	136	35	40	59	103	46	70	91	101	127	5	47	46	7	50
	73	135	73	77	112	119	127	128	120	42	87	55	112	47	2	32	121	83
	33	52	69	117	70	65	47	113	117	57	97	45	33	14	43	30	43	76
	89	88	72	61	110	123	105	61	10	19	37	88	15	46	79	111	104	124
5	116	102	112	68	20	63	114	34	55	31	71	118	139	41	119	106	120	107
	46	79	123	116	8	44	138	113	47	21	124	138	45	115	83	119	25	52
	113	136	122	131	76	48	54	22	132	39	1	76	5	114	138	63	92	93
	106	39	37	70	103	7	66	41	59	104	82	95	19	28	76	56	104	63
	65	121	136	22	44	39	126	9	112	40	75	12	128	82	34	98	56	96

Table - A-2.3 (contd.)

Sample Size	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
2	44	115	81	54	10	46	95	138	28	2	132	107	85	37	18	93	4	96
	129	112	4	53	11	81	139	122	71	127	128	108	115	56	99	19	82	74
3	51	55	45	76	5	80	82	110	80	12	101	138	53	62	112	17	125	48
	139	128	48	127	125	116	39	35	64	44	1	65	122	61	66	2	22	55
	133	109	30	49	18	94	94	139	55	132	123	43	58	127	12	43	78	110
4	111	76	142	35	11	59	121	28	95	41	30	17	9	98	24	132	130	
	56	59	60	46	118	7	107	20	16	54	43	27	33	61	50	30	25	
	142	34	10	77	114	86	28	1	100	142	24	134	122	69	138	110	49	
	135	28	129	40	142	53	117	129	6	72	70	73	115	95	14	48	22	
5	104	141	89	110	85	128	54	54	62	99	125	15						
	22	125	82	66	113	49	19	124	78	110	40	131						
	46	7	15	100	17	63	11	41	40	9	13	135						
	134	61	106	62	62	3	132	22	36	8	87	85						
	15	121	140	29	66	40	43	56	139	107	33	78						

contd. ...

Table - A-2.3 (contd.)

Sample Size	Sample Number																	
	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54
2	112	133	126	111	7	89	121	56	71	68	27	41	98	105	81	81	102	16
	60	126	60	73	112	139	104	108	46	98	120	140	133	134	94	59	12	55
3	49	92	36	11	67	135	24	38	103	134	9	59	23	72				
	109	90	107	63	115	137	72	104	117	94	7	140	5	48				
	123	88	17	133	10	28	76	107	137	46	22	130	122	63				

contd...

Sample Size	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70
2	139	65	46	38	85	73	93	12	94	65	8	106	130	25	67	60
	28	102	102	8	94	124	49	85	100	49	11	93	57	15	20	109

APPENDIX A-2 (contd.)

POPULATION II

TABLE - A-2.4

Independent simple random samples without replacement (Sample Units)

Sample Size	Sample Number 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2	7 14	11 17	2 9	19 3	7 4	5 4	4 2	12 5	6 2	3 6	14 8	2 9	20 2	13 5	8 10
3	14 1 18	18 2 7	16 14 9	11 3 16	5 15 3	6 18 11	11 18 4	12 13 11	14 20 16	13 3 7	3 2 15	3 14 10	5 1 16	6 18 19	7 20 16

contd. ..

Sample Size	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
2	7 20	12 7	10 19	5 17	10 18	10 20	3 18	7 19	16 6	17 20	11 5	15 9	16 12	1 20	15 6.
3	17 5 7	12 20 19	4 17 16	19 11	12 16 5										

APPENDIX A-2 (contd.)

POPULATION 7

TABLE - A-2.5

Independent simple random samples without replacement (Sample Units)

Sample Size	Sample Number 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2	42	16	26	43	40	33	43	44	28	29	15	43	1	15	22
	36	24	32	18	18	30	7	7	11	15	36	19	32	33	9
3	12	6	3	2	3	24	44	1	25	38	10	23	27	9	3
	20	19	41	28	16	44	19	5	41	45	13	38	16	41	5
	44	23	33	9	22	26	27	44	39	36	23	33	44	45	13
4	32	43	24	19	45	33	20	10	39	28	20	42	6	30	8
	40	22	28	6	32	22	43	43	24	37	14	26	3	42	1
	15	19	13	27	7	7	40	44	12	33	32	1	44	28	21
	5	12	6	28	4	5	18	14	2	3	21	20	26	17	31
5	20	14	24	7	44	4	10	44	3	23	9	12	14	31	20
	17	36	19	22	43	42	11	36	23	45	6	1	34	37	14
	1	11	33	17	12	6	39	22	7	43	4	2	24	12	33
	19	35	17	15	41	34	21	1	5	25	40	18	22	33	7
	23	21	10	1	23	1	32	12	12	42	23	15	8	5	22

Table - A.-2.5 (contd.)

