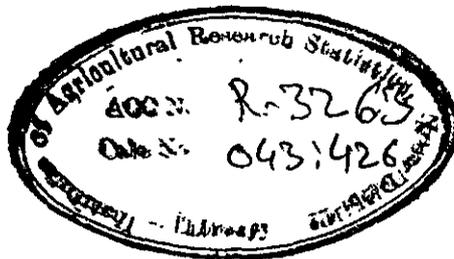


84

195

**SOME GENERALISED RATIO - TYPE
ESTIMATORS**

KRISHNA KUMAR SINGH BHATT



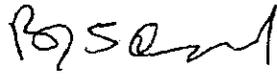
**Dissertation submitted for the award of Post-Graduate
Diploma in Agricultural Statistics of Institute
of Agricultural Research Statistics
(I.C.A.R.)
New Delhi - 110012**

**INSTITUTE OF AGRICULTURAL RESEARCH STATISTICS
(I.C.A.R.)
LIBRARY AVENUE, NEW DELHI - 12**

1977

C E R T I F I C A T E

THIS is to certify that the work incorporated in the dissertation entitled " SOME GENERALISED RATIO TYPE ESTIMATORS " by KRISHNA KUMAR SINGH BHATT and submitted for the award of Post-Graduate Diploma in Agricultural Statistics of the Institute of Agricultural Research Statistics, New Delhi was done under my guidance.



(B. B. P. S. GOEL)
Associate Professor of Statistics
Institute of Agricultural Research Statistics
(I. G. A. R.)
Library Avenue, New Delhi -12

ACKNOWLEDGEMENTS

I have great pleasure in expressing my deep sense of gratitude to Dr. B. B. P. S. Goyal, Associate Professor, Institute of Agricultural Research Statistics (I.C.A.R.), Supervisor for my Diploma Course, for his inspired guidance, keen interest and constant encouragement throughout the period of investigation and immense help in writing - up this thesis.

It is my earnest and pleasant duty to place on record my gratefulness to Dr. Daroga Singh, Director, Institute of Agricultural Research Statistics, who provided the necessary facilities for the work.

I am thankful to all of my friends who extended every kind of help.

My acknowledgement will not be complete unless I mention the Indian Council of Agricultural Research for extending the financial assistance in the form of fellowship; Mr. Gaya Prasad Pal (I.A.R.S.) who typed this dissertation intine inspite of several problems.

K. K. Singh Bisht
(KRISHNA KUMAR SINGH BISHT)

CONTENTS

CHAPTER		PAGE
	NOTATIONS	1 - 11
I	INTRODUCTION, REVIEW OF LITERATURE AND ORIENTATION OF PROBLEM	1 - 8
II	THE FIRST ESTIMATOR	9 - 53
	SINGLE PHASE SAMPLING	9
	TWO PHASE SAMPLING	22
	SAMPLING WITH VARYING PROBABILITIES	32
	CASE OF NEGATIVELY CORRELATED CHARACTERS	46
	ANNEXURE	51
III	THE SECOND ESTIMATOR	54 - 62
	SINGLE PHASE SAMPLING	54
	TWO PHASE SAMPLING	61
IV	EMPIRICAL INVESTIGATION ON THE PROPOSED ESTIMATOR	63 - 74
	SUMMARY AND CONCLUSION	79 - 81
	REFERENCES	

- 1 -

NOTATIONS

Suppose that to each unit in a bivariate population of size N there correspond two characters, namely, main character or character under study designated by y and auxiliary character designated by x . Suppose a sample of size n ($n < N$) drawn in order to estimate the mean of main character in the population. Let us make the following denotations -

$$\bar{y} - \text{sample mean of main character} = \frac{1}{n} \sum_{i=1}^n y_i$$

$$\bar{x} - \text{sample mean of auxiliary character} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$\bar{y} - \text{population mean of main character} = \frac{1}{N} \sum_{i=1}^N y_i$$

$$\bar{x} - \text{population mean of auxiliary character} = \frac{1}{N} \sum_{i=1}^N x_i$$

$$s_y^2 - \text{sample mean square of main character} = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2$$

$$s_x^2 - \text{sample mean square of auxiliary character} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

$$S_y^2 - \text{population mean square of main character} = \frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{y})^2$$

$$S_x^2 - \text{population mean square of auxiliary character} = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2$$

$$s_{xy} - \text{sample covariance between } x \text{ and } y = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})$$

$$S_{xy} - \text{population covariance between } x \text{ and } y = \frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{y})(x_i - \bar{x})$$

$$r - \text{population correlation coefficient between } x \text{ and } y = S_{xy} / S_x \cdot S_y$$

$$B - \text{population coefficient of regression of } y \text{ on } x = S_{xy} / S_x^2$$

$$R - \text{population ratio of mean of main character to mean of auxiliary character} = \bar{y} / \bar{x}$$

C_y - population coefficient of variation of main character = σ_y / \bar{y}

C_x - population coefficient of variation of auxiliary character = σ_x / \bar{x}

\hat{R} - sample ratio of mean of main character to mean of auxiliary character
= \bar{y} / \bar{x}

$\hat{\beta}$ - sample coefficient of regression of y on x = σ_{xy} / σ_x^2

\bar{y}_R - the usual ratio estimator = $\bar{y} \frac{\bar{X}}{\bar{x}}$

\bar{y}_{lr} - the regression estimator = $\bar{y} + \hat{\beta} (\bar{X} - \bar{x})$

\bar{y}_p - the product estimator = $\bar{y} \bar{x} / \bar{X}$

\bar{y}_{Rd} - the usual ratio estimator in double sampling = $\bar{x}' \frac{\bar{y}}{\bar{x}}$ (where \bar{x}' denote the mean of auxiliary character based upon larger sample.)

Note : x and y are assumed to be positively correlated unless otherwise mentioned specifically.

**INTRODUCTION, REVIEW OF LITERATURE
AND
ORIENTATION OF PROBLEM**

Since times immemorial man has been subjectively utilizing the concept of generalisation, about certain population characteristics, based upon a part, the sample, of the whole population, known as inductive logic. It was only during the later half of nineteenth century that objective and systematic methods of such generalisation started spreading up which helped in evolution of modern or random sampling theory.

Sampling theory aims at devising - (i) suitable methods of selecting a part from the whole population, termed as selection or sampling procedure and, (ii) the methods of generalisation, about the desired characteristic, from the sample to the population, termed as estimation procedure. The random sampling theory deals with the determination of a combination of procedures of sampling and estimation so as to infer about the population with minimum error and consequently minimise the risk of loss which might be associated with such an error. The error in the estimate in turn depends upon the size of the sample which is directly related to the funds available for the survey. More explicitly, sampling and estimation procedures are so chosen that either the highest precision of the estimate can be attained for a given cost or a given level of precision can be achieved at the minimum cost.

The earlier developments in the field of sampling theory centered

around the procedure of drawing a representative sample is the sense that each unit in the population gets an equal chance of being included in the sample, known as equal probability sampling or simple random sampling. This is, perhaps, the simplest and the most standard technique of sampling. However, the procedure is not applicable as such in many situations, and has, therefore, to be used in conjunction with other sampling designs. A serious drawback of this method is that it does not give due importance to the sizes of the units in the population, which unduly affect both of the mean and the variability in the population. To overcome this drawback, several modifications, like stratification of the population or selection of units with unequal probabilities, are introduced. These techniques require the knowledge of the sizes of the units or information on some auxiliary character which bears a high positive correlation with the character under study.

Much of the development in the theory of sampling took place during the decade 1940-50. The works of Mahalanobis (1940, 44 and 46), Cochran (1942), Hansen and Hurvitz (1943, 46) and Madoc and Madoc (1944) are not only fundamental but also are the pillars on which the modern sampling theory rests.

Cochran (1942) gave a new turn to the modern sampling theory by suggesting methods of using the auxiliary information for the purpose of estimation in order to improve the precision of the estimates. He introduced

ratio and regression methods of estimation, which make use of supplementary information. For using these methods, information on each unit in the sample is collected not only with regard to the character under study but also to the auxiliary character. These methods of estimation take advantage of the correlation between the two characters, viz, the main character and the auxiliary character.

If y and x denote the main and auxiliary character respectively and simple random sampling is adopted for selecting the sample, the ratio estimator of population mean, say \bar{y}_R , is given by

$$\bar{y}_R = \bar{x} \frac{\bar{y}}{\bar{x}}$$

where \bar{y} and \bar{x} are the sample means of x and y values and \bar{X} is the population mean of x . Under the assumptions:

$$(I) \quad \left| \frac{\bar{x} - \bar{X}}{\bar{X}} \right| < 1 \text{ and } .$$

(II) Neither \bar{x} nor \bar{X} is zero.

the bias and the variance of \bar{y}_R , to $O(n^{-1})$ (n being the sample size), are given by

$$B(\bar{y}_R) = \left(\frac{1}{n} - \frac{1}{N} \right) \bar{y} \left| \frac{S_x^2}{\bar{x}^2} - \frac{S_{xy}}{\bar{x}\bar{y}} \right|$$

$$\text{and } V(\bar{y}_R) = \left(\frac{1}{n} - \frac{1}{N} \right) \left(S_y^2 + R^2 S_x^2 - 2R S_{xy} \right)$$

respectively. \bar{y}_R is more efficient than the simple mean \bar{y} if ρ , the coefficient of correlation between x and y , is greater than half the ratio of coefficient of variation of x to coefficient of variation of y , i.e., if

$$\rho > \frac{1}{2} \frac{C_x}{C_y}$$

Thus, if the information on auxiliary variable is either already available or can be obtained at no extra cost and it has a high positive correlation with the main character, one would certainly prefer ratio estimator to simple mean. This encouraged researchers to develop more and more superior techniques to reduce bias and also to obtain unbiased estimator by modifying either the sampling scheme or the estimation procedure or both.

The modification of sampling scheme, in order to obtain unbiased ratio estimator, basically exploits the fact that the estimator of the population ratio ($= \frac{\bar{y}}{\bar{x}}$, in the case of simple random sampling) obtained from a sample will be unbiased if the probability of selection of the sample is proportional to the product of the sample total of auxiliary variable and the probability of selecting that sample if the original sampling procedure were used, e.g., probability of selecting sample should be proportional

to $\sum_{i=1}^n x_i \cdot \frac{1}{N C_x}$ in order to make $\frac{\bar{y}}{\bar{x}}$ unbiased for population ratio $\frac{\bar{Y}}{\bar{X}}$ (In this case the sampling method is, of course, simple random sampling). It is obvious that if the probability of selection, according to

the original method of sampling, is same for all the samples, as in the case of simple random sampling, the modified sampling scheme should be such that the probability of obtaining a given sample is proportional to the total of auxiliary variable in that sample, i.e., probability $\propto \sum_{i=1}^n x_i$. Lahiri (1951), Misra (1952), and Sen (1952) independently gave the sampling procedures for obtaining such a sample. Based upon these sampling procedures, Das Raj (1956) suggested modified sampling schemes in relation to multistage, stratified, multistage and multiphase sampling designs, so as to provide unbiased ratio estimators. Murthy, Nayama and Sethi (1959) proposed sampling procedures similar to that of Misra and also some others with the advantage of achieving unbiasedness while retaining the form of usual biased ratio estimator.

In the direction of modifying estimation procedure so that less biased or unbiased estimators can be obtained, the work by Hartley and Ross (1954) deserves special mention. They gave a ratio type estimator which is unbiased under simple random sampling without replacement. Their estimator for estimating population mean is given by

$$\bar{y}_R' = \bar{y}_n \bar{x} + \frac{n(N-1)}{N(n-1)} (\bar{y} - \bar{y}_n \bar{x})$$

where \bar{y}_n is the mean of the ratios, y_i/x_i ($y_i = \frac{Y_i}{C_i}$), of the main variable to the auxiliary variable for each unit in the sample. Hsueh (1957) derived an exact expression for the variance of \bar{y}_R' by the

application of multivariate polynomials. He also showed that variance of \bar{y}_R' for large samples is given by

$$V(\bar{y}_R') = \frac{1}{n} (S_y^2 + \bar{y}_R^2 S_x^2 - 2 \bar{y}_R S_{xy}) .$$

Queneville (1956) proposed a method for reducing bias in ratio estimator. According to him the sample is split into two sub-samples of equal size and the following estimator is used to estimate the population mean

$$\bar{y}_{RQ} = 2 \bar{y}_{R1} - \frac{1}{2} (\bar{y}_{R2} + \bar{y}_{R3})$$

where \bar{y}_{R1} is the usual ratio estimator based upon the whole sample and \bar{y}_{R2} and \bar{y}_{R3} are the usual ratio estimators based upon the two splits. To the first order of approximation the estimator is unbiased and has the same variance as the usual ratio estimator has. Beale (1952) suggested the following estimator for population mean

$$\bar{y}_{RB} = \bar{y}_R \left[1 + \left(\frac{1}{n} - \frac{1}{N} \right) \frac{s_{xy}}{\bar{y}_R \bar{x}} \right] / \left[1 + \left(\frac{1}{n} - \frac{1}{N} \right) \frac{s_x^2}{\bar{x}^2} \right]$$

This estimator is also unbiased to the first order of approximation and, in fact, even to the second order of approximation it is less biased than either \bar{y}_R or \bar{y}_{RQ} . To the first order of approximation its variance is the same as that of \bar{y}_R . Tin (1965) compared usual ratio estimator, Queneville's estimator, Beal's estimator and a modified ratio estimator with regard to their biases, variances and approach to normality, etc.

Mickey (1959) developed a general theory for the construction of unbiased ratio type estimators in simple random sampling without

replacement, using information on the population means of several auxiliary variables. Shastri (1964) made a critical study of Mickey's work and attempted some of the problems not covered by him.

Most of the methods developed so far in order to reduce bias or to obtain unbiased ratio estimators, increase problem of one sort or the other. In some cases computational difficulty is enhanced, e.g., Hartley and Ross estimator is certainly more difficult to compute as compared to the usual ratio estimator; in some cases selection procedure becomes complicated, e.g., selection with varying probabilities and ; in some others the theory becomes very cumbersome. Furthermore, all of them require the existence of a high correlation between the main and the auxiliary character, which may not always be the case. This problem becomes more serious when the auxiliary information is not already available and a part of the budget has to be spent on collecting it. Generally, it is seen that it is cheaper and easier to collect information on an auxiliary character with a lower correlation with the main character than on the one with a higher correlation. For example - (i) If highly trained personnel and sophisticated measuring instruments are utilized, no doubt, the measurement errors will be negligible and consequently the coefficient of correlation will be higher than that in the case when ordinary devices are used (The tendency of measurement errors is to lower the value of correlation coefficient) ; 'but a very high cost will be involved in the former case ; (ii) The correlation between the yield of a certain crop,

any wheat, and the corresponding yield of its straw may be higher than the correlation between the production of and the area under the crop, but the data on area under the crop is much easier to obtain as compared to that on straw. Further, sometimes there is a time constraint, e. g., the estimate of wheat production is required within a very short period after harvesting, in such a case data on area under wheat crop can be obtained very soon whereas obtaining the data on straw will require a relatively very long period.

All the above mentioned factors emphasize the need for some modified ratio type estimators which can be used in situations where ordinary ratio estimator fails and which retain the plus points of ordinary ratio estimator, like the simplicity and the efficiency. The present investigation aims at developing some such modified ratio estimators. These estimators may be less biased and more efficient than the usual ratio estimator. Moreover, these are simple to compute and more flexible to use as compared to the other modified ratio estimators available, like Queenville's estimator and Beal's estimator.

CHAPTER - II

THE FIRST ESTIMATOR

SINGLE PHASE SAMPLING

(\bar{X} is known)

2.1.1 The first proposed estimator of population mean when the population mean of the auxiliary character is known and the main variable and the auxiliary variable are positively correlated is

$$T_1 = p \bar{X} \frac{\bar{Y}}{\bar{X} + (p-1) \bar{X}} \quad (p > 1) \quad (2.1.1)$$

where p is an arbitrary constant greater than unity, with, of course, certain restrictions imposed on it to achieve greater efficiency. Like the usual ratio estimator, T_1 is a consistent estimator of population mean \bar{Y} but with many advantages over the former and involves least computational problems among all modified ratio estimators.

2.1.2 Upper bound to the bias in T_1 :

$$\begin{aligned} \bar{Y} - E(\bar{Y}) &= E \left[\bar{Y}_1 \{ \bar{X} + (p-1) \bar{X} \} / p \bar{X} \right] \\ &= E \left[\bar{Y}_1 \{ \bar{X} + (p-1) \bar{X} \} \right] / p \bar{X} \end{aligned}$$

Now

$$\begin{aligned} \text{Bias in } T_1 &= E(T_1) - \bar{Y} = E(T_1) - E \left[\bar{Y}_1 \{ \bar{X} + (p-1) \bar{X} \} \right] / p \bar{X} \\ &= \frac{E(T_1) p \bar{X} - E \left[\bar{Y}_1 \{ \bar{X} + (p-1) \bar{X} \} \right]}{p \bar{X}} \end{aligned}$$

$$= \frac{E(T_1) E[\bar{X} + (p-1)\bar{X}] - E[T_1 \{ \bar{X} + (p-1)\bar{X} \}]}{p\bar{X}}$$

$$= -\text{Cov}[T_1, \{ \bar{X} + (p-1)\bar{X} \}] / p\bar{X} = -\text{Cov}(T_1, \bar{X}) / p\bar{X}$$

$$\therefore |\text{Bias in } T_1| \leq \sigma_{T_1} \sigma_{\bar{X}} / p\bar{X} \quad (2.1.2.1)$$

$$\text{Also } |\text{Bias in } \bar{Y}_R| \leq \sigma_{\bar{Y}_R} \sigma_{\bar{X}} / \bar{X}$$

Hence, the upper bound to the bias in T_1 will be less than that in \bar{Y}_R unless the standard deviation of T_1 is greater than p times the standard deviation of \bar{Y}_R .

