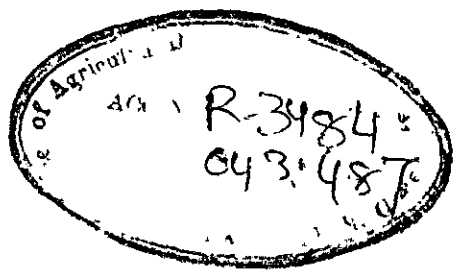
V84

MULTIPLE CHARACTERS IN MULTIPHASE SAMPLING

F.K. TYAGI



Dissertation submitted in fulfilment of the requirements for the award of Diploma in Agricultural Statistics of the 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

1975

ACKNOWLEDGEMENTS

I wish to express my deep sense of gratitude to Shri M. Rajagopalan, Statistician-cum-Associate Professor, Institute of Agricultural Research Statistics (I.C.A.R.), New Delhi, for his valuable guidance, keen interest and constant encouragement throughout the course of investigation and of the preparation of the thesis.

I am thankful to Dr. D. Singh, Director, Institute of Agricultural Research Statistics (I.C.A.R.), New Delhi for providing me the adequate facilities for this work.

My thanks are also due to the Indian Council of Agricultural Research for providing the financial assistance in the form of a fellowship.

Krishan Kant Tyagi

I. A. R. S.
NEW DELHI - 12.

(KRISHAN KANT TYAGI)

CONTENTS

CHAPTER - I

INTRODUCTION

History of learning about population by using sampling methods could be traced out even to very early stages of primitive life of mankind. It was more than 250 years ago that Bernoulli developed the theory of independent random sampling of elements from a population when the unit of sampling and the unit of analysis was the same (Hansen and Hurvitz, 1943). A century later, Poisson indicated the use of stratification in sampling to get possibly more precise estimate of the parameters. Later on the techniques such as cluster sampling, sub-sampling, double sampling and successive sampling were developed with an idea of making the best use of the available resources.

Research in the theory of sampling for surveys has been mainly concerned with the development of more efficient sampling systems, the systemincluding both, the sampling design and the method of sampling. One sampling system is said to be more efficient than the other, if the variance or the mean square error of the estimate with the former system is less than that of the later, provided the cost of obtaining the data is the same. The development of stratified, multistage, multiphase, systematic and other sampling designs over simple random sampling, have all resulted in increased efficiency in specific circumstances.

With an idea to build up more precise estimators of the

population parameters, the information collected on auxiliary variates are used. The auxiliary information is obtained in addition to the character under study and some extra information can be derived from it. The use of auxiliary information can be made on two occasions, one at the time of designing the survey and the other at the time of estimating the parameters. Sometimes the sampling units are selected with probability proportional to some measure of size, size being the value of auxiliary variable. The estimators which make use of auxiliary information include ratio and regression methods of estimation, which are by far the most common, though the estimators are biased.

In sample surveys it happens sometimes that several auxiliary variables are available which are highly correlated with the variable under study. It is, therefore, of interest to investigate the effect of using all or some of them in building up an estimate of the population mean or total of the character under study.

Ingram Olkin (1958) has discussed the extension of ratio method of estimation to the case where multi-auxiliary variables are used to increase precision. Here it is assumed that the population means corresponding to the different auxiliary variables are available before the sample is drawn. In the univariate case a simple random sample $(x_1 , y_1) \ldots (x_n , y_n)$ from a finite population $(X_1 , Y_1) \ldots (X_N , Y_N)$ is observed. The mean $\bar{X}$ is known and $\bar{Y}$ is to be estimated. The estimator

$$\bar{y}_R = (\bar{y} / \bar{x}) \bar{X} = r \bar{X}$$

is called the ratio estimate of $\bar{Y}$. In general $\bar{y}_R$ is biased and for large n, approximation for $E(\bar{y}_R)$ and $V(\bar{y}_R)$ are given by

$$E(\bar{y}_R) = \bar{Y} + \frac{N-n}{N} \frac{\bar{Y}}{n}(C_x^2 - \rho_{xy} C_x C_y)$$

and

$$V(\bar{y}_R) = \frac{N-n}{N} \cdot \frac{\bar{Y}^2}{n}(C_x^2 + C_y^2 - 2\rho_{xy} C_x C_y)$$

where $C_x = \frac{S_x}{\bar{X}}$, coefficient of variation of x,

$C_y = \frac{S_y}{\bar{Y}}$, coefficient of variation of y, and

ρ_{xy} is the correlation coefficient between x and y (Cochran, 1953, pp. 115 - 16).

It is easily shown that $\bar{y}_R$ is a consistent estimator of $\bar{Y}$ in the sense of Cochran (1953, p.13) i.e. $\bar{y}_R \longrightarrow \bar{Y}$ as $n \longrightarrow N$ and also in the sense of Hansen, Hurvitz and Madow (1953, p.74) i.e. $\operatorname{plim}_{n \to \infty} \bar{y}_R = \bar{Y}$ with the restrictions

i) as n increases, N increases with $n < \Theta N$, $0 < \Theta < 1$ and

ii) $\bar{Y}$ remains constant as N increases

The ratio method can be made unbiased by changing the sampling procedure. Midzuno gave this method under which ratio estimate is unbiased. In this method the unit at the first draw is selected with unequal probability proportional to x and in the remaining (n - 1) subsequent draws they are selected with equal

probability without replacement. The other method is Lahiri's Method of sample selection which provides unbiased ratio estimate. It consists in selecting a pair of random numbers say (i, j) such that $1 \leq i \leq N$ and $1 \leq j \leq M$, where M is the maximum of the sizes of the N units in the population. If $j \leq x_i$, the i-th unit is selected; otherwise it is rejected and another pair of random numbers is chosen. For selecting a sample of n units with probability proportional to size and with replacement, the procedure is to be repeated till n units are selected.

Hartley and Ross (1954) were first to build up an unbiased ratio type estimator in sampling with equal probability. In simple random sampling without replacement they found an unbiased estimator of the bias and hence an unbiased estimator of the population parameter viz.

$$\bar{y}_R^* = \bar{r}\,\bar{X} + \frac{n(N-1)}{N(n-1)}(\bar{y} - \bar{r}\,\bar{x})$$

Robson (1957) gave an exact formula for its variance. Goodman and Hartley (1958) studied the precision of biased and unbiased ratio estimator and it is shown that, when the ratio of y to x decreases as x increases, the unbiased ratio estimator is more precise than the ratio estimator ($\bar{y} = \bar{r}\,\bar{X}$). It is also shown that neglecting certain negligible terms, the criterion as to whether $V(\bar{y}_R^*) < V(\bar{y}_R)$ or vice versa will depend upon whether or not $\beta \lessgtr \frac{1}{2\bar{X}}$, where

$$\beta = \frac{E(\bar{r} - \bar{R})^2 (\bar{x} - \bar{X})}{\mathrm{Var}(\bar{r} - \bar{R})(\bar{x} - \bar{X})} .$$

Ramachandran (1969) extended the usual Hartley and Ross type unbiased ratio estimator for the use of multi-auxiliary variables in simple random sampling without replacement.

Olkin (1958) extended the ratio method of estimation to the use of several auxiliary variables. In the multivariate extension we have the following model:

Population : $Y_1, \ldots, Y_N$, unknown

$X_{11}, \ldots, X_{1N} \neq 0$ known, $R_1 = \bar{Y} / \bar{X}_1$,

. . . . . .

$X_{p1}, \ldots, X_{pN} \neq 0$ known, $R_p = \bar{Y} / \bar{X}_p$

and the $(p + 1) \times (p + 1)$ covariance matrix S is known. The subscripts $0, 1, \ldots, p$, refer to $Y, X_1, \ldots, X_p$ respectively e.g. ρ_{02} is the correlation between Y and X_2. Higher moments will have superscripts referring to the variables and subscripts to the powers, e.g.

$$\mu_{12}^{ij} = \sum_k (X_{ik} - \bar{X}_i) (X_{jk} - \bar{X}_j)^2 / N ,$$

$$\mu_{111}^{0ij} = \sum_k (Y_k - \bar{Y}) (X_{ik} - \bar{X}_i) (X_{jk} - \bar{X}_j) / N .$$

Finally, $S_{ij} = N \mu_{11}^{ij} / (N-1)$ denotes the covariance and $C_i = S_i / \bar{X}_i$, the coefficient of variation. The later development will be considerably simplified if we have a notation for moments

divided by means, thus

$$w_{12}^{ij} = \mu_{12}^{ij} / \bar{X}_i \bar{X}_j^2 \quad \text{etc.}$$

A simple random sample $(y_j , x_{1j} , \ldots , x_{pj})(j = 1, \ldots, n)$, from the population is observed. The proposed ratio estimate of $\bar{Y}$ is

$$\bar{y}_{wR} = w_1 r_1 \bar{X}_1 + \ldots + w_p r_p \bar{X}_p ,$$

where $w = (w_1 , \ldots , w_p)$, $\sum_1 w_i = 1$ is a weighing function and $r_i = (\bar{y} / \bar{x}_i)$.

The Hartley - Ross estimator can be generalised so that

$$\bar{y}^* = \sum_1^p w_i \bar{r}_i \bar{X}_i + \frac{n(N-1)}{N(n-1)} \left(\bar{y} - \sum_1^p w_i \bar{r}_i \bar{x}_i \right)$$

is an unbiased estimator of $\bar{Y}$, where $n \bar{r}_i = \sum_{j=1}^n y_j / x_{ij}$.

Consistency in the multivariate case follows from the fact that we have a linear combination of consistent estimators.

Olkin gave the formula for optimum weights and for the variance of multivariate estimator $\bar{y}_{wR}$. He also showed that if $V (\bar{y} / p)$ and $V (\bar{y} / p, q)$ denote the variances of $\bar{y}$ based on the auxiliary variables $x_1 , x_2 , \ldots , x_p$ and $x_1 , x_2 , \ldots , x_q$ $(q > p)$ respectively, using optimum weights $V (\bar{y}/p)$ will always be greater than or equal to $V (\bar{y} / p, q)$.

The linear regression estimator $\bar{y}_{lr}$ of population mean of y when y is of the form $y = \alpha + \beta x + e$ (where x is an auxiliary

variable and e is the error term due to random causes) was given by

$$\bar{y}_{lr} = \bar{y} + b\,(\bar{X} - \bar{x})$$

where b is the sample regression coefficient $\Sigma(y-\bar{y})(x-\bar{x})/\Sigma(x-\bar{x})^2$. And in samples in which the x's remain fixed, the sampling variance of $\bar{y}_{lr}$ is

$$V(\bar{y}_{lr}) = \sigma_y^2\,(1-\rho^2)\left[\frac{1}{n} + \frac{(\bar{X}-\bar{x})^2}{\Sigma(x-\bar{x})^2}\right]$$

n being the sample size and ρ is the correlation coefficient between y and x. The distribution of $\bar{y}_{lr}$ tends to normality as n increases. And if the correct form of the regression is used, population estimates derived from regression remain unbiased.

