MULTIVARIATE UNBIASED RATIO-TYPE, PRODUCT-TYPE, AND RATIO-CUM-PRODUCT-TYPE ESTIMATORS IN FINITE POPULATIONS.

V. RAMACHANDRAN

Institute of Agricultural Research Statistics - Library - I.A.R.S.
Acc. 2600
C. No. 043:123

Dissertation submitted in fulfilment of the requirements for the award of Diploma in Agricultural and Animal Husbandry Statistics of the Institute of Agricultural Research Statistics (I.C.A.R.)

NEW DELHI
1969

ACKNOWLEDGEMENTS

I have great pleasure in expressing my deep sense of grattitude to Shri. S. S. Pillai, Senior Statistician, Institute of Agricultural Research Statistics for his valuable guidance, keen interest and constant encouragement throughout the course of investigation and immense help in writing the thesis.

I also express my grattitude to Dr. M. N. Das, Senior Professor of Statistics for his constant encouragement throughout the preparation of this thesis.

In the end I am highly thankful to Dr. G. R. Seth, Director, Institute of Agricultural Research Statistics, (I. C. A. R.) for providing me adequate facilities and to the Indian Council of Agricultural Research for providing financial assistance in the form of a fellowship.

Dated- 29 December, 1969. V. Ramachandran (V. Ramachandran)

CONTENTS

CHAPTER I
INTRODUCTION

From times immemorial the concept of generalising from a part of the population to the whole has been used more or less subjectively in daily life. But not until the later half of the nineteenth century objective methods of generalising from a part to the whole seem to have received much attention . In this case two questions arise, i) how to select the part from the whole and ii) how to generalise from the selcted part to the whole. The problem is one of finding that combination of selection and estimation proceeedures which would minimise the risk involved in generalising from the part to the whole per unit of cost. Alternatively the problem may be viewed as that of finding that combination of selection and estimation proceeedures which would minimise the cost, ensuring at the same time a specified precision for the inference from a part to the whole.

The earlier developments in this field relative to the second question posed above and the result has been a fairly well developed theory of estimation and statistical inference based on the simplest of selection proceedures, namely equal probability sampling with replacement. Improvements in selection proceedures may be considered to have been initiated by Bowly (1926) who used stratified simple random sampling with proportional allocation. Neyman (1934) considered the question of optimum allocation in stratified sampling. During the decade 1940-50, considerable developments in sampling theory have taken place and the works of Mahalonobis (1940, 1944 and 1946), Cochran (1942), Hansen and

Hurwitz(1943,1946) and Madow (1944) need special mention. Cochran considered the question of supplementary information at the estimation stage by using ratio and regression estimators. Cochran (1953), Deming (1950), Sukhatme (1953), Hansen, Hurwitz and Madow (1953) and Yates (1953) have covered in detail the earlier developments in the theory of sampling in respective books.

Simple random sampling is by far the most commonly used method of sampling in surveys. It is simple, operationally convenient and gives equal chance of selection for all the units in the population. When however, the units vary considerably in size as is often the case, simple random sampling does not take into account the possible importance of larger units in the population. Under such circumstances without foregoing the operational convenience of simple random sampling, it is desirable to use auxiliary information, such as size of unit at the estimation stage for obtaining more efficient estimators of the population value in the sense of giving estimators with smaller standard errors. Two examples of such estimation procedures are the ' Ratio and Regression' methods of estimation.

The two classical ratio estimators are the ratio of the mean estimator $\bar{y}_R = \dfrac{\bar{y}_n}{\bar{x}_n} \bar{X}$ or equivalantly the weighted mean of the ratios estimator

$$\bar{r}_w \bar{X} = \frac{\sum_{i=1}^{n} x_i r_i}{\sum_{i=1}^{n} x_i} \bar{X}$$

and the mean of the ratios estimator $\bar{y}_r = \bar{r}_n \bar{X}$, where $\bar{y}_n$ and $\bar{x}_n$ are the sample means, $\bar{r}_n$, the mean of the individual ratios $r = (y/x)$

in the sample and $\bar{X}$ is known population mean of the auxiliary variable x. Both the estimators are known to be biased. The latter is not even consistent. An exact expression for the bias in $\bar{y}_r$ is available which does not depend upon the sample size n. Since the unweighted mean $\bar{r}_n$ may be seriously biased if r_n tends to be larger (or smaller) for large x than for small x, the estimator $\bar{y}_r$ is likely to be more biased than the estimator $\bar{y}_R$ based on weighted mean $\bar{r}_w$. No exact expression for the bias in $\bar{y}_R$ is available, but for samples of moderate size, from populations in which the coefficient of variation of x is not too large, the bias in $\bar{y}_R$ is negligible. But the problem how large the sample should be to make the bias negligible has not yet been solved satisfactorily for all types of populations.

The available bias expressions and variance formula for both the regression estimator $\bar{y}_{lr}$ and the ratio estimator $\bar{y}_R$ are only approximate, the approximations assuming the sample size n to be sufficiently large. For small samples nothing is known about the nature of their bias and precision. This situation has led the research workers in the field to explore ways and means of obtaining ' Ratio and Regression ' type estimators which are either completely free from bias or subject to a smaller bias than the customary one.