Sample Size	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
2	14	41	41	36	32	27	4	41	15	17	19	36	33	25	31
	43	17	43	38	29	4	41	25	40	16	34	15	45	41	27
3	33	41	2	31	18	14	44	44	2	3	41	17	21	23	27
	42	31	24	10	4	6	4	36	25	26	45	26	23	34	9
	16	24	29	25	24	9	40	27	9	22	4	44	9	32	36
4	36	42	31	35	42	29	35	14	10	27					
	45	11	41	23	36	32	17	19	37	9					
	22	4	21	28	18	7	12	3	24	17					
	32	44	36	42	14	11	11	23	2	14					
5	9	38	32	28	13										
	19	14	18	30	15										
	25	34	41	5	1										
	39	37	30	42	30										
	35	27	6	43	19										

contd...

Sample Size	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
2	41	11	45	43	37	5	9	34	15	25	3	10	41	28	43
	42	9	14	21	35	17	11	42	10	29	16	4	16	23	26

APPENDIX A-3

POPULATION 1

TABLE - A-3.1

Independent Midzuno-Sen scheme samples (Units)

Sample Size	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2	14	8	22	42	41	17	38	11	34	14	6	21	13	29	40	32	18	17	14	11
	1	10	5	16	33	36	32	12	29	8	25	11	3	16	9	21	31	23	13	38
3	8	1	14	16	33	14	27	33	12	34	9	30	6	40	37	6	6	38	27	11
	42	41	29	26	8	9	1	35	25	6	15	2	18	20	33	16	15	40	14	34
	33	14	9	42	13	26	40	43	13	11	16	41	14	10	12	26	30	10	30	19
4	7	16	22	26	39	42	32	29	36	19	36	11	23	25	14	13	42	39	19	4
	20	10	24	22	15	19	15	32	2	29	35	6	10	34	34	41	6	30	41	19
	12	9	40	6	9	39	22	28	34	21	17	34	8	42	30	20	20	36	6	26
	4	41	38	10	10	41	3	3	39	40	29	36	39	35	18	8	21	41	27	36
5	14	7	12	13	8	8	34	25	39		9	40	21	7	19	39	17	39	24	11
	21	24	8	11	21	40	31	22	40	10	35	36	37	32	9	14	23	8	31	42
	1	37	3	43	16	36	33	27	17	9	38	33	19	27	6	2	28	11	5	34
	38	36	22	37	20	16	40	14	41	33	36	30	42	25	32	42	26	5	6	18
	25	29	28	42	15	19	24	7	19	30	20	43	23	38	14	30	36	9	25	30

contd...

Table - A-3.1 (contd.)

Sample Size	Sample Number 21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
2	37 12	38 10	36 8	8 39	9 28	19 42	43 28	19 27	38 5	13 6	34 22	12 13	17 28	13 34	34 9
3	38 30 25	9 18 23	10 12 19	35 17 2	26 15 5	34 3 6	34 37 43	38 34 10	19 12 24	10 11 32	32 20 42	38 24 9	6 25 5	1 24 18	14 8 39
4	39 26 29 14	31 39 7 24	37 28 3 43	34 10 5 12	11 3 23 38										

contd...

Sample Size	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
2	2 15	1 4	33 17	17 12	9 30	14 19	20 43	21 26	17 38	26 37	23 13	4 41	20 27	8 34	6 14

APPENDIX A-3 (contd.)

POPULATION 2

TABLE - A-3.2

Independent Midzuno-Sen scheme Samples (Units)

Sample Size	Sample Number 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2	19	4	7	1	2	4	18	19	7	3	5	4	8	1	1
	3	15	6	14	22	19	15	3	11	20	9	3	20	13	19
3	5	1	4	7	1	13	3	5	14	10	8	12	6	19	2
	4	5	22	13	16	6	5	13	20	2	17	6	20	13	6
	10	18	10	2	22	4	15	9	17	21	21	18	18	15	20
4	1	13	21	4	11	3	9	2	4	21	2	1	11	3	10
	3	3	19	12	21	8	7	22	6	9	16	5	5	11	6
	22	9	17	5	19	19	19	11	17	11	15	12	18	15	18
	18	20	7	1	10	15	14	14	15	14	19	18	21	14	3
5	12	1	13	2	4	2	1	8	2	1	8	5	1	4	7
	3	15	17	4	17	22	14	11	19	3	11	17	17	9	3
	1	9	21	16	2	19	20	16	5	7	16	21	19	10	19
	15	4	16	20	6	5	21	5	21	22	5	8	6	18	2
	22	20	2	19	10	16	10	1	17	4	20	16	8	11	9

Table - A-3.2 (contd.)

Sample Size	Sample Number														
	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
2	7 15	1 17	3 11	4 21	9 4	2 10	14 9	20 12	3 6	1 3	14 20	5 17	5 8	21 10	1 14
3	2 12 11	1 20 3	8 16 7	7 18 8	9 22 17	12 3 6	12 6 14	5 3 22	6 13 4	15 1 11					
4	4 8 16 1	20 16 21 3	3 1 13 11	22 14 8 4	9 8 13 6										

contd...

Sample Size	Sample Number									
	31	32	33	34	35	36	37	38	39	40
2	8 14	1 7	4 15	1 6	8 16	6 10	5 9	7 16	6 15	1 7

APPENDIX A-3 (contd.)