In the Olkin's method the population means corresponding to different auxiliary variables are available before the sample is drawn. In some situation these population values are not available, in which case Multiphase Sampling is resorted to. The technique consists in taking a larger sample of size n' to estimate the population mean $\bar{X}_N$ while a sub-sample of size n is drawn from n' to observe the characteristic under study, Y. Several estimates of the population mean $\bar{Y}_N$ can be formed. The simplest is the usual biased ratio estimate $\bar{y}_R$, with $\bar{X}_N$ replaced by its estimate $\bar{x}_{n'}$ based on a sample of size n', given by

$$\bar{y}_{Rd} = \frac{\bar{y}_n}{\bar{x}_n}\,\bar{x}_{n'}$$

The variance of the above estimate is given by

$$V(\bar{y}_{Rd}) = \left(\frac{1}{n} - \frac{1}{n'}\right)\left(S_y^2 + R_N^2 S_x^2 - 2R_N S_{yx}\right) + \left(\frac{1}{n'} - \frac{1}{N}\right) S_y^2$$

from here it follows that the estimate $\bar{y}_{Rd}$ based on double sampling is more efficient than the estimate $\bar{y}_n$ based on simple random sampling when no auxiliary variable is used, if

$$R_N^2 S_x^2 - 2R_N S_{yx} \quad 0$$

i.e. if $\rho_{yx} > \frac{1}{2} \frac{C_x}{C_y}$.

Sometimes it may not be worthwhile to collect information on all auxiliary characters from the same larger sample due to varying degrees of cost of enumerating the different characters. On some characters the cost of enumeration is higher than on a few other. So in this study, this aspect is taken into consideration in drawing a multiphase sample. We take smaller sample on the auxiliary character for which the cost of enumeration is high while the sample size will be large for low cost of enumeration.

CHAPTER - II

COMPARISON OF DIFFERENT ESTIMATES USING MULTIPLE AUXILIARY CHARACTERS

2.1 Introduction

In sample surveys sometimes we have multiple auxiliary characters highly correlated with the study variable Y. The cost of enumeration of units for each auxiliary character will be different for different characters. In some cases it may be very high and in some cases it may be very low. So it is useless to collect information on all auxiliary characters for the same large sample because of varying degrees of cost of enumerating the different characters. So here in our study we draw the sample in multiphases i.e. we use multiphase sampling. We take the largest sample on the character for which the cost of enumeration is lowest. Then out of this larger sample we take a sub-sample and observe another character for which the cost of enumeration is little higher than the previous one and so on. On the character, for which the cost of enumeration is maximum we take the smallest sample and out of which we take a smaller sample and observe the study character Y.

Let there be N units in the population $Y_1, \ldots, Y_N$ and $X_1, \ldots, X_p$ are the p auxiliary characters highly correlated with the study variable Y. Let the cost of enumeration for X_p is the minimum, than that of $X_{p-1}, \ldots$ and the cost of enumeration of X_1 is the maximum. Then according to our sampling scheme,

we observe the character X_p on the largest number of units in the sample, n_p, and observe the character X_{p-1} on n_{p-1} units, . . . and observe the character X_1 on n_1 units where ($0 < n_1 < n_2 < \ldots < n_{p-1} < n_p < N$). These $n_1, \ldots, n_p$ units are such that the n_{p-1} units are out of n_p units, n_{p-2} units are out of n_{p-1} units, . . . , n_1 units are out of n_2 units. From these n_1 units we take n units and observe the character Y under study so that ($0 < n < n_1 < \ldots < n_p < N$).

2.2 Notations

Let

Y : character under study

X_1 : auxiliary character correlated with Y having maximum cost of enumeration.

X_2 : auxiliary character correlated with Y having (lesser than X_1) cost of enumeration.

.
.
.

X_p : auxiliary character correlated with Y having minimum cost of enumeration.

Again let,

N : total number of units in the population

n_p : number of sample units on which X_p is observed

n_{p-1} : number of sample units out of n_p on which X_{p-1} is observed

.
.
.

n_1 : number of sample units out of n_2 on which X_1 is observed

n : number of sample units out of n_1 on which Y is observed.

Now

$$\bar{y}_n = \frac{1}{n} \sum_{1}^{n} y_i \,, \qquad \text{sample mean of } Y$$

$$\bar{x}_{1n_1} = \frac{1}{n_1} \sum_{1}^{n_1} x_{1i} \,, \qquad \text{sample mean of } X_1$$

.
.
.

$$\bar{x}_{pn_p} = \frac{1}{n_p} \sum_{1}^{n_p} x_{pi} \,, \qquad \text{sample mean of } X_p$$

2.3 Multivariate Ratio and Regression Estimates of $\bar{Y}$

First Estimate ($\bar{y}_{wR}$)

The ratio estimates for $X_1, \ldots, X_p$ are as follows:

$$\bar{y}_{Rx_1} = \frac{\bar{y}_n}{\bar{x}_{1n}} \bar{x}_{1n_1}, \ldots, \bar{y}_{Rx_p} = \frac{\bar{y}_n}{\bar{x}_{pn}} \bar{x}_{pn_p} \,.$$

The combined ratio estimate, $\bar{y}_{wR}$ of $\bar{Y}$ is given by

$$\bar{y}_{wR} = \sum_{i=1}^{p} w_i \, \bar{y}_{Rx_i} \qquad (2.3.1)$$

where, $w_1, w_2, \ldots, w_p$ are the weights.

Second Estimate ($\bar{y}_{w\,lr}$)

This estimate is a regression estimate, as follows:

$$\bar{y}_{w\,lr} = \bar{y}_n + \sum_{i=1}^{p} b_i (\bar{x}_{in_i} - \bar{x}_{in}) \qquad (2.3.2)$$

where $b_1, b_2, \ldots, b_p$ are the sample regression coefficients of y on $x_1, x_2, \ldots, x_p$ respectively.

Third Estimate ($\bar{y}_{u\,lr}$)

This estimate is also a regression estimate as follows:

$$\bar{y}_{u\,lr} = \bar{y}_n + \sum_{i=1}^{p} b_i (\bar{x}_{in_i} - \bar{x}_{in_{i-1}}) \qquad (2.3.3)$$

Now we shall find the bias and variance of these three estimates.

2.4 Bias and Variances of the Estimates

First Estimate ($\bar{y}_{wR}$)

The combined ratio estimate is

$$\bar{y}_{wR} = \sum_{i=1}^{p} w_i \bar{y}_{Rx_i}$$

as $E(\bar{y}_{wR}) \neq \bar{Y}$, so $\bar{y}_{wR}$ is a biased estimate of $\bar{Y}$.

The bias is given by

$$\text{Bias in } (\bar{y}_{wR}) = E (\bar{y}_{wR}) - \bar{Y}$$

where $$E (\bar{y}_{wR}) = \sum_{i=1}^{p} w_i E (\bar{y}_{Rx_i}) \qquad (2.4.1)$$

So first we will find $E(\bar{y}_{Rx_i})$ $\forall$ $i = 1, 2, \ldots, p$

$$\bar{y}_{Rx_i} = \frac{\bar{y}_n}{\bar{x}_{in}} \bar{x}_{in_i} \qquad \text{Let} \quad \bar{y}_n = \bar{Y} + e_0$$
$$\bar{x}_{in} = \bar{X}_i + e_i$$
$$\bar{x}_{in_i} = \bar{X}_i + e_i'$$

$$= \frac{(\bar{Y} + e_0)}{(\bar{X}_i + e_i)} (\bar{X}_i + e_i')$$

$$= \bar{Y}\left(1 + \frac{e_0}{\bar{Y}}\right)\left(1 + \frac{e_i'}{\bar{X}_i}\right)\left(1 - \frac{e_i}{\bar{X}_i} + \frac{e_i^2}{\bar{X}_i^2}\right) \quad \text{neglecting higher powers of } (e_i/\bar{X}_i)$$

$$= \bar{Y}\left[1 + \frac{e_0}{\bar{Y}} - \frac{e_i}{\bar{X}_i} + \frac{e_i'}{\bar{X}_i} + \frac{e_0 e_i'}{\bar{Y}\bar{X}_i} - \frac{e_0 e_i}{\bar{Y}\bar{X}_i} - \frac{e_i e_i'}{\bar{X}_i^2} + \frac{e_i^2}{\bar{X}_i^2}\right]$$

Now

$$E(e_0 e_i') = E(\bar{y}_n - \bar{Y})(\bar{x}_{in_i} - \bar{X}_i)$$

$$= \mathrm{Cov}(\bar{y}_n, \bar{x}_{in_i})$$

$$= \mathrm{Cov}\left[E(\bar{y}_n/n_i), E(\bar{x}_{in_i}/n_i)\right] + E\left[\mathrm{Cov}(\bar{y}_n, \bar{x}_{in_i}/n_i)\right]$$

$$= \mathrm{Cov}(\bar{y}_{n_i}, \bar{x}_{in_i}) \quad \text{as} \quad \mathrm{Cov}(\bar{y}_n, \bar{x}_{in_i}/n_i) = 0 \text{ as } n_i \text{ is fixed.}$$

$$= \left(\frac{1}{n_i} - \frac{1}{N}\right) S_{yx_i}$$

Similarly,

$$E(e_0 e_i) = \mathrm{Cov}(\bar{y}_n, \bar{x}_{in}) = \left(\frac{1}{n} - \frac{1}{N}\right) S_{yx_i}$$

$$E(e_i e_i') = V(\bar{x}_{in_i}) = \left(\frac{1}{n_i} - \frac{1}{N}\right) S_{x_i}^2$$

$$E(e_i^2) = V(\bar{x}_{in}) = \left(\frac{1}{n} - \frac{1}{N}\right) S_{x_i}^2$$

Hence

$$E(\bar{y}_{Rx_i}) = \bar{Y}\left[1 + \left(\frac{1}{n} - \frac{1}{n_i}\right)\left(C_{x_i}^2 - \rho_{yx_i} C_y C_{x_i}\right)\right] \quad \ldots (2.4.2)$$

where $C_y = \frac{S_y}{\bar{Y}}$, coefficient of variation of Y

$C_{x_i} = \frac{S_{x_i}}{\bar{X}_i}$, coefficient of variation of X_i

and $\rho_{y x_i}$ is the correlation coefficient between y and x_i.

Now substituting the value of $E(\bar{y}_{Rx_i})$ in (2.4.1), we get

$$E(\bar{y}_{wR}) = \sum_{i=1}^{p} w_i \bar{Y}\left[1 + \left(\frac{1}{n} - \frac{1}{n_i}\right)\left(C_{x_i}^2 - \rho_{yx_i} C_y C_{x_i}\right)\right] \quad \ldots (2.4.3)$$

Hence

$$\text{Bias in } (\bar{y}_{wR}) = E(\bar{y}_{wR}) - \bar{Y}$$

$$= \bar{Y}\left[\sum_{i}^{p} w_i - 1\right] + \bar{Y} \sum_{i}^{p} w_i \left(\frac{1}{n} - \frac{1}{n_i}\right)\left(C_{x_i}^2 - \rho_{yx_i} C_y C_{x_i}\right) \quad \ldots (2.4.4)$$

So the bias will vanish if the following conditions are satisfied

(1) $\sum_{i}^{p} w_i = 1$ i.e. sum of weights should be equal to 1.

(2) $C_{x_i}^2 - \rho_{yx_i} C_y C_{x_i} = 0 \;\forall\; i$, i.e. the regression line of y on x_i $(i = 1, \ldots, p)$ should pass through the origin.

or $C_{x_i} = \rho_{yx_i} C_y$

To find the variance of $\bar{y}_{wR}$, first of all we will find $V(\bar{y}_{Rx_1})$ and $Cov(\bar{y}_{Rx_i}, \bar{y}_{Rx_j})$.