2.1.3 Expected value of T_1 :

$$\begin{aligned} T_1 &= p\bar{X} \frac{\bar{Y}}{\bar{X} + (p-1)\bar{X}} \\ &= p\bar{X} \frac{\bar{Y}(1 + \frac{\bar{Y} - \bar{Y}}{\bar{Y}})}{p\bar{X}(1 + \frac{\bar{X} - \bar{X}}{p\bar{X}})} \end{aligned}$$

Assuming $\left| \frac{\bar{X} - \bar{X}}{p\bar{X}} \right| < 1$, expanding $(1 + \frac{\bar{X} - \bar{X}}{p\bar{X}})^{-1}$ and

simplifying, we get

$$T_1 = \bar{Y} \left[1 + \frac{\bar{Y} - \bar{Y}}{\bar{Y}} - \frac{\bar{X} - \bar{X}}{p\bar{X}} + \frac{(\bar{X} - \bar{X})^2}{p^2\bar{X}^2} - \frac{(\bar{X} - \bar{X})(\bar{Y} - \bar{Y})}{p\bar{X}\bar{Y}} \dots \right]$$

$$\therefore E(T_1) = \bar{y} \left[1 + \left(\frac{1}{n} - \frac{1}{N} \right) \left(\frac{S_x^2}{p^2 \bar{X}^2} - \frac{S_{xy}}{p \bar{X} \bar{Y}} \right) \right] \text{ to } O(n^{-1}) \quad (2.13.1)$$

2.1.4 Contention regarding bias with usual ratio estimator:

(A) In finite populations -

The expected value of usual ratio estimator to $O(n^{-1})$ is

$$E(\bar{y}_R) = \bar{y} \left[1 + \left(\frac{1}{n} - \frac{1}{N} \right) \left(\frac{S_x^2}{\bar{X}^2} - \frac{S_{xy}}{\bar{X} \bar{Y}} \right) \right]$$

Case 1: When \bar{y}_R is positively biased

(a) T_1 is also positively biased

In this case the following inequalities will hold

$$(1) \frac{S_x^2}{p^2 \bar{X}^2} - \frac{S_{xy}}{p \bar{X} \bar{Y}} > 0 \text{ i.e. } \frac{S_x^2}{p \bar{X}^2} > \frac{S_{xy}}{\bar{X} \bar{Y}}$$

and

$$(2) \frac{S_x^2}{\bar{X}^2} - \frac{S_{xy}}{\bar{X} \bar{Y}} > 0 \text{ i.e. } \frac{S_x^2}{\bar{X}^2} > \frac{S_{xy}}{\bar{X} \bar{Y}}$$

Obviously, the difference between $\frac{S_x^2}{p \bar{X}^2}$ and $\frac{S_{xy}}{\bar{X} \bar{Y}}$ (and, hence,

between $\frac{S_x^2}{\bar{X}^2}$ and $\frac{S_{xy}}{p \bar{X} \bar{Y}}$) will be less than the difference between

$\frac{S_x^2}{\bar{X}^2}$ and $\frac{S_{xy}}{\bar{X} \bar{Y}}$. Therefore the bias in T_1 will be less than

that in \bar{y}_R .

This can alternatively be proved as follows :

If $B(\bar{Y}_R)$ and $B(T_1)$ denote the biases in \bar{Y}_R and T_1 respectively,

then

$$B(\bar{Y}_R) = \left(\frac{1}{n} - \frac{1}{N} \right) \left(\frac{S_x^2}{K^2} - \frac{S_{xy}}{K\bar{Y}} \right)$$

$$B(T_1) = \left(\frac{1}{n} - \frac{1}{N} \right) \left(\frac{S_x^2}{p^2 K^2} - \frac{S_{xy}}{p K \bar{Y}} \right)$$

Now

$$B(\bar{Y}_R) - B(T_1) > B(\bar{Y}_R) - p B(T_1) \quad (\because p > 1)$$

$$\text{i.e.} \quad > \left(\frac{1}{n} - \frac{1}{N} \right) \left(\frac{S_x^2}{K^2} - \frac{S_{xy}}{K\bar{Y}} \right) - p \left(\frac{1}{n} - \frac{1}{N} \right) \left(\frac{S_x^2}{p^2 K^2} - \frac{S_{xy}}{p K \bar{Y}} \right)$$

$$\text{i.e.} \quad > \left(\frac{1}{n} - \frac{1}{N} \right) \frac{S_x^2}{K^2} \left(1 - \frac{1}{p} \right)$$

$$\text{i.e.} \quad > 0 \quad (\because p > 1)$$

$$\therefore B(\bar{Y}_R) > B(T_1)$$

Hence, if both \bar{Y}_R and T_1 are positively biased, T_1 will always be less biased.

(b) T_1 is negatively biased

(In this case (and also in the following cases) the previous type of comparison will not be always valid, for the difference (absolute)

between $\frac{S_x^2}{p K^2}$ and $\frac{S_{xy}}{K \bar{Y}}$ may be greater than the difference (absolute) between $\frac{S_x^2}{K^2}$ and $\frac{S_{xy}}{K \bar{Y}}$, though the

difference (absolute) between $\frac{S_x^2}{p^2 \bar{X}^2}$ and $\frac{S_{xy}}{p \bar{X} \bar{Y}}$ is not

necessarily greater than the difference (absolute) between $\frac{S_x^2}{\bar{X}^2}$ and

$\frac{S_{xy}}{\bar{X} \bar{Y}}$. e.g., $\frac{S_x^2}{\bar{X}^2} = 10$, $\frac{S_{xy}}{\bar{X} \bar{Y}} = 8$ and $p = 2$, which does not

happen in previous case because $(\frac{S_x^2}{p \bar{X}^2} - \frac{S_{xy}}{\bar{X} \bar{Y}})$ (which has to be

positive for bias in T_1 to be positive) cannot be greater than

$(\frac{S_x^2}{\bar{X}^2} - \frac{S_{xy}}{\bar{X} \bar{Y}})$ (which is also positive because \bar{Y}_H is positively

baised). So we go for more general type of comparison)

In this case the following inequalities will hold

$$(i) \quad \frac{S_x^2}{p^2 \bar{X}^2} - \frac{S_{xy}}{p \bar{X} \bar{Y}} < 0 \quad \text{i.e.} \quad \frac{S_x^2}{p^2 \bar{X}^2} < \frac{S_{xy}}{p \bar{X} \bar{Y}}$$

and

$$(ii) \quad \frac{S_x^2}{\bar{X}^2} - \frac{S_{xy}}{\bar{X} \bar{Y}} > 0 \quad \text{i.e.} \quad \frac{S_x^2}{\bar{X}^2} > \frac{S_{xy}}{\bar{X} \bar{Y}}$$

T_1 will be less biased than \bar{Y}_H if

$$\frac{S_{xy}}{p \bar{X} \bar{Y}} - \frac{S_x^2}{p^2 \bar{X}^2} < \frac{S_x^2}{\bar{X}^2} - \frac{S_{xy}}{\bar{X} \bar{Y}}$$

$$\text{i.e.} \quad \frac{S_{xy}}{\bar{X} \bar{Y}} \left(1 + \frac{1}{p} \right) < \frac{S_x^2}{\bar{X}^2} \left(1 + \frac{1}{p^2} \right)$$

$$\text{i.e. } \rho \frac{C_Y}{C_X} < \frac{p^2 + 1}{p(p+1)}$$

T_1 will be more biased otherwise, save one possibility when the inequality becomes an equality, in which case both T_1 and \bar{Y}_R will be equally biased.

Case II: When \bar{Y}_R is negatively biased

(a) T_1 is positively biased

This case can never arise because positive bias in T_1 implies positive bias in \bar{Y}_R , as is clear from Case I (a)

(b) T_1 is negatively biased

T_1 will be less biased than \bar{Y}_R , when both are negatively biased, if

$$\frac{S_{xy}}{p \bar{X} \bar{Y}} - \frac{S_x^2}{p^2 \bar{X}^2} < \frac{S_{xy}}{\bar{X} \bar{Y}} - \frac{S_x^2}{\bar{X}^2}$$

$$\text{i.e. } \frac{S_x^2}{\bar{X}^2} \left(1 - \frac{1}{p^2}\right) < \frac{S_{xy}}{\bar{X} \bar{Y}} \left(1 - \frac{1}{p}\right)$$

$$\text{i.e. } \rho \frac{C_Y}{C_X} > \frac{p+1}{p}$$

If the inequality holds in another direction T_1 will be more biased. T_1 and \bar{Y}_R will be equally biased if the inequality becomes equality.

(B) In infinite populations in which x and y have a bivariate normal distribution.

Following Sukhatme (1944), the expected value of T_1 and \bar{Y}_R to $O(n^{-2})$, when the population is bivariate normal, are as follows :

$$E(T_1) = \bar{Y} \left[1 + \frac{1}{n} \left(\frac{S_x^2}{p^2 K^2} - \frac{S_{xy}}{p K Y} \right) \left(1 + \frac{p}{n} \frac{S_x^2}{p^2 K^2} \right) \right]$$

$$E(\bar{Y}_R) = \bar{Y} \left[1 + \frac{1}{n} \left(\frac{S_x^2}{K^2} - \frac{S_{xy}}{K Y} \right) \left(1 + \frac{p}{n} \frac{S_x^2}{K^2} \right) \right]$$

Case I: When \bar{Y}_R is positively biased

(a) T_1 is also positively biased

Proceeding exactly as in Case I (a) under (A) we can infer that T_1 will be less biased in this case also. Moreover the bias will be further reduced because the factor $\left(1 + \frac{p}{n} \frac{S_x^2}{p^2 K^2} \right)$ is less than $\left(1 + \frac{p}{n} \frac{S_x^2}{K^2} \right)$.

(b) T_1 is negatively biased

T_1 will be less biased than \bar{Y}_R if

$$\left(\frac{S_{xy}}{p K Y} - \frac{S_x^2}{p^2 K^2} \right) \left(1 + \frac{p}{n} \frac{S_x^2}{p^2 K^2} \right) < \left(\frac{S_x^2}{K^2} - \frac{S_{xy}}{K Y} \right) \left(1 + \frac{p}{n} \frac{S_x^2}{K^2} \right)$$

Obviously, T_1 will be less biased whenever $\frac{S_{xy}}{p K Y} - \frac{S_x^2}{p^2 K^2} < \frac{S_x^2}{K^2} - \frac{S_{xy}}{K Y}$.

In general, T_1 will be less biased if

$$\frac{p C_y}{C_x} < \frac{p(p+1) \left[p^2 + \frac{p}{n} (p^2 + p + 1) C_x^2 \right]}{\left[p^2 (p^2 + 1) + \frac{p}{n} (p^4 + 1) C_x^2 \right]}$$

Case II: When \bar{Y}_R is negatively biased

(a) T_1 is positively biased.

Just as in Case I (a) under (A), this case cannot arise.

(b) T_1 is also negatively biased

T_1 will be less biased than \bar{Y}_R if

$$\left(\frac{S_{xy}}{p\bar{X}\bar{Y}} - \frac{S_x^2}{p^2\bar{X}^2} \right) \left(1 + \frac{b}{a} \frac{S_x^2}{p^2\bar{X}^2} \right) < \left(\frac{S_{xy}}{\bar{X}\bar{Y}} - \frac{S_x^2}{\bar{X}^2} \right) \left(1 + \frac{b}{a} \frac{S_x^2}{\bar{X}^2} \right)$$

This inequality will hold whenever $\left(\frac{S_{xy}}{p\bar{X}\bar{Y}} - \frac{S_x^2}{p^2\bar{X}^2} \right) < \left(\frac{S_{xy}}{\bar{X}\bar{Y}} - \frac{S_x^2}{\bar{X}^2} \right)$.

In general, T_1 will be less biased if

$$p \cdot \frac{C_y}{C_x} > \frac{p+1}{p} \frac{p^2 + \frac{b}{a}(p^2+1)C_x^2}{p^2 + \frac{b}{a}(p^2+p+1)C_x^2}$$

$$\text{i.e. } p \frac{C_y}{C_x} > \frac{p+1}{p} - \frac{\frac{b}{a}(p+1)C_x^2}{p^2 + \frac{b}{a}(p^2+p+1)C_x^2}$$

2.1.5 Variance of T_1 :

$$\begin{aligned} V(T_1) &= E \left[T_1 - E(T_1) \right]^2 \\ &= E \left[\bar{Y} \left(\frac{\bar{y} - \bar{Y}}{\bar{Y}} - \frac{\bar{x} - \bar{X}}{p\bar{X}} \right) \right]^2 \quad \text{to } O(n^{-1}) \\ &= \bar{Y}^2 \left(\frac{1}{n} - \frac{1}{N} \right) \left(\frac{S_y^2}{\bar{Y}^2} + \frac{S_x^2}{p^2\bar{X}^2} + \frac{2S_{xy}}{p\bar{X}\bar{Y}} \right) \\ &= \left(\frac{1}{n} - \frac{1}{N} \right) \left(S_y^2 - \frac{R^2 S_x^2}{p^2} - \frac{2RS_{xy}}{p} \right) \quad (2.1.5.1) \end{aligned}$$

2.1.6 Comparison of efficiencies:

(A) Finite populations

(a) With simple mean \bar{y}

The variance of \bar{y} is equal to $(\frac{1}{n} - \frac{1}{N}) S_y^2$.

T_1 will be more efficient than \bar{y} if

$$\frac{R^2 S_x^2}{p^2} - \frac{2RS_{xy}}{p} < 0$$

i.e.
$$p \frac{C_y}{C_x} > \frac{1}{2p} \quad (2.1.21)$$

$\bar{\Delta}$ N.B. : The condition for \bar{y}_R to be more efficient than simple mean is $p \frac{C_y}{C_x} > \frac{1}{2}$. So, if $p \frac{C_y}{C_x} < \frac{1}{2}$, \bar{y}_R will be less efficient than \bar{y} , thus making \bar{y}_R useless, whereas T_1 can still be safely used by adjusting p accordingly. This leads us to a useful conclusion - if there are two auxiliary characters having same coefficient of variation and the one with higher correlation (with main character) involves a higher cost, we can safely use the other auxiliary character with a lower correlation (with the main character) and involving lesser cost, by making use of T_1 by simply choosing p and retaining the sample size same, thus obtaining a saving in cost without loss in efficiency; or increasing efficiency by increasing sample size (since cheaper auxiliary character is used) without additional cost. $\bar{\Delta}$

(b) With ratio estimator:

The variance of \bar{y}_R to $O(n^{-1})$ is given by

$$V(\bar{y}_R) = (\frac{1}{n} - \frac{1}{N}) (S_y^2 + R^2 S_x^2 - 2RS_{xy})$$

T_1 will be more efficient than \bar{y}_R if

$$\frac{R^2 S_R^2}{p^2} - \frac{2RS_{xy}}{p} < R^2 S_R^2 - 2RS_{xy}$$

i.e.
$$p \frac{C_y}{C_x} < \frac{p+1}{2p} \quad (2.1.6.2)$$

The above inequality shows that whenever C_y is of same or of lower order than C_x , the proposed estimator can be made more efficient than \bar{y}_R . But if C_y is of a higher order than C_x , T_1 can be made more efficient than \bar{y}_R if p is small ($< \frac{C_x}{C_y} \frac{p+1}{2p}$).

(c) WJH regression estimator $\bar{y}_{1R} [\bar{y} = \bar{y} + \rho (\bar{X} - \bar{x})]$:
The variance of \bar{y}_{1R} to $O(n^{-1})$ is $(\frac{1}{N} - \frac{1}{N}) S_y^2 (1 - \rho^2)$

T_1 will be more efficient than \bar{y}_{1R} if

$$\frac{R^2 S_R^2}{p^2} - \frac{2RS_{xy}}{p} < -\rho^2 S_y^2$$

i.e.
$$(\rho S_y - \frac{RS_R}{p})^2 < 0 \quad (2.1.6.3)$$

which is impossible. Hence T_1 can never be more efficient than the regression estimator. However, T_1 and \bar{y}_{1R} will be equally efficient if

$$p = \frac{1}{\rho} \frac{C_x}{C_y} \quad (2.1.6.4)$$

This is the optimum value of p , as we will see later.

(d) WJH bivariate ratio estimator :

Under a particular case (for simplicity sake) when the coefficients of correlation of auxiliary characters with the main character are equal ($= \rho$), and the coefficients of variation of both the auxiliary characters are equal ($= C_x$, say). In such a case the bivariate ratio estimator \bar{y}_{BR} takes the form $\bar{y}_{BR} = \frac{1}{2} \bar{y}_{R1} + \frac{1}{2} \bar{y}_{R2}$, where \bar{y}_{R1} and \bar{y}_{R2} are the usual ratio estimator's based upon the two auxiliary variables respectively and ; its variance is equal to

$$\left(\frac{1}{2} - \frac{1}{N} \right) \bar{y}^2 \left[C_y^2 - 2\rho C_x C_y + \frac{C_x^2}{2} (1 + \rho') \right]$$

where ρ' is the coefficient of correlation between the two auxiliary characters.

It is clear from the above variance expression that the usual univariate ratio estimator can never be more efficient than a bivariate ratio estimator ($\because \frac{1+\rho'}{2} \leq 1$) ; they will, however, will be equally efficient if the two auxiliary characters are perfectly correlated.

Now, T_1 will be more efficient than \bar{y}_{BR} if

$$\frac{C_x^2}{p^2} + \frac{2\rho C_x C_y}{p} < \frac{C_x^2}{2} (1 + \rho') + 2\rho C_x C_y$$

$$\text{i.e. } \rho \frac{C_y}{C_x} < \frac{1}{4} \frac{p^2 (1 + \rho') + 2}{p(p-1)} \quad (2.1.6.5)$$

which shows that T_1 can be more efficient than the bivariate ratio estimator.