POPULATION 3

TABLE - A-3.3

Independent Midzuno-Sen scheme samples (Units)

Sample Size	Sample Number														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2	7	1	8	15	1	3	1	13	1	5	1	15	1	5	3
	9	5	13	2	6	18	2	18	22	19	15	22	13	14	10
3	7	1	1	1	3	3	22	1	3	1	5	7	11	13	3
	18	17	7	9	19	21	18	11	15	9	1	12	20	1	13
	16	16	10	4	13	4	6	15	5	4	15	4	4	20	6
4	8	1	9	11	5	17	3	1	4	2	6	8	4	20	5
	2	17	12	9	3	22	15	10	13	9	8	7	6	3	1
	11	13	18	4	17	10	22	14	11	19	17	3	7	16	14
	5	16	13	12	10	12	2	18	9	4	22	6	15	13	19
5	3	1	10	14	10	1	1	2	9	10	5	9	19	4	5
	10	18	20	6	13	9	10	13	10	1	22	14	13	3	21
	20	17	14	13	22	15	6	14	4	20	10	6	18	1	19
	22	20	9	18	8	13	22	19	3	15	4	16	15	21	15
	6	16	21	15	21	10	9	11	5	2	18	12	16	2	3

Table – A-3.3 (contd.)

Sample Size	Sample Number 16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
2	1 22	3 19	5 20	1 22	15 3	8 20	2 16	4 20	19 10	8 18	6 1	5 12	13 20	16 1	7 6
3	4 18 21	6 16 10	5 6 3	3 16 22	1 19 18	3 21 16	19 2 22	6 18 3	5 10 20	12 20 18					
4	6 19 20 18	1 9 16 14	17 8 1 9	10 16 4 17	2 15 18 13										

contd. . .

Sample Size	31	32	33	34	35	36	37	38	39	40
2	4 14	6 18	3 6	2 15	4 9	5 18	2 15	3 21	4 14	1 12

APPENDIX A-3 (contd.)

POPULATIONS 4 and 5

TABLE - A-3.4

Independent Midzuno-Sen scheme samples (Units)

Sample Size	Sample Number 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2	5	12	6	6	16	18	5	7	5	10	3	5	6	15	1
	2	1	4	20	12	5	12	17	7	16	12	11	9	20	12
3	2	1	8	9	3	8	10	9	6	4	7	16	15	15	1
	11	18	2	16	17	22	9	13	12	6	13	2	6	17	15
	15	3	13	3	1	20	19	8	2	9	14	21	2	7	6
4	17	5	5	1	16	1	18	10	2	18	1	15	21	22	4
	9	11	8	21	7	22	13	19	11	22	19	6	18	10	3
	22	6	13	14	19	21	4	22	12	3	8	20	11	16	22
	8	15	21	5	17	18	15	11	15	19	13	16	10	5	18
5	4	1	15	1	1	13	11	1	4	3	17	9	1	10	14
	2	11	16	3	11	3	16	14	3	1	18	14	13	5	19
	13	14	13	8	12	14	1	20	17	15	16	21	8	17	20
	3	9	18	11	9	17	10	9	21	2	4	15	16	13	4
	11	18	1	20	2	4	15	16	2	19	19	5	2	7	21

contd...

Table - A-3.4 (contd.)

Sample Size	Sample Number 16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
2	12	20	8	5	4	19	13	5	10	3	1	1	14	19	1
	1	15	5	18	9	2	8	14	2	19	5	18	8	3	6
3	1	8	7	9	16	6	8	10	1	11					
	15	16	14	6	8	7	12	8	12	9					
	5	17	8	18	6	13	19	5	20	18					
4	16	2	6	4	10										
	2	8	14	22	3										
	22	15	20	14	6										
	3	5	7	20	14										

contd...

Sample Size	31	32	33	34	35	36	37	38	39	40
2	5	1	14	5	1	2	1	14	17	6
	21	19	16	16	21	1	13	17	10	15

APPENDIX A -3 (contd.)

POPULATION 6

TABLE - A-3.5

Independent Midzuno-Sen Scheme Samples (Units)

Sample Size	Sample Number 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
2	5	53	104	53	27	135	53	5	135	109	40	96	135	10	70	117	141	135
	29	87	26	33	14	44	11	46	10	40	79	52	11	100	21	104	142	142
3	96	89	108	118	1	26	135	26	39	135	72	99	17	104	127	104	1	135
	141	113	142	102	92	131	21	24	134	72	9	7	58	129	135	51	16	2
	21	126	112	20	123	58	26	120	83	127	84	79	130	6	96	61	140	47
4	135	43	42	123	26	1	104	112	91	39	94	53	42	1	88	58	113	53
	99	9	35	91	93	107	68	90	67	55	45	14	61	25	84	103	10	45
	81	40	128	72	88	4	102	119	70	64	54	133	131	58	41	117	112	19
	59	25	129	27	15	113	33	22	131	5	25	47	45	67	141	10	127	88
5	35	29	135	96	119	141	141	74	135	135	97	113	142	53	89	135	135	66
	28	1	100	59	4	118	15	117	116	72	58	127	128	112	40	32	118	84
	36*	91	110	91	40	18	119	89	55	11	100	21	8	135	62	50	53	124
	32	101	53	140	23	27	91	100	123	60	48	48	133	28	21	6	115	25
	127	117	8	124	83	43	4	4	112	128	46	109	86	78	114	66	18	40

contd...