Now

$$V(\bar{y}_{Rx_1}) = E(\bar{y}_{Rx_1} - \bar{Y})^2$$

$$= \bar{Y}^2 E\left[\frac{e_0^2}{\bar{Y}^2} + \frac{e_1^2}{\bar{X}_1^2} + \frac{e_1'^2}{\bar{X}_1^2} - 2\frac{e_0 e_1}{\bar{Y}\bar{X}_1} + 2\frac{e_0 e_1'}{\bar{Y}\bar{X}_1} - 2\frac{e_1 e_1'}{\bar{X}_1^2}\right]$$

$$= \bar{Y}^2\left[\left(\frac{1}{n} - \frac{1}{N}\right)C_y^2 + \left(\frac{1}{n} - \frac{1}{N}\right)C_{x_1}^2 + \left(\frac{1}{n_1} - \frac{1}{N}\right)C_{x_1}^2\right.$$

$$- 2\left(\frac{1}{n} - \frac{1}{N}\right)\rho_{yx_1} C_y C_{x_1} + 2\left(\frac{1}{n_1} - \frac{1}{N}\right)\rho_{yx_1} C_y C_{x_1}$$

$$\left. - 2\left(\frac{1}{n_1} - \frac{1}{N}\right)C_{x_1}^2\right]$$

$$= \bar{Y}^2\left[\left(\frac{1}{n} - \frac{1}{N}\right)C_y^2 + \left(\frac{1}{n} - \frac{1}{n_1}\right)\left(C_{x_1}^2 - 2\rho_{yx_1} C_y C_{x_1}\right)\right] \quad \ldots(2.4.5)$$

for $i = 1, \ldots, p$

and

$$Cov(\bar{y}_{Rx_i}, \bar{y}_{Rx_j}) = \bar{Y}^2 E\left[\frac{e_0^2}{\bar{Y}^2} - \frac{e_0 e_j}{\bar{Y}\bar{X}_j} + \frac{e_0 e_j'}{\bar{Y}\bar{X}_j} - \frac{e_0 e_i}{\bar{Y}\bar{X}_i} + \frac{e_i e_j}{\bar{X}_i\bar{X}_j} - \frac{e_i e_j'}{\bar{X}_i\bar{X}_j}\right.$$

$$\left. + \frac{e_0 e_i'}{\bar{Y}\bar{X}_i} - \frac{e_i' e_j}{\bar{X}_i\bar{X}_j} + \frac{e_i' e_j'}{\bar{X}_i\bar{X}_j}\right]$$

$$= \bar{Y}^2\left[\left(\frac{1}{n} - \frac{1}{N}\right)C_y^2 - \left(\frac{1}{n} - \frac{1}{N}\right)\rho_{yx_j} C_y C_{x_j} + \left(\frac{1}{n_j} - \frac{1}{N}\right)\rho_{yx_j} C_y C_{x_j}\right.$$

$$- \left(\frac{1}{n} - \frac{1}{N}\right)\rho_{yx_i} C_y C_{x_i} + \left(\frac{1}{n} - \frac{1}{N}\right)\rho_{x_i x_j} C_{x_i} C_{x_j}$$

$$- \left(\frac{1}{n_j} - \frac{1}{N}\right)\rho_{x_i x_j} C_{x_i} C_{x_j} + \left(\frac{1}{n_i} - \frac{1}{N}\right)\rho_{yx_i} C_y C_{x_i}$$

$$\left. - \left(\frac{1}{n_i} - \frac{1}{N}\right)\rho_{x_i x_j} C_{x_i} C_{x_j} + \left(\frac{1}{n_j} - \frac{1}{N}\right)\rho_{x_i x_j} C_{x_i} C_{x_j}\right]$$

$$= \bar{Y}^2 \left[\left(\frac{1}{n} - \frac{1}{N}\right) C_y^2 + \left(\frac{1}{n} - \frac{1}{n_i}\right)\left(\rho_{x_i x_j} C_{x_i} C_{x_j} - \rho_{y x_i} C_y C_{x_i}\right) \right.$$

$$\left. - \left(\frac{1}{n} - \frac{1}{n_j}\right) \rho_{y x_j} C_y C_{x_j} \right] \quad i \neq j \quad \ldots (2.4.6)$$

Using Olkin's method, multivariate ratio estimator, $\bar{y}_{wR(M.V.)}$ of population mean could be given as below

$$\bar{y}_{wR(M.V.)} = \sum_{i=1}^{p} w_i \, \bar{y}_{R x_i} \quad \ldots (2.4.7)$$

where w_i's $(i = 1, 2, \ldots, p)$ are so obtained as to minimise the variance of multivariate estimator and $\sum_{i=1}^{p} w_i = 1$.

Let A be the $p \times p$ variance covariance matrix

$$A = \begin{bmatrix} V_{11} & V_{12} & \cdots & V_{1p} \\ V_{21} & V_{22} & \cdots & V_{2p} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ V_{p1} & V_{p2} & \cdots & V_{pp} \end{bmatrix}$$

where V_{ii} is given by (2.4.5) and V_{ij} by (2.4.6).

As shown by Olkin, the weight w_i, which will minimise the $V\left[\bar{y}_{wR(M.V.)}\right]$ would be T_i / T where T_i is the total of the elements in the i-th row of the inverse matrix (A^{-1}) and T is the total of all the elements in A^{-1} matrix. The variance of $\bar{y}_{wR(M.V.)}$ is given by $1/T$.

It can be shown that with the increase in number of auxiliary variables, the variance of multivariate estimator using optimum weights decreases steadily.

A Particular Case (when $p = 2$)

When there are only two auxiliary characters, then the matrix A will be of the form

$$A = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}$$

where

$$V_{11} = \bar{Y}^2 \left[\left(\frac{1}{n} - \frac{1}{N} \right) C_y^2 + \left(\frac{1}{n} - \frac{1}{n_1} \right) \left(C_{x_1}^2 - 2 \rho_{y x_1} C_y C_{x_1} \right) \right]$$

$$V_{22} = \bar{Y}^2 \left[\left(\frac{1}{n} - \frac{1}{N} \right) C_y^2 + \left(\frac{1}{n} - \frac{1}{n_2} \right) \left(C_{x_2}^2 - 2 \rho_{y x_2} C_y C_{x_2} \right) \right]$$

$$V_{12} = \bar{Y}^2 \left[\left(\frac{1}{n} - \frac{1}{N} \right) C_y^2 + \left(\frac{1}{n} - \frac{1}{n_1} \right) \left(\rho_{x_1 x_2} C_{x_1} C_{x_2} - \rho_{y x_1} C_y C_{x_1} \right) \right.$$

$$\left. - \left(\frac{1}{n} - \frac{1}{n_2} \right) \rho_{y x_2} C_y C_{x_2} \right]$$

The inverse of the matrix A is given by

$$A^{-1} = \frac{1}{|A|} \begin{bmatrix} V_{22} & -V_{12} \\ -V_{12} & V_{11} \end{bmatrix}$$

where

$$|A| = (V_{11})(V_{22}) - (V_{12})^2$$

The weights w_1 and w_2 are given by

$$w_1 = \frac{T_1}{T} = \frac{(V_{22} - V_{12})/|A|}{(V_{22} - 2V_{12} + V_{11})/|A|} = \frac{V_{22} - V_{12}}{V_{22} - 2V_{12} + V_{11}}$$

$$w_2 = \frac{T_2}{T} = \frac{(V_{11} - V_{12})/|A|}{(V_{22} - 2V_{12} + V_{11})/|A|} = \frac{V_{11} - V_{12}}{V_{22} - 2V_{12} + V_{11}}$$

The variance of the estimate $\bar{y}_{wR(M.V.)}$ is given by

$$V[\bar{y}_{wR(M.V.)}] = \frac{1}{T} = \frac{1}{(V_{22} - 2V_{12} + V_{11})/A} = \frac{|A|}{(V_{22} - 2V_{12} + V_{11})}$$

Now

$$V_{22} - V_{12} = \bar{Y}^2\left[\left(\frac{1}{n} - \frac{1}{n_2}\right)\left(C_{x_2}^2 - \rho_{yx_2} C_y C_{x_2}\right) - \left(\frac{1}{n} - \frac{1}{n_1}\right)\left(\rho_{x_1x_2} C_{x_1} C_{x_2} - \rho_{yx_1} C_y C_{x_1}\right)\right]$$

$$V_{11} - V_{12} = \bar{Y}^2\left[\left(\frac{1}{n} - \frac{1}{n_1}\right)\left(C_{x_1}^2 - \rho_{yx_1} C_y C_{x_1} - \rho_{x_1x_2} C_{x_1} C_{x_2}\right) + \left(\frac{1}{n} - \frac{1}{n_2}\right)\rho_{yx_2} C_y C_{x_2}\right]$$

$$V_{22} - 2V_{12} + V_{11} = \bar{Y}^2\left[\left(\frac{1}{n} - \frac{1}{n_1}\right)\left(C_{x_1}^2 - 2\rho_{x_1x_2} C_{x_1} C_{x_2}\right) + \left(\frac{1}{n} - \frac{1}{n_2}\right)C_{x_2}^2\right]$$

$$|A| = V_{11} \cdot V_{22} - (V_{12})^2$$

$$= \bar{Y}^4 C_y^4 \left[\left(\frac{1}{n} - \frac{1}{N}\right)\left(\frac{1}{n} - \frac{1}{n_1}\right)\rho_{yx_1}\left(\rho_{yx_1} - 2\rho_{x_1x_2}\rho_{yx_2}\right)\right.$$

$$+ \left(\frac{1}{n} - \frac{1}{N}\right)\cdot\left(\frac{1}{n} - \frac{1}{n_2}\right)\rho_{yx_2}^2 + \left(\frac{1}{n} - \frac{1}{n_1}\right)\cdot\left(\frac{1}{n} - \frac{1}{n_2}\right)\rho_{yx_1}\rho_{yx_2}\left(\rho_{yx_1}\rho_{yx_2} - 2\rho_{x_1x_2}\rho_{yx_2}\right)$$

$$\left. - \left(\frac{1}{n} - \frac{1}{n_1}\right)^2 \rho_{yx_1}^2\left(\rho_{x_1x_2}\rho_{yx_2} - \rho_{yx_1}\right)^2 - \left(\frac{1}{n} - \frac{1}{n_2}\right)^2 \rho_{yx_2}^4\right]$$

Hence

$$w_1 = \frac{-\left(\frac{1}{n}-\frac{1}{n_1}\right)\left(\rho_{x_1x_2}C_{x_1}C_{x_2}-\rho_{yx_1}C_yC_{x_1}\right)+\left(\frac{1}{n}-\frac{1}{n_2}\right)\left(C_{x_2}^2-\rho_{yx_2}C_yC_{x_2}\right)}{\left(\frac{1}{n}-\frac{1}{n_1}\right)\left(C_{x_1}^2-2\rho_{x_1x_2}C_{x_1}C_{x_2}\right)+\left(\frac{1}{n}-\frac{1}{n_2}\right)C_{x_2}^2}$$

$$w_2 = \frac{\left(\frac{1}{n}-\frac{1}{n_1}\right)\left(C_{x_1}^2-\rho_{yx_1}C_yC_{x_1}-\rho_{x_1x_2}C_{x_1}C_{x_2}\right)+\left(\frac{1}{n}-\frac{1}{n_2}\right)\rho_{yx_2}C_yC_{x_2}}{\left(\frac{1}{n}-\frac{1}{n_1}\right)\left(C_{x_1}^2-2\rho_{x_1x_2}C_{x_1}C_{x_2}\right)+\left(\frac{1}{n}-\frac{1}{n_2}\right)C_{x_2}^2}$$