Unbiased ratio type estimators have been evolved in recent years. Hartley and Ross (1954) have been the pioneers in the group of authors who obtained unbiased estimators under the commonly adopted sampling designs. In simple random sampling without replacement they have given an elegant expression for the bias in $\bar{y}_r$ and an unbiased estimator of that bias, thereby arriving at an unbiased ratio estimator

$$\bar{y}_r' = \bar{r}_n \bar{X} + \frac{(N-1)\,n}{N\,(n-1)} \left(\bar{y}_n - \bar{r}_n \bar{x}_n \right)$$

Robson (1957) has derived the exact formula for its variance and an unbiased estimate of the variance. In large samples, more simple estimators of the variance of $\bar{y}_r'$ and an extensive discussion of the relative efficiencies of $\bar{y}_R$, $\bar{y}_r$, $\bar{y}_r'$ have been given by Goodman and Hartley (1958).

Occasionally situations arise when the coefficient of correlation is negative. In such cases the product estimator $(\bar{y}_n \bar{x}_n / \bar{X}_N)$ proposed by Robson (1957) can be used with advantage. It can be seen that the product estimator will be more efficient than the simple mean estimator $\bar{y}_n$ if the correlation coefficient between x and y (ρ) is less than $-\frac{1}{2}\,\frac{C_y}{C_x}$. Robson has also derived an unbiased product type estimator corresponding to $\frac{\bar{y}_n \bar{x}_n}{\bar{X}_N}$ and he derived the exact variance of the estimator by using multivariate symmetric means.

Olkin (1958) has considered the use of multi-supplementary variables in building up ratio estimator and this estimator has been found to be more efficient. However when the supplementary variables are negatively correlated with the variable under consideration, the ratio method of estimation cannot be efficiently used. In such cases, M. P. Singh (1967 a) has considered the use of multivariate product estimator. He has obtained the exact expressions for the bias and the mean square error.

The next question arises is, can we make the best use of the combination of ratio and product methods of estimation. M. P. Singh (1967a) has considered the estimation procedure which is the combination of ratio and product methods. He has derived the conditions for their efficient use in sample surveys.

The method of estimation suggested by Olkin (1954) is extended to the case where some or all of the auxiliary variables are positively correlated and some or all are negatively correlated with y is dealt by S.K.Srivasthava (1965) and S.R.S.Rao and Govind.S.Mudholkar (1967). M.P. Singh (1967b) has considered the case of Multivariate product estimators (biased) and he extended it to Double sampling.

The present investigation can be summarised as follows.

Chapter II deals with the unbiased multivariate ratio type estimator. Exact variance and unbiased estimate of variance are given by using multivariate symmetric means. Comparisons have been made between different type of estimators with the help of some emperical data.

Chapter III deals with the unbiased ratio - cum - product type of estimate. The derivations of the exact variance and unbiased estimate of the variance are given.

Chapter IV deals with the multivariate unbiased product type estimator. The exact variance and unbiased estimate of the variance are obtained with the help of multivariate symmetric means.

Chapter V deals with the generalized multivariate unbiased estimation where some of the supplemenatary variables having + ve correlation and the rest having - ve correlation with the variable under study.

As in the above cases, the concept of multivariate symmetric means are used in finding the variance and the estimate of the variance.

In Chapter VI, an emperical stydy has been made by using these estimators and results are included as illustration.

CHAPTER II

MULTIVARIATE UNBIASED RATIO TYPE ESTIMATOR

2.1. INTRODUCTION

The unbiased ratio type estimator was proposed by Hartley and Ross (1954) in the case of simple random sampling without repacement. They have given an elegant expression for the bias in $\bar{y}_r$ and an unbiased estimate of that bias there by arriving at an unbiased ratio estimator,

$$\bar{y}_r' = \bar{r}_n \bar{X} + \frac{n\,(N-1)}{N\,(n-1)} \left(\bar{y}_n - \bar{r}_n \bar{x}_n \right) .$$

Robson (1957) has derived the exact formula for its variance and unbiased estimator of the variance by using the multi-variate symmetric means. In large samples more simple estimators of the variance of $\bar{y}_r'$ and an extensive discussion of the relative efficiencies of $\bar{y}_R$, $\bar{y}_r$, and $\bar{y}_r'$ have been given by Goodman and Hartley (1958). Olkin has considered the use of multi-supplementary variables in building up ratio estimator and this estimator has been found to be more efficienct

The present investigation is an extension of the usual Hartley and Ross unbiased ratio type estimator to the multivariate case. In section 2.2 exact expression for the variance and an unbiased estimate of the variance have been obtained and then comparisons have been made between different type of estimators in Section 2.3 with the help of emporical data.

2.2. MULTIVARIATE UNBIASED RATIO TYPE ESTIMATION.

Let there be 'p' supplementary variables $x_1, x_2, x_3, \ldots, x_p$; information which is available for each unit of the population. Let Y and X_i, $i = 1, 2, \ldots p$ be the population totals of the variable under study and the p supplementary variables. Let the population size be N, and a random sample of size n be drawn, the values of the character under study and the supllementary variables be

$(y_i, x_{1i}, x_{2i}, \ldots x_{pi}, r_{1i}, r_{2i}, \ldots r_{pi})$ $(i = 1, 2, \ldots n)$, where

$r_{ji} = y_{ji}/x_{ji}$ $(i = 1, 2, \ldots n, j = 1, 2, \ldots p)$.

Let $\bar{r}_j = (1/n) \sum_{i=1}^{n} y_{ji}/x_{ji}$ $(j = 1, 2, \ldots p)$

$\bar{y} = (1/n) \sum_{i=1}^{n} y_i$ and $\bar{x}_j = (1/n) \sum_{i=1}^{n} x_{ji}$ $(j = 1, 2, \ldots p)$

Let $\bar{y}_j' = \bar{r}_j \bar{X} - \frac{n(N-1)}{N(n-1)} (\bar{x}_j \bar{r}_j - \bar{y})$, $j = 1, 2, \ldots p$. (2.2.1.)

be the Hartley and Ross's unbiased ratio type estimators.