(B) Infinite population in which x and y have a bivariate normal distribution :

Using the results given by Sukhatme (1944), the variances of T_1 and

\bar{Y}_R to $O(n^{-2})$. under this situation, are as follows :

$$V(T_1) = \frac{1}{n} \bar{Y}^2 \left(C_Y^2 + \frac{C_X^2}{p^2} - \frac{2\rho C_X C_Y}{p} \right) \left(1 + \frac{3}{n} \frac{C_X^2}{p^2} \right) + \frac{8C_X^2}{n^2 p^2} \left(\frac{C_X}{p} - \rho C_Y \right)^2$$

$$V(\bar{Y}_R) = \frac{1}{n} \bar{Y}^2 (C_Y^2 + C_X^2 - 2\rho C_X C_Y) \left(1 + \frac{3}{n} C_X^2 \right) + \frac{8C_X^2}{n^2} (C_X - \rho C_Y)^2$$

For T_1 to be more efficient than \bar{Y}_R

$$V(T_1) - V(\bar{Y}_R) < 0$$

which after simplification leads to the following condition

$$\frac{8}{n} \left(1 + \frac{1}{p} \right) \left(\rho \frac{C_Y}{C_X} \right)^2 - 2 \left[1 + \frac{8C_X^2}{n} \left(1 + \frac{1}{p} + \frac{1}{p^2} \right) \right] \rho \frac{C_Y}{C_X}$$

$$+ \left(1 + \frac{1}{p} \right) \left[\frac{3}{n} C_Y^2 + \frac{8}{n} C_X^2 \left(1 + \frac{1}{p^2} \right) + 1 \right] > 0 \quad (2.1.6.6)$$

There does not seem to be a simple way to find out the nature of this inequality, that is, how it behaves with changing values of ρ , C_X , C_Y or p , since ρ , C_X and C_Y are interdependent and the above inequality is of third degree in p , which makes it extremely difficult to predict the limiting points beyond which the inequality will be satisfied or cease to hold good. To find out the approximate limit points, let us assume that C_X and C_Y do not change significantly with the change in $\rho \frac{C_Y}{C_X}$ so that $\frac{\partial C_X^2}{\partial \left(\rho \frac{C_Y}{C_X} \right)} = \frac{\partial C_Y^2}{\partial \left(\rho \frac{C_Y}{C_X} \right)} = 0$. Further, let $f \left(\rho \frac{C_Y}{C_X} \right)$ denote the following polynomial

$$\begin{aligned} E\left(\rho \frac{C_Y}{C_X}\right) &= \frac{\sigma}{\sigma} \left(1 + \frac{1}{p}\right) \left(\rho \frac{C_Y}{C_X}\right)^2 = 2 \sqrt{1 + \frac{\sigma C_X^2}{\sigma} \left(1 + \frac{1}{p} + \frac{1}{p^2}\right)} \sqrt{\rho \frac{C_Y}{C_X}} \\ &+ \left(1 + \frac{1}{p}\right) \sqrt{\frac{\sigma}{\sigma} C_Y^2 + \frac{\sigma}{\sigma} C_X^2 \left(1 + \frac{1}{p^2}\right) + 1} \sqrt{\quad} \end{aligned}$$

then, the above inequality will be satisfied beyond or within the limiting points for $\rho \frac{C_Y}{C_X}$.

$$\begin{aligned} &\sqrt{\left[1 + \frac{\sigma}{\sigma} C_X^2 \left(1 + \frac{1}{p} + \frac{1}{p^2}\right)\right]} \pm \sqrt{\left[1 + \frac{\sigma}{\sigma} C_X^2 \left(1 + \frac{1}{p} + \frac{1}{p^2}\right)\right]^2} \\ &\cdot \frac{\sigma}{\sigma} \left(1 + \frac{1}{p}\right)^2 \left\{ \frac{\sigma}{\sigma} C_Y^2 + \frac{\sigma}{\sigma} C_X^2 \left(1 + \frac{1}{p^2}\right) + 1 \right\} \sqrt{\quad} / \frac{\sigma}{\sigma} \left(1 + \frac{1}{p}\right) \end{aligned}$$

(2.1.6.7)

2.1.7 Estimate of variance:

A consistent estimator of variance of T_1 is given by

$$\hat{V}(T_1) = \left(\frac{1}{N} + \frac{1}{n}\right) \left(s_y^2 + \frac{\hat{R}^2 s_x^2}{p^2} - \frac{2 \hat{R} s_{xy}}{p} \right) \quad (2.1.7.1)$$

2.1.8 Optimum value of p:

Differentiating the variance expression (2.1.8.1) with respect to p and equating it to zero, we get the optimum value of p as

$$p = \frac{1}{\rho} \frac{C_X}{C_Y} \quad (2.1.8.1)$$

This is exactly that is obtained on comparing the efficiencies of T_1 and regression estimator. Hence the minimum value that the variance of T_1 can attain is the variance of regression estimator, i.e., $\left(\frac{1}{N} + \frac{1}{n}\right) s_y^2 (1 - \rho^2)$. Also with optimum value of p , T_1 becomes unbiased. (2.1.8.1) gives us a guide line to choose p which can be chosen to be the one that is obtained on substituting the sample estimate of ρ , C_X and C_Y for their true values.

TWO PHASE SAMPLING

(\bar{X} is not known)

2.2.1 In the previous section, we considered the case when the population mean of auxiliary character, viz., \bar{X} , is known. In case, however, information on \bar{X} is not available in advance, the usual procedure in such a situation, known as two phase or double sampling, can be adopted. This sampling design consists in taking a large preliminary sample, of size n' , say, in order to estimate \bar{X} and, taking a sub-sample, of size n , say, from the preliminary sample to observe the character under study. The usual ratio estimator in double sampling is given by

$$\bar{y}_{RD} = \bar{n}' \frac{\bar{y}}{\bar{X}}$$

where \bar{y} is the mean of character under study as obtained from the sample of size n and, \bar{X} and \bar{n}' are the means of auxiliary character as obtained from samples of sizes n and n' respectively.

The expected value and the variance of \bar{y}_{RD} are given by

$$E(\bar{y}_{RD}) = \bar{Y} \left[1 + \left(\frac{1}{n} - \frac{1}{n'} \right) \left(\frac{S_X^2}{\bar{X}^2} - \frac{S_{XY}}{\bar{X}\bar{Y}} \right) \right]$$

and

$$V(\bar{y}_{RD}) = \left(\frac{1}{n} - \frac{1}{n'} \right) (S_Y^2 + R^2 S_X^2 - 2RS_{XY}) + \left(\frac{1}{n} - \frac{1}{n'} \right) S_Y^2$$

Following two phase sampling approach the following estimator is proposed :

$$T_M = p \bar{n}' \frac{\bar{y}}{\bar{n} + (p-1)\bar{X}} \quad (p > 1) \quad (2.2.1.1)$$

2.2.2 Expected value of T_{14} :

$$T_{14} = p \bar{X}' \frac{\bar{Y}}{\bar{X} + (p-1)\bar{X}'}$$

$$= \frac{p \bar{X} \left[1 + \frac{\bar{X}' - \bar{X}}{\bar{X}} \right] \bar{Y} \left[1 + \frac{\bar{Y} - \bar{Y}}{\bar{Y}} \right]}{p \bar{X} \left[1 + \frac{\bar{X}' - \bar{X}}{p \bar{X}} \right] + \frac{(p-1)(\bar{X}' - \bar{X})}{p \bar{X}} \bar{Y}}$$

$$= \bar{Y} \left[1 + \frac{c}{\bar{X}} \right] \left[1 + \frac{q}{\bar{Y}} \right] \left[1 + \frac{c'}{p \bar{X}} + \frac{(p-1)c'}{p \bar{X}} \right]^{-1}$$

where $q = (\bar{Y} - \bar{Y})$, $c = (\bar{X} - \bar{X})$ and $c' = (\bar{X}' - \bar{X})$. Assuming

$$\left| \frac{c}{p \bar{X}} + \frac{(p-1)c'}{p \bar{X}} \right| < 1 \text{ and expanding, we get}$$

$$T_{14} = \bar{Y} \left(1 + \frac{q}{\bar{Y}} \right) \left(1 + \frac{c}{\bar{X}} \right) \left(1 - \frac{c}{p \bar{X}} - \frac{(p-1)c'}{p \bar{X}} + \frac{c^2}{p^2 \bar{X}^2} + \frac{(p-1)^2 c'^2}{p^2 \bar{X}^2} + \frac{2(p-1)cc'}{p^2 \bar{X}^2} \dots \right)$$

Simplifying, we get

$$T_{14} = \bar{Y} \left(1 + \frac{q}{\bar{Y}} + \frac{c}{p \bar{X}} - \frac{c}{p \bar{X}} - \frac{p-1}{p^2} \frac{c'^2}{\bar{X}^2} + \frac{c^2}{p^2 \bar{X}^2} + \frac{p-2}{p^2} \frac{cc'}{\bar{X}^2} + \frac{c'q}{p \bar{X} \bar{Y}} + \frac{q^2}{p \bar{X} \bar{Y}} \dots \right)$$

Taking expectation on both sides and using

$$E(q) = E(c) = E(c') = 0;$$

$$E(c^2) = V(\bar{X}) = \sigma^2 S_{XX}^{-1};$$

$$E(c'^2) = V(\bar{X}') = \sigma^2 S_{XX'}^{-1};$$

$$E(c|c') = E \left[\bar{c}' E(c|c') \right] = E(c'^2) = \sigma' S_{\bar{c}}^2 ;$$

$$E(c|q) = \text{Cov}(\bar{c}, \bar{y}) = \sigma S_{\bar{c}y} ;$$

$$E(c'q) = E \left[\bar{c}' E(q|c') \right] = E(c'q') = \text{Cov}(\bar{c}', \bar{y}') = \sigma' S_{\bar{c}y'} ;$$

where $\sigma = \left(\frac{1}{n} - \frac{1}{N} \right)$, $\sigma' = \left(\frac{1}{n_1} - \frac{1}{N} \right)$, \bar{y}' is mean of character

under study as would be obtained from the preliminary sample of size n'

and $q' = \bar{y}' - \bar{y}$; we get, to $O(n^{-1})$

$$\begin{aligned} E(T_{14}) &= \bar{y} \left[1 - \frac{p-1}{p^2} \sigma' \frac{S_{\bar{c}}^2}{\bar{X}^2} + \sigma \frac{S_{\bar{c}}^2}{p^2 \bar{X}^2} + \frac{p-1}{p^2} \sigma \frac{S_{\bar{c}}^2}{\bar{X}^2} \right. \\ &\quad \left. + \sigma' \frac{S_{\bar{c}y}}{p \bar{X} \bar{Y}} - \sigma \frac{S_{\bar{c}y}}{p \bar{X} \bar{Y}} \right] \\ &= \bar{y} \left[1 + (\sigma - \sigma') \left(\frac{S_{\bar{c}}^2}{p^2 \bar{X}^2} - \frac{S_{\bar{c}y}}{p \bar{X} \bar{Y}} \right) \right] \\ &= \bar{y} \left[1 + \left(\frac{1}{n} - \frac{1}{n_1} \right) \left(\frac{S_{\bar{c}}^2}{p^2 \bar{X}^2} - \frac{S_{\bar{c}y}}{p \bar{X} \bar{Y}} \right) \right] \quad (2.2.2.1) \end{aligned}$$

2.2.3 Comparison of biases in T_{14} and \bar{y}_{R4} :

It is evident from the expressions of expected values of T_{14} and \bar{y}_{R4} that the conditions regarding the comparison of biases in this case will be identical to those obtained in the case of single sampling, i.e., the case where \bar{X} is known.

2.2.4 Variance of T_{14} :

$$\begin{aligned} V(T_{14}) &= E \left[\bar{y} T_{14} - E(T_{14}) \right]^2 \\ &= E \left[\bar{y} \left(\frac{y}{\bar{y}} + \frac{y'}{p \bar{X}} - \frac{y}{p \bar{X}} + \dots \right) \right]^2 \end{aligned}$$

$$\begin{aligned}
 &= \bar{Y}^2 \left(0 \frac{S_y^2}{Y^2} + 0' \frac{S_x^2}{p^2 \bar{X}^2} + 0 \frac{S_x^2}{p^2 \bar{X}^2} + 2 \cdot 0' \frac{S_{xy}}{p \bar{X} \bar{Y}} - 2 \cdot 0 \frac{S_{xy}}{p \bar{X} \bar{Y}} \right. \\
 &\quad \left. + 2 \cdot 0'^2 \frac{S_x^2}{p^2 \bar{X}^2} \right) \left[\text{to } o(n^{-1}) \right] \\
 &= \bar{Y}^2 \left[0 \frac{S_y^2}{Y^2} + (0 - 0') \left(\frac{S_x^2}{p^2 \bar{X}^2} - \frac{2 S_{xy}}{p \bar{X} \bar{Y}} \right) \right] \\
 &= \bar{Y}^2 (0 - 0') \left(\frac{S_y^2}{Y^2} + \frac{S_x^2}{p^2 \bar{X}^2} - \frac{2 S_{xy}}{p \bar{X} \bar{Y}} \right) + 0' S_y^2 \\
 &= \left(\frac{1}{n} - \frac{1}{n'} \right) \left(S_y^2 + \frac{n^2 S_x^2}{p^2} - \frac{2 R S_{xy}}{p} \right) + \left(\frac{1}{n'} - \frac{1}{n} \right) S_y^2 \tag{2.2.4.1}
 \end{aligned}$$

2.2.5 Comparison of efficiency:

(a) With disbursements (\bar{X}) based upon a ratio

T_{10} will be more efficient than $\bar{Y} U$

$$\frac{R^2 S_x^2}{p^2} - \frac{2 R S_{xy}}{p} < 0$$

i.e.
$$p \frac{C_y}{C_x} > \frac{1}{2 p} \tag{2.2.5.1}$$

Although this comparison is not very relevant and justified because of the involvement of cost in obtaining information on auxiliary character, yet it brings forth a fact that can be exploited in choosing an auxiliary character, when several auxiliary characters are available, in order to reduce cost or to increase efficiency or to choose a strategy to get an optimum combination of cost and efficiency. This we shall attempt after the comparison of efficiency with \bar{Y}_{Rd} .

(b) With usual ratio estimator in double sampling:

T_{14} will be more efficient than \bar{y}_{RD} if

$$\frac{R^2 S_y^2}{p^2} - \frac{2RS_{xy}}{p} < R^2 S_y^2 - 2RS_{xy}$$

i.e.
$$p \frac{C_y}{C_x} < \frac{p+1}{2p} \quad (2.2.5.2)$$

(2.2.5.1) indicates that if an auxiliary character does not conform to the condition $p \frac{C_y}{C_x} > \frac{1}{2}$, which is the condition for \bar{y}_{RD} to be more efficient than simple mean, it can still be used by making use of T_{14} by choosing p so that (2.2.5.1) is satisfied and thus obtaining an estimate which will be more efficient than the simple mean. (2.2.5.1) together with (2.2.5.2) indicates that if an auxiliary character satisfies $\frac{1}{2p} < p \frac{C_y}{C_x} < \frac{p+1}{2p}$, by making use of T_{14} we will get an estimate that will be more efficient than simple mean as well as usual ratio estimate in double sampling for a given cost or conversely obtain a reduction in cost for a given level of precision (by taking a smaller sample). Further, if two or more auxiliary characters are available with the same coefficients of variation and which are such that the one with lower correlation coefficient with the main character involves less cost than the one with higher correlation coefficient, we can use the cheaper one and make use of T_{14} to obtain an equally precise estimate while retaining the same sample size and thus cut down the cost or alternatively increase the precision of the estimate by increasing the sample size with the available funds.

2.2.6 Estimate of variance :

A consistent estimator of variance is obtained by the following expression :

$$\hat{V}(T_M) = \left(\frac{1}{n} - \frac{1}{N}\right) \left(s_y^2 + \frac{\hat{R}^2 s_x^2}{p^2} - \frac{2\hat{R} s_{xy}}{p}\right) + \left(\frac{1}{n'} - \frac{1}{N}\right) s_y^2 \quad (2.2.6.1)$$

2.2.7 Optimum combination of n and n' :

As pointed out earlier the comparison of efficiencies of T_M and \bar{y} as such is not very relevant and justified. One would certainly like to know whether the extra cost, incurred on the additional sample taken for observing the auxiliary character, would increase the precision of the estimate. We, therefore, choose that sampling procedure which results in minimum variance for a given cost. Let c and c' be the cost per unit for observing the main character and the auxiliary character respectively and C_0 be the total cost of the survey. Then

$$C_0 = cn + c'n' \quad (2.2.7.1)$$

Assuming N large, we have

$$V(T_M) = \frac{V_1}{n} + \frac{V_2}{n'} \quad (2.2.7.2)$$

where $V_1 = \left(s_y^2 + R^2 \frac{s_x^2}{p^2} - 2R \frac{s_{xy}}{p}\right)$ and $V_2 = -\left(R^2 \frac{s_x^2}{p^2} - 2R \frac{s_{xy}}{p}\right)$

We will now determine n and n' such that the variance is minimum subject to the budget constraint (2.2.7.1). Consider, therefore, the expression

$$\psi = V(T_M) + \lambda (cn + c'n' - C_0) \quad (2.2.7.3)$$

Case I : p is fixed

Differentiating ϕ with respect to n and n' , equating the results to zero, eliminating λ and, using (2.2.7.1), we get the optimum values of n and n' as follows :

$$n = \frac{C_0 \sqrt{V_1}}{\sqrt{c} (\sqrt{c} V_1 + \sqrt{c'} V_2)} \quad (2.2.7.4)$$

and

$$n' = \frac{C_0 \sqrt{V_2}}{\sqrt{c'} (\sqrt{c} V_1 + \sqrt{c'} V_2)} \quad (2.2.7.5)$$

Substituting the values of n and n' in (2.2.7.2), we get the optimum variance of T_{1d} as

$$V(T_{1d})_{opt} = \frac{1}{C_0} (\sqrt{c} V_1 + \sqrt{c'} V_2)^2 \quad (2.2.7.6)$$

If the simple random sampling is adopted, a sample of size C_0/c can be drawn, therefore

$$V(\bar{y})_{opt} = \frac{S_y^2}{C_0/c} = \frac{c S_y^2}{C_0} \quad (2.2.7.7)$$

Hence, the relative efficiency of T_{1d} as compared to simple mean under optimum conditions is given by

$$RE = \frac{c S_y^2}{(\sqrt{c} V_1 + \sqrt{c'} V_2)^2} \quad (2.2.7.8)$$

Similarly, the relative efficiency of T_{1d} as compared to \bar{V}_{Rd} using the same auxiliary character in both the cases, (if T_{1d} can be used, i.e., (2.2.5.2) is satisfied), is given by

$$RE = \frac{(\sqrt{c} V_{1R} + \sqrt{c'} V_{2R})^2}{(\sqrt{c} V_1 + \sqrt{c'} V_2)^2} \quad (2.2.7.9)$$

where V_{1R} and V_{2R} are the counterparts of V_1 and V_2 for the usual ratio estimator and may be obtained by putting $p=1$ in V_1 and V_2 respectively (since \bar{Y}_{Rd} is a special case of T_{1d} for $p=1$).