Now

$$A = V_{11}\cdot V_{22}-\left(V_{12}\right)^2$$

$$= \bar{Y}^4\Big[\Big\{\left(\frac{1}{n}-\frac{1}{N}\right)C_y^2+\left(\frac{1}{n}-\frac{1}{n_1}\right)\left(C_{x_1}^2-2\rho_{yx_1}C_yC_{x_1}\right)\Big\}\Big\{\left(\frac{1}{n}-\frac{1}{N}\right)C_y^2+\left(\frac{1}{n}-\frac{1}{n_2}\right)\left(C_{x_2}^2-2\rho_{yx_2}C_yC_{x_2}\right)\Big\}-\Big\{\left(\frac{1}{n}-\frac{1}{N}\right)C_y^2+\left(\frac{1}{n}-\frac{1}{n_1}\right)\left(\rho_{x_1x_2}C_{x_1}C_{x_2}-\rho_{yx_1}C_yC_{x_1}\right)-\left(\frac{1}{n}-\frac{1}{n_2}\right)\rho_{yx_2}C_yC_{x_2}\Big\}^2\Big]$$

$$= \bar{Y}^4\Big[\left(\frac{1}{n}-\frac{1}{N}\right)\left(\frac{1}{n}-\frac{1}{n_2}\right)C_y^2C_{x_2}^2+\left(\frac{1}{n}-\frac{1}{N}\right)\left(\frac{1}{n}-\frac{1}{n_1}\right)C_y^2\left(C_{x_1}^2-2\rho_{x_1x_2}C_{x_1}C_{x_2}\right)+\left(\frac{1}{n}-\frac{1}{n_1}\right)\left(\frac{1}{n}-\frac{1}{n_2}\right)\left(C_{x_1}^2C_{x_2}^2-2\rho_{yx_2}C_yC_{x_1}^2C_{x_2}-2\rho_{yx_1}C_yC_{x_1}C_{x_2}^2+2\rho_{yx_1}\rho_{yx_2}C_y^2C_{x_1}C_{x_2}+2\rho_{yx_2}\rho_{x_1x_2}C_yC_{x_1}C_{x_2}^2\right)-\left(\frac{1}{n}-\frac{1}{n_1}\right)^2\left(\rho_{x_1x_2}C_{x_1}C_{x_2}-\rho_{yx_1}C_yC_{x_1}\right)^2-\left(\frac{1}{n}-\frac{1}{n_2}\right)^2\rho_{yx_2}^2C_y^2C_{x_2}^2\Big]$$

Applying the condition of unbiasedness, $C_{x_i} = \rho_{yx_i}C_y$, we get

$$A = \bar{Y}^4 C_y^4 \Big[\left(\frac{1}{n} - \frac{1}{N}\right)\left(\frac{1}{n} - \frac{1}{n_1}\right)\rho_{yx_1}\left(\rho_{yx_1} - 2\rho_{yx_2}\rho_{x_2x_1}\right)$$

$$+ \left(\frac{1}{n} - \frac{1}{N}\right)\left(\frac{1}{n} - \frac{1}{n_2}\right)\rho_{yx_2}^2 - \left(\frac{1}{n} - \frac{1}{n_1}\right)\left(\frac{1}{n} - \frac{1}{n_2}\right)\rho_{yx_1}\rho_{yx_2}^2\left(\rho_{yx_1} -$$

$$2\rho_{yx_2}\rho_{x_2x_1}\right) - \left(\frac{1}{n} - \frac{1}{n_1}\right)^2\rho_{yx_1}^2\left(\rho_{yx_1} - \rho_{yx_2}\rho_{x_2x_1}\right)^2 + \left(\frac{1}{n} - \frac{1}{n_2}\right)^2\rho_{yx_2}^4 \Big]$$

Hence

$$V\left[\bar{y}_{wR(M.V.)}\right] = \frac{Nr.}{Dr.}, \qquad \ldots (2.4.8)$$

where

$$Nr. = \bar{Y}^4 C_y^4 \Big[\left(\frac{1}{n} - \frac{1}{N}\right)\left(\frac{1}{n} - \frac{1}{n_1}\right)\rho_{yx_1}\left(\rho_{yx_1} - 2\rho_{yx_2}\rho_{x_2x_1}\right)$$

$$+ \left(\frac{1}{n} - \frac{1}{N}\right)\left(\frac{1}{n} - \frac{1}{n_2}\right)\rho_{yx_2}^2 - \left(\frac{1}{n} - \frac{1}{n_1}\right)\left(\frac{1}{n} - \frac{1}{n_2}\right)\rho_{yx_1}\rho_{yx_2}^2\left(\rho_{yx_1} -$$

$$2\rho_{yx_2}\rho_{x_2x_1}\right) - \left(\frac{1}{n} - \frac{1}{n_1}\right)^2\rho_{yx_1}^2\left(\rho_{yx_1} - \rho_{yx_2}\rho_{x_2x_1}\right)^2 + \left(\frac{1}{n} - \frac{1}{n_2}\right)^2\rho_{yx_2}^4 \Big]$$

and $Dr. = \bar{Y}^2 C_y^2 \Big[\left(\frac{1}{n} - \frac{1}{n_1}\right)\rho_{yx_1}\left(\rho_{yx_1} - 2\rho_{yx_2}\rho_{x_2x_1}\right) + \left(\frac{1}{n} - \frac{1}{n_2}\right)\rho_{yx_2}^2 \Big]$

Second Estimate $(\bar{y}_{w/r})$

The multivariate regression estimate is given by

$$\bar{y}_{w/r} = \bar{y}_n + \sum_{i=1}^{p} b_i \left(\bar{x}_{in_i} - \bar{x}_{in}\right) \quad \text{where} \quad b_i = \frac{s_{yx_i}}{s_{x_i}^2}$$

Let

$$\bar{x}_{in} = \bar{X}_i + \epsilon_i \quad , \quad s_{yx_i} = S_{yx_i} + k_i$$

$$\bar{x}_{in_i} = \bar{X}_i + \epsilon_i' \quad , \quad s_{x_i}^2 = S_{x_i}^2 + k_i'$$

Then

$$\bar{y}_{w/s} = \bar{y}_n + \sum_{i=1}^{p} \left(\frac{S_{yx_i} + k_i}{S^2_{x_i} + k'_i} \right) (\epsilon'_i - \epsilon_i)$$

$$= \bar{y}_n + \sum_{i=1}^{p} \beta_i \left(1 + \frac{k_i}{S_{yx_i}} \right) \left(1 - \frac{k'_i}{S^2_{x_i}} \right) (\epsilon'_i - \epsilon_i)$$

where $\beta_i = \frac{S_{yx_i}}{S^2_{x_i}}$.

$$= \bar{y}_n + \sum_{i=1}^{p} \beta_i \left(1 + \frac{k_i}{S_{yx_i}} - \frac{k'_i}{S^2_{x_i}} \right) (\epsilon'_i - \epsilon_i)$$

neglecting the term $\left(\frac{k_i}{S_{yx_i}} \cdot \frac{k'_i}{S^2_{x_i}} \right)$

Now

$$E(\bar{y}_{w/s}) = \bar{Y} + \sum_{i=1}^{p} \beta_i \left[\frac{E(k_i \epsilon'_i) - E(k_i \epsilon_i)}{S_{yx_i}} - \frac{E(k'_i \epsilon'_i) - E(k'_i \epsilon_i)}{S^2_{x_i}} \right]$$

$$= \bar{Y} + \sum_{i=1}^{p} \beta_i \left[\frac{\mathrm{Cov}(s_{yx_i}, \bar{x}_{in_i}) - \mathrm{Cov}(s_{yx_i}, \bar{x}_{in})}{S_{yx_i}} - \right.$$

$$\left. \frac{\mathrm{Cov}(s^2_{x_i}, \bar{x}_{in_i}) - \mathrm{Cov}(s^2_{x_i}, \bar{x}_{in})}{S^2_{x_i}} \right]$$

$$E(\bar{y}_{w/x}) = \bar{Y} - \sum_{i=1}^{p} \beta_i \left[\frac{\mathrm{Cov}(s_{yx_i}, \bar{x}_{in})}{S_{yx_i}} - \frac{\mathrm{Cov}(s^2_{x_i}, \bar{x}_{in})}{S^2_{x_i}} \right]$$

as $\mathrm{Cov}(s_{yx_i}, \bar{x}_{in_i}) = 0 = \mathrm{Cov}(s^2_{x_i}, \bar{x}_{in_i})$

because n_i is fixed here.

Using the method of symmetric functions or polykays and neglecting terms of order $\frac{1}{n^a}$, $a > 1$, it can be shown that

$$\mathrm{Cov}(\bar{x}_n, s_{yx}) = \frac{N-n}{Nn} \mu_{21}$$

where $\mu_{21} = E\left[(x - \bar{x}_N)^2 (y - \bar{Y})\right]$

and

$$\mathrm{Cov}(\bar{x}_n, s^2_x) = \frac{N-n}{Nn} \mu_{30}$$

where $\mu_{30} = E(x - \bar{x}_N)^3$.

Hence

$$E(\bar{y}_{w/x}) = \bar{Y} - \sum_{i=1}^{p} \beta_i \left(\frac{1}{n} - \frac{1}{N}\right) \left[\frac{\mu_{21}^{x_i}}{S_{yx_i}} - \frac{\mu_{30}^{x_i}}{S^2_{x_i}} \right]$$

So

$$\text{Bias in } (\bar{y}_{w/x}) = E(\bar{y}_{w/x}) - \bar{Y}$$

$$= - \sum_{i=1}^{p} \beta_i \left(\frac{1}{n} - \frac{1}{N}\right) \left[\frac{\mu_{21}^{x_i}}{S_{yx_i}} - \frac{\mu_{30}^{x_i}}{S^2_{x_i}} \right] \quad \ldots (2.4.9)$$

The bias in $(\bar{y}_{w\,lr})$ will be negligible if the sample size, n is sufficiently large and it will vanish if

$$\frac{\mu_{21}^{x_i}}{S_{y x_i}} = \frac{\mu_{30}^{x_i}}{S^2_{x_i}} \quad \forall \; i = 1, 2, \ldots, p$$

or

$$\beta_i = \frac{S_{yx_i}}{S^2_{x_i}} = \frac{\mu_{21}^{x_i}}{\mu_{30}^{x_i}} \quad \forall \; i = 1, 2, \ldots, p$$

When the sample size is sufficiently large, the multivariate regression estimate $\bar{y}_{w\,lr}$ behaves like the estimator given by

$$\bar{y}'_{w\,lr} = \bar{y}_n + \sum_{i=1}^{p} \beta_i \, (\bar{x}_{in_i} - \bar{x}_{in})$$

which is an unbiased estimate of $\bar{Y}$.