In terms of the symmetric means we can write,

$\bar{r}_j = \langle (\overset{1}{0} \overset{2}{0} 0 \ldots \overset{p}{0} \overset{1}{0} \ldots \overset{j}{1} \overset{j+1}{0} \ldots \overset{p}{0}) \rangle$ $(j = 1, 2, \ldots p)$

$\bar{x}_j = \langle (\overset{1}{0} \overset{2}{0} 0 \ldots \overset{j}{1} \overset{j+1}{0} \ldots \overset{p}{0} \overset{1}{0} \ldots \overset{p}{0}) \rangle$ $(j = 1, 2, \ldots p)$

$\bar{y} = \langle (1 \overset{1}{0} \overset{2}{0} \ldots \overset{p}{0} \overset{1}{0} \ldots \overset{p}{0}) \rangle$

Here the first element in the vector represents the y unit, the succeding p units for $x_1, x_2, \ldots x_p$ and the remaining p units for $r_1, r_2, \ldots r_p$ respectively.

An angle bracket with a prime $\langle \; \rangle'$ is used to denote a population symmetric mean while $\langle \; \rangle$ without prime denotes a sample symmetric mean. Adetailed method of multiplication of symmetric means has been given by Robson (1957).

We can write (2.2.1.) with symmetric mean notations but it will be tedius to write the expression in this form. Also to derive $V_{ij} = \text{cov}(\bar{y}_1', \bar{y}_2')$ keeping $\bar{y}_j'$ in symmetric mean notation in the form given above will be laborious. Hence we shall derive $V_{12} = (\bar{y}_1', \bar{y}_2')$, assuming that we have two supplementary information x_1, x_2 with y without loss of generality.

In terms of symmetric means notation, we have,

$$\bar{y}_1' = \langle(01000)\rangle' \langle(00010)\rangle' - \frac{(N-1)}{N}\left\{\langle(01000)\rangle\langle(00010)\rangle - \langle(10000)\rangle\right\}$$

$$\bar{y}_2' = \langle(00100)\rangle' \langle(00001)\rangle' - ((N-1)/N)\left\{\langle(00100)\rangle\langle(00001)\rangle - \langle(10000)\rangle\right\}$$

$$\text{Cov}(\bar{y}_1', \bar{y}_2') = E(\bar{y}_1' \bar{y}_2') - \bar{Y}^2 \quad \ldots\ldots \quad (2.2.2.)$$

Taking the expectations after multiplying the respective symmetric means, we have the exact expression of $\text{Cov}(\bar{y}_1', \bar{y}_2')$ as

$$\text{Cov}(\bar{y}_1' \bar{y}_2') = \frac{(N-1)}{N^2 n(n-1)}\Big\{ \langle(01100),(00011)\rangle' + \langle(01001),(00110)\rangle'$$

$$+ (n-2)\langle(01100),(00010),(00001)\rangle' + (n-2)\langle(01001),(00100),(00010)\rangle'$$

$$+ (n-2)\langle(01000),(00110),(00001)\rangle' + (n-2)\langle(01000),(00101),(00010)\rangle'$$

$$+ (n-2)(n-3)\langle(01000),(00100),(00010),(00001)\rangle'\Big\}$$

$$+ \frac{(N-1)^2}{nN^2}\Big\{ \langle(20000)\rangle' + (n-1)\langle(10000),(10000)\rangle'$$

$$- \langle(11000),(00100)\rangle' - \langle(01000),(10010)\rangle'$$

$$- (n-2)\langle(10000),(01000),(00010)\rangle' - \langle(10100),(00001)\rangle'$$

$$- \langle(00100),(10001)\rangle' - (n-2)\langle(10000),(00100),(00001)\rangle'\Big\}$$

$$+ \frac{1}{N^2 n} \Big\{ \langle (01111) \rangle' + (N-1) \langle (01100), (00011) \rangle'$$

$$+ (n-1) \langle (01110), (00001) \rangle' + (n-1) \langle (01101), (00010) \rangle'$$

$$+ (n-1)(N-2) \langle (01100), (00010), (00001) \rangle' + (N-1) \langle (00111), (01000) \rangle'$$

$$+ (N-1) \langle (01011), (00100) \rangle' + (N-1)(N-2) \langle (00011), (01000), (00100) \rangle'$$

$$+ (n-1) \langle (01010), (00101) \rangle' + (n-1) \langle (01001), (00110) \rangle'$$

$$+ (n-1)(N-2) \langle (01010), (00100), (00001) \rangle'$$

$$+ (n-1)(N-2) \langle (01001), (00100), (00010) \rangle'$$

$$+ (n-1)(N-2) \langle (01000), (00010), (00101) \rangle'$$

$$+ (n-1)(N-2) \langle (01000), (00110), (00001) \rangle'$$

$$+ (n-1)(N-2)(N-3) \langle (01000), (00100), (00010), (00001) \rangle' \Big\}$$

$$- \frac{(N-1)}{N^2 n} \Big\{ \langle (02010), (00010) \rangle' + \langle (01010), (01010) \rangle'$$

$$+ (N-2) \langle (01010), (01000), (00010) \rangle' + \langle (00201), (00001) \rangle'$$

$$+ \langle (00101), (00101) \rangle' + (N-2) \langle (00101), (00100), (00100) \rangle'$$