Hence, there will be a gain in efficiency by using T_{1d} if

$$\frac{c}{c'} > \frac{\sqrt{V_2} - \sqrt{V_{2R}}}{\sqrt{V_{1R}} - \sqrt{V_1}} \quad (2.2.7.10)$$

which will be true, in particular, when either the numerator or the denominator will be negative.

Further, the relative efficiency of T_{1d} when it is used with an auxiliary character that has a lower correlation with the main character and involves less per unit cost, compared to \bar{Y}_{Rd} being used with another auxiliary character having higher correlation with character under study and involving a higher cost per unit, is given by

$$RE = \frac{\left[\sqrt{c} \sqrt{(S_y^2 + R_1^2 S_{x1}^2 - 2R_1 S_{x1y})} + \sqrt{c_1'} \sqrt{(2R_1 S_{x1y} - R_1^2 S_{x1}^2)} \right]^2}{\left[\sqrt{c} \sqrt{\left(S_y^2 + \frac{R_2^2 S_{x2}^2}{p^2} - \frac{2R_2 S_{x2y}}{p} \right)} + \sqrt{c_2'} \sqrt{\left(\frac{2R_2 S_{x2y}}{p} - \frac{R_2^2 S_{x2}^2}{p^2} \right)} \right]^2} \quad (2.2.7.11)$$

where all the symbols bear their usual meanings and suffix '1' is used to denote the costlier character and suffix '2' for the cheaper one.

Hence, T_{1d} with a cheaper character will lead to a gain in precision over using \bar{Y}_{Rd} with the costlier character if

$$\frac{c_1}{c_2} > \frac{\sqrt{c_1}}{\sqrt{c_2}} \left\{ \sqrt{ \left(S_y^2 + \frac{R_2^2 S_{M2}^2}{p^2} - \frac{2R_2 S_{M2Y}}{p} \right) - \left(R_1^2 + R_1^2 S_{M1}^2 - 2R_1 S_{M1Y} \right) + \sqrt{ \left(\frac{2R_2 S_{M2Y}}{p} - \frac{R_2^2 S_{M2}^2}{p^2} \right) } \right\} / \left(2R_1 S_{M1Y} - R_1^2 S_{M1}^2 \right) \quad (2.2.7.12)$$

Case II: p is not fixed.

Differentiating ϕ with respect to n , n' and p and proceeding as in Case I we find that the optimum values of n and n' are same as given by (2.2.7.4) and (2.2.7.5) and that of p is given by

$$p = \frac{1}{p} \frac{C_M}{C_Y} \quad (2.2.7.13)$$

Substitution of the values of n , n' and p gives the optimum variance of

$$T_M \text{ as } V(T_M)_{opt} = \frac{S_Y^2}{C_0} \left[\sqrt{c(1-p^2)} + p\sqrt{c'} \right]^2 \quad (2.2.7.14)$$

giving the relative efficiency of T_M as compared to simple mean as

$$RE = \frac{c}{\left[\sqrt{c(1-p^2)} + p\sqrt{c'} \right]^2} \quad (2.2.7.15)$$

Hence the double sampling will lead to a gain in precision if

$$\frac{c}{c'} > \frac{p^2}{\left[1 - \sqrt{(1-p^2)} \right]^2} \quad (2.2.7.16)$$

Similarly, the relative efficiency of T_M as compared to \bar{y}_{M1} , using the same auxiliary character, provided T_M can be used, i.e., (2.2.5.2)

is satisfied, is given by

$$RE = \frac{\left[\sqrt{c} \sigma_{IR} + \sqrt{c'} \sigma_{IR} \right]^2}{\sigma_y^2 \left[\sqrt{c(1-\rho^2)} + \rho \sqrt{c'} \right]^2} \quad (2.2.7.17)$$

which will always be greater than unity, since individually $\sqrt{c} \sigma_{IR} > \sigma_y \sqrt{c(1-\rho^2)}$ and $\sqrt{c'} \sigma_{IR} > \sigma_y \rho \sqrt{c'}$. Hence under optimum conditions, if T_{14} can be used, it will always be better than \bar{V}_{R4} .

Again, the relative efficiency of T_{14} , when it is used with an auxiliary character with lower correlation with the main character (the cheaper one) as compared to \bar{V}_{R4} , being used with another auxiliary character with higher correlation with main character (the costlier one) is given by

$$RE = \frac{\left[\sqrt{c} \sigma_y \left(\sigma_y^2 + R_1^2 \sigma_{x1}^2 - 2R_1 \sigma_{x1y} \right) + \sqrt{c'} \sigma_1 \left(2R_1 \sigma_{x1y} - R_1^2 \sigma_{x1}^2 \right) \right]^2}{\sigma_y^2 \left[\sigma \left(1 - \rho_2^2 \right) + \rho_2 \sigma_2 \right]^2} \quad (2.2.7.18)$$

Using T_{14} , with a cheaper v character will lead to a gain in precision, under optimum conditions, will be advantageous in comparison to using \bar{V}_{R4} , with a costlier character, if

$$\frac{c'}{c_2} > \frac{\left[\sqrt{\frac{c}{c_2}} \left\{ \sigma_y \sqrt{1 - \rho_2^2} - \sqrt{\left(\sigma_y^2 + R_1^2 \sigma_{x1}^2 - 2R_1 \sigma_{x1y} \right)} \right\} + \sigma_y \rho_2 \right]^2}{\left(2R_1 \sigma_{x1y} - R_1^2 \sigma_{x1}^2 \right)} \quad (2.2.7.19)$$

SAMPLING WITH VARYING PROBABILITIES

The proposed estimator T_1 provides an unbiased estimate of the population mean under the following varying probability sampling schemes. We shall, first, take up the case of single phase sampling and then extend it to stratified and two phase sampling.

2.3.1 Single-phase sampling

2.3.1.1 Sampling Scheme:

Select the first unit with probability proportional to the value of the auxiliary character plus $(p - 1)$ times the population mean of the auxiliary character. I.e., i -th unit will be selected with probability proportional to $[x_i + (p-1)\bar{X}]$ and select the remaining $(n - 1)$ units with equal probability and without replacement from the remaining $(N-1)$ units of the population.

2.3.1.2 Probability of obtaining a specified sample:

Probability of selecting a specified sample y_1, y_2, \dots, y_n of n units under the above scheme is given by

$$\begin{aligned}
 P(y_1, y_2, \dots, y_n) &= \sum_{i=1}^n \left[\text{Probability that } y_i \text{ is selected at first draw} \right] \times \left[\text{Probability of selecting remaining } n-1 \text{ units, given that } y_i \text{ is selected at first draw} \right] \\
 &= \sum_{i=1}^n \frac{x_i + (p-1)\bar{X}}{\sum_{i=1}^N [x_i + (p-1)\bar{X}]} \cdot \frac{1}{N-1 C_{n-1}} \\
 &= \frac{\bar{X} + (p-1)\bar{X}}{p\bar{X}} \cdot \frac{1}{N C_n} \\
 &\qquad\qquad\qquad (2.3.1.2.1)
 \end{aligned}$$

2.3.1.3 Expected value of T_1 :

$$T_1 = p \bar{X} \frac{\bar{y}}{\bar{X} + (p-1)\bar{X}}$$

$$E(T_1) = p \bar{X} E \left[\frac{\bar{y}}{\bar{X} + (p-1)\bar{X}} \right]$$

$$= p \bar{X} E' \frac{\bar{y}}{\bar{X} + (p-1)\bar{X}} \cdot \frac{\bar{X} + (p-1)\bar{X}}{p \bar{X}} \cdot \frac{1}{N C_n}$$

$$= \bar{y} \quad (2.3.1.3.1)$$

where Σ' denotes summation extended over all $N C_n$ possible samples.

Hence T_1 provides an unbiased estimate of population mean under above sampling scheme.

2.3.1.4 Variance of T_1 :

$$V(T_1) = E(T_1^2) - \left[E(T_1) \right]^2$$

$$= E' \left[p \bar{X} \frac{\bar{y}}{\bar{X} + (p-1)\bar{X}} \right]^2 \cdot \frac{\bar{X} + (p-1)\bar{X}}{p \bar{X}} \cdot \frac{1}{N C_n} - \bar{y}^2$$

$$= \frac{p \bar{X}}{N C_n} E' \frac{\bar{y}^2}{\bar{X} + (p-1)\bar{X}} - \bar{y}^2 \quad (2.3.1.4.1)$$

2.3.1.5 Estimate of variance :

The variance of T_1 can be estimated unbiasedly if \bar{y}^2 can be estimated unbiasedly.

$$\text{Now } \bar{y}^2 = \left(\frac{1}{N} \sum_{i=1}^N y_i \right)^2$$

$$= \frac{1}{N^2} \sum_{i=1}^N y_i^2 + \frac{1}{N^2} \sum_{(i,j)=1}^N y_i y_j$$

Consider the statistic

$$\begin{aligned}
 & \sum_{i=1}^n y_i^2 \frac{p \bar{X}}{Nn \left[\bar{X} + (p-1) \bar{X} \right]} \\
 E \left[\sum_{i=1}^n y_i^2 \frac{p \bar{X}}{Nn \left[\bar{X} + (p-1) \bar{X} \right]} \right] \\
 &= E \left[\sum_{i=1}^n y_i^2 \frac{p \bar{X}}{Nn \left[\bar{X} + (p-1) \bar{X} \right]} \cdot \frac{\bar{X} + (p-1) \bar{X}}{p \bar{X}} \cdot \frac{1}{N C_n} \right] \\
 &= E \left[\sum_{i=1}^n \frac{y_i^2}{Nn} \cdot \frac{1}{N C_n} \cdot N-1 C_{n-1} \sum_{i=1}^n \frac{y_i^2}{Nn} \cdot \frac{1}{N C_n} \right] \\
 &= \frac{1}{N^2} \sum_{i=1}^n y_i^2
 \end{aligned}$$

Now, consider another statistic

$$\begin{aligned}
 & \sum_{i \neq j}^n y_i y_j \frac{(N-1) p \bar{X}}{Nn(n-1) \left[\bar{X} + (p-1) \bar{X} \right]} \\
 E \left[\sum_{i \neq j}^n y_i y_j \frac{(N-1) p \bar{X}}{Nn(n-1) \left[\bar{X} + (p-1) \bar{X} \right]} \right] \\
 &= E \left[\sum_{i \neq j}^n y_i y_j \frac{N-1}{Nn(n-1)} \cdot \frac{1}{N C_n} \right] \\
 &= N-2 C_{n-2} \sum_{i \neq j}^n y_i y_j \frac{N-1}{Nn(n-1)} \cdot \frac{1}{N C_n} \\
 &= \frac{1}{N^2} \sum_{i \neq j}^n y_i y_j
 \end{aligned}$$

Thus, the unbiased estimator of \bar{Y}^2 is given by

$$\sum_{i=1}^n y_i^2 \frac{p \bar{X}}{Nn \left[\bar{X} + (p-1) \bar{X} \right]} + \sum_{i \neq j}^n y_i y_j \frac{(N-1) p \bar{X}}{Nn(n-1) \left[\bar{X} + (p-1) \bar{X} \right]}$$

$$= \frac{p\bar{X}}{Np\left[\bar{X} + (p-1)\bar{X}\right]} \left[\sum_{i=1}^n y_i^2 + \frac{N-1}{n-1} \sum_{i \neq j=1}^n y_i y_j \right]$$

Hence, an unbiased estimate of variance of T_1 is given by

$$\hat{V}(T_1) = T_1^2 - \frac{p\bar{X}}{Np\left[\bar{X} + (p-1)\bar{X}\right]} \left[\sum_{i=1}^n y_i^2 + \frac{N-1}{n-1} \sum_{i \neq j=1}^n y_i y_j \right] \quad (2.3.1.5.1)$$

2.3.2 Extension to Stratified Sampling

In the case of stratified sampling there are three methods of sampling in which the population mean can be estimated unbiasedly.

2.3.2.1 Method 1 :

Consider each stratum as a population and draw samples from each stratum in a manner stated in the previous case $\left[\text{vide (2.3.1.1)} \right]$; prepare an estimator similar to T_1 from each of the sample and; finally, combine them with weights proportional to the sizes of respective strata to yield a single estimator of population mean. Thus -

let y_{ij} denote the value of the character under study for the j -th unit of the i -th stratum ($j = 1, 2, \dots, N_i$), ($i = 1, 2, \dots, L$), then T_{1st} the estimator of population mean is given by

$$T_{1st} = \sum_{i=1}^L \frac{N_i}{N} T_{1i} \quad (2.3.2.1.1)$$

where $T_{1i} = p \bar{X}_i \frac{\bar{y}_i}{\bar{X}_i + (p-1)\bar{X}_i}$ and, all the notations carry

their usual meanings and suffix 'i' stands for i-th stratum.

It follows from the previous case that

$$E(T_{1st}) = \sum_{i=1}^L \frac{N_i}{N} E(T_{1i}) = \sum_{i=1}^L \frac{N_i}{N} \bar{Y}_i$$

$$= \bar{Y} \quad (2.3.2.1.2)$$

thus making T_{1st} unbiased for population mean :

and

$$V(T_{1st}) = \sum_{i=1}^L \sum_{C_{N_i}} \frac{N_i^2}{N^2} \frac{p \bar{X}_i}{N_i} \frac{\bar{Y}_i^2}{\bar{X}_i + (p-1) \bar{X}_i} = \bar{Y}^2$$

(2.3.2.1.3)

where summation $\sum_{C_{N_i}}$ is extended over $N_i C_{N_i}$ samples from the i -th stratum; n_i being the number of units selected from the i -th stratum.

Similarly, an unbiased estimate of variance is given by

$$\hat{V}(T_{1st}) = \sum_{i=1}^L \frac{N_i^2}{N^2} \left[(T_{1i})^2 - \frac{p \bar{X}_i}{N_i n_i \left[\bar{X}_i + (p-1) \bar{X}_i \right]} \left(\sum_{j=1}^{n_i} y_{ij}^2 \right) \right. \\ \left. + \frac{N_i - 1}{n_i - 1} \sum_{j \neq k}^{n_i} y_{ij} y_{ik} \right]$$

(2.3.2.1.3)

3.2.2 Method II:

3.2.2.1 Sampling Scheme:

Select first unit from the whole population with probability proportional to the value of auxiliary character plus $(p-1)$ times the population mean of auxiliary character, i.e., j -th unit of the i -th stratum will be selected with probability proportional to $\left[\sum_{j=1}^{n_i} x_{ij} + (p-1) \bar{X}_i \right]$. If this unit happens to belong to i -th stratum, select remaining $(n_i - 1)$ units, from this stratum, with equal probability and without replacement from the remaining $(N_i - 1)$ units. From the other strata, select specified number of units with equal probability and without replacement. It may be mentioned

here that n_1 's can be determined on the basis of optimum or proportional allocation.

3.2.2.2 Probability of selecting a specified sample :

Probability of selecting a specified sample is given by

$$P = \sum_{i=1}^L \sum_{j=1}^{n_i} \left[\left(\text{Probability that } y_{ij} \text{ is selected at first draw} \right) \times \left(\text{Probability of selecting remaining } n_i - 1 \text{ units from } i\text{-th stratum, given that } y_{ij} \text{ is selected at first draw} \right) \times \left(\text{Probability of selecting remaining units from the other strata, given that the first unit belongs to the } i\text{-th stratum} \right) \right]$$

$$= \sum_{i=1}^L \sum_{j=1}^{n_i} \left[\frac{n_{ij} + (p-1)\bar{X}}{\sum_{i=1}^L \sum_{j=1}^{n_i} \{n_{ij} + (p-1)\bar{X}\}} \cdot \frac{1}{N_i - 1} \cdot \frac{1}{\sum_{h \neq i}^L \frac{N_h}{N} C_{N_h}} \right]$$

$$= \frac{\sum_{i=1}^L \sum_{j=1}^{n_i} \{n_{ij} + (p-1)\bar{X}\}}{p N \sum_{i=1}^L \frac{N_i}{N} C_{N_i}} \quad (2.2.2.2.1)$$

3.2.2.3 Estimator and its expected value :

An unbiased estimator of population mean is given by

$$T'_{1st} = p \bar{X} \frac{\sum_{i=1}^L \frac{N_i}{N} \bar{y}_i}{\sum_{i=1}^L \frac{N_i}{N} \{ \bar{X}_i + (p-1)\bar{X} \}} \quad (2.2.2.3.1)$$

$$E(T'_{1st}) = E_0' p \bar{X} \frac{\sum_{i=1}^L \frac{N_i}{N} \bar{y}_i}{\sum_{i=1}^L \frac{N_i}{N} \{ \bar{X}_i + (p-1)\bar{X} \}} \cdot \frac{\sum_{i=1}^L \frac{N_i}{N} \{ \bar{X}_i + (p-1)\bar{X} \}}{p N \sum_{i=1}^L \frac{N_i}{N} C_{N_i}}$$

$$= E_0' \sum_{i=1}^L \frac{N_i}{N} \bar{y}_i \cdot \frac{1}{\sum_{i=1}^L \frac{N_i}{N} C_{N_i}}$$

$$= \bar{Y}$$

(2.3.2.2.2)

where Σ_0' is extended over all $\prod_{l=1}^L N_l C_{Nl}$ samples.