Now the variance of $\bar{y}_{w\,lr}$, when it satisfies the condition of unbiasedness is given by $V(\bar{y}_{w\,lr}) = E(\bar{y}'_{w\,lr} - \bar{Y})^2$

$$= E\left[(\bar{y}_n - \bar{Y}) + \sum_{i=1}^{p} \beta_i (\bar{x}_{in_i} - \bar{x}_{in})\right]^2$$

$$= E\left[(\bar{y}_n - \bar{Y}) + \sum_{i=1}^{p} \beta_i (\epsilon'_i - \epsilon_i)\right]^2$$

$$= E(\bar{y}_n - \bar{Y})^2 + \sum_{i=1}^{p} \beta_i^2 E(\epsilon'_i - \epsilon_i)^2 + 2\sum_{i=1}^{p} \beta_i E(\bar{y}_n - \bar{Y}) \cdot (\epsilon'_i - \epsilon_i) + 2\sum_{i=1}^{p} \sum_{j=1}^{p} \beta_i \beta_j E(\epsilon'_i - \epsilon_i)(\epsilon'_j - \epsilon_j)$$

$$(i < j)$$

Now $E(\bar{y}_n - \bar{Y})^2 = V(\bar{y}_n) = (\frac{1}{n} - \frac{1}{N}) S_y^2$

$$E(\epsilon_i' - \epsilon_i)^2 = E(\epsilon_i'^2) + E(\epsilon_i^2) - 2E(\epsilon_i' \epsilon_i)$$

$$= (\frac{1}{n_i} - \frac{1}{N}) S_{x_i}^2 + (\frac{1}{n} - \frac{1}{N}) S_{x_i}^2 - 2(\frac{1}{n_i} - \frac{1}{N}) S_{x_i}^2$$

$$= (\frac{1}{n} - \frac{1}{n_i}) S_{x_i}^2$$

$$E(\bar{y}_n - \bar{Y})(\epsilon_i' - \epsilon_i) = E[\epsilon_i' (\bar{y}_n - \bar{Y})] - E[\epsilon_i (\bar{y}_n - \bar{Y})]$$

$$= [(\frac{1}{n_i} - \frac{1}{N}) - (\frac{1}{n} - \frac{1}{N})] S_{yx_i}$$

$$= -(\frac{1}{n} - \frac{1}{n_i}) S_{yx_i}$$

$$E(\epsilon_i' - \epsilon_i)(\epsilon_j' - \epsilon_j) = E(\epsilon_i' \epsilon_j' - \epsilon_i' \epsilon_j - \epsilon_i \epsilon_j' + \epsilon_i \epsilon_j)$$

$$= [(\frac{1}{n_j} - \frac{1}{N}) - (\frac{1}{n_i} - \frac{1}{N}) - (\frac{1}{n_j} - \frac{1}{N}) + (\frac{1}{n} - \frac{1}{N})] S_{x_i x_j}$$

$$(i < j)$$

$$= (\frac{1}{n} - \frac{1}{n_i}) S_{x_i x_j} \qquad (i < j)$$

Substituting these values in the variance expression of $\bar{y}_{w\,l\,x}$, we get

$$V(\bar{y}_{w\,l\,x}) = \left(\frac{1}{n} - \frac{1}{N}\right)S_y^2 + \sum_{i=1}^{p} \beta_i^2 \left(\frac{1}{n} - \frac{1}{n_i}\right)S_{x_i}^2 - 2\sum_{i=1}^{p} \beta_i \left(\frac{1}{n} - \frac{1}{n_i}\right)S_{yx_i}$$

$$+ 2 \sum_{i=1}^{p} \sum_{j=1}^{p} \beta_i \beta_j \left(\frac{1}{n} - \frac{1}{n_i}\right) S_{x_i x_j} \qquad (i < j)$$

Now

$$\beta_i = \frac{S_{y x_i}}{S_{x_i}^2} = \frac{\rho_{y x_i} S_y}{S_{x_i}}$$

So $\beta_i^2 S_{x_i}^2 = \rho_{y x_i}^2 S_y^2$, and

$$\beta_i S_{y x_i} = \rho_{y x_i}^2 S_y^2 .$$

Hence

$$V(\bar{y}_{w\,l\,x}) = S_y^2 \Bigg[\left(\frac{1}{n} - \frac{1}{N}\right) - \sum_{i=1}^{p} \left(\frac{1}{n} - \frac{1}{n_i}\right) \rho_{y x_i}^2$$

$$+ 2 \sum_{i=1}^{p} \sum_{j=1}^{p} \left(\frac{1}{n} - \frac{1}{n_i}\right) \rho_{x_i x_j} \rho_{y x_i} \rho_{y x_j} \Bigg]$$

$$(i < j)$$

... (2.4.10)

A Particular Case (when $p = 2$)

When there are only two auxiliary characters, in that case the variance of $\bar{y}_{w\,/\,r}$ is given by

$$V(\bar{y}_{w\,/\,r}) = S_y^2 \Big[(\frac{1}{n} - \frac{1}{N}) - (\frac{1}{n} - \frac{1}{n_1}) \rho_{yx_1}^2 - (\frac{1}{n} - \frac{1}{n_2}) \rho_{yx_2}^2 + 2 (\frac{1}{n} - \frac{1}{n_1}) \rho_{yx_1} \rho_{yx_2} \rho_{x_1x_2} \Big] \quad \ldots (2.4.11)$$

Third Estimate ($\bar{y}_{u\,/\,r}$)

It is also a multivariate regression estimate given by

$$\bar{y}_{u\,/\,r} = \bar{y}_n + \sum_{i=1}^{p} b_i (\bar{x}_{in_i} - \bar{x}_{in_{i-1}})$$

Let

$$\bar{x}_{in_{i-1}} = \bar{X}_i + \epsilon_i \quad ; \qquad s_{yx_i} = S_{yx_i} + k_i$$

$$\bar{x}_{in_i} = \bar{X}_i + \epsilon_i' \quad ; \qquad s_{x_i}^2 = S_{x_i}^2 + k_i'$$

Then

$$\bar{y}_{u\,/\,r} = \bar{y}_n + \sum_{i=1}^{p} \Big(\frac{S_{yx_i} + k_i}{S_{x_i}^2 + k_i'} \Big) (\epsilon_i' - \epsilon_i)$$

$$= \bar{y}_n + \sum_{i=1}^{p} \beta_i \Big(1 + \frac{k_i}{S_{yx_i}} \Big) \Big(1 - \frac{k_i'}{S_{x_i}^2} \Big) (\epsilon_i' - \epsilon_i)$$

where $\beta_i = (S_{yx_i} / S_{x_i}^2)$

$$= \bar{y}_n + \sum_{i=1}^{p} \beta_i \Big(1 + \frac{k_i}{S_{yx_i}} - \frac{k_i'}{S_{x_i}^2} \Big) (\epsilon_i' - \epsilon_i)$$

Now

$$E(\bar{y}_{u/x}) = \bar{Y} + \sum_{i=1}^{p} \beta_i E\left(1 + \frac{k_i}{S_{yx_i}} - \frac{k_i'}{S_{x_i}^2}\right)(\epsilon_i' - \epsilon_i)$$

$$= \bar{Y} + \sum_{i=1}^{p} \beta_i \Bigg[- \frac{Cov(s_{yx_i}, \bar{x}_{in_i}) - Cov(s_{yx_i}, \bar{x}_{in_{i-1}})}{S_{yx_i}}$$

$$- \frac{Cov(s_{x_i}^2, \bar{x}_{in_i}) - Cov(s_{x_i}^2, \bar{x}_{in_{i-1}})}{S_{x_i}^2} \Bigg]$$

$$= \bar{Y} - \sum_{i=1}^{p} \beta_i \Bigg[\frac{Cov(s_{yx_i}, \bar{x}_{in_{i-1}})}{S_{yx_i}} - \frac{Cov(s_{x_i}^2, \bar{x}_{in_{i-1}})}{S_{x_i}^2} \Bigg]$$

as $Cov(s_{yx_i}, \bar{x}_{in_i}) = 0 = Cov(s_{x_i}^2, \bar{x}_{in_i})$

because n_i is fixed here.

Using the same method of symmetric functions or polyknays and neglecting terms of order $\frac{1}{n^a}$ $(a > 1)$, it can be shown that

$$Cov(\bar{x}_n, s_{yx}) = \frac{N-n}{Nn} \mu_{21} \text{ where } \mu_{21} = E(x - \bar{x}_N)^2 (y - \bar{Y})$$

$$Cov(\bar{x}_n, s_x^2) = \frac{N-n}{Nn} \mu_{30} \text{ where } \mu_{30} = E(x - \bar{x}_N)^3$$

$$\text{So } E(\bar{y}_{u/x}) = \bar{Y} - \sum_{i=1}^{p} \beta_i \Bigg[\frac{\mu_{21}^{x_i}}{S_{yx_i}} - \frac{\mu_{30}^{x_i}}{S_{x_i}^2} \Bigg] \left(\frac{1}{n_{i-1}} - \frac{1}{N}\right)$$

So

$$\text{Bias in } (\bar{y}_{u\,/r}) = E(\bar{y}_{u\,/r}) - \bar{Y}$$

$$= -\sum_{i=1}^{p} \beta_i \left(\frac{1}{n_{i-1}} - \frac{1}{N}\right)\left[\frac{\mu_{21}^{x_i}}{S_{yx_i}} - \frac{\mu_{30}^{x_i}}{S_{x_i}^2}\right] \quad \ldots \quad (2.4.12)$$

The bias in $(\bar{y}_{u\,/r})$ will be negligible if the sample size n_{i-1} is sufficiently large and it will vanish if

$$\frac{\mu_{21}^{x_i}}{S_{yx_i}} = \frac{\mu_{30}^{x_i}}{S_{x_i}^2} \quad \forall \quad i = 1, 2, \ldots, p$$

or

$$\beta_i = \frac{S_{yx_i}}{S_{x_i}^2} = \frac{\mu_{21}^{x_i}}{\mu_{30}^{x_i}} \quad \forall \quad i = 1, 2, \ldots, p$$

When the sample size is sufficiently large, the multivariate regression estimate $\bar{y}_{u\,/r}$ behaves like the estimator given by

$$\bar{y}'_{u\,/r} = \bar{y}_n + \sum_{i=1}^{p} \beta_i \left(\bar{x}_{in_i} - \bar{x}_{in_{i-1}}\right)$$

which is an unbiased estimate of $\bar{Y}$.

Now the variance of $\bar{y}_{u\,/r}$, when it satisfies the condition of unbiasedness, is given by

$$V(\bar{y}_{u\,lr}) = E(\bar{y}'_{u\,lr} - \bar{Y})^2$$

$$= E\left[(\bar{y}_n - \bar{Y}) + \sum_{i=1}^{p} \beta_i (\bar{x}_{in_i} - \bar{x}_{in_{i-1}})\right]^2$$

$$= E\left[(\bar{y}_n - \bar{Y}) + \sum_{i=1}^{p} \beta_i (\epsilon'_i - \epsilon_i)\right]^2$$

$$= E\left[(\bar{y}_n - \bar{Y})^2 + \sum_{i=1}^{p} \beta_i^2 (\epsilon'_i - \epsilon_i)^2 + 2\sum_{i=1}^{p} \beta_i (\bar{y}_n - \bar{Y})(\epsilon'_i - \epsilon_i)\right.$$
$$\left. + 2\sum_{i=1}^{p}\sum_{j=1}^{p} \beta_i \beta_j (\epsilon'_i - \epsilon_i)(\epsilon'_j - \epsilon_j)\right]$$
$$(i<j)$$

Now

$$E(\bar{y}_n - \bar{Y})^2 = V(\bar{y}_n) = \left(\frac{1}{n} - \frac{1}{N}\right) S_y^2$$

$$E(\epsilon'_i - \epsilon_i)^2 = E(\epsilon'^2_i) + E(\epsilon_i^2) - 2E(\epsilon'_i \epsilon_i)$$

$$= \left(\frac{1}{n_i} - \frac{1}{N}\right) S_{x_i}^2 + \left(\frac{1}{n_{i-1}} - \frac{1}{N}\right) S_{x_i}^2 - 2\left(\frac{1}{n_i} - \frac{1}{N}\right) S_{x_i}^2$$

$$= \left(\frac{1}{n_{i-1}} - \frac{1}{n_i}\right) S_{x_i}^2$$

$$E(\bar{y}_n - \bar{Y})(\epsilon'_i - \epsilon_i) = E\left[\epsilon'_i (\bar{y}_n - \bar{Y})\right] - E\left[\epsilon_i (\bar{y}_n - \bar{Y})\right]$$

$$= \left(\frac{1}{n_i} - \frac{1}{N}\right) S_{yx_i} - \left(\frac{1}{n_{i-1}} - \frac{1}{N}\right) S_{yx_i}$$

$$= -\left(\frac{1}{n_{i-1}} - \frac{1}{n_i}\right) S_{yx_i}$$

$$E(\epsilon_i' - \epsilon_i)(\epsilon_j' - \epsilon_j) = E(\epsilon_i' \epsilon_j') - E(\epsilon_i' \epsilon_j) - E(\epsilon_i \epsilon_j') + E(\epsilon_i \epsilon_j)$$

$$= (\frac{1}{n_j} - \frac{1}{N}) S_{x_i x_j} - (\frac{1}{n_{j-1}} - \frac{1}{N}) S_{x_i x_j} - (\frac{1}{n_j} - \frac{1}{N}) S_{x_i x_j}$$

$$+ (\frac{1}{n_{j-1}} - \frac{1}{N}) S_{x_i x_j}$$

$$= 0$$

Substituting these values in the variance expression of $\bar{y}_{u\,\ell\,r}$.

we get

$$V(\bar{y}_{u\,\ell\,r}) = (\frac{1}{n} - \frac{1}{N}) S_y^2 + \sum_{i=1}^{p} (\frac{1}{n_{i-1}} - \frac{1}{n_i}) \beta_i^2 S_{x_i}^2 - 2 \sum_{i=1}^{p} (\frac{1}{n_{i-1}} - \frac{1}{n_i}) \beta_i S_{yx_i}$$

Now

$$\beta_i = \frac{S_{y x_i}}{S_{x_i}^2} = \frac{\rho_{yx_i} S_y}{S_{x_i}}$$

So $\beta_i^2 S_{x_i}^2 = \rho_{yx_i}^2 S_y^2$, and

$\beta_i S_{yx_i} = \rho_{yx_i}^2 S_y^2$.