$$+ \langle (02000), (00020) \rangle' + \langle (01020), (01000) \rangle'$$

$$+ (N-2) \langle (00020), (01000), (01000) \rangle' + \langle (00102), (00100) \rangle'$$

$$+ \langle (00200), (00002) \rangle' + (N-2) \langle (00002), (00100), (00100) \rangle'$$

$$+ (n-2) \langle (02000), (00100), (00010) \rangle'$$

$$+ 2(n-2) \langle (01000), (01010), (00010) \rangle'$$

$$+ (N-3)(n-2) \langle (01000), (01000), (00010), (00010) \rangle'$$

$$+ (n-2) \langle (00200), (00001), (00001) \rangle'$$

$$+ 2(n-2) \langle (00101), (00100), (00001) \rangle'$$

$$+ (n-2)(N-3) \langle (00100), (00100), (00001), (00001) \rangle' - \langle (11010) \rangle'$$

$$- (N-1) \langle (10010), (01000) \rangle - (n-1) \langle (11000), (00010) \rangle$$
$$- (n-1) \langle (10000), (01010) \rangle - (n-1)(N-2) \langle (01000), (00010), (10000) \rangle$$
$$- \langle (10101) \rangle - (N-1) \langle (00100), (10001) \rangle$$
$$- (n-1) \langle (10100), (00001) \rangle - (n-1) \langle (10000), (00101) \rangle$$
$$- (n-1)(N-2) \langle (00100), (10000), (00001) \rangle \Big\}$$
$$- \frac{1}{N} \langle (20000) \rangle - \frac{(N-1)}{N} \langle (10000), (10000) \rangle \qquad (2.2.3)$$

An unbiased estimate of cov $(\bar{y}_1' \bar{y}_2')$ can be obtained by writing the same expression, replacing the population symmetric means with the respective sample symmetric means. We assume that N is very large. Then we can omit terms of the order of $(1/N)$. Further we can use the approximation

$$\langle (a_1), (a_2), \ldots\ldots (a_n) \rangle = \langle (a_1) \rangle \langle (a_2) \rangle \ldots \langle (a_n) \rangle$$

Under these assumptions (2.2.3) becomes

$$\mathrm{Cov}(\bar{y}_1', \bar{y}_2') = \frac{1}{n} \langle (20000) \rangle - \langle (10000) \rangle \langle (10000) \rangle$$
$$- \langle (11000) \rangle \langle (00010) \rangle + \langle (10000) \rangle \langle (01000) \rangle \langle (00010) \rangle$$
$$- \langle (10100) \rangle \langle (00001) \rangle + \langle (10000) \rangle \langle (00100) \rangle \langle (00001) \rangle$$
$$- \langle (00100) \rangle \langle (00001) \rangle \{ \langle (00101) \rangle - \langle (00100) \rangle \langle (00001) \rangle \}$$
$$- \langle (00100) \rangle \langle (00100) \rangle \{ \langle (00002) \rangle - \langle (00001) \rangle \langle (00001) \rangle \}$$
$$- \langle (01000) \rangle \langle (01000) \rangle \{ \langle (00020) \rangle - \langle (00010) \rangle \langle (00010) \rangle \}$$
$$- \langle (01000) \rangle \langle (00010) \rangle \{ \langle (01010) \rangle - \langle (01000) \rangle \langle (00010) \rangle \}$$
$$+ 2 \langle (01000) \rangle \langle (00001) \rangle \{ \langle (00110) \rangle - \langle (00100) \rangle \langle (00010) \rangle \}$$
$$+ \langle (01000) \rangle \langle (00100) \rangle \{ \langle (00011) \rangle - \langle (00010) \rangle \langle (00001) \rangle \}$$
$$+ \langle (00010) \rangle \langle (00001) \rangle \{ \langle (01100) \rangle - \langle (01000) \rangle \langle (00100) \rangle \}$$

$$+ \langle(00100)\rangle' \langle(00010)\rangle' \left\{ \langle(01001)\rangle' - \langle(01000)\rangle' \langle(00001)\rangle' \right\} \Big] \Big]$$

$$- \frac{1}{n(n-1)} \Big\{ \langle(01100)\rangle' \langle(00011)\rangle' - \langle(01100)\rangle' \langle(00010)\rangle' \langle(00001)\rangle'$$

$$- \langle(00011)\rangle' \langle(01000)\rangle' \langle(00100)\rangle'$$

$$+ \langle(01000)\rangle' \langle(00100)\rangle' \langle(00010)\rangle' \langle(00001)\rangle'$$

$$+ \langle(01001)\rangle' \langle(00110)\rangle' - \langle(01001)\rangle' \langle(00100)\rangle' \langle(00010)\rangle'$$

$$- \langle(00110)\rangle' \langle(01000)\rangle' \langle(00001)\rangle'$$

$$+ \langle(01000)\rangle' \langle(00100)\rangle' \langle(00010)\rangle' \langle(00001)\rangle' \Big\} \Big] \qquad (2.2.4)$$

Expressing the symmetric means interms of the usual notations,

we have

$$\text{Cov}(y_1\, y_2) = (1/n) \Big(\sigma_{yy} - \bar{R}_1 \sigma_{x_1 y} - \bar{R}_2 \sigma_{x_2 y} - \bar{X}_2 \bar{R}_2 \sigma_{x_1 r_1}$$

$$- \bar{X}_1^2 \sigma_{r_1 r_1} - \bar{X}_2^2 \sigma_{r_2 r_2} + \bar{R}_2 \bar{X}_1 \sigma_{x_2 r_1}$$

$$+ \bar{R}_1 \bar{R}_2 \sigma_{x_1 x_2} + X_2 R_1 \sigma_{x_1 r_2} + 2 X_1 X_2 \sigma_{r_1 r_2} \Big)$$

$$+ \frac{1}{n(n-1)} \left(\sigma_{x_1 x_2} \sigma_{r_1 r_2} + \sigma_{x_1 r_2} \sigma_{x_2 r_1} \right) \qquad (2.2.5)$$

Where the symbols have the usual meaning.