2.3.2.2.4 Variance of the estimator:

$$V(T'_{1st}) = E(T'_{1st})^2 - [E(T'_{1st})]^2$$

$$= E_0' \left[\frac{p \bar{X}}{\sum_{l=1}^L N_l \{ \bar{N}_l + (p-1) \bar{X} \}} \frac{\sum_{l=1}^L N_l \bar{Y}_l}{\sum_{l=1}^L N_l \{ \bar{N}_l + (p-1) \bar{X} \}} - \frac{\sum_{l=1}^L N_l \{ \bar{N}_l + (p-1) \bar{X} \}}{p \bar{X} \prod_{l=1}^L N_l C_{Nl}} \right]^2$$

$$= \frac{p \bar{X}}{N} \frac{1}{\prod_{l=1}^L N_l C_{Nl}} E_0' \frac{(\sum_{l=1}^L N_l \bar{Y}_l)^2}{\sum_{l=1}^L N_l \{ \bar{N}_l + (p-1) \bar{X} \}} - \bar{Y}^2 \quad (2.3.2.2.4.1)$$

2.3.2.2.5 Estimate of variance:

Consider the statistic

$$t_1 = \frac{\sum_{l=1}^L \frac{N_l}{N C_{Nl}} \sum_{j=1}^{N_l} v_{lj}^2}{\sum_{l=1}^L N_l \{ \bar{N}_l + (p-1) \bar{X} \}} \frac{p \bar{X}}{\sum_{l=1}^L N_l \{ \bar{N}_l + (p-1) \bar{X} \}}$$

$$E(t_1) = E_0' \frac{\sum_{l=1}^L \frac{N_l}{N C_{Nl}} \sum_{j=1}^{N_l} v_{lj}^2}{\sum_{l=1}^L N_l \{ \bar{N}_l + (p-1) \bar{X} \}} \frac{p \bar{X}}{\sum_{l=1}^L N_l \{ \bar{N}_l + (p-1) \bar{X} \}} \frac{\sum_{l=1}^L N_l \{ \bar{N}_l + (p-1) \bar{X} \}}{p \bar{X} \prod_{l=1}^L N_l C_{Nl}}$$

$$= E_0' \frac{\sum_{l=1}^L \frac{N_l}{N^2 C_{Nl}} \sum_{j=1}^{N_l} v_{lj}^2}{\prod_{l=1}^L N_l C_{Nl}}$$

$$= \frac{\sum_{l=1}^L (\prod_{l=1}^L N_l^{(l-1)} C_{Nl} - 1)}{\prod_{l=1}^L N_l^2 C_{Nl}} \sum_{j=1}^{N_l} v_{lj}^2 \frac{1}{\prod_{l=1}^L N_l C_{Nl}}$$

$$= \frac{1}{N} \sum_{i=1}^L \left(\prod_{j=1}^p N_j C_{N_j} \right) \frac{1}{N^2} \sum_{j=1}^p N_j^2 v_j^2 \frac{1}{\prod_{i=1}^L N_i C_{N_i}}$$

$$= \frac{1}{N^2} \sum_{i=1}^L \sum_{j=1}^p N_j^2 v_j^2 \quad (2.3.2.2.8.1)$$

Consider another statistic

$$t_2 = \frac{1}{N} \sum_{i=1}^L \frac{N_i(N_i - 1)}{N n_i(n_i - 1)} \sum_{j \neq k=1}^p v_j v_k \frac{p \bar{X}}{\sum_{i=1}^L N_i \sqrt{N_i + (p-1) \bar{X}}}$$

$$E(t_2) = \sum_0^1 \frac{1}{N} \sum_{i=1}^L \frac{N_i(N_i - 1)}{N^2 n_i(n_i - 1)} \sum_{j \neq k=1}^p v_j v_k \frac{1}{\prod_{i=1}^L N_i C_{N_i}}$$

$$= \frac{1}{N} \sum_{i=1}^L \left(\prod_{i=1}^L N_i - 2 C_{N_i - 2} \right) \frac{N_i(N_i - 1)}{N^2 n_i(n_i - 1)} \sum_{j \neq k=1}^p v_j v_k \frac{1}{\prod_{i=1}^L N_i C_{N_i}}$$

$$= \frac{1}{N^2} \sum_{i=1}^L \sum_{j \neq k=1}^p N_i v_j v_k \quad (2.3.2.2.8.2)$$

Finally, consider the statistic

$$t_3 = \frac{1}{N} \sum_{i \neq h=1}^L \frac{N_i N_h}{N n_i n_h} \sum_{j=1}^p \sum_{k=1}^p v_j v_k \frac{p \bar{X}}{\sum_{i=1}^L N_i \sqrt{N_i + (p-1) \bar{X}}}$$

$$E(t_3) = \sum_0^1 \frac{1}{N} \sum_{i \neq h=1}^L \frac{N_i N_h}{N^2 n_i n_h} \sum_{j=1}^p \sum_{k=1}^p v_j v_k \frac{1}{\prod_{i=1}^L N_i C_{N_i}}$$

$$= \frac{1}{N} \sum_{i \neq h=1}^L \left(N_i - 1 C_{N_i - 1} N_h - 1 C_{N_h - 1} \right) \sum_{j \neq k=1}^p N_i C_{N_i} \frac{N_i N_h}{N^2 n_i n_h}$$

$$= \sum_{j=1}^p \sum_{k=1}^p v_j v_k \frac{1}{\prod_{i=1}^L N_i C_{N_i}}$$

$$= \frac{1}{N^2} \sum_{i=1}^L \sum_{j=1}^{N_i} \sum_{k=1}^{N_j} \sum_{h=1}^{N_k} y_{ij} y_{hk} \quad (2.3.2.2.5.3)$$

Thus, \bar{Y}^2 , which can be expressed as

$$\bar{Y}^2 = \frac{1}{N^2} \left(\sum_{i=1}^L \sum_{j=1}^{N_i} y_{ij}^2 + \sum_{i=1}^L \sum_{j \neq k=1}^{N_i} y_{ij} y_{ik} + \sum_{i \neq j=1}^L \sum_{k=1}^{N_i} \sum_{h=1}^{N_j} y_{ij} y_{jh} \right)$$

can be estimated unbiasedly by $\hat{\bar{Y}}^2$, where $\hat{\bar{Y}}^2$ is given by

$$\hat{\bar{Y}}^2 = t_1 + t_2 + t_3$$

Hence, an unbiased estimate of variance of T'_{1st} is given by

$$\begin{aligned} \hat{V}(T'_{1st}) &= (T'_{1st})^2 - \frac{p \bar{X}}{N \sum_{i=1}^L N_i \left[\bar{N}_i + (p-1) \bar{X} \right]} \left[\sum_{i=1}^L \frac{N_i}{n_i} \sum_{j=1}^{N_i} y_{ij}^2 \right. \\ &\quad \left. + \sum_{i=1}^L \frac{N_i(N_i-1)}{n_i(n_i-1)} \sum_{j \neq k=1}^{N_i} y_{ij} y_{ik} + \sum_{i \neq j=1}^L \frac{N_i N_j}{n_i n_j} \sum_{k=1}^{N_i} \sum_{h=1}^{N_j} y_{ij} y_{jh} \right] \end{aligned} \quad (2.3.2.2.5.4)$$

2.3.2.3 Method III:

2.3.2.3.1 Sampling Scheme:

Select first unit from the whole population with probability proportional to the value of auxiliary character plus $(p-1)$ times the value of the mean of auxiliary character in the stratum to which the unit belongs. i.e., j -th unit of the i -th stratum is selected with probability proportional to $\left[z_{ij} + (p-1) \bar{X}_i \right]$, and select the remaining units of the sample from their respective strata with equal probability without replacement.

2.3.2.3.2 Estimator, its variance and its estimate of variance:

Under above sampling scheme the following estimator provides an unbiased.

estimate of the population mean

$$T_{1st}^u = p \bar{X} \frac{\sum_{i=1}^L N_i \bar{y}_i}{\sum_{i=1}^L N_i [\bar{X}_i + (p-1) \bar{X}_i]} \quad (2.3.2.3.2.1)$$

Proceeding in a manner similar to the one adopted in method II, we can easily find that T_{1st}^u is unbiased for population mean; its variance

is given by

$$V(T_{1st}^u) = \frac{p \bar{X}}{N} E_0' \frac{(\sum_{i=1}^L N_i \bar{y}_i)^2}{\sum_{i=1}^L N_i [\bar{X}_i + (p-1) \bar{X}_i]} \frac{1}{\sum_{i=1}^L N_i C_{ui}} - \bar{Y}^2 \quad (2.3.2.3.2.2)$$

and an unbiased estimate of variance is given by

$$\hat{V}(T_{1st}^u) = (T_{1st}^u)^2 \cdot \frac{p \bar{X}}{N \sum_{i=1}^L N_i [\bar{X}_i + (p-1) \bar{X}_i]} \left[\sum_{i=1}^L \frac{N_i}{n_i} \sum_{j=1}^E y_{ij}^2 + \sum_{i=1}^L \frac{N_i(N_i-1)}{n_i(n_i-1)} \sum_{j \neq h=1}^E y_{ij} y_{ih} + \sum_{i \neq h=1}^L \frac{N_i N_h}{n_i n_h} \sum_{j=1}^E \sum_{u=1}^E y_{ij} y_{hu} \right] \quad (2.3.2.3.2.3)$$

Note - I: The estimates of variances in all the above three methods are not necessarily non-negative.

Note - II: 'p' can be taken to be varying from stratum to stratum. In above cases 'p' is taken to be fixed for the sake of simplicity.

Note - III: The optimum value of 'p' can be obtained by minimising the variance expressions with respect to p. In practical situations p can be estimated by substituting the sample

values for their respective population values in the optimum value of p .

2.3.3 Extension to Two-stage or Double Sampling

T_{11} defined in the section 2.2.1 turns out to be an unbiased estimate of population mean under the following sampling scheme.

2.3.3.1 Sampling Scheme :

Draw a sample of size n' from the whole population with equal probability and without replacement. From this sample observe only the auxiliary character. Let K' be the mean of the auxiliary character in the sample. Draw a sub sample of size n from the sample of size n' such that the first unit is selected with probability proportional to the value of auxiliary character plus $(p - 1)$ times the mean of auxiliary character as estimated from the first sample, i. e., i -th unit from the first sample is selected with probability proportional to $x_i + (p - 1) K'$, and remaining $(n - 1)$ units with equal probability and without replacement. Observe both main and auxiliary characters from this sample.

2.3.3.2 Probability of obtaining a specified sample :

Probability of obtaining a sample $y_1 + y_2 + \dots + y_n$ of n units from a given larger sample of n' units is

$$P(y_1 + y_2 + \dots + y_n | y_1 + y_2 + \dots + y_{n'})$$

$$= \sum_{i=1}^{n'} \left[\text{Probability of selecting } y_i \text{ at first draw} \right] \times \left[\text{Probability of selecting remaining } n-1 \text{ units, given that } y_i \text{ is selected at first draw} \right] | y_1 + y_2 + \dots + y_{n'}$$

$$= \sum_{i=1}^n \frac{x_i + (p-1)\bar{x}'}{p n' \bar{x}'} \cdot \frac{1}{n^{n-1} C_{n-1}}$$

$$= \frac{\bar{x} + (p-1)\bar{x}'}{p \bar{x}'} \cdot \frac{1}{n' C_n}$$

Therefore, the unconditional probability of obtaining the sample $y_1 \cdot y_2 \dots y_n$ is equal to

$$\frac{\bar{x} + (p-1)\bar{x}'}{p \bar{x}'} \cdot \frac{1}{n' C_n} \cdot \frac{1}{N_{C_n}} \quad (2.3.3.1)$$

2.3.3.3 Expected value of T_{1d} :

$$T_{1d} = p \bar{x}' \cdot \frac{\bar{y}}{\bar{x} + (p-1)\bar{x}'}$$

$$E(T_{1d}) = \sum' \sum'' p \bar{x}' \cdot \frac{\bar{y}}{\bar{x} + (p-1)\bar{x}'} \cdot \frac{\bar{x} + (p-1)\bar{x}'}{p \bar{x}'} \cdot \frac{1}{n' C_n} \cdot \frac{1}{N_{C_n}}$$

$$= \frac{1}{n' C_n} \cdot \frac{1}{N_{C_n}} \cdot \sum' \sum'' \bar{y}$$

$$= \bar{y} \quad (2.3.3.3.1)$$

where \sum'' denotes summation over all the $n' C_n$ samples of size n from the first sample of size n' and, \sum' denotes summation over all N_{C_n} samples of size n' from the population.

2.3.3.4 Variance of T_{1d} :

$$V(T_{1d}) = E(T_{1d})^2 - [E(T_{1d})]^2$$

$$= \sum' \sum'' \left[\frac{p \bar{x}' \bar{y}}{\bar{x} + (p-1)\bar{x}'} \right]^2 \cdot \frac{\bar{x} + (p-1)\bar{x}'}{p \bar{x}'} \cdot \frac{1}{n' C_n} \cdot \frac{1}{N_{C_n}} - \bar{y}^2$$

$$= \frac{1}{n' C_n} \cdot \frac{1}{N_{C_n}} \cdot \sum' p \bar{x}' \sum'' \frac{\bar{y}^2}{\bar{x} + (p-1)\bar{x}'} - \bar{y}^2 \quad (2.3.3.4.1)$$

2.3.3.3 Estimate of variance:

Consider the statistic

$$\begin{aligned}
 & \sum_{i=1}^n y_i^2 \frac{p \bar{x}'}{N n \sqrt{\bar{x} + (p-1) \bar{x}'}} \\
 & E \left[\sum_{i=1}^n y_i^2 \frac{p \bar{x}'}{N n \{ \bar{x} + (p-1) \bar{x}' \}} \right] \\
 & = E' E'' \sum_{i=1}^n y_i^2 \frac{p \bar{x}'}{N n \sqrt{\bar{x} + (p-1) \bar{x}'}} \frac{\bar{x} + (p-1) \bar{x}'}{p \bar{x}'} \frac{1}{n C_n} \frac{1}{N C_n} \\
 & = E' E'' \sum_{i=1}^n y_i^2 \frac{1}{N n} \frac{1}{n C_n} \frac{1}{N C_n} \\
 & = E' \sum_{i=1}^{n-1} C_{n-1} \sum_{i=1}^n y_i^2 \frac{1}{N n} \frac{1}{n C_n} \frac{1}{N C_n} \\
 & = E' \sum_{i=1}^n y_i^2 \frac{1}{N n} \frac{1}{N C_n} \\
 & = \sum_{i=1}^n \sum_{j=1}^{n-1} C_{n-1} y_i^2 \frac{1}{N n} \frac{1}{N C_n} \\
 & = \frac{1}{N^2} \sum_{i=1}^n y_i^2
 \end{aligned}$$

Consider another statistic

$$\begin{aligned}
 & \sum_{i \neq j=1}^n y_i y_j \frac{(N-1)}{N n (n-1)} \frac{p \bar{x}'}{\bar{x} + (p-1) \bar{x}'}, \\
 & E \left[\sum_{i \neq j=1}^n y_i y_j \frac{(N-1)}{N n (n-1)} \frac{p \bar{x}'}{\bar{x} + (p-1) \bar{x}'} \right] \\
 & = E' E'' \sum_{i \neq j=1}^n y_i y_j \frac{(N-1)}{N n (n-1)} \frac{p \bar{x}'}{\bar{x} + (p-1) \bar{x}'}
 \end{aligned}$$

$$= \sum' \sum'' \frac{N}{(p-j)h} y_i y_j \frac{(N-1)}{Nn(n-1)} \frac{1}{n C_n} \frac{1}{n C_n}$$

$$= \sum' n' \cdot \frac{1}{n C_n} \sum'' \frac{N}{(p-j)h} y_i y_j \frac{(N-1)}{Nn(n-1)} \frac{1}{n C_n} \frac{1}{n C_n}$$

$$= \sum' \frac{N}{(p-j)h} y_i y_j \frac{(N-1)}{Nn'(n'-1)} \frac{1}{n C_n}$$

$$= \sum'' \frac{N-1}{n C_n} \sum' y_i y_j \frac{(N-1)}{Nn'(n'-1)} \frac{1}{n C_n}$$

$$= \frac{1}{N^2} \sum'' y_i y_j$$

That

$$V^2 = \frac{1}{N^2} \sum_{i=1}^N y_i^2 + \frac{1}{N^2} \sum'' y_i y_j$$

can be estimated unbiasedly by

$$\begin{aligned} & \frac{N}{n} \sum_{i=1}^n y_i^2 \frac{p\bar{N}'}{Nn \sqrt{\bar{N} + (p-1)\bar{N}'}} + \frac{N}{(p-j)h} y_i y_j \frac{(N-1)}{Nn(n-1)} \frac{p\bar{N}'}{\bar{N} + (p-1)\bar{N}'} \\ &= \frac{p\bar{N}'}{Nn \sqrt{\bar{N} + (p-1)\bar{N}'}} \left[\sum_{i=1}^N y_i^2 + \frac{N-1}{n-1} \sum'' y_i y_j \right] \end{aligned}$$

Hence, an unbiased estimate of variance of T_{10} is given by

$$\hat{V}(T_{10}) = (T_{10})^2 \cdot \frac{p\bar{N}'}{Nn \sqrt{\bar{N} + (p-1)\bar{N}'}} \left[\sum_{i=1}^N y_i^2 + \frac{N-1}{n-1} \sum'' y_i y_j \right] \quad (2.5.3.5.1)$$

CASE OF NEGATIVELY CORRELATED CHARACTERS

2.4.1 The proposed estimator T_1 can be used also if the auxiliary character is negatively correlated with the character under study. In such a case, p is required to undergo a slight modification, viz., p is required to be negative. It is more useful than the conventional estimator used under such a situation, viz., the product estimator, in the sense that it can reduce bias and increase efficiency, if p is suitably chosen. The conditions on the value of p are that: p should be such that (1) $\left| \frac{\bar{X} - \bar{Y}}{p \bar{X}} \right|$ is less than unity, so that the expansion given in section 2.1.3 remains valid and, (2) $\bar{X} + (p-1)\bar{X}$ is a negative quantity, so that T_1 may remain positive.