Hence

$$V(\bar{y}_{u\,\ell\,r}) = (\frac{1}{n} - \frac{1}{N}) S_y^2 + \sum_{i=1}^{p} (\frac{1}{n_{i-1}} - \frac{1}{n_i}) \rho_{yx_i}^2 S_y^2 - 2 \sum_{i=1}^{p} (\frac{1}{n_{i-1}} - \frac{1}{n_i}) \rho_{yx_i}^2 S_y^2$$

$$= S_y^2 \left[(\frac{1}{n} - \frac{1}{N}) - \sum_{i=1}^{p} (\frac{1}{n_{i-1}} - \frac{1}{n_i}) \rho_{yx_i}^2 \right] \quad \ldots (2.4.13)$$

A Particular Case (when p = 2)

When there are only two auxiliary characters, in that case the variance of $\bar{y}_{u\,/\,r}$ is given by

$$V(\bar{y}_{u/r}) = S_y^2 \Big/ \left(\frac{1}{n} - \frac{1}{N}\right) + \left(\frac{1}{n} - \frac{1}{n_1}\right) \rho^2_{yx_1} - \left(\frac{1}{n_1} - \frac{1}{n_2}\right) \rho^2_{yx_2} \Big/ \quad \ldots (2.4.14)$$

2.5 Comparison between First, Second and Third Estimate

Second Estimate ($\bar{y}_{w\,/\,r}$) and Third Estimate ($\bar{y}_{u\,/\,r}$) .

From (2.4.10) and (2.4.13) it is clear that

$$V(\bar{y}_{w/r}) - V(\bar{y}_{u/r}) = 2 S_y^2 \sum_{i=1}^{p} \sum_{\substack{j=1 \\ (i<j)}}^{p} \left(\frac{1}{n} - \frac{1}{n_i}\right) \rho_{x_i x_j} \rho_{yx_i} \rho_{yx_j} \quad \ldots (2.5.1)$$

On the R.H.S. $S_y^2 \left(\frac{1}{n} - \frac{1}{n_i}\right)$ is positive always (as $n < n_i$) so $V(\bar{y}_{w\,/\,r})$ will be greater than $V(\bar{y}_{u\,/\,r})$ if $\rho_{x_i x_j}$, ρ_{yx_i} and ρ_{yx_j} are positive.

Hence $\bar{y}_{u\,/\,r}$ is more efficient than $\bar{y}_{w\,/\,r}$ if $\rho_{x_i x_j}$, ρ_{yx_i} and ρ_{yx_j} are all positive.

First Estimate ($\bar{y}_{wR(M.V.)}$) and Second Estimate ($\bar{y}_{w\,/\,r}$) .

As it is very difficult to calculate the variance of ($\bar{y}_{wR(M.V.)}$) for p auxiliary characters, so here we will

consider the particular case when $p = 2$.

The estimate $(\bar{y}_{w\ell x})$ is more efficient than $(\bar{y}_{wR(M.V.)})$ if

$$V\left[\bar{y}_{wR(M.V.)}\right] - V\left(\bar{y}_{w\ell x}\right) > 0$$

i.e.

$$\bar{Y}^4 C_y^4 \Big/ \Big\{ \left(\frac{1}{n} - \frac{1}{N}\right)\left(\frac{1}{n} - \frac{1}{n_1}\right)\rho_{yx_1}\left(\rho_{yx_1} - 2\rho_{yx_2}\rho_{x_2x_1}\right)$$
$$+\left(\frac{1}{n} - \frac{1}{N}\right)\left(\frac{1}{n} - \frac{1}{n_2}\right)\rho_{yx_2}^2 - \left(\frac{1}{n} - \frac{1}{n_1}\right)\left(\frac{1}{n} - \frac{1}{n_2}\right)\rho_{yx_1}\rho_{yx_2}^2\left(\rho_{yx_1} - 2\rho_{yx_2}\rho_{x_2x_1}\right)$$
$$-\left(\frac{1}{n} - \frac{1}{n_1}\right)^2\rho_{yx_1}^2\left(\rho_{yx_1} - \rho_{yx_2}\rho_{x_2x_1}\right)^2 - \left(\frac{1}{n} - \frac{1}{n_2}\right)^2\rho_{yx_2}^4 \Big\}$$
$$-\Big\{\left(\frac{1}{n} - \frac{1}{n_1}\right)\rho_{yx_1}\left(\rho_{yx_1} - 2\rho_{yx_2}\rho_{x_2x_1}\right) + \left(\frac{1}{n} - \frac{1}{n_2}\right)\rho_{yx_2}^2\Big\} \cdot$$
$$\Big\{\left(\frac{1}{n} - \frac{1}{N}\right) - \left(\frac{1}{n} - \frac{1}{n_1}\right)\rho_{yx_1}\left(\rho_{yx_1} - 2\rho_{yx_2}\rho_{x_2x_1}\right) - \left(\frac{1}{n} - \frac{1}{n_2}\right)\rho_{yx_2}^2\Big\}$$

is greater than 0 .

or

$$\bar{Y}^4 C_y^4 \left(\frac{1}{n} - \frac{1}{n_1}\right)\left(\frac{1}{n} - \frac{1}{n_2}\right)\rho_{yx_1}\rho_{yx_2}^2\left(\rho_{yx_1} - 2\rho_{yx_2}\rho_{x_2x_1}\right) > 0$$

Now $\bar{Y}^4 C_y^4 \rho_{yx_2}^2 \left(\frac{1}{n} - \frac{1}{n_1}\right)\left(\frac{1}{n} - \frac{1}{n_2}\right)$ is positive $(n < n_1 < n_2)$.

So $\bar{y}_{w\ell x}$ will be more efficient than $\bar{y}_{wR(M.V.)}$ if

$$\rho_{yx_1}\left(\rho_{yx_1} - 2\rho_{yx_2}\rho_{x_2x_1}\right) > 0$$

which $\Rightarrow$ that ρ_{yx_1} should be positive and $\rho_{yx_1} - 2\rho_{yx_2}\rho_{x_2x_1} > 0$

$$\text{or } \rho_{x_2x_1} < \frac{1}{2}\frac{\rho_{yx_1}}{\rho_{yx_2}} = \frac{1}{2}\frac{C_{x_1}}{C_{x_2}}$$

hence the multivariate regression estimate $\bar{y}_{w\ell r}$ is more efficient than the multivariate ratio estimate $\bar{y}_{wR(M.V.)}$ if

(i) the correlation coefficient ρ_{yx_1} is positive,

(ii) the correlation coefficient $\rho_{x_2x_1}$ is less than half of the ratio of coefficient of variation of x_1 (C_{x_1}) and x_2 (C_{x_2}).

So in general

$$V(\bar{y}_{u\ell r}) < V(\bar{y}_{w\ell r}) < V(\bar{y}_{wR}) \text{ if}$$

(i) the correlation coefficients ρ_{yx_i} ($i = 1, \ldots, p$) are positive,

(ii) the correlation coefficients $\rho_{x_ix_j}$ ($i \neq j$) are positive

(iii) the correlation coefficients $\rho_{x_ix_j}$ is less than half of the ratio of coefficients of variation of x_i (C_{x_i}) and x_j (C_{x_j}).

CHAPTER - III

OPTIMUM SAMPLE SIZES FOR DIFFERENT CHARACTERS

3.1 Improved Estimates of the Mean of Auxiliary Characters

Now we will find the optimum values of sample sizes, n, n_i ($i = 1, \ldots, p$) for which the variance is minimum. For calculating these values, we first of all find the improved estimate of $\bar{X}_1$ depending upon $X_2, \ldots, X_p$; improved estimate of $\bar{X}_2$ depending upon $X_3, \ldots, X_p$; ...; improved estimate of $\bar{X}_{p-1}$ depending on X_p. We can only find the improved estimates of $\bar{X}_i$ depending on $X_{i+1}, \ldots, X_p$; but not on $X_{i-1}, \ldots, X_1$ as these characters are being observed on the lesser number of sample units on which X_i is observed. So we can not find an improved estimate of $\bar{X}_p$.

The improved estimates of $\bar{X}_1, \bar{X}_2, \bar{X}_3, \ldots, \bar{X}_{p-1}$ are given as follows:

$$\bar{x}_{1\,lr} = \bar{x}_1 + \beta_{12}(\bar{x}_{2n_2} - \bar{x}_{2n_1}) + \beta_{13}(\bar{x}_{3n_3} - \bar{x}_{3n_2}) + \ldots + \beta_{1p}(\bar{x}_{pn_p} - \bar{x}_{pn_{p-1}})$$

$$\bar{x}_{2\,lr} = \bar{x}_2 + \beta_{23}(\bar{x}_{3n_3} - \bar{x}_{3n_2}) + \beta_{24}(\bar{x}_{4n_4} - \bar{x}_{4n_3}) + \ldots + \beta_{2p}(\bar{x}_{pn_p} - \bar{x}_{pn_{p-1}})$$

$$\bar{x}_{3\,lr} = \bar{x}_3 + \beta_{34}(\bar{x}_{4n_4} - \bar{x}_{4n_3}) + \beta_{35}(\bar{x}_{5n_5} - \bar{x}_{5n_4}) + \ldots + \beta_{3p}(\bar{x}_{pn_p} - \bar{x}_{pn_{p-1}})$$

$$\vdots \qquad \vdots \qquad \vdots \qquad \vdots \qquad \vdots \qquad \ldots \qquad \vdots \qquad \vdots \qquad \vdots$$

$$\bar{x}_{p-1\,lr} = \bar{x}_{p-1} + \beta_{p\,p-1}(\bar{x}_{pn_p} - \bar{x}_{pn_{p-1}})$$

So in general the improved estimate of $\bar{X}_i$ depending upon $X_{i+1}, \ldots, X_p$ is given by

$$\bar{x}_{i\,f_p} = \bar{x}_i + \beta_{i,i+1}(\bar{x}_{i+1\,n_{i+1}} - \bar{x}_{i+1\,n_i}) + \beta_{i,i+2}(\bar{x}_{i+2\,n_{i+2}} - \bar{x}_{i+2\,n_{i+1}}) + \ldots + \beta_{ip}(\bar{x}_{pn_p} - \bar{x}_{pn_{p-1}})$$

$$= \bar{x}_i + \sum_{k=1}^{p-i} \beta_{i,i+k}(\bar{x}_{i+k\,n_{i+k}} - \bar{x}_{i+k\,n_{i+k-1}}) \quad \ldots(3.1.1)$$

Now taking expectation, we get

$E(\bar{x}_{i\,f_p}) = \bar{X}_i$ as $E(\bar{x}_i) = \bar{X}_i$ and the expectation of the other term is 0.