In general

$$\text{Cov}(\bar{y}_i'\, \bar{y}_j') = (1/n) \Big(\sigma_{yy} - \bar{R}_i \sigma_{x_i y} - \bar{R}_j \sigma_{x_j y} - \bar{X}_j \bar{R}_j \sigma_{x_j r_j}$$

$$- \bar{X}_i \bar{R}_i \sigma_{x_i r_i} - \bar{X}_j^2 \sigma_{r_j r_j} - \bar{X}_i^2 \sigma_{r_i r_i}$$

$$+ \bar{R}_j \bar{X}_i \sigma_{x_j r_i} + \bar{X}_j \bar{R}_i \sigma_{r_j x_i} + 2 X_i X_j \sigma_{r_i r_j}$$

$$+ \bar{R}_i \bar{R}_j \sigma_{x_i x_j} \Big)$$

$$+ \frac{1}{n(n-1)} \left(\sigma_{x_i x_j} \sigma_{r_i r_j} + \sigma_{x_i r_j} \sigma_{x_j r_i} \right) \qquad (2.2.6)$$

$$\mathrm{Var}(\bar{y}_i') = (1/n)\left(\sigma_{yy} - 2\bar{R}_i\,\sigma_{x_i y} + \bar{R}_i^2\,\sigma_{x_i x_i}\right) + \frac{1}{n(n-1)}\left(\sigma_{x_i x_i} + \sigma_{r_i r_i} + \sigma_{x_i r_j}\ \sigma_{x_j r_i}\right) \qquad (2.2.7)$$

An unbiased estimate of (2.2.7) is explained in the paper by Goodman and Hartley (1958). Using the same technique the unbiased estimate of (2.2.7) can be obtained and is omitted here.

Using Olkin's method, we can define the multivariate unbiased ratio type estimate of the population mean as

$$\bar{y}'_{r(M.V.)} = \sum_{j=1}^{p} w_j\,\bar{y}'_j \qquad \ldots\ldots \qquad (2.2.8)$$

where $\bar{y}'_j$ is given by (2.2.1) and weights w_j are chosen such that their sum is equal to 1.

(2.2.8) can be written in the matrix form as $\bar{y}'_{r(M.V.)} = W'\ \bar{y}'_r$

where $W = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_p \end{bmatrix}$ and $\bar{y}'_r = \begin{bmatrix} \bar{y}'_1 \\ \bar{y}'_2 \\ \vdots \\ \bar{y}'_p \end{bmatrix}$

and W' is the transpose of W. If $e' = (1, 1, \ldots 1)$ ($1 \times p$ matrix)

then $e'W = 1$. $\qquad \ldots\ldots\ldots\ldots\ldots \qquad$ (2.2.9)

If V is the Variance, Covariance matrix then V is given by

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

where V_{ij} is given by (2.2.6) and V_{ii} by (2.2.7)

The weights w_j ($j = 1, 2, \ldots p$) are chosen so as to minimise the variance of $\bar{y}'_{r(M.V.)}$.

Again follwing the algebra of Olkin, we have $w_i = R_i/D$ where R_i is the sum of all the elements in the ith row of the matrix V^{-1} and D is the sum of all the elements in the matrix V^{-1}, V^{-1} being the inverse of V. The minimum variance of $\bar{y}'_{r(M.V.)}$ is equal to 1/D. For estimating the weights and the variance of the estimated mean using all the auxiliary variables in the multivariate unbiased estimator, the estimates of V_{ij} at (2.2.6) and (2.2.7) are substituted for V_{ij} in the matrix V.

Comparisons between Multi-variate unbiased ratio type estimators with different number of supplementary variables:

Theorem. Let $\bar{y}'_{rq}$ and $\bar{y}'_{rk}$ denote respectively the multivariate unbiased ratio type estimators of $\bar{Y}$ with optimum weights based on the supplementary variables (x_1), (x_2), . . . (x_q), and (x_1), (x_2), . . . (x_k) where k is greater than q.

Then $V(\bar{y}'_{rq}) \geq V(\bar{y}'_{rk})$

Proof of the theorem is similar to Olkin and is omitted here.

2.3. Numerical illustration.

The results obtained in the preceding section will now be illustrated with the help of data collected from a sample survey conducted on pepper in the State of Kerala during the years 1966-68. The data relates to the number of pepper standards and area under pepper enumerated completely in each of a random sample of 120 villages. Information on two auxiliary characters

1. Geographical area of the village and

2. Garden area of the village,

were available for all villages in the state.

Notations and Definitions.

N - Total number of villages in the population.

n - Total number of villages in the sample.

y_{1i} - Number of pepper standards enumerated in the ith selected village.

y_{2i} - Area under pepper (in hectacres) enumerated in the ith selected village.

Y_{1i}, Y_{2i} the true values of the number of pepper standrads and area under pepper.

$\bar{Y}_1 = (1/N) \sum_{i=1}^{N} Y_{1i}$ - Average number of pepper standards per village in the population.

$\bar{Y}_2 = (1/N) \sum_{i=1}^{n} Y_{2i}$ - Average area under pepper per village in the population.