Let $p = -q$, where q is a positive quantity, then T_1 reduces to T_1' given by

$$\begin{aligned} T_1' &= -q \bar{X} \frac{\bar{Y}}{\bar{X} + (-q-1)\bar{X}} \\ &= q \bar{X} \frac{\bar{Y}}{(q+1)\bar{X} - \bar{X}} \end{aligned} \quad (2.4.1.1)$$

Using (2.1.3.1) and (2.1.5.1), the expected value and the variance of T_1' , to $O(n^{-1})$, are as follows:

$$E(T_1') = \bar{Y} \left[1 + \left(\frac{1}{n} - \frac{1}{N} \right) \left(\frac{B_x^2}{q \bar{X}^2} + \frac{\rho_{xy}}{q \bar{X} \bar{Y}} \right) \right] \quad (2.4.1.2)$$

and

$$V(T_1') = \left(\frac{1}{n} - \frac{1}{N} \right) \left(B_y^2 + \frac{R^2 S_x^2}{q^2} + \frac{2RS_{xy}}{q} \right) \quad (2.4.1.3)$$

Also, the expected value and the variance of $\bar{y}_p (= \frac{\bar{X}\bar{Y}}{\bar{X}})$ are given

by

$$E(\bar{Y}_P) = \bar{Y} \left[1 + \left(\frac{1}{n} - \frac{1}{N} \right) \frac{S_{xy}}{R \bar{Y}} \right]$$

and

$$V(\bar{Y}_P) = \left(\frac{1}{n} - \frac{1}{N} \right) (S_y^2 + R^2 S_x^2 + 2R S_{xy})$$

2.4.2 Comparison regarding biases in T_1' and \bar{Y}_P :

$$\text{Bias in } T_1' = \bar{Y} \left(\frac{1}{n} - \frac{1}{N} \right) \left(\frac{S_x^2}{q^2 \bar{X}^2} + \frac{S_{xy}}{q \bar{X} \bar{Y}} \right)$$

$$\text{Bias in } \bar{Y}_P = \bar{Y} \left(\frac{1}{n} - \frac{1}{N} \right) \frac{S_{xy}}{R \bar{Y}}$$

It is obvious that \bar{Y}_P will always be negatively biased whereas T_1' can be negatively as well as positively biased. So, the following two cases arise :

Case 1 : When T_1' is negatively biased

T_1' will be less biased than \bar{Y}_P if

$$\bar{Y} \left(\frac{1}{n} - \frac{1}{N} \right) \left(\frac{S_x^2}{q^2 \bar{X}^2} + \frac{S_{xy}}{q \bar{X} \bar{Y}} \right) > \bar{Y} \left(\frac{1}{n} - \frac{1}{N} \right) \frac{S_{xy}}{R \bar{Y}}$$

$$\text{i.e. } \frac{S_x^2}{q^2 \bar{X}^2} > \frac{S_{xy}}{R \bar{Y}} \left(1 - \frac{1}{q} \right)$$

$$\text{i.e. } p \frac{C_y}{C_x} < \frac{1}{q(q-1)} \quad (2.4.2.1)$$

This inequality will always hold when $q > 1$, i.e., $p < -1$, since in such a case the right hand side will be positive while left hand side is negative because p is negative.

Case II : When T_1' is positively biased

T_1' will be less biased than \bar{Y}_P if

$$V \left(\frac{1}{n} - \frac{1}{N} \right) \left(\frac{S_x^2}{q^2 N^2} + \frac{S_{xy}}{q N Y} \right) < - V \left(\frac{1}{n} - \frac{1}{N} \right) \frac{S_{xy}}{N Y}$$

i.e. $\frac{S_x^2}{q^2 N^2} < - \frac{S_{xy}}{N Y} \left(1 + \frac{1}{q} \right)$

i.e. $\rho \frac{C_y}{C_x} < - \frac{1}{q(q+1)} \quad (2.4.2.2)$

and \bar{Y}_P will be less biased, otherwise.

2.4.3 Comparison of efficiencies of T_1' and \bar{Y}_P :

T_1' will be more efficient than \bar{Y}_P if

$$V(T_1') < V(\bar{Y}_P)$$

i.e. $\left(\frac{1}{n} - \frac{1}{N} \right) \left(S_y^2 + \frac{R^2 S_x^2}{q^2} + \frac{2 R S_{xy}}{q} \right) < \left(\frac{1}{n} - \frac{1}{N} \right) \left(S_y^2 + R^2 S_x^2 + 2 R S_{xy} \right)$

i.e. $R^2 S_x^2 \left(\frac{1}{q^2} - 1 \right) < - 2 R S_{xy} \left(\frac{1}{q} - 1 \right) \quad (2.4.3.1)$

This condition is equivalent to the following two conditions :

$| \rho | \frac{C_y}{C_x} > \frac{q}{q+1} \quad \text{if } q > 1 \text{ (i.e., } \rho < -1)$
(2.4.3.1a)

$| \rho | \frac{C_y}{C_x} < \frac{q}{q+1} \quad \text{if } q < 1 \text{ (i.e., } -1 < \rho < 0)$
(2.4.3.1b)

N. B. Inequalities (2.4.2.1), (2.4.2.2), (2.4.3.1a) and (2.4.3.1b)

Suggest that stronger the correlation between the auxiliary and the main characters lower should be the value of q .

2.4.4 A special case (q = 1) :

The preceding discussion excludes the case when q is unity. In this case the following results will be obtained ;

$$E(T_1') = \bar{Y} \left[1 + \left(\frac{1}{n} - \frac{1}{N} \right) \left(\frac{S_x^2}{\bar{X}^2} + \frac{S_{xy}}{\bar{X}\bar{Y}} \right) \right]$$

Obviously, T_1' will be less biased than \bar{Y}_D whenever T_1' will have a negative bias and T_1' will be less biased than \bar{Y}_D when T_1' is positively biased, if $\rho \frac{C_y}{C_x} < -\frac{1}{2}$. It is interesting to notice that the condition for \bar{Y}_D to be more efficient than \bar{y} , the simple mean, is also the same (i.e. $\rho \frac{C_y}{C_x} < -\frac{1}{2}$). Hence, whenever it is advantageous to use \bar{Y}_D , it will be more advantageous to use T_1' with $q = 1$, since apart from gain in efficiency, compared to \bar{y} , there will be a reduction in bias too. Moreover, the case when $q = 1$ is the simplest case from the computational point of view which makes it more worthwhile to use.

2.4.5 Optimum value of q :

Differentiating the variance expression (2.4.1.3) with respect to q and equating the result to zero, we get the optimum value of q as follows :

$$q = -\frac{1}{\rho} \frac{C_x}{C_y} \quad (2.4.5.1)$$

To choose q in practice, the sample values of ρ , C_x and C_y may be used to serve as a guide line.

2.4.6 Estimate of variance :

A consistent estimator of variance of T_1 can be given by

$$\hat{V}(T_1) = \left(\frac{1}{n} - \frac{1}{N} \right) \left(\sigma_y^2 + \frac{\hat{\Sigma}^2 \sigma_x^2}{\sigma^2} + \frac{2 \hat{\Sigma} \sigma_{xy}}{\sigma} \right) \quad (2.4.6.1)$$

ANNEXURE

Earlier we noticed that T_1 is more efficient than usual ratio estimator \bar{y}_R

if

$$p \frac{C_y}{C_x} < \frac{p+1}{2p}$$

In case $p C_y / C_x$ exceeds $(p+1) / 2p$, we can use a more general form of T_1 , say T_{1A} , given by

$$T_{1A} = \frac{p\bar{X}}{q\bar{X} + (p-q)\bar{X}} \bar{Y}$$

Its expected value and variance are given by

$$E(T_{1A}) = \bar{Y} \left[1 + \left(\frac{1}{q} - \frac{1}{N} \right) \left(\frac{q^2 S_x^2}{p^2 \bar{X}^2} - \frac{q^2 S_{xy}}{p \bar{X} \bar{Y}} \right) \right]$$

and

$$V(T_{1A}) = \bar{Y}^2 \left(\frac{1}{q} - \frac{1}{N} \right) \left(\frac{S_y^2}{\bar{Y}^2} + \frac{q^2 S_x^2}{p^2 \bar{X}^2} - \frac{2q S_{xy}}{p \bar{X} \bar{Y}} \right)$$

T_{1A} will be more efficient than \bar{y}_R if

$$p \frac{C_y}{C_x} < \frac{p+q}{2p}$$

Thus, by taking $q > 1$, we can widen our range of gaining efficiency (but, it will, of course, lessen on lower correlation side).

A still more general form of T_1 , say T'_{1A} , that can be used in such a case is

$$T'_{1A} = \frac{a\bar{X} + b\bar{Y}}{c\bar{X} + d\bar{Y}} \bar{Y}$$

where $a+b = c+d = p$ (say).

Its expected value and variance are as follows :

$$E(T'_{1A}) = \bar{Y} \left[1 + \left(\frac{1}{a} - \frac{1}{b} \right) \left\{ \frac{(c-a)^2}{p^2} \frac{S_x^2}{K^2} + \frac{(c-a)}{p} \frac{S_{xy}}{K\bar{Y}} \right\} \right]$$

and

$$V(T'_{1A}) = \bar{Y}^2 \left(\frac{1}{a} - \frac{1}{b} \right) \left[\frac{S_y^2}{p^2} + \frac{(c-a)^2}{p^2} \frac{S_x^2}{K^2} + \frac{2(c-a)S_{xy}}{pK\bar{Y}} \right]$$

The expectation and variance terms show that c must be greater than a .

Again, T'_{1A} will be more efficient than \bar{y}_R if

$$p \frac{C_x}{C_y} < \frac{p+c-a}{2p}$$

i.e. $< \frac{b+c}{2p}$ (since $p = a+b$)

Example: Suppose $p = 2$ and $C_x = C_y$, then for T_1 to be more efficient than \bar{y}_R , $p < 0.75$; for T_{1A} , $p < 0.74$ (c chosen to be 1.75); and for T'_{1A} , $p < 0.875$ (b and c chosen to be equal to 1.75 each).

Thus, if correlation is very high (or variability in y is very high or variability in x is low), instead of T_1 one can use T_{1A} or T'_{1A} .

Even the above mentioned estimators fail when $p C_y / C_x$ exceeds unity. This is the case when the variability in the auxiliary character is low. More explicitly, when $C_x < p C_y$. In such a case T_1 can be used by taking p less than unity and choosing p in such a manner that $\left| \frac{\bar{x} - \bar{X}}{p \bar{X}} \right| < 1$; this is not unrealistic because of the two reasons (1) the

optimum value of p ($= C_x / p C_y$) in such a case will be less than unity and, (ii) the variability in x , the auxiliary character, is low.

CHAPTER - III

THE SECOND ESTIMATOR

SINGLE PHASE SAMPLING

3.1.1 The second proposed estimator of population mean is

$$T_2 = \frac{b\bar{y} + c\bar{X}}{b\bar{y} + c\bar{X}} \bar{y} \quad (b > 0, c > 0) \quad (3.1.1)$$

This estimator is also, like the first estimator, a consistent estimator with b and c arbitrary within certain limits. Later it will be found that both b and c are not arbitrary but are interdependent to yield optimum results, thus, leaving only one of them arbitrary. Here it may be noted that T_2 can be reduced to a form involving only one constant, by dividing the numerator and denominator by either b or c . But, the form given above is retained as it is more general.

3.1.2 Expected value of T_2 :

$$\begin{aligned} T_2 &= \frac{b\bar{y} + c\bar{X}}{b\bar{y} + c\bar{X}} \bar{y} \\ &= \frac{[b\bar{y} + c\bar{X} + b(\bar{y} - \bar{Y})]}{[b\bar{y} + c\bar{X} + b(\bar{y} - \bar{Y}) + c(\bar{X} - \bar{X})]} [\bar{y} + (\bar{y} - \bar{Y})] \\ &= \bar{y} \left[1 + \frac{b(\bar{y} - \bar{Y})}{b\bar{y} + c\bar{X}} \right] \left[1 + \frac{\bar{y} - \bar{Y}}{\bar{y}} \right] \left[1 + \frac{b(\bar{y} - \bar{Y})}{b\bar{y} + c\bar{X}} + \frac{c(\bar{X} - \bar{X})}{b\bar{y} + c\bar{X}} \right]^{-1} \end{aligned}$$

Assuming $\left| \frac{b(\bar{y} - \bar{Y}) + c(\bar{X} - \bar{X})}{b\bar{y} + c\bar{X}} \right| < 1$, to make the expansion of last factor

on right hand side ; expanding and simplifying and taking expectation we get

$$E(T_2) = \bar{Y} \sqrt{1 + \left(\frac{1}{n} - \frac{1}{N}\right) \left\{ \frac{c^2 S_X^2}{(b\bar{Y} + c\bar{X})^2} + \frac{bc S_{XY}}{(b\bar{Y} + c\bar{X})^2} - \frac{c S_{XY}}{\bar{Y}(b\bar{Y} + c\bar{X})} \right\}} \sqrt{}$$

to $O(n^{-1})$

$$= \bar{Y} \sqrt{1 + \left(\frac{1}{n} - \frac{1}{N}\right) \frac{c^2}{(b\bar{Y} + c\bar{X})^2} \left(S_X^2 - \frac{S_{XY}}{\bar{Y}} \right) \sqrt{}} \quad (3.1.2.1)$$

3.1.3 Comparison of biases in T_2 and \bar{Y}_R :

The expected value of \bar{Y}_R , to $O(n^{-1})$, can be written in the following form

$$E(\bar{Y}_R) = \bar{Y} \sqrt{1 + \left(\frac{1}{n} - \frac{1}{N}\right) \frac{1}{N^2} \left(S_X^2 - \frac{S_{XY}}{\bar{Y}} \right) \sqrt{}}$$

We, therefore, have

$$\text{Bias in } T_2 = \left(\frac{1}{n} - \frac{1}{N}\right) \frac{c^2 \bar{Y}}{(b\bar{Y} + c\bar{X})^2} \left(S_X^2 - \frac{S_{XY}}{\bar{Y}} \right)$$

and

$$\text{Bias in } \bar{Y}_R = \left(\frac{1}{n} - \frac{1}{N}\right) \frac{\bar{Y}}{N^2} \left(S_X^2 - \frac{S_{XY}}{\bar{Y}} \right)$$

$$\therefore \frac{\text{Bias in } T_2}{\text{Bias in } \bar{Y}_R} = \frac{c^2}{(b\bar{Y} + c\bar{X})^2} / \frac{1}{N^2}$$

$$= \frac{c^2}{(bN + c)^2}$$

Since right hand side is always less than unity, the bias in T_2 will always be less than the bias in \bar{Y}_R , except when $b = 0$, in which case T_2 reduces to \bar{Y}_R .

3.1.4 Variance of T_2 :

$$V(T_2) = E \left[T_2 - E(T_2) \right]^2$$

$$\begin{aligned}
 &= E \left[\bar{Y} \left(\frac{\bar{Y} - \bar{Y}}{\bar{Y}} - \frac{c(\bar{X} - \bar{X})}{b\bar{Y} + c\bar{X}} \dots\dots\dots \right) \right]^2 \\
 &= \bar{Y}^2 \left(\frac{1}{\bar{Y}^2} - \frac{1}{\bar{X}^2} \right) \left[\frac{S_Y^2}{\bar{Y}^2} + \frac{c^2 S_X^2}{(b\bar{Y} + c\bar{X})^2} - \frac{2c S_{XY}}{\bar{Y}(b\bar{Y} + c\bar{X})} \right] \\
 &\hspace{15em} \text{to } O(n^{-1}) \\
 &= \left(\frac{1}{\bar{Y}^2} - \frac{1}{\bar{X}^2} \right) \left[S_Y^2 + \frac{c^2 R^2 S_X^2}{(bR + c)^2} - \frac{2c R S_{XY}}{(bR + c)} \right] \quad (3.1.4.1)
 \end{aligned}$$

3.1.5 Comparison of efficiencies :

It can be easily seen from the variance expressions of T_1 and T_2 , i.e., from (2.1.5.1) and (3.1.4.1), that if $p = (bR + c)/c$, the variance expression of T_1 is the same as that of T_2 . Hence, all the comparisons regarding efficiencies related to T_1 will be valid in the case of T_2 also, if p is replaced by $(bR + c)/c$. Thus, we have the following results :

(a) T_2 will be more efficient than simple mean \bar{Y} if

$$p \frac{C_Y}{C_X} > \frac{1}{2} \frac{c}{bR + c} \quad (3.1.5.1)$$

(b) T_2 will be more efficient than usual ratio estimator \bar{Y}_R if

$$p \frac{C_Y}{C_X} < \frac{bR + 2c}{2(bR + c)} \quad (3.1.5.2)$$

(c) T_2 can never be more efficient than the regression estimator \bar{Y}_{LR} .

However, T_2 and \bar{Y}_{LR} will be equally efficient if

$$\frac{bR + c}{c} = \frac{1}{p} R \frac{S_X}{S_Y}$$

$$\text{or } \frac{b}{c} = \frac{1}{\rho} \frac{S_x}{S_y} = \frac{1}{R} \quad (3.1.5.3)$$

Again, as observed in the case of T_1 , this value of $\frac{b}{c}$ is the same as the optimum value of $\frac{b}{c}$.

(d) T_2 will be more efficient than the bivariate ratio estimator under the previously stated, $\sqrt{\quad}$ 2.1.5 - (d) $\sqrt{\quad}$ circumstances if

$$\rho \frac{C_y}{C_x} < \frac{1}{4} \frac{\left(\frac{bR+c}{c}\right)^2 (1+\rho^2) - 2}{\frac{bR+c}{c} \left(\frac{bR+c}{c} - 1\right)}$$

$$\text{i.e. } < \frac{1}{4} \frac{(bR+c)^2 (1+\rho^2) - 2}{bR(bR+c)} \quad (3.1.5.4)$$

3.1.6 Optimum set of values of b and c :

Differentiating the variance expression (3.1.4.1) with respect to either b or c or $\frac{b}{c}$, equating the result to zero, we obtain that a set of values of b and c will be optimum for which

$$\frac{b}{c} = \frac{1}{\rho} \frac{S_x}{S_y} = \frac{1}{R} \quad (3.1.6.1)$$

Thus, both b and c are not arbitrary to obtain optimum results; having chosen either of them the another has to be determined from (3.1.6.1).