Table - A-3.5 (contd.)

Sample Size	Sample Number 19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
2	135	1	15	141	114	135	134	53	135	97	1	15	63	1	135	135	73	27
	12	15	38	59	36	52	40	43	10	80	87	95	8	9	21	36	15	7
3	87	53	16	141	87	98	50	81	70	141	1	70	26	43	42	96	131	96
	73	78	105	94	38	73	9	34	131	12	56	1	50	69	73	1	62	105
	44	88	60	87	127	109	53	111	134	128	103	117	65	8	102	104	127	8
4	40	72	8	19	26	15	104	135	53	57	114	98	86	102	54	26	42	
	35	138	93	3	97	99	36	140	60	106	11	103	122	20	113	15	81	
	21	91	57	51	10	109	107	54	21	122	132	8	76	10	137	19	75	
	137	84	94	13	64	39	38	4	99	66	12	85	77	37	133	132	141	
5	45	46	117	15	96	135	54	90	75	53	115	53						
	36	123	4	39	53	20	102	38	70	13	100	115						
	47	135	41	47	65	69	122	49	42	74	139	87						
	65	17	13	118	23	46	98	109	58	45	14	33						
	60	39	61	28	16	56	29	119	40	134	73	56						

contd...

Table - A-3.5 (contd.)

Sample Size	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54
2	53	53	96	26	135	113	137	116	12	141	44	27	90	135	137	135	135	16
	112	104	69	42	137	42	75	124	125	121	94	25	94	84	122	57	16	131
3	96	142	141	55	53	97	141	128	38	113	74	135	135	10				
	2	34	131	105	20	56	93	15	54	32	116	73	80	35				
	51	117	121	47	120	7	33	34	33	106	72	105	49	104				

contd...

Sample Size	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70
2	55	45	127	85	101	53	16	42	72	135	56	3	44	38	135	108
	77	29	5	97	34	25	85	105	25	123	52	81	17	10	79	44

APPENDIX A-3 (contd.)

POPULATION 8

TABLE - A -3.6

Independent Midzuno-Sen Scheme Samples (Units)

Sample Size	Sample Number 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
2	130	35	18	53	96	100	94	42	83	27	70	37	13	53	30	54	91	58
	28	87	27	33	14	44	11	46	10	41	79	53	11	100	21	105	142	142
3	135	96	114	82	135	135	53	115	135	127	96	26	135	83	25	95	100	9
	141	113	142	103	91	130	21	24	133	72	9	7	57	129	135	51	15	2
	21	126	111	20	122	57	26	120	82	128	83	80	129	6	97	61	140	48
4	3	119	87	56	135	135	74	135	70	100	26	37	17	27	135	31	141	141
	100	9	35	92	92	106	68	90	67	54	46	14	61	24	84	103	10	45
	82	40	128	73	87	3	103	118	71	63	55	133	131	58	41	117	112	19
	60	25	129	27	15	112	33	22	131	5	25	48	45	67	141	10	126	87
5	1	96	60	96	72	118	96	53	133	135	1	15	141	73	55	126	44	71
	29	1	101	59	4	119	15	117	116	72	59	127	128	112	40	32	119	64
	36	90	111	91	40	18	120	56	55	11	100	22	8	135	63	50	54	124
	33	101	53	140	23	27	91	100	123	60	49	49	133	28	21	6	116	25
	127	117	8	124	84	43	4	4	112	128	47	110	86	78	114	66	18	40

Table - A-3.6 (contd.)

Sample Size	Sample Number 19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
2	135	45	112	42	36	27	114	76	118	79	134	71	19	24	121	3	115	53
	12	14	37	60	37	53	40	43	10	81	86	95	8	8	21	37	16	7
3	57	135	113	53	2	135	113	135	66	121	60	102	101	85	28	8	55	135
	74	77	104	95	39	73	9	34	131	12	55	1	49	68	73	1	63	104
	44	87	59	88	127	108	52	110	134	129	103	117	64	8	102	104	128	8
4	90	17	21	68	119	96	141	113	53	121	86	62	45	31	17	71	91	
	35	138	93	3	96	99	36	140	60	105	11	103	122	20	113	15	80	
	21	91	57	50	10	109	106	54	21	122	132	8	77	10	137	19	74	
	137	84	94	13	63	38	38	4	99	65	12	86	78	38	133	132	141	
5	115	98	114	135	21	132	53	113	114	53	31	141						
	36	123	4	38	54	20	102	38	70	13	101	114						
	46	135	41	46	66	69	122	49	42	74	139	86						
	64	17	13	117	24	46	98	108	58	45	14	33						
	59	39	61	27	16	56*	29	119	40	134	74	55						

contd...

Table - A - 3.6 (contd.)