X_1 - Total Geographical area in the population.

X_2 - Total Garden area in the population.

$\bar{X}_1$, $\bar{X}_2$ - the respective population means of X_1 and X_2.

$R_{11} = \bar{Y}_1 / \bar{X}_1$

$R_{22} = \bar{Y}_2 / \bar{X}_2$

$R_{12} = \bar{Y}_1 / \bar{X}_2$

$R_{21} = \bar{Y}_2 / \bar{X}_1$

$\hat{\bar{Y}}_1$ - Estimate of the average number of pepper standards based on the sample mean.

$\hat{\bar{Y}}_2$ - Estimate of the average area under pepper per village based on the sample mean.

The tables 2.1, 2.2, 2.3, 2.4 given at the end give the different estimates of the number of pepper standards and area under pepper, obtained by using various procedures.

Thus $\hat{\bar{Y}}_{1R11}$, $\hat{\bar{Y}}_{1R12}$, $\hat{\bar{Y}}_{2R21}$, $\hat{\bar{Y}}_{2R22}$ are the estimated mean number of pepper standards and area under pepper, the estimates are based on the biased ratios which are indicated by the subscripts. $\bar{y}'_{1r11}$, $\bar{y}'_{1r12}$, $\bar{y}'_{2r21}$, $\bar{y}'_{2r22}$ represent the estimated mean number of pepper standards and area under pepper, based on the unbiased ratio type estimates given by Hartley and Ross.

$\hat{\bar{Y}}_{1R(M.V.)}$, $\hat{\bar{Y}}_{2R(M.V.)}$ are the estimated mean of the multivariate biased ratio.

$\bar{y}'_{1r(M.V.)}$, $\bar{y}'_{2r(M.V.)}$ are the estimated mean of the Multivariate unbiased ratio type estimate.

From the tables it may be seen that the multivariate unbiased ratio type estimate while being unbiased is equally efficient as the multivariate biased ratio estimate and is always more efficient than each of the single unbiased ratio estimates. It is always more efficient than the estimate based on the simple mean.

CHAPTER III

UNBIASED RATIO CUM PRODUCT TYPE ESTIMATORS.

3.1. INTRODUCTION.

Use of information on one or more auxiliary variables at the estimation stage to increase the precision of the sample results, is frequently resorted to by survey practitioners. When two auxiliary variables are available and if one of them is having positive correlation and the other - negative correlation with the variable under study then a ratio cum product estimator will be more efficient than

i) Simple unbiased estimate(estimator with no supplementary information)

ii) Estimators which utilise one supplementary character (ratio and product)

iii) Estimators which utilise more than one supplementary characters(multivariate ratio and multivariate product)

has been dealt by M. P. Singh (1967). In his study he has considered the estimators which are all biased. In the present investigation an unbiased Ratio cum product type estimator is obtained, along with the exact variance by using the concept of multivariate symmetric means.

3.2. Unbiased Ratio cum Product type estimator.

Let there be two supplementary variables x, z information on which is available for each unit of the population. Let Y, X and Z be the unbiased estimates of the parameters corresponding to the variable under study and the supplementary variables x and z, and the population totals X (total of x), Y (total of y) be known in advance. Let the correlation between y and x be posivive and that of y and z be negative.

Let a random sample of n observations (x_i, y_i, z_i, rp_i) $(i = 1, 2, \ldots n)$ be obtained from a population size N where $rp_i = \frac{y_i z_i}{x_i}$ $(i = 1, 2, \ldots n)$

Let $\overline{rp} = \frac{1}{n} \sum_{i=1}^{n} rp_i$

$\bar{y} = \frac{1}{N} \sum_{i=1}^{n} y_i$, $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$, and $\bar{z} = \frac{1}{n} \sum_{i=1}^{n} z_i$

Then the ratio cum product estimator will be

$\bar{y}_{rp} = \overline{rp} \; \frac{\bar{X}}{\bar{Y}}$ where $\bar{X}$ and $\bar{Z}$ are the population means of the variates x and z respectively.

Let $\overline{RP} = \frac{1}{N} \sum_{i=1}^{N} rp_i$.

When we express the above expressions in terms of symmetric means,

$\bar{x} = \langle(1000)\rangle$

$\bar{y} = \langle(0100)\rangle$

$\bar{z} = \langle(0010)\rangle$

$\overline{rp} = \langle(0001)\rangle$

The corresponding populations means can be expressed as

$\bar{X} = \langle(1000)\rangle'$

$\bar{Y} = \langle(0100)\rangle'$

$\bar{Z} = \langle(0010)\rangle'$

$\overline{RP} = \langle(0001)\rangle'$

As usual an angle bracket with a prime $\langle\ \rangle'$ is usedto denote a symmetric population mean while $\langle\ \rangle$ without prime denotes a sample symmetric mean.