In practice, one can assign any value to either b or c and determine the other by substituting the sample values for the true values in (3.1.6.1).

The optimum variance of T_2 will obviously be equal to that of the regression estimator i.e. $\left(\frac{1}{R} - \frac{1}{R}\right) S_y^2 (1 - \rho^2)$.

3.1.7 T_2 and the product estimator :

Like T_1 , T_2 can also be used if the variables are negatively correlated as is clear from its variance expression (3.1.4.1), by choosing a negative value for c in such a way that both $|c\bar{X}|$ as well as $|c\bar{Y}|$ are either greater or smaller than $b\bar{Y}$ so that $\frac{b\bar{Y} + c\bar{X}}{b\bar{Y} + c\bar{Y}}$ may remain positive; and for obvious reasons b will be positive. Alternatively, one could choose b to be negative such that $|b\bar{Y}|$ is either greater than or smaller than both of $c\bar{X}$ and $c\bar{Y}$. But, choosing b or c is not as easy as choosing p in T_1 because it can be seen from the variance expression of T_2 that if c is negative, it must satisfy $(bR + c) > 0$ (for $\frac{cRS_{xy}}{bR + c}$ to be negative), and if b is negative, it must satisfy $(bR + c) < 0$ (for the same reason). In practice, one can use the sample ratio $(= \frac{\bar{Y}}{\bar{X}})$ as a guide to decide upon the magnitude of b or c , whichever is chosen, arbitrarily, to be negative. In the following discussion we shall assume that c is negative ($= -d$, say).

3.1.7.1 Comparison of biases :

$$\text{Bias in } \bar{Y}_P = \left(\frac{1}{n} - \frac{1}{N}\right) \frac{S_{xy}}{\bar{X}} = \left(\frac{1}{n} - \frac{1}{N}\right) \frac{\bar{Y}}{\bar{X}R} \frac{S_{xy}}{R}$$

$$\text{Bias in } T_2 = \left(\frac{1}{n} - \frac{1}{N}\right) \frac{d^2\bar{Y}}{(b\bar{Y} - d\bar{X})^2} \left(S_R^2 - \frac{S_{xy}}{R}\right)$$

Obviously, \bar{Y}_P is negatively biased whereas T_2 is positively biased.

Now, T_1 will be less biased than \bar{Y}_P if

$$\left(\frac{1}{n} - \frac{1}{N}\right) \frac{\bar{Y}}{\bar{X}^2} \left| \frac{S_{xy}}{R} \right| > \left(\frac{1}{n} - \frac{1}{N}\right) \frac{d^2 \bar{Y}}{(b\bar{Y} - d\bar{X})^2} \left(S_x^2 - \frac{S_{xy}}{R} \right)$$

$$\text{i.e.} \quad -\left(\frac{1}{n} - \frac{1}{N}\right) \frac{\bar{Y}}{\bar{X}^2} \frac{S_{xy}}{R} > \left(\frac{1}{n} - \frac{1}{N}\right) \frac{d^2 \bar{Y}}{(b\bar{Y} - d\bar{X})^2} \left(S_x^2 - \frac{S_{xy}}{R} \right)$$

$$\text{i.e.} \quad -\frac{S_{xy}}{R} \left[1 - \frac{d^2}{(bR - d)^2} \right] > \frac{d^2}{(bR - d)^2} S_x^2$$

$$\text{i.e.} \quad S_{xy} (b^2 R - 2bd) > d^2 S_x^2$$

This condition is equivalent to the following two conditions :

$$\left| \rho \frac{S_y}{S_x} \right| > \frac{d^2}{(b^2 R - 2bd)}$$

and

$$\left| \rho \frac{S_y}{S_x} \right| > \frac{d^2}{(2bd - b^2 R)}$$

according as $(b^2 R - 2bd)$ is positive or negative. T_2 will be more biased, otherwise, except when $\left| \rho \frac{S_y}{S_x} \right| = \frac{d^2}{|b^2 R - 2bd|}$, in which case both T_2 and \bar{Y}_P will be equally biased.

3.1.7.2 Comparison of efficiencies :

$$V(\bar{Y}_P) = \left(\frac{1}{n} - \frac{1}{N}\right) (S_y^2 + R^2 S_x^2 + 2R S_{xy})$$

$$V(T_2) = \left(\frac{1}{n} - \frac{1}{N}\right) \left(S_y^2 + \frac{d^2 R^2}{(bR - d)^2} S_x^2 + \frac{2dR}{(bR - d)} S_{xy} \right)$$

T_2 will be more efficient than \bar{Y}_P if

$$R^2 S_x^2 + 2R S_{xy} > \frac{d^2 R^2}{(bR - d)^2} S_x^2 + \frac{2dR}{(bR - d)} S_{xy}$$

$$\text{i.e. } -2R\hat{\sigma}_{xy} \left[\frac{d}{bR-d} - 1 \right] > R^2 \hat{\sigma}_x^2 \left[\frac{d^2}{(bR-d)^2} - 1 \right]$$

This condition can be resolved into the following two conditions :

$$|\rho| \frac{C_y}{C_x} > \frac{1}{2} \frac{bR}{bR-d} \quad \text{if } \frac{d}{bR-d} > 1$$

$$|\rho| \frac{C_y}{C_x} < \frac{1}{2} \frac{bR}{bR-d} \quad \text{if } \frac{d}{bR-d} < 1$$

3.1.8 Estimate of variance of T_2 :

A consistent estimator of variance of T_2 is given by

$$\hat{V}(T_2) = \left(\frac{1}{n} - \frac{1}{N} \right) \left[\hat{\sigma}_y^2 + \frac{\hat{\sigma}_x^2 \hat{R}^2 \hat{\sigma}_x^2}{(b\hat{R}+c)^2} - \frac{2c\hat{R}\hat{\sigma}_{xy}}{(b\hat{R}+c)} \right]$$

(3.1.8.1)

TWO PHASE SAMPLING

3.2.1 In two phase sampling (described in section - 2.2.1) T_2 takes the following form:

$$T_{2d} = \frac{b\bar{y} + c\bar{x}'}{b\bar{y} + c\bar{x}} \bar{y} \quad (3.2.1.1)$$

3.2.2 Expected value of T_{2d} :

$$E(T_{2d}) = \frac{b\bar{y} + c\bar{x}'}{b\bar{y} + c\bar{x}} \bar{y}$$

$$= \frac{\bar{y} \left[1 + \frac{b(\bar{y} - \bar{Y})}{b\bar{y} + c\bar{x}} + \frac{c(\bar{x}' - \bar{X})}{b\bar{y} + c\bar{x}} \right]}{\bar{y} \left[1 + \frac{b(\bar{y} - \bar{Y})}{b\bar{y} + c\bar{x}} + \frac{c(\bar{x}' - \bar{X})}{b\bar{y} + c\bar{x}} \right]} \left[1 + \frac{\bar{y} - \bar{Y}}{\bar{y}} \right]$$

Simplifying the above expression and using the results of section 2.2.2 we finally obtain, to $O(n^{-1})$,

$$E(T_{2d}) = \bar{y} \left[1 + \left(\frac{1}{n} - \frac{1}{n'} \right) \frac{c^2}{(b\bar{y} + c\bar{x})^2} \left(S_x^2 - \frac{S_{xy}}{R} \right) \right] \quad (3.2.2.1)$$

3.2.3 Comparison of biases in T_{2d} and \bar{y}_{Rd} :

$$E(\bar{y}_{Rd}) = \bar{y} \left[1 + \left(\frac{1}{n} - \frac{1}{n'} \right) \frac{1}{R^2} \left(S_x^2 - \frac{S_{xy}}{R} \right) \right]$$

Thus, we see once again the bias is reduced $c^2 / (bR + c)^2$ times, by making use of T_{2d} , as was the case in single phase sampling.

3.2.4 Variance of T_{2d} :

Variance of T_{2d} can be obtained easily in a manner similar to the one

adopted for T_{1d} . The variance of T_{2d} is given by

$$V(T_{2d}) = \left(\frac{1}{n} - \frac{1}{n'}\right) \left[S_y^2 + \frac{c^2 R^2 S_x^2}{(bR+c)^2} - \frac{2cRS_{xy}}{(bR+c)} \right] + \left(\frac{1}{n'} - \frac{1}{N}\right) S_y^2 \quad (3.2.4.1)$$

3.2.5 Comparison of efficiencies and, optimum values of n and n' under different situations can be worked out simply by replacing p by $(bR+c)/c$ in the results obtained for T_{1d} in sections 2.2.5 and 2.2.6.

3.2.6 Estimate of variance :

A consistent estimator of the variance of T_{2d} is given by

$$\hat{V}(T_{2d}) = \left(\frac{1}{n} - \frac{1}{n'}\right) \left[\hat{S}_y^2 + \frac{c^2 \hat{R}^2 \hat{S}_x^2}{(b\hat{R}+c)^2} - \frac{2c\hat{R}\hat{S}_{xy}}{(b\hat{R}+c)} \right] + \left(\frac{1}{n'} - \frac{1}{N}\right) \hat{S}_y^2 \quad (3.2.6.1)$$

CHAPTER - IV

EMPERICAL INVESTIGATIONS ON THE PROPOSED ESTIMATOR

The proposed estimator T_1 was tested on a few natural populations. All the villages falling under the three Police Stations, namely, Bunki ; Mahulpada and ; Tilayatpali, of Sundergarh District in Orissa State were considered as the units constituting the populations. For each of these villages, information on six characters, viz., area of village ; number of house holds ; total population of village ; number of workers* of all types ; number of cultivators and ; number of agricultural labourers, were obtained from District Census Handbook (1971).

Three characters, namely, number of workers of all types ; number of cultivators and ; number of agricultural labourers, are considered as characters under study and T_1 is tested for each of them. The character that is considered as main character at one instance serves as an auxiliary character at other instances, i. e., when some other character is considered as character under study.

Tables for each population depicting mean, variance etc. for each character are presented. The relative efficiency of the proposed

* Worker - A worker is a person whose main activity is participation in any economically productive work by his physical or mental activity. (This category includes cultivators and agricultural labourers too).

estimator vis-a-vis that of simple mean and usual ratio estimator is also tabulated for various characters. As pointed out earlier the efficiency of T_1 depends upon the value of ρ , so only those values of ρ are considered that are near the optimum value. Here it may be noted that the tables have not been presented for the characters for which $\rho C_y / C_x$ is very high because in those cases it would be better to use the usual ratio estimator as even the optimum variance of T_1 will not be much different from that of usual ratio estimator.

Explanation of tables :

The tables under the head Table for Mean, Variance, etc., will maintain the following convention :

Characters are numbered from 1 to 6 such that '1' stands for area under village ; '2' for number of households ; '3' for total population of village ; '4' for number of workers of all types ; '5' for number of cultivators and ; '6' for agricultural labourers ; but when considered as main character , '1' stands for number of workers of all types ; '2' for number of cultivators and ; '3' for number of agricultural labourers. Thus , ρ_{11} means the coefficient of correlation between number of workers of all types and 1-th auxiliary character ($i = 1, \dots, 6$). Similarly , $\rho_{11} C_{y1} / C_{x1}$ denotes the ratio of the product of ρ_{11} and coefficient of variation of number of workers of all types to the coefficient of variation of 1-th auxiliary character. Further , when a character which is one of the three main characters , is being used as an auxiliary character it will retain its usual number.

Population - I
(Bundi Police Station)

Total number of villages - 39

Table for Mean, Variance, SE.

Character	1	2	3	4	5	6
Mean	933.94871	56.666667	318.66667	97.651538	64.923076	19.256410
S^2	517037.32	1776.0702	74352.120	5956.8870	3022.1781	278.87989
C_1	0.7699053	0.7637078	0.8557720	0.7919110	0.8467611	1.0946020
S_{11}	45274.790	3158.1579	14725.159	-	3961.3320	974.51010
P_{11}	0.8153039	0.9709438	0.6996267	-	0.9336277	0.7560806
$P_{11} C_{y1}/C_{x1}$	0.8391202	1.0338751	0.6475450	-	0.8731506	0.5470010
S_{21}	32124.8650	1991.3684	11149.025	3961.3320	-	560.30970
P_{21}	0.8126814	0.8595310	0.7437536	0.9336277	-	0.6103227
$P_{21} C_{y2}/C_{x1}$	0.8938060	0.9786335	0.7360057	0.9922934	-	0.6721327
S_{31}	7482.0660	524.32450	2919.6930	974.51010	560.30970	-
P_{31}	0.6230342	0.7457194	0.6411828	0.7560806	0.6103227	-
$P_{31} C_{y3}/C_{x1}$	0.8898600	1.0975626	0.8202157	1.0450761	0.7892393	-

Variable under study - Number of workers of all types

(1) Auxiliary character - area of village

$$p \frac{C_Y}{C_X} = 0.8391208$$

Here usual ratio estimator and T_1 both can be used. T_1 will be more efficient than simple mean as well as usual ratio estimator if

$1 < p < 1.4744036$. The optimum value of p is 1.1917249.

$$E(\bar{Y}_R / \bar{Y}) = 2.7861439$$

p	1.05	1.1	1.15	1.1917249	1.2
$E(T_1 / \bar{Y})$	2.8833298	2.9691208	2.9821059	2.9898586	2.9896304
$E(T_1 / \bar{Y}_R)$	1.6385998	1.6384958	1.6703344	1.6731170	1.6730351

p	1.25	1.3	1.4
$E(T_1 / \bar{Y})$	2.9770390	2.9488678	2.8637761
$E(T_1 / \bar{Y}_R)$	1.6385137	1.6504047	1.6378036

(2)

(2) Auxiliary character - Number of households

$$p \frac{C_Y}{C_X} = 1.633875$$

T_1 can not be used as the value of $p C_Y / C_X$ is greater than unity.

(3) Auxiliary character - Total population of village

$$p \frac{C_Y}{C_X} = 0.647548 \quad 1 < p < 3.2837066 \quad p_{opt} = 1.5442942$$

$$E(\bar{Y}_R / \bar{Y}) = 1.8286158$$

p	1.3	1.4	1.5	1.5442942	1.6
$E(T_1 / \bar{Y})$	1.8949234	1.9893498	1.9587522	1.9590994	1.9368268
$E(T_1 / \bar{Y}_R)$	1.2370481	1.2711067	1.2839036	1.2841962	1.2826463

P	1.7	2	3
$E(T_1/\bar{Y})$	1.9434646	1.8661772	1.8981913
$E(T_1/\bar{Y}_R)$	1.2738880	1.2232281	1.0475708

(4) Auxiliary character - Number of cultivators

$$\rho \frac{C_Y}{C_X} = 0.8731506 \quad 1 < \rho < 1.3899418 \quad \rho_{opt} = 1.1482778$$

$$E(\bar{Y}_R/\bar{Y}) = 6.8149509$$

P	1.05	1.1	1.1482778	1.15	1.2
$E(T_1/\bar{Y})$	7.3791752	7.7032067	7.7918394	7.7909470	7.6833241
$E(T_1/\bar{Y}_R)$	1.0827921	1.1303889	1.1033449	1.1482139	1.1274217

(5) Auxiliary character - Number of Agricultural Labourers

$$\rho \frac{C_Y}{C_X} = 0.5470010 \quad 1 < \rho < 10.638071 \quad \rho_{opt} = 1.8281502$$

$$E(\bar{Y}_R/\bar{Y}) = 1.2189160$$

P	1.5	1.75	1.8281502	2	2.5
$E(T_1/\bar{Y})$	2.1964203	2.3288849	2.3345819	2.5118030	2.1293449
$E(T_1/\bar{Y}_R)$	1.6003081	1.9102094	1.9132923	1.8966086	1.7469168

P	?	?
$E(T_1/\bar{Y})$	1.9396136	1.7895210
$E(T_1/\bar{Y}_R)$	1.5912610	1.4682073

Variable under study - Number of Cultivators

(1) Auxiliary character - Area of village

$$\rho \frac{C_Y}{C_X} = 0.8938060 \quad \rho_{opt} = 1.1188110 \quad E(\bar{Y}_R/\bar{Y}) = 2.8664942$$

$$E_{opt}(T_1/\bar{Y}) = 2.9451784 \quad E_{opt}(T_1/\bar{Y}_R) = 1.0274496$$

which shows that \bar{Y}_R is preferable to T_1 .

(2) Auxiliary character - Number of households

$$p \frac{C_Y}{C_R} = 0.9786335$$

T_1 can be used but \bar{Y}_R is preferable.

(3) Auxiliary character - Total population of village.

$$p \frac{C_Y}{C_R} = 0.7360057 \quad 1 \quad p \quad 2.1185927 \quad p_{opt} = 1.3586851$$

$$E(\bar{Y}_R/\bar{Y}) = 1.9205197$$

p	1.2	1.3	1.3586851	1.4	1.5
$E(T_1/\bar{Y})$	2.1905792	2.2323698	2.2380017	2.2355913	2.2136777
$E(T_1/\bar{Y}_R)$	1.1347095	1.1563569	1.1592742	1.1580256	1.1466744

(4) Auxiliary character - Number of workers of all types

$$p \frac{C_Y}{C_R} = 0.9982934$$

\bar{Y}_R is preferable.

(5) Auxiliary character - Number of Agricultural Labourers

$$p \frac{C_Y}{C_R} = 0.4721327$$

Since $p C_Y/C_R$ is less than half, \bar{Y}_R can not be used but T_1 can be used by taking p greater than 1.0590241. The optimum value of p is 2.1180485.

p	1.5	2	2.1180485	2.5	3	3.5
$E(T_1/\bar{Y})$	1.4677133	1.5903209	1.5936097	1.5718205	1.5158425	1.4586225

Variable under study - Number of Agricultural Labourers

(1) Auxiliary character - Area of village

$$\rho \frac{C_Y}{C_X} = 0.8358600 \quad P_{opt} = 1.1288485 \quad E(\bar{Y}_R / \bar{Y}) = 1.6178039$$

$$E_{opt}(T_1 / \bar{Y}) = 1.6348003 \quad E_{opt}(T_1 / \bar{Y}_R) = 1.6105060$$

(2) Auxiliary character - Number of households

$$\rho \frac{C_Y}{C_X} = 1.0975626$$

T_1 cannot be used since $\rho C_Y / C_X$ is greater than unity.