So $\bar{x}_{i\,f_p}$ is an improved unbiased estimate of $\bar{X}_i$.

3.2 Variance of the Improved Estimate $(\bar{x}_{i\,f_p})$

The variance of $\bar{x}_{i\,f_p}$ is given by

$$V(\bar{x}_{i\,f_p}) = E(\bar{x}_{i\,f_p} - \bar{X}_i)^2$$

$$= E\left[(\bar{x}_i - \bar{X}_i) + \sum_{k=1}^{p-i} \beta_{i,i+k}(\bar{x}_{i+k\,n_{i+k}} - \bar{x}_{i+k\,n_{i+k-1}})\right]^2$$

$$= E\left[(\bar{x}_i - \bar{X}_i) + \sum_{k=1}^{p-i} \beta_{i,i+k}(\epsilon'_{i+k} - \epsilon_{i+k})\right]^2$$

where $\epsilon_{i+k} = \bar{x}_{i+k\,n_{i+k-1}} - \bar{X}_{i+k}$ and $E(\epsilon_{i+k}) = 0$

$\epsilon'_{i+k} = \bar{x}_{i+k\,n_{i+k}} - \bar{X}_{i+k}$ and $E(\epsilon'_{i+k}) = 0$.

$$V(\bar{x}_{l/x}) = E(\bar{x}_l - \bar{X}_l)^2 + \sum_{k=1}^{p-l} \beta_{l,l+k}^2 E(\epsilon'_{l+k} - \epsilon_{l+k})^2 + 2 \sum_{k=1}^{p-l} \beta_{l,l+k} E(\bar{x}_l - \bar{X}_l)(\epsilon'_{l+k} - \epsilon_{l+k})$$

$$+ \sum_{k=1}^{p-1} \sum_{\ell=1}^{p-l} \beta_{l,l+k} \beta_{l,l+\ell} E(\epsilon'_{l+k} - \epsilon_{l+k})(\epsilon'_{l+\ell} - \epsilon_{l+\ell}) \quad \ldots (3.2.1)$$

$(k < \ell)$

Now

$$E(\bar{x}_l - \bar{X}_l)^2 = V(\bar{x}_l) = \left(\frac{1}{n_l} - \frac{1}{N}\right) S_{x_l}^2$$

$$E(\epsilon'_{l+k} - \epsilon_{l+k})^2 = E(\epsilon'^2_{l+k}) + E(\epsilon^2_{l+k}) - 2E(\epsilon'_{l+k}\ \epsilon_{l+k})$$

$$= \left(\frac{1}{n_{l+k}} - \frac{1}{N}\right) S^2_{x_{l+k}} + \left(\frac{1}{n_{l+k-1}} - \frac{1}{N}\right) S^2_{x_{l+k}} - 2\left(\frac{1}{n_{l+k}} - \frac{1}{N}\right) S^2_{x_{l+k}}$$

$$= \left(\frac{1}{n_{l+k-1}} - \frac{1}{n_{l+k}}\right) S^2_{x_{l+k}}$$

$$E(\bar{x}_l - \bar{X}_l)(\epsilon'_{l+k} - \epsilon_{l+k}) = E[(\bar{x}_l - \bar{X}_l)\epsilon'_{l+k}] - E[(\bar{x}_l - \bar{X}_l)\epsilon_{l+k}]$$

$$= \left(\frac{1}{n_{l+k}} - \frac{1}{N}\right) S_{x_l x_{l+k}} - \left(\frac{1}{n_{l+k-1}} - \frac{1}{N}\right) \cdot S_{x_l x_{l+k}}$$

$$= -\left(\frac{1}{n_{l+k-1}} - \frac{1}{n_{l+k}}\right) S_{x_l x_{l+k}}$$

$$E\left(\epsilon'_{i+k}-\epsilon_{i+k}\right)\left(\epsilon'_{i+\ell}-\epsilon_{i+\ell}\right)=E\left(\epsilon'_{i+k}\epsilon'_{i+\ell}-\epsilon'_{i+k}\epsilon_{i+\ell}-\epsilon_{i+k}\epsilon'_{i+\ell}+\epsilon_{i+k}\,\epsilon_{i+\ell}\right)$$

$$=\left(\frac{1}{n_{i+\ell}}-\frac{1}{N}\right)-\left(\frac{1}{n_{i+\ell-1}}-\frac{1}{N}\right)-\left(\frac{1}{n_{i+\ell}}-\frac{1}{N}\right)+\left(\frac{1}{n_{i+\ell-1}}-\frac{1}{N}\right)S_{x_{i+\ell}x_{i+k}}$$

$$(k<\ell)$$

$$=0$$

Substituting these values in (3.2.1), we get

$$V(\bar{x}_{i\ell x})=\left(\frac{1}{n_i}-\frac{1}{N}\right)S^2_{x_i}+\sum_{k=1}^{p-i}\left(\frac{1}{n_{i+k-1}}-\frac{1}{n_{i+k}}\right)\beta^2_{i,i+k}\,S^2_{x_{i+k}}$$

$$-2\sum_{k=1}^{p-i}\left(\frac{1}{n_{i+k-1}}-\frac{1}{n_{i+k}}\right)\beta_{i,i+k}\,S_{x_i x_{i+k}}$$

where $$\beta_{i,i+k}=\frac{S_{x_i x_{i+k}}}{S^2_{x_{i+k}}}=\frac{\rho_{x_i x_{i+k}}\,S_{x_i}}{S_{x_{i+k}}}$$

so $$\beta^2_{i,i+k}\,S^2_{x_{i+k}}=\rho^2_{x_i x_{i+k}}\,S^2_{x_i}$$

and $$\beta_{i,i+k}\,S_{x_i x_{i+k}}=\rho^2_{x_i x_{i+k}}\,S^2_{x_i}$$

Substituting the above values in the variance expression, we get

$$V(\bar{x}_{i\ell x})=\left(\frac{1}{n_i}-\frac{1}{N}\right)S^2_{x_i}-\sum_{k=1}^{p-i}\left(\frac{1}{n_{i+k-1}}-\frac{1}{n_{i+k}}\right)\rho^2_{x_i x_{i+k}}\,S^2_{x_i} \qquad \ldots (3.2.2)$$

3.3 Precision for estimating $\bar{X}_i$

Suppose if we want to estimate, $\bar{X}_i$ with λ_i per cent precision then $\lambda_i = \frac{\sigma_{\bar{x}_i}/x}{\bar{X}_i}$. Now dividing both the sides of (3.2.2) by $\bar{X}_i^2$, we get

$$\frac{V(\bar{x}_i/x)}{\bar{X}_i^2} = \lambda_i^2 = \left[\left(\frac{1}{n_i} - \frac{1}{N}\right) - \sum_{k=1}^{p-i}\left(\frac{1}{n_{i+k-1}} - \frac{1}{n_{i+k}}\right)\rho^2_{x_i x_{i+k}}\right] C^2_{x_i}$$

where $C_{x_i} = \frac{S_{x_i}}{\bar{X}_i}$, coefficient of variation of x_i.

The above equation can be written as follows:

$$\frac{\lambda_i^2}{C^2_{x_i}} = \left(\frac{1}{n_i} - \frac{1}{N}\right) - \sum_{k=1}^{p-i}\left(\frac{1}{n_{i+k-1}} - \frac{1}{n_{i+k}}\right)\rho^2_{x_i x_{i+k}} \quad \ldots (3.3.1)$$

$i = 1, 2, \ldots, p.$

3.4 To find the value of optimum sample sizes for different character

By putting $i = 1, 2, \ldots, p$ in (3.3.1) we will get p equations containing $n_1, n_2, \ldots, n_p$. The variance expression of the estimate of $\bar{Y}$ will contain $n, n_1, \ldots, n_p$. So in all we will get $(p+1)$ equations in terms of $n, n_1, \ldots, n_p$. Solving these $(p+1)$ equations, we can get the values of the optimum sample sizes in terms of C_y, C_{x_i} $(i = 1, \ldots, p)$ and λ, λ_i $(i = 1, \ldots, p)$ and correlation coefficient between y and x_i i.e. ρ_{yx_i} $(i = 1, \ldots, p)$ and the correlation coefficient between x_i

and x_j i.e. $\rho_{x_i x_j}$ $(i \neq j)$. So knowing all these values, we can find the values of optimum sample sizes n, n_i $(i = 1, 2, \ldots, p)$.

3.5 Particular Case (when $p = 3$)

When there are only three auxiliary characters X_1, X_2 and X_3, to find the values of n, n_1, n_2 and n_3, we proceed as follows.

From (3.3.1) we have

$$\left(\frac{\lambda_1}{C_1}\right)^2 = \left(\frac{1}{n_1} - \frac{1}{N}\right) - \left(\frac{1}{n_1} - \frac{1}{n_2}\right)\rho_{12}^2 - \left(\frac{1}{n_2} - \frac{1}{n_3}\right)\rho_{13}^2 \qquad \ldots(3.5.1)$$

$$\left(\frac{\lambda_2}{C_2}\right)^2 = \left(\frac{1}{n_2} - \frac{1}{N}\right) - \left(\frac{1}{n_2} - \frac{1}{n_3}\right)\rho_{23}^2 \qquad \ldots(3.5.2)$$

$$\left(\frac{\lambda_3}{C_3}\right)^2 = \left(\frac{1}{n_3} - \frac{1}{N}\right) \qquad \ldots(3.5.3)$$

where $\rho_{ij} = \rho_{x_i x_j}$ and $C_i = C_{x_i}$.

Suppose N is large so that we can neglect $\frac{1}{N}$.

From equation (3.5.3), we get

$$n_3 = \frac{1}{(\lambda_3/C_3)^2} \qquad \ldots(3.5.4)$$

From equation (3.5.2), we get

$$(\lambda_2/C_2)^2 = \frac{1}{n_2}(1 - \rho_{23}^2) + (\lambda_3/C_3)^2 \rho_{23}^2$$

$$\text{or} \quad n_2 = \frac{(1-\rho_{23}^2)}{(\lambda_2/C_2)^2 - \rho_{23}^2(\lambda_3/C_3)^2} \qquad \ldots(3.5.5)$$

From equation (3.5.1),we get

$$(\lambda_1/C_1)^2 = \frac{1}{n_1}(1-\rho_{12}^2) + \frac{1}{n_2}(\rho_{12}^2 - \rho_{13}^2) + \frac{1}{n_3}\rho_{13}^2$$

$$= \frac{1}{n_1}(1-\rho_{12}^2) + \frac{(\rho_{12}^2-\rho_{13}^2)}{(1-\rho_{23}^2)}\left[\left(\frac{\lambda_2}{C_2}\right)^2 - \rho_{23}^2\left(\frac{\lambda_3}{C_3}\right)^2\right] + \left(\frac{\lambda_3}{C_3}\right)^2\rho_{13}^2$$

So

$$n_1 = \frac{(1-\rho_{12}^2)}{\left(\frac{\lambda_1}{C_1}\right)^2 - \frac{(\rho_{12}^2-\rho_{13}^2)}{(1-\rho_{23}^2)}\left[\left(\frac{\lambda_2}{C_2}\right)^2 - \rho_{23}^2\left(\frac{\lambda_3}{C_3}\right)^2\right] - \left(\frac{\lambda_3}{C_3}\right)^2\rho_{13}^2} \qquad \ldots(3.5.6)$$

Now we have to take into consideration a variance expression of the estimate of population mean $\bar{Y}$.