Sample Size	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54
2	47	135	53	37	96	135	41	53	70	53	135	85	114	137	33	120	48	86
	112	103	70	42	137	42	76	124	125	122	93	25	93	84	123	57	15	131
3	121	75	37	1	108	135	119	53	53	53	52	135	48	136				
	2	34	132	105	20	56	93	15	54	32	116	73	81	34				
	51	118	122	48	120	7	33	34	33	107	73	105	50	103				

contd...

Sample Size	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70
2	96	53	87	85	66	53	141	135	57	122	55	12	92	74	9	5
	76	29	5	97	34	25	84	104	25	124	52	81	17	10	80	45

APPENDIX A-3 (contd.)

POPULATIONS 9 AND 10

TABLE - A-3.7

Independent Midzuno-Sen Scheme Samples (Units)

Sample Size	Sample Number 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
2	55	39	99	72	53	61	27	96	54	135	139	86	70	57	53	42	45	15
	28	87	26	33	14	44	11	45	10	40	78	52	11	100	21	105	142	142
3	107	57	30	135	83	43	114	97	124	70	53	53	104	19	72	50	91	72
	141	113	142	102	92	131	21	24	134	73	9	7	57	129	135	52	15	2
	21	126	112	20	123	58	26	120	82	128	84	80	136	6	97	62	140	47
4	113	6	93	116	70	53	3	53	4	96	53	36	59	114	91	135	70	136
	99	10	35	91	93	107	69	91	68	54	45	14	61	24	84	102	10	45
	81	41	128	72	88	3	103	119	71	63	55	133	131	57	41	116	113	19
	59	26	129	27	15	113	34	22	131	5	25	48	44	66	141	10	127	87
5	26	112	45	119	29	18	9	23	8	141	141	135	30	35	46	42	1	22
	29	1	101	59	4	119	16	117	117	72	58	126	129	112	40	32	119	65
	36	90	111	91	41	19	120	89	56	11	99	21	8	135	63	51	54	124
	33	100	54	140	23	28	92	100	124	60	48	48	134	28	21	6	116	26
	127	117	8	124	84	44	4	4	113	128	46	109	87	78	114	67	19	41

Table - A-3.7 (contd.)

Sample Size	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
2	113	91	129	59	70	96	96	55	93	119	141	115	136	135	27	26	22	135
	12	14	37	60	36	52	40	43	10	80	86	94	8	8	21	37	16	7
3	135	26	34	94	93	44	114	96	1	141	54	16	135	61	1	139	135	86
	73	78	105	95	38	74	9	34	131	12	56	1	49	69	73	1	62	105
	44	88	60	87	127	109	52	111	134	128	103	117	64	8	102	103	127	8
4	131	51	2	70	96	33	99	112	116	35	29	141	59	53	43	81	96	
	35	138	93	3	97	99	36	140	59	106	11	102	122	20	113	15	80	
	21	91	57	50	10	109	107	54	21	122	132	8	77	10	137	19	74	
	137	84	94	13	63	39	38	4	98	66	12	85	78	37	133	132	141	
5	112	44	53	53	15	26	74	53	42	109	21	9						
	36	123	4	38	54	20	102	38	71	13	101	115						
	46	135	41	46	66	70	122	49	43	73	139	87						
	64	17	13	118	24	47	98	109	59	45	14	34						
	59	39	62	27	17	57	29	119	40	134	74	56						

contd....

Table - A-3.7 (contd.)

Sample Size	Sample Number																	
	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54
2	53	45	26	135	135	11	141	141	53	26	135	61	1	135	100	133	95	93
	112	104	70	41	137	43	75	123	125	122	93	25	94	84	123	57	16	131
3	52	141	60	1	141	53	124	123	53	95	41	116	53	113				
	2	34	132	105	20	57	93	15	54	32	116	73	81	34				
	51	117	122	48	119	7	33	34	33	107	73	105	49	103				

contd...

	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70
2	16	135	53	64	38	53	54	39	43	141	116	89	54	37	2	141
	77	29	5	97	34	25	85	105	25	123	52	80	17	10	80	44

APPENDIX - A-3 (contd.)

POPULATION - 11

TABLE - A-3.8

Independent Midzuno - Sen Scheme Samples (Units)

Sample Size	Sample Number														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2	5	8	13	13	3	13	6	17	12	5	15	8	12	13	7
	12	18	17	5	15	13	5	19	9	19	2	18	19	5	15
3	1	4	18	15	14	4	9	3	5	18	18	13	8	7	15
	10	6	17	5	1	9	12	11	16	2	4	16	3	8	3
	15	11	10	20	18	5	3	14	3	5	20	18	2	6	20

contd. . .

Sample Size	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
2	5	6	11	5	13	2	12	17	20	6	10	13	2	13	11
	7	15	9	19	10	16	3	18	12	20	8	14	16	2	3
3	18	4	2	10	8										
	8	5	17	15	2										
	3	19	8	5	19										

CONCLUDING REMARKS

We now indicate further lines of research arising out of our work.