(3) Auxiliary character - Total population of village

$$\rho \frac{C_Y}{C_X} = 0.8202187 \quad P_{opt} = 1.2191914 \quad E(\bar{Y}_R / \bar{Y}) = 1.643077$$

$$E_{opt}(T_1 / \bar{Y}) = 1.6981685 \quad E_{opt}(T_1 / \bar{Y}_R) = 1.0330937$$

(4) Auxiliary character - Number of workers of all types

$$\rho \frac{C_Y}{C_X} = 1.0430761$$

T_1 cannot be used.

(5) Auxiliary character - Number of cultivators

$$\rho \frac{C_Y}{C_X} = 0.7892393 \quad P_{opt} = 1.2670428 \quad E(\bar{Y}_R / \bar{Y}) = 1.6286829$$

$$E_{opt}(T_1 / \bar{Y}) = 1.5936107 \quad E_{opt}(T_1 / \bar{Y}_R) = 1.042473$$

Population - II
(Mahalpada Police Station)
Total number of villages - 35
Table for Mean, Variance, etc.

Character	1	2	3	4	5	6
Mean	1472.0000	60.228571	306.17142	89.514285	63.285714	18.457142
S_1^2	1352380.6	7562.9290	54891.566	5170.7870	2562.7984	731.7840
C_1	0.7900848	1.4638604	0.7652232	0.8033151	0.7999291	1.4666397
S_{11}	42270.210	3681.1145	16539.234	-	3352.5584	1208.4933
P_{11}	0.5054457	0.5886717	0.9813562	-	0.9209591	0.6212612
$P_{11} C_{y1}/C_{x1}$	0.5134135	0.3275169	1.0302068	-	0.9248573	0.5405124
S_{21}	35684.390	2509.4622	10958.037	3352.5548	-	358.68910
P_{21}	0.6060933	0.5700273	0.9238953	0.9209591	-	0.2619203
$P_{21} C_{y2}/C_{x1}$	0.6136450	0.3158071	0.9657975	0.9170771	-	0.1429530
S_{31}	2459.7050	827.53930	3686.0664	1203.4933	358.68910	-
P_{31}	0.0781825	0.3817784	0.5815923	0.6212612	0.2619203	-
$P_{31} C_{y3}/C_{x1}$	0.1450316	0.3570845	1.1139295	1.1334842	0.4798934	-

Variable under study - Total number of workers of all types

(1) Auxiliary character - Area of village

$$p \frac{C_Y}{C_X} = 0.5134135 \quad 1 < p < 37.275878 \quad P_{opt} = 1.9477477$$

$$E(\bar{y}_R / \bar{y}) = 1.0276590$$

P	1.25	1.5	1.75	1.9477477	2
$E(T_1/\bar{y})$	1.2140402	1.3036189	1.3392924	1.3431601	1.3428030
$E(T_1/\bar{y}_R)$	1.1813648	1.2685325	1.3013970	1.3069900	1.306662

P	2.5	3	3.5	4
$E(T_1/\bar{y})$	1.3208720	1.2885483	1.2580257	1.2314827
$E(T_1/\bar{y}_R)$	1.2833212	1.2538680	1.2241664	1.1985324

(2) Auxiliary character - Number of households

$$p \frac{C_Y}{C_X} = 0.5275169$$

The usual ratio estimator can not be used whereas T_1 can be used by taking p greater than 1.5266388. The optimum value of p is 3.0532775.

P	2	2.5	3	3.0532775	3.5	4
$E(T_1/\bar{y})$	1.3340877	1.491562	1.5300467	1.5303022	1.5171953	1.4861541

(3) Auxiliary character - Total population of village

$$p \frac{C_Y}{C_X} = 1.0308068$$

T_1 cannot be used.

(4) Auxiliary character - Number of cultivators

$$p \frac{C_Y}{C_X} = 0.9248573 \quad P_{opt} = 1.0812478 \quad E(\bar{y}_R / \bar{y}) = 6.351981$$

$$E_{opt}(T_1/\bar{y}) = 6.5861237 \quad E_{opt}(T_1/\bar{y}_R) = 1.0368612$$

(5) Auxiliary character - Number of Agricultural Labourers

$$p \frac{C_Y}{C_X} = 0.2405124$$

The usual ratio estimator cannot be used whereas T_1 can be used by taking p greater than 1.4685753. The optimum value of p is 2.9367506.

p	1.5	2	2.5	2.9367506	3
$E(T_1/\bar{y})$	1.0329121	1.4312175	1.5979183	1.6265727	1.6281179

p	3.5	4	4.5
$E(T_1/\bar{y})$	1.6024855	1.5593194	1.5137469

Character under study - Number of cultivators

(1) Auxiliary character - Area of village

$$p \frac{C_Y}{C_X} = 0.6136480 \quad 1 < p < 4.3996656 \quad p_{opt} = 1.6296066$$

$$E(\bar{y}_R/\bar{y}) = 1.2849874$$

p	1	1.5	1.6296066	2
$E(T_1/\bar{y})$	1.2849874	1.5728853	1.5806992	1.5498176
$E(T_1/\bar{y}_R)$	1.0000000	1.2248253	1.2390123	1.2060954

(2) Auxiliary character - Number of households

$$p \frac{C_Y}{C_X} = 0.3138071$$

\bar{y}_R cannot be used whereas T_1 can be used for any value of p greater than

1.5832449 . The optimum value of p is 3.1664899 .

p	2	2.5	3	3.1664899	3.5	4
$E(T_1/\bar{y})$	1.2573492	1.4197154	1.4697718	1.4728933	1.4680317	1.4459261

(3) Auxiliary character - Total population of village

$$p \frac{C_y}{C_x} = 0.9657978$$

\bar{y}_R is preferable because of high value of $p C_y / C_x$.

(4) Auxiliary character - Number of workers of all types

$$p \frac{C_y}{C_x} = 0.9170771$$

This case is also similar to the last case.

(5) Auxiliary character - No. of Agricultural Labourers

$$p \frac{C_y}{C_x} = 0.1429830$$

\bar{y}_R cannot be used. T_1 will be more efficient than \bar{y} if $p > 3.4976550$.

The optimum value of p is 6.9953061 and the optimum efficiency of T_1 is 1.0055675.

Character under study - Number of Agricultural Labourers

(1) Auxiliary character - Area of village

$$p \frac{C_y}{C_x} = 0.1540316$$

The ordinary ratio estimator cannot be used but T_1 can be used by taking p greater than 3.4475245. The optimum value of p is 6.8950408 and optimum efficiency is 1.0061510.

(2) Auxiliary character - Number of households

$$\rho \frac{C_y}{C_x} = 0.3870845$$

Ordinary ratio estimator cannot be used whereas T_1 can be used by choosing any value of p greater than 1.4002287. The optimum value of p is 2.8004578.

p	1.5	2	2.5	2.8004578	3	3.5
$E(T_1/\bar{y})$	1.0317093	1.1159793	1.1389116	1.1412246	1.1405119	1.1348223

(3) Auxiliary character - Total population of village

$$\rho \frac{C_y}{C_x} = 1.5443013$$

T_1 cannot be used because $\rho C_y / C_x$ is greater than unity. So is the case when number of workers of all types is used as auxiliary character.

(4) Auxiliary character - Number of cultivators

$$\rho \frac{C_y}{C_x} = 0.4798934$$

Ordinary ratio estimator cannot be used but T_1 can be used for the values of p greater than 1.041898. The optimum value of p is 2.037961.

p	1.25	1.50	1.75	2	2.037961	2.25
$E(T_1/\bar{y})$	1.039586	1.0618088	1.0707859	1.0735164	1.0735549	1.0732239

Population - III
 (Tibhainpali Police Station)
 Total number of villages - 60
Table for Mean, Variance, Etc.

Character	Mean	52.92500	319.48000	97.00000	74.175000	15.775000
S_1^2	696191.65	1663.8103	100428.05	9490.0000	5637.3790	627.50709
C_1	1.0235190	0.9987981	0.9920295	1.0042950	1.0122342	1.0030509
S_{11}	58601.210	5692.0870	30761.231	-	7000.7180	1724.9087
P_{11}	0.6354377	0.9928682	0.9960215	-	0.9571298	0.7067114
$P_{11} C_{Y1} / C_{X1}$	0.6211057	0.9523323	1.0067012	-	0.9496227	0.4753635
S_{21}	64728.970	0197.3981	22757175	7000.7180	-	1113.0408
P_{21}	0.6292285	0.9698701	0.9564279	0.9571298	-	0.5917899
$P_{21} C_{Y2} / C_{X1}$	0.6225531	0.9626480	0.9759074	0.9646961	-	0.4022092
S_{31}	6673.2640	1030.4699	5038.1850	1724.9087	1113.0408	-
P_{31}	0.2810511	0.6989671	0.6339259	0.7067114	0.5917899	-
$P_{31} C_{Y3} / C_{X1}$	0.4105693	1.0008571	0.9840943	1.0506505	0.6728990	-

Character under study - Number of workers of all types

Auxiliary character - Area of village

$$p \frac{C_Y}{C_X} = 0.6211057 \quad 1 < p < 4.12316248 \quad p_{opt} = 1.6100319$$

p	1.25	1.5	1.6100319	1.75	2	2.5
$E(T_1/\bar{Y})$	1.5909077	1.6710121	1.6772360	1.6693900	1.6532220	1.6429627
$E(T_2/\bar{Y}_R)$	1.1327684	1.2428962	1.2469225	1.2410650	1.2146527	1.1671000

(2) Auxiliary character - Number of households

$$p \frac{C_Y}{C_X} = 0.9983333$$

Ordinary ratio estimator is preferable.

(3) Auxiliary character - Total population of village

$$p \frac{C_Y}{C_X} = 1.007412$$

T_1 cannot be used.

(4) Auxiliary character - Number of cultivators

$$p \frac{C_Y}{C_X} = 0.9096227$$

\bar{Y}_R is preferable.

(5) Auxiliary character - Number of Agricultural Labourers

$$p \frac{C_Y}{C_X} = 0.6733633$$

Obviously, the ordinary ratio estimator cannot be used whereas T_1 can be used by choosing p greater than 1.051266. The optimum value

of ρ is 3.2036533

ρ	1.25	1.5	1.75	2	2.2036533
$E(T_1/\bar{y})$	1.9648042	1.9213283	1.9211891	1.9939735	1.9977664
ρ	2.25	2.5	2.75		
$E(T_1/\bar{y})$	1.9907394	1.9500753	1.8944108		

Character under study - Number of cultivators

(1) Auxiliary character - Area of village

$$\rho \frac{C_y}{C_x} = 0.6223531 \quad \rho_{opt} = 1.6068013 \quad E(\bar{y}_R/\bar{y}) = 1.9326840$$

$$E_{opt}(T_1/\bar{y}) = 1.6556461 \quad E_{opt}(T_1/\bar{y}_R) = 1.2414080$$

(2) Auxiliary character - Number of households

$$\rho \frac{C_y}{C_x} = 0.9626480$$

\bar{y}_R is preferable.

(3) Auxiliary character - Total population of village

$$\rho \frac{C_y}{C_x} = 0.9739074$$

\bar{y}_R is preferable.

(4) Auxiliary character - Number of workers of all types

$$\rho \frac{C_y}{C_x} = 0.9546951$$

\bar{y}_R is preferable.

(5) Auxiliary character - Number of Agricultural Labourers

$$p \frac{C_Y}{C_X} = 0.4012092$$

The usual ratio estimator cannot be used but T_1 can be used by choosing

p greater than 1.2462326. The optimum value of p is 2.4924652.

p	1.5	2	2.4924652	2.5	3	3.5
$E(T_1/\bar{y})$	1.2454038	1.4905166	1.5391202	1.5391721	1.5157599	1.4733108

Character under study - Number of Agricultural Labourers.

(1) Auxiliary character - Area of village

$$p \frac{C_Y}{C_X} = 0.4105930 \quad p_{opt} = 2.4994240 \quad p > 1.2178212$$

$$E_{opt} (T_1/\bar{y}) = 1.0860378$$

(2) Auxiliary character - Number of households

$$p \frac{C_Y}{C_X} = 1.0448571$$

T_1 cannot be used.

(3) Auxiliary character - Total population of village

$$p \frac{C_Y}{C_X} = 0.9540945$$

Both T_1 and \bar{y}_R can be used but \bar{y}_R is preferable.

(4) Auxiliary character - Number of workers of all types

$$p \frac{C_Y}{C_X} = 1.0505050$$

T_1 cannot be used.

(5) Auxiliary character - Number of cultivators

$$\rho \frac{C_y}{C_x} = 0.8728990$$

\bar{y}_R is preferable.

SUMMARY AND CONCLUSIONS

Usual ratio estimator has limited utility in the sense that in populations where the condition

$$\rho \frac{C_Y}{C_X} > \frac{1}{2}$$

is not satisfied, it falls also in such a case it becomes less efficient than the simple mean of the sample as an estimate of the population mean. In practice, we do come across many situations in which the above condition is not satisfied. This condition generally requires the existence of a high correlation between the main and the auxiliary character. Sometimes an auxiliary character bears a high correlation with the main character but the information on it is either not available or involves a high cost for its collection. On the other hand, there may be some other auxiliary characters which are somewhat poorly correlated with the main character but information on them is either available or can be collected at a small cost. To deal with such situations, two modified ratio type estimators have been proposed. The biases and efficiencies for these estimators have been obtained and compared with that of simple mean, usual ratio estimator and regression estimator. The use of these estimator has also been extended to the case of double sampling and to the case of negatively correlated characters. The case of sampling with varying probabilities is also discussed for the first estimator. Some empirical studies have also been conducted to investigate the various aspects of the proposed

estimator.

The advantage of using the proposed estimator in place of usual ratio estimator can be summarised as follows :

(i) If there are two auxiliary characters available and one of them involves a higher cost and has a higher correlation with the main character than the other one (assuming the coefficients of variation to be same for both the auxiliary characters), the cheaper one (having a lower correlation with the main character) can be used without any loss of efficiency and without increasing the cost by making use of the proposed estimator and simply adjusting the values of certain constants involved in the estimator.

(ii) These estimators are simple from the computational point of view like usual ratio estimator.

(iii) They can be used where usual ratio estimator fails, examples are

$$(a) \quad \rho \frac{C_y}{C_x} > \frac{1}{2}$$

$$(b) \quad \text{if } \frac{\bar{y} - \bar{Y}}{\bar{Y}} > 1, \text{ the expansion of } \left(1 + \frac{\bar{y} - \bar{Y}}{\bar{Y}} \right)^{-1}$$

is invalid, whereas for T_1 the condition is $\frac{\bar{y} - \bar{Y}}{\bar{Y}} > p$ and p being greater than unity widens the range of $\frac{\bar{y} - \bar{Y}}{\bar{Y}}$ for

the validity of the expansions involving such terms.

(c) Even if the variables are negatively correlated, the proposed estimators can be used by giving negative values to certain constants involved in the estimator. Moreover, they possess the advantage of possible reduction in bias.

(iv) Instead of using a bivariate ratio estimator, the proposed estimator can be used with single auxiliary character by suitably choosing the constants.

REFERENCES

- Beale, E.M.L. (1962) : Some uses of computers in operational research. *Industrielle Organisation* 31, 51-52.
- Cochran, W.G. (1942) : Sampling theory when the sampling units are of unequal sizes. *JASA* 37, 199 - 212.
- Des Raj (1956) : Ratio estimation in Sampling with equal and unequal probabilities. *JIAS* 6, 127 - 138.
- Hansen, M.H and Hurwitz, W.N. (1943) : On the theory of sampling from finite populations. *AMS* 14, 333 - 352.
- Hansen, M.H and Hurwitz, W.N. (1946) : The problem of non-response in sample surveys. *JASA* 41, 517 - 529.
- Hartley, H.O. and Ross, A. (1954) : Unbiased ratio estimators. *Nature* 174, 270 - 271.
- Lahiri, D. B. (1951) : A method of sample selection providing unbiased estimates. *HISI* 33, 193 - 196.
- Madow, W.G. and Madow, L.H. (1944) : On the theory of systematic sampling - I. *AMS* 15, 1-24.
- Mahalanobis, P.C. (1940) : A sample survey of acreage under jute in Bengal. *Sankhya* 4, 511 - 530.
- Mahalanobis, P.C. (1946) : Sample Surveys of crop yields in India. *Sankhya* 7, 269 - 280.
- Mickey, M.R. (1959) : Some finite population unbiased ratio and regression estimators. *JASA* 54, 594 - 612.
- Midsuno, H. (1932) : On the sampling system with probability proportional to the sum of sizes. *AIJM* 3, 99 - 103.
- Murthy, M.N., Nanjama, N.S. and Sethi, V.K. (1959) : Some sampling systems providing unbiased ratio estimators. *Sankhya* 21, 299-316.
- Quenecillo, M.H. (1956) : Notes on bias in estimation. *Biometrics* 43, 353 - 360.
- Robson, D.S. (1957) : Application of multivariate polynomials to the theory of unbiased ratio type estimators. *JASA* 52, 511 - 522.

**Shastri, N.S. (1964) : Unbiased ratio and regression type estimators .
Diploma Dissertation submitted to IARS (ICAR).**

**Sen, A.R. (1952) : Present status of probability sampling and its use
in estimation in farm characteristics (abstract).
Econometrica 20, 103.**

**Sukhatme, P.V. (1944) : Moments and product moments of moment
statistics for samples of finite and infinite populations.
Sankhya 6, 363 - 382.**

**Tin, M. (1965) : Comparison of some ratio estimators. JASA 60,
294 - 307.**