Let us consider the third estimate, $\bar{y}_{u\,lr}$, which is the most efficient out of the three estimates.

Nowfrom equation (3.4.13), we have

$$V(\bar{y}_{u\,lr}) = \left(\frac{1}{n} - \frac{1}{N}\right)S_y^2 - \sum_{i=1}^{p}\left(\frac{1}{n_{i-1}} - \frac{1}{n_i}\right)\rho_{yx_i}^2 S_y^2$$

so

$$\frac{V(\bar{y}_{u\,lr})}{\bar{Y}^2} = \lambda_y^2 = \left[\left(\frac{1}{n} - \frac{1}{N}\right) - \sum_{i=1}^{p}\left(\frac{1}{n_{i-1}} - \frac{1}{n_i}\right)\rho_{yx_i}^2\right]C_y^2$$

where $\lambda_y = \dfrac{\sigma_{\bar{y}_{u\,lr}}}{\bar{Y}}$, percentage of precision for estimating $\bar{Y}$,

and $C_y = \dfrac{S_y}{\bar{Y}}$, coefficient of variation of y.

From the above equation, we get (for p = 3)

$$\left(\frac{\lambda_y}{C_y}\right)^2 = \left(\frac{1}{n} - \frac{1}{N}\right)\left[\left(\frac{1}{n} - \frac{1}{n_1}\right)\rho_{01}^2 + \left(\frac{1}{n_1} - \frac{1}{n_2}\right)\rho_{02}^2 + \left(\frac{1}{n_2} - \frac{1}{n_3}\right)\rho_{03}^2\right]$$

...(3.5.7)

where $\rho_{01} = \rho_{y z_1}$

Neglecting $\frac{1}{N}$, equation (3.5.7) gives

$$\left(\frac{\lambda_y}{C_y}\right)^2 = \frac{1}{n}(1 - \rho_{01}^2) + \frac{1}{n_1}(\rho_{01}^2 - \rho_{02}^2) + \frac{1}{n_2}(\rho_{02}^2 - \rho_{03}^2) + \frac{1}{n_3}\cdot\rho_{03}^2$$

...(3.5.8)

Substituting the values of n_1, n_2 and n_3 from equations (3.5.6), (3.5.5) and (3.5.4) in (3.5.8), we get

$$\left(\frac{\lambda_y}{C_y}\right)^2 = \frac{1}{n}(1-\rho_{01}^2) + \frac{(\rho_{01}^2 - \rho_{02}^2)}{(1-\rho_{12}^2)}\left[\left(\frac{\lambda_1}{C_1}\right)^2 - \left\{\left(\frac{\lambda_2}{C_2}\right)^2 - \rho_{23}^2\left(\frac{\lambda_3}{C_3}\right)^2\right\}\cdot\right.$$

$$\left.\cdot\left\{\frac{(\rho_{12}^2 - \rho_{13}^2)}{(1-\rho_{23}^2)}\right\} - \left(\frac{\lambda_3}{C_3}\right)^2\rho_{13}^2\right] + \frac{(\rho_{02}^2 - \rho_{03}^2)}{(1-\rho_{23}^2)}\cdot$$

$$\cdot\left[\left(\frac{\lambda_2}{C_2}\right)^2 - \left(\frac{\lambda_3}{C_3}\right)^2\rho_{23}^2\right] + \left(\frac{\lambda_3}{C_3}\right)^2\rho_{03}^2$$

So n is given by

$$n = \frac{(1-\rho_{01}^2)}{A_3} \qquad \ldots(3.5.9)$$

where $A_3 = \left(\frac{\lambda_y}{C_y}\right)^2 - \frac{(\rho_{01}^2 - \rho_{02}^2)}{(1-\rho_{12}^2)}\left[\left(\frac{\lambda_1}{C_1}\right)^2 - \left\{\left(\frac{\lambda_2}{C_2}\right)^2 - \rho_{23}^2\left(\frac{\lambda_3}{C_3}\right)^2\right\}\right.$

$$\left. - \left(\frac{\lambda_3}{C_3}\right)^2 \rho_{13}^2\right] - \frac{(\rho_{02}^2 - \rho_{03}^2)}{(1-\rho_{23}^2)}\left[\left(\frac{\lambda_2}{C_2}\right)^2 - \rho_{23}^2\left(\frac{\lambda_3}{C_3}\right)^2\right]$$

$$- \left(\frac{\lambda_3}{C_3}\right)^2 \rho_{03}^2$$

So from equations (3.5.9), (3.5.6), (3.5.5) and (3.5.4), we can find the values of optimum sample sizes n, n_1, n_2 and n_3.

3.6 Another Particular Case (when $p = 2$)

When there are only two auxiliary characters X_1 and X_2, in that case we can very easily show that

$$n_2 = \frac{1}{(\lambda_2/C_2)^2} \qquad \ldots(3.6.1)$$

$$n_1 = \frac{(1-\rho_{12}^2)}{(\lambda_1/C_1)^2 - \rho_{12}^2 (\lambda_2/C_2)^2} \qquad \ldots(3.6.2)$$

and $$n = \frac{(1-\rho_{01}^2)}{A_2} \qquad \ldots(3.6.3)$$

where $A_2 = \left(\frac{\lambda_y}{C_y}\right)^2 - \frac{(\rho_{01}^2 - \rho_{02}^2)}{(1-\rho_{12}^2)}\left[\left(\frac{\lambda_1}{C_1}\right)^2 - \rho_{12}^2 \left(\frac{\lambda_2}{C_2}\right)^2\right] - \rho_{02}^2\left(\frac{\lambda_2}{C_2}\right)^2$

3.7 Generalization

Seeing the results of Section 3.5 and 3.6, we can generalise these results for the case when there are p auxiliary characters.

In this case n_i is given by

$$n_i = \frac{(1-\rho^2_{i,i+1})}{D_i} \quad ; \quad i = 1, \ldots, p \qquad \ldots (3.7.1)$$

where

$$D_i = \left(\frac{\lambda_i}{C_i}\right)^2 - \left[\left(\frac{\lambda_{i+1}}{C_{i+1}}\right)^2 - \rho^2_{i+1,i+2}\left(\frac{\lambda_{i+2}}{C_{i+2}}\right)^2\right] \frac{(\rho^2_{i,i+1} - \rho^2_{i,i+2})}{(1-\rho^2_{i+1,i+2})}$$

$$- \left[\left(\frac{\lambda_{i+2}}{C_{i+2}}\right)^2 - \rho^2_{i+2,i+3}\left(\frac{\lambda_{i+3}}{C_{i+3}}\right)^2\right] \frac{(\rho^2_{i,i+2} - \rho^2_{i,i+3})}{(1-\rho^2_{i+2,i+3})}$$

$$\vdots \qquad \vdots \qquad \vdots \qquad \vdots \qquad \vdots \qquad \vdots$$

$$- \left(\frac{\lambda_p}{C_p}\right)^2 \rho^2_{i,p}$$

$$= \left(\frac{\lambda_i}{C_i}\right)^2 - \sum_{k=1}^{p} \left[\left(\frac{\lambda_{i+k}}{C_{i+k}}\right)^2 - \rho^2_{i+k,i+k+1}\left(\frac{\lambda_{i+k+1}}{C_{i+k+1}}\right)^2\right] \cdot$$

$$\cdot \left[\frac{(\rho^2_{i,i+k} - \rho^2_{i,i+k+1})}{(1-\rho^2_{i+k,i+k+1})}\right]$$

and n is given by

$$n = \frac{(1 - \rho_{01}^2)}{A_p} \qquad \ldots \quad (3.7.2)$$

where

$$A_p = \left(\frac{\lambda_y}{C_y}\right)^2 \Bigg/ \left(\frac{\lambda_1}{C_1}\right)^2 - \left\{\left(\frac{\lambda_2}{C_2}\right)^2 - \left(\frac{\lambda_3}{C_3}\right)^2 \rho_{23}^2\right\} \frac{(\rho_{12}^2 - \rho_{13}^2)}{(1-\rho_{23}^2)} - \left(\frac{\lambda_3}{C_3}\right)^2 \rho_{12}^2 \Bigg/ \cdot \frac{(\rho_{01}^2 - \rho_{02}^2)}{(1 - \rho_{12}^2)}$$

$$- \left(\frac{\lambda_2}{C_2}\right)^2 - \left\{\left(\frac{\lambda_3}{C_3}\right)^2 - \left(\frac{\lambda_4}{C_4}\right)^2 \rho_{34}^2\right\} \frac{(\rho_{23}^2 - \rho_{24}^2)}{(1-\rho_{34}^2)} - \left(\frac{\lambda_4}{C_4}\right)^2 \rho_{24}^2 \Bigg/ \cdot \frac{(\rho_{02}^2 - \rho_{03}^2)}{(1 - \rho_{23}^2)}$$

$$\vdots \qquad \vdots \qquad \vdots \qquad \vdots \qquad \vdots \qquad \vdots \qquad \vdots$$

$$- \left(\frac{\lambda_p}{C_p}\right)^2 \rho_{0p}^2$$

$$= \left(\frac{\lambda_y}{C_y}\right)^2 - \sum_{k=1}^{p} \Bigg/ \left(\frac{\lambda_k}{C_k}\right)^2 - \left\{\left(\frac{\lambda_{k+1}}{C_{k+1}}\right)^2 - \left(\frac{\lambda_{k+2}}{C_{k+2}}\right)^2 \rho_{k+1,k+2}^2\right\} \cdot$$

$$\cdot \frac{(\rho_{k,k+1}^2 - \rho_{k,k+2}^2)}{(1 - \rho_{k+1,k+2}^2)} - \left(\frac{\lambda_{k+2}}{C_{k+2}}\right)^2 \rho_{k,k+2}^2 \Bigg/ \cdot \frac{(\rho_{0,k}^2 - \rho_{0,k+1}^2)}{(1 - \rho_{k,k+1}^2)}$$

So the equations (3.7.2) and (3.7.1) gives the values of n, n_i ($i = 1, \ldots, p$), respective optimum sample sizes for y, x_i ($i = 1, \ldots, p$) when there are p auxiliary characters.

REFERENCES

1. Cochran, W.G. (1953). Sampling Techniques, New York: John Wiley and Sons.

2. Kathuria, O.P. (1959). "On Alternative Replacement Procedures in Sampling on Successive Occasions with a Two-stage Design and on Use of Multi-Auxiliary Information in such designs". Thesis submitted for the award of Ph.D. degree to I.A.R.I. (I.C.A.R.)

3. Lahiri, D.B. (1951). A method of sample selection providing unbiased ratio estimates, Bull. Inter. Stat. Inst. 33(2), 133-140.

4. Midzuno, H. (1950). An outline of the theory of sampling system. Ann. Inst. Stat. Math., Japan, 1, 149-156.

5. Olkin, I. (1958). Multi-variate ratio estimation for finite populations. Biometrika, 45, 154-165.

6. Robson, D.S. (1957). Application of multivariate polykays to the theory of unbiased ratio type estimators. Jour. Amer. Stat. Assoc., 52, 511-522.

7. Singh, B.D. (1962). Use of double sampling in repeated surveys. Unpublished thesis submitted towards fulfilment of requirements for Diploma, I.C.A.R., New Delhi.

8. Sukhatme, B.V. (1962). Some ratio type estimators in two-phase sampling. Jour. Amer. Stat. Assoc., 57, 628-632.

9. Sukhatme, P.V. (1954). Sampling Theory of Surveys with Applications, Iowa State College Press. Ames., U.S.A.