For Mickey's general unbiased estimator G_m^*, we have derived the exact variance only for the choice $m = 1$ (c.f. Ch. III). The problem remains unsolved for other choices of m in the range 1 to n - 1. A meaningful expression for the exact variance of G_m^* valid for all choices of m, if available, may have interesting applications. For example, it may be useful for determining optimum choices of the coefficient functions $a_i (Z_m)$. Using the result available on the exact variance of G_1^*, especially when $N \gg n$, one may determine optimum choices of $a_i (Z_1)$. It may be noted that we have solved the problem of determining optimum choices of $a_i (Z_m)$ for G_m^* in large samples (c.f. Ch. VII).

The unbiased estimator of variance developed for G_{n-1}^* (c.f. Ch. III) has not performed satisfactorily, as evident from our computations on unbiased ratio, product and regression type estimators in Chapters - IV, VI and VII respectively. At the same time we have shown that the choice $m = n - 1$ is important for the unbiased ratio type estimator T_m^* and the unbiased regression type estimator D_m^* (c.f. Chs. IV, V, VII and VIII). Consequently, the problem of evolving a suitable unbiased estimator of variance for G_{n-1}^* assumes significance.

Under super-population model Δ we have obtained a comparison of the unbiased ratio type estimator T_m^* and the classical biased ratio estimator $\hat{\bar{Y}}_R$ (c.f. Ch. V). One may examine the possibility of including Quenouille's (1956) ratio estimator also in this comparison. Under super-population model $\triangle'$ we have derived the expected exact variance of T^*_m (c.f. Ch. V), which is valid for any sample size n, but attempted a comparison of the same with the expected exact mean square error of $\overline{\hat{Y}}_R$ only for h = 3, n = 2 (m = 1), n = 3 (m = 1,2) and n = 4 (m = 1,2,3), because of limited computational facilities. It is worthwhile to extend this comparison for a range of values of h, n and m using high speed electronic computers.

An indication of the relative efficiency of the unbiased product type estimators $S^*_{(q_0' < o)}$, H^*_m and L^*_m is available only from our empirical studies (c.f. Chs. VI and VIII). Theoretical studies on this problem, at least under a suitable super-population model, are of interest. We have derived exact variances for the unbiased ratio-cum-product type estimators $S^*_{(\pm 1)}$ and TH^*_1 (when N >> n) and large sample variances for the estimators TH^*_m and TL^*_m ; and, making use of the variances, determined the optimum choices for the weight vector W (c.f. Ch. VI). With the optimum choices for W, the relative efficiencies of (i) $S^*_{(\pm 1)}$ and TH^*_1 and (ii) TH^*_m and TL^*_m may be investigated.

Under two super-population models $\triangle_1$ and $\triangle_2$ involving quadratic regression of y on x, we have shown that the unbiased regression type estimator D^*_1 is less efficient than the classical biased linear regression estimator $\hat{\overline{Y}}_r$ (c.f. Ch. VII). It is of interest to know how the unbiased regression type estimator D^*_m ($2 \leq m \leq n-1$) compares with $\hat{\overline{Y}}_r$ under the same super-population models. Our empirical studies <u>in small samples</u> (c.f. Chs. VII and VIII) have indicated that the regression estimators $\hat{\bar{Y}}_r$ and D_m^* ($2 \leq m \leq n-1$) are less efficient than the ratio estimators $\hat{\bar{Y}}_R$ and T_m^* It is, therefore, worthwhile to examine the possibility of making a theoretical comparison of these estimators in small samples under an appropriate super-population set-up.

In small samples, Lahiri's (1951) unbiased ratio estimator has performed better than Mickey's unbiased ratio type estimator T_m^* ($1 \leq m \leq n-1$) in our empirical study (c.f. Ch. VIII). A theoretical comparison of these ratio estimators in small samples under an appropriate super-population model may, therefore, be attempted.

Our empirical study in small samples on (i) ratio and regression estimators and (ii) product and regression estimators has been conducted on the basis of samples selected from unstratified populations (c.f. Ch. VIII). Basic data of these populations have been given in the thesis. The same populations may be stratified and an empirical study conducted on 'seperate estimators' computed on the basis of stratified samples.

BIBLIOGRAPHY

BASU, D. (1958): On sampling with and without replacement, Sankhya, 20, 287 - 294

BEALE, E.M.L. (1962): Some uses of computers in operational research, Industrielle Organisation, 31, 51 - 52.

BOHRNSTEDT, G.W. and GOLDBERGER, A.S. (1969): On the exact covariance of products of random variables, Jour. Amer. Stat. Assn., 64, 1439 - 1442.

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 - 177.

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. (1962): Recent advances in sample survey theory and methods, Ann. Math. Stat., 33, 325 - 349.

DES RAJ (1954): Ratio estimation in sampling with equal and unequal probabilities, 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 and KHAMIS, S.H. (1958): Some remarks on sampling with replacement, Ann. Math. Stat., 29, 550 - 557.

DURBIN, J. (1953): Some results in sampling theory when the units are selected with unequal probabilities, Jour. Roy. Stat. Soc., Ser. B, 15, 262 - 269.

DURBIN, J. (1959): A note on the application of Quenouille's method of bias reduction to the estimation of ratios., Biometrika, 46, 477 - 480.