Hence $$\bar{y}_{rp} = \frac{\langle(0001)\rangle \quad \langle(1000)\rangle'}{\langle(0010)\rangle'} \qquad (3.2.1)$$

The bias of $\bar{y}_{rp}$ is obtained as

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

$$= \frac{\langle(0001)\rangle' \langle(1000)\rangle' - \langle(0100)\rangle' \langle(0010)\rangle'}{\langle(0010)\rangle'}$$

Using the multiplication rule of symmetric means,

$$(\text{Bias } \bar{y}_{rp}) = \frac{1}{\bar{Z}} \left\{ \begin{array}{l} (1/N)\,\langle(1001)\rangle' + ((N-1)/N)\langle(0001),\ (1000)\rangle' \\ -(1/N)\,\langle(0110)\rangle' + ((N-1)/N)\,\langle(0010),\ (0100)\rangle' \end{array} \right\}$$

$$= \frac{1}{\bar{Z}} \left\{ \begin{array}{l} ((N-1)/N)\langle(0001),\ (1000)\rangle' \\ -((N-1)/N)\,\langle(0010),\ (0100)\rangle' \end{array} \right\} \qquad (3.2.2)$$

for $(0110) = (1001)$

An unbiased estimate of (3.2.2) will be

$$\text{Est.}(\text{Bias } \bar{y}_{rp}) = \frac{(N-1)}{N\bar{Z}} \left\{ \langle(0001),(1000)\rangle - \langle(0010),(0100)\rangle \right\}$$

Hence we can obtain an unbiased estimate of $\bar{y}_{rp}$ by subtracting the estimated bias. Let it be $\bar{y}'_{rp}$. Then

$$\bar{y}'_{rp} = \bar{y}_{rp} - \frac{(N-1)}{N\bar{Z}} \left\{ \langle(0001),(1000)\rangle - \langle(0010),(0100)\rangle \right\} \qquad (3.2.3)$$

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

Hence an unbiased estimate is given by

$$\bar{y}'_{rp} = \frac{1}{\bar{Z}} \left\{ \overline{rp}\,\bar{X} - \frac{(N-1)\,n}{N\,(n-1)} \left(\overline{rp}\,\bar{x} - \bar{y}\,\bar{x} \right) \right\} \qquad (3.2.4)$$

The variance of $\bar{y}'_{rp}$ may then be calculated as a straight forward application of the multiplication formula for symmetric means.

Using (3.2.1.) and (3.2.3.)

$$y'_{rp} = \frac{1}{Z}\left\{ \langle(1000)\rangle\langle(0001)\rangle - ((N-1)/N)\left[\langle(0001),(1000)\rangle - \langle(0010),(0100)\rangle \right] \right\}$$

$$\begin{aligned} y'^{2}_{rp} = {} & \frac{1}{Z^2}\left[\langle(1000)\rangle' \langle(1000)\rangle' \langle(0001)\rangle \langle(0001)\rangle \right] \\ & + \frac{(N-1)^2}{N^2}\frac{1}{Z^2}\left[\langle(0001),(1000)\rangle \langle(0001),(1000)\rangle \right] \\ & + \frac{(N-1)^2}{N^2}\frac{1}{Z^2}\left[\langle(0010),(0100)\rangle \langle(0010)\rangle \langle(0100)\rangle \right] \\ & - 2\frac{(N-1)}{N Z^2}\left[\langle(1000)\rangle' \langle(0001)\rangle \langle(0001),(1000)\rangle \right] \\ & + \frac{2(N-1)}{N Z^2}\left[\langle(1000)\rangle\langle(0001)\rangle \langle(0010),(0100)\rangle \right] \\ & - \frac{2(N-1)^2}{N^2 Z^2}\left[\langle(1000),(0001)\rangle \langle(0010),(0100)\rangle \right] \end{aligned} \tag{3.2.5.}$$

$$V(y'_{rp}) = E(y'_{rp})^{22} - Y^2 \tag{3.2.6.}$$

$$\bar{Y}^2 = \frac{1}{Z^2 N^2}\left[\langle(0200)\rangle' + (N-1)\langle(0100),(0100)\rangle' \right] . \left[\langle(0020)\rangle' + (N-1)\langle(0010),(0010)\rangle \right] \tag{3.2.7.}$$

Expanding (3.2.5) and (3.2.7.), using the multiplication rules of symmetric means, and using (3.2.6.), we have after simplifications,

$$
\begin{aligned}
V(\bar{y}'_{rp}) = \frac{1}{\bar{Z}^2} \Big[& \langle(2002)\rangle \left\{ (1/N^2 n) - (1/N^3) \right\} \\
& + \langle(2000),(0002)\rangle \left\{ (N-1)/N^2 n + (N-1)^2/n N^2 (n-1) - 2(N-1)/n N^2 \right\} \\
& + \langle(2001),(0001)\rangle \left\{ 2(n-1)/N^2 n - 2(N-1)/n N^2 \right\} \\
& + \langle(1002),(1000)\rangle \left\{ \frac{2(n-1)}{N^2 n} - \frac{2(N-1)}{n N^2} +- \frac{(N-1)^2}{n N^2 (n-1)} \right\} \\
& + \langle(2000),(0001),(0001)\rangle \left\{ \frac{(n-1)(N-2)}{n N^2} + \frac{(N-1)^2 (n-2)}{n(n-1) N^2} - \frac{2(N-1)\,8n-2)}{n N^2} \right\} \\
& + \langle(0002),(1000),(1000)\rangle \left\{ \frac{(N-1)(N-2)}{N^2 n} + \frac{(N-1)^2 (n-2)}{n(n-1) N^2} - \frac{2(N-1)(n-2)}{n N^2} \right\} \\
& + \langle(1001),(1000),(0001)\rangle \left\{ \frac{4(N-1)(N-2)(n-1)}{n(N-1) N^2} + \frac{2(n-2)(N-1)^2}{n(n-1) N^2} - \frac{4(N-1)(n-2)}{n N^2} - \frac{2(N-1)(N-2)}{n N^2} \right\} \\
& + \langle(0110),(0110)\rangle \left\{ \frac{(N-1)^2}{n(n-1) N^2} - \frac{2(N-1)^2}{N^3} \right\} \\
& + \langle(0200),(0020)\rangle \left\{ \frac{(N-1)^2}{n(n-1) N^2} - \frac{(N-1)}{N^3} \right\}
\end{aligned}
$$