FELLIGI, I.M. (1963): Sampling with varying probabilities without replacement: Rotating and non-rotating samples., Jour. Amer. Stat. Assn., 58, 183 - 201.

GODAMBE, V.P. (1955): A unified theory of sampling from finite populations, Jour. Roy. Stat. Soc., Ser. 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, R. and KISH, L. (1950): Controlled selection - a technique in probability sampling, Jour. Amer. Stat. Assn., 45, 350-372.

GOODMAN, L.A. and HARTLEY, H.O. (1958): The precision of unbiased ratio type estimators, Jour. Amer. Stat. Assn., 53, 491-508.

GOODMAN, L.A. (1960): On the exact variance of products, Jour. Amer. Stat. Assn., 55, 708-713.

GRUNDY, P.M. (1954): A method of sampling with probability exactly proportional to size, Jour. Roy. Stat. Soc., Ser. B, 16, 236-238.

HAJEK, J. (1958): Some contributions to the theory of probability sampling Bull. Int. Stat. Inst., 36, 127-134.

HÁJEK, J. (1959): Optimum strategy and other problems in probability sampling, Casipis pest. Math., 84, 387-428.

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 Survey Methods and Theory, Wiley, New York.

HANURAV, T.V. (1962): Some sampling schemes in probability sampling Sankhya, Ser. A, 24, 421-428.

HANURAV, T.V. (1962 a): An existence theorem in sampling, Sankhya, Ser, A, 24, 227-230.

HANURAV, T.V. (1965): Optimum sampling strategies and some related problems, Ph.D. Thesis submitted to the Indian Statistical Institute.

HANURAV, T.V. (1967): Optimum utilization of auxiliary information - πPS sampling of two units from a stratum, Jour. Roy. Stat. Soc., Ser. B, 29, 374-391.

HARTLEY, H.O. and ROSS, A. (1954): Unbiased ratio estimators, Nature, 174, 270-271.

HARTLEY, H.O. and RAO, J.N.K. (1962): Sampling with unequal probabilities and without replacement, Ann. Math. Stat., 33, 350-374.

HORVITZ, D.G. and THOMPSON, D.J. (1952): A generalization [illegible] without replacement from a finite universe, Jour. Am[illegible] 47, 663-685.

HUTCHISON, M.C. (1971): A Monte Carlo comparison of some ratio estimators, Biometrika, 58, 313-321.

JOHNSON, N.L. (1950): On the comparison of estimates, Biometrika, 37, 281-287.

JONES, H.L. (1963): The jack-knife method, Proc. IBM Scientific Computing Symposium on Statistics, IBM.

JOSHI, V.M. (1965a): Admissibility and Bayes estimation in sampling finite populations - II, Ann. Math. Stat., 36, 1723-1729.

JOSHI, V.M. (1965 b): 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, 141-146.

LAHIRI, D.B. (1951): A method of sample selection providing unbiased ratio estimates, Bull. Int. Stat. Inst., 33 (2), 133-140.

MADOW, W.G. and MADOW, L.H. (1944): On the theory of systematic sampling-1, Ann. Math. Stat., 15, 1-24.

MAHALANOBIS, P.C. (1940): A sample survey of the acreage under jute in Bengal, Sankhya, 4, 511-530.

Md. ANWAR HOSSAIN TALUKDAR (1968): Multivariate regression estimates of finite populations, Ann. Inst. Stat. Math., 20, 441-455.

MICKEY, M.R. (1959): Some finite population unbiased ratio and regression estimators, Jour. Amer. Stat. Assn., 54, 594-612.

MIDZUNO, H. (1952): On the sampling system with probability proportional to the sum of sizes, Ann. Inst. Stat. Math., 3, 99-108.

MURTHY, M.N. (1957): Ordered and unordered estimators in sampling without replacement, Sankhya, 18, 378-390.

MURTHY, M.N. and NANJAMMA, N.S. (1959): Almost unbiased ratio estimates based on interpenetrating sub-sample estimates, Sankhya, 21, 381-392.

MURTHY, M.N. (1962): Almost unbiased estimators based on interpenetrating sub-samples, Sankhya, Ser. A, 24, 303-314.

MURTHY, M.N. (1963): Some recent advances in sampling theory, Jour. Amer. Stat. Assn., 58, 737-755.

MURTHY, M.N. (1964): Product method of estimation, Sankhya, Ser. A, 26, 69-74.

MURTHY, M.N. (1967): Sampling Theory and Methods, Statistical Publishing Society, Calcutta.

NAGI REDDY, V. (1972): On ratio and product methods of estimation, To appear in Sankhya, Ser. B.

NANJAMMA, N.S., MURTHY, M.N. and SETHI, V.K. (1959): Some sampling systems providing unbiased ratio estimators, Sankhya, 21, 299-314.

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.

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.

QUENOUILLE, M.H. (1956): Notes on bias in estimation, Biometrika, 43, 353-360.

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., Ser. B., 24, 482-491.

RAO, J.N.K. (1964): Unbiased ratio and regression estimators in multi-stage sampling, Jour. Ind. Soc. Agr. Stat., 14, 175-188.

RAO, J.N.K. (1965): A note on estimation of ratios by Quenouille's method, Biometrika, 52, 647-649.

RAO, J.N.K. and WEBSTER, J.T. (1966): On two methods of bias reduction in the estimation of ratios, Biometrika, 53, 571-577.

RAO, J.N.K. and BEEGLE, L.D. (1967): A Monte Carlo study of some estimators, Sankhya, Ser. B, 29, 47-56.

RAO, J.N.K. (1967): Precision of Mickey's unbiased ratio estimator, Biometrika, 54, 321 - 324.

RAO, J.N.K. and PEREIRA, N.P. (1968): Unbiased double ratio estimators, Sankhya, Ser. A, 30, 83 - 90.

RAO, J.N.K. (1969): Ratio and regression estimators, New Developments in Survey Sampling, Wiley, New York, 213 - 234.

RAO, J.N.K. and KUZIK, R.A. (1973): Sampling errors in ratio estimation, To appear in Sankhya, Ser. A.

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. (1968): On three procedures of sampling from finite populations, Biometrika, 55, 438 - 441.

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 RAO, J.N.K. (1971 a): On the jack-knife estimator in finite populations, Proc. of 38-th session of Int. Stat. Inst., 325-329.

RAO, P.S.R.S. and RAO, J.N.K. (1971b): Small sample results for ratio estimators, Biometrika, 58, 625-630.

RAO, T.J. (1966): On certain unbiased ratio estimators, Ann. Inst. Stat. Math., 18, 117 - 121.

RAO, T.J. (1966 a): On the variance of the ratio estimator for Midzuno-Sen sampling scheme, Metrika, 10, 89 - 91.

RAO, T.J. (1967): Contributions to the theory of sampling strategies, Ph.D. Thesis submitted to the Indian Statistical Institute.

RAO, T.J. (1967a): On the choice of a strategy for ratio method of estimation Jour. Roy. Stat. Soc., Ser. B, 29, 392 -397.

ROBSON, D.S. (1957): Application of multivariate polykays to the theory of unbiased ratio type estimators, Jour. Amer. Stat. Assn., 52, 511 - 522.

ROBSON, D.S. and VITHAYASAI, C. (1961): Unbiased component-wise ratio estimation, Jour. Amer. Stat. Assn., 56, 350 -358.

ROY, J. and CHAKRAVARTY, I.M. (1960): Estimating the mean of a finite population, Ann. Math. Stat., 31, 392-398.

SAMPFORD, M.R. (1969): A comparison of some possible methods of sampling from smallish populations, with units of unequal size, New Developments in Survey Sampling, Wiley, New York, 170-187.

SASTRY, N.S. (1964): Unbiased ratio and regression type estimators, Diploma Dissertation submitted to the Institute of Agricultural Research Statistics (Indian Council of Agricultural Research).

SEN, A.R. (1952): Present status of probability sampling and its use in estimation of farm characteristics (abstract), Econometrica, 20, 103.

SENG, YON POH (1951): Historical survey of the development of sampling theories and practice, Jour. Roy. Stat. Soc., Ser. A, 114, 214-231.

SINGH, M.P. (1965): On the estimation of ratio and product of the population parameters, Ser. B, Sankhya, 27, 321-328.

SINGH, M.P. (1967): Multivariate product method of estimation for finite populations, Jour. Ind. Soc. Agr. Stat., 19, 1-10.

SINGH, M.P. (1967 a): Ratio-cum-product method of estimation, Metrika, 12, 34-42.

SINGH, M.P. (1969): Some aspects of estimation in sampling from finite populations, Ph.D. Thesis submitted to the Indian Statistical Institute.

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.

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, P.V. (1944): Moments and product moments of moment statistics for samples of the finite and infinite populations, Sankhya, 6, 363-382.

SUKHATME, P. V. (1947): The problem of plot size in large scale yield surveys, Jour. Amer. Stat. Assn., 42, 297-310.

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.

SUKHATME, P. V. and SUKHATME, B. V. (1970): Sampling Theory of Surveys with Applications, 2nd Revised Ed. Asia Publishing House, Bombay.

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.

TIN, M. (1965): Comparison of some ratio estimators, Jour. Amer. Stat. Assn., 60, 294-307.

VIJAYAN, K. (1967): On an exact πPS sampling (abstract), Sankhya, Ser. A, 29, 113.

WILLIAMS, W. H. (1961): Generating unbiased ratio and regression estimators, Biometrics, 17, 267-274.

WILLIAMS, W. H. (1962): On two methods of unbiased estimation with auxiliary variates, Jour. Amer. Stat. Assn., 57, 184-186.

WILLIAMS, W. H. (1963): The precision of some unbiased regression estimators, Biometrics, 19, 352-361.

YATES, F. (1946): A review of recent statistical developments in sampling and sampling surveys, Jour. Roy. Stat. Soc., 109, 12-43.

YATES, F. and GRUNDY, P. M. (1953): Selection without replacement from within strata with probability proportional to size, Jour. Roy. Stat. Soc, Ser. B, 15, 253-261.