$$+\langle(0200),(0010),(0010)\rangle\left\{\frac{(N-1)^2(n-2)}{N^2 n(n-1)} - \frac{(N-1)(N-2)}{N^3}\right\}$$

$$+\langle(0020),(0100),(0100)\rangle\left\{\frac{(N-1)^2(n-2)}{N^2 n(n-1)} - \frac{(N-1)(N-2)}{N^3}\right\}$$

$$+\langle(0110),(0100),(0010)\rangle\left\{\frac{2(N-1)^2(n-2)}{N^2 n(n-1)} - \frac{4(N-1)^2(N-2)}{(N-1)N^3} + \frac{2(N-1)(n-2)}{n N^2}\right\}$$

$$+\langle(0100),(0100),(0010),(0010)\rangle\left\{\frac{(n-2)(n-3)(N-1)^2}{N^2 n(n-1)} - \frac{(N-1)^2(N-2)(N-3)^2}{(N-1)N^3}\right\}$$

$$+\langle(1101),(0010)\rangle\left\{\frac{2(N-1)}{n N^2} - \frac{2(N-1)}{N^3}\right\}$$

$$+\langle(1011),(0010)\rangle\left\{\frac{2(N-1)}{n N^2} - \frac{2(N-1)}{N^3}\right\}$$

$$+\langle(1010),(0101)\rangle\left\{\frac{2(N-1)}{n N^2} - \frac{2(N-1)^2}{N^2 n(n-1)}\right\}$$

$$+\langle(1100),(0011)\rangle\left\{\frac{2(N-1)}{n N^2} - \frac{2(N-1)^2}{N^2 n(n-1)}\right\}$$

$$+\langle(1000),(0011),(0100)\rangle\left\{\frac{2(N-1)(N-2)}{n N^2} - \frac{2(n-2)(N-1)^2}{N^2 n(n-1)}\right\}$$

$$+\left\{\langle(0001),(1100),(0010)\rangle + \langle(1010),(0100),(0001)\rangle\right\}\left\{\frac{2(N-1)(n-2)}{n N^2} - \frac{2(n-2)(N-1)^2}{N^2 n(n-1)}\right\}$$

+

$$+ \langle(1000),(0100),(0010),(0001)\rangle' \left\{ \frac{2\,(N-1)(n-2)(N-3)}{N^2\,n} - \frac{2\,(N-1)^2(n-2)(n-3)}{N^2\,n\,(n-1)} \right\} \Big] \qquad (3.2.8)$$

An unbiased estimator of Var ($\bar{y}'_{rp}$) is therefore constructed simply by substituting the sample symmetric means for population symmetric means in (3.2.8).

When population size is very large then in the limiting case (3.2.8) takes the form

$$\begin{aligned}
\text{Var}\,(\bar{y}'_{rp}) = \frac{1}{\bar{Z}^2 n} \Big[& \langle(2000,(0001),(0001)\rangle' \\
& - \langle(1000),(1000),(0001),(0001)\rangle' \\
& + \langle(0020),(0100),(0100)\rangle' - \langle(0100),(0100),(0010),(0010)\rangle' \\
& + \langle(0200),(0010),(0010)\rangle' - \langle(0100),(0100),(0010),(0010)\rangle' \\
& + 2\,\langle(0110),(0100),(0010)\rangle' - \langle(0100),(0100),(0010),(0010)\rangle' \\
& - 2\,\langle(1100),(0010),(0001)\rangle' - \langle(1000),(0100),(0010),(0001)\rangle' \\
& - 2\,\langle(1010),(0100),(0001)\rangle' + \langle(1000),(0100),(00010),(0001)\rangle' \Big] \\
+ \frac{1}{n\,(n-1)\,\bar{Z}^2} \Big[& \langle(2000),(0002)\rangle' - \langle(2000),(0100),(0001)\rangle' \\
& - \langle(0002),(1000),(1000)\rangle' + \langle(1000),(1000),(0001),(0001)\rangle' \\
& + \langle(1001),(1001)\rangle' - 2\,\langle(1001),(1000),(0001)\rangle' \\
& + \langle(1000),(1000),(0001),(0001)\rangle' + \langle(0200),(0020)\rangle' \\
& - \langle(0020),(0100),(0100)\rangle' - \langle(0200),(0010),(0010)\rangle' \\
& + \langle(0100),(0100),(0010),(0010)\rangle'
\end{aligned}$$

$$- 2\langle(1100),(0011)\rangle' - \langle(1000),(0100),(0011)\rangle'$$
$$- \langle(1100),(0010),(0001)\rangle' + \langle(1000),(0100),(0010),(0001)\rangle'$$
$$+ 2\langle(1010),(0101)\rangle' - \langle(1010),(0100),(0001)\rangle'$$
$$- \langle(1000),(0101),(0010)\rangle' + \langle(1000),(0100),(0010),(0001)\rangle' \Big] \qquad (3.2.9.)$$

Now we can use the approximation

$$\langle(a_1),(a_2),\ldots(a_r)\rangle' = \langle(a_1)\rangle'\langle(a_2)\rangle'\ldots\langle(a_r)\rangle'$$

when N is very large. Expressing the symmetric means in terms of the usual notations, (3.2.9.) can be simplified and we have,

$$V(\bar{y}'_{rp}) = \frac{1}{n\bar{Z}^2}\Big[\overline{RP}^2\sigma_{xx} + \bar{Z}^2\sigma_{yy} + \bar{Y}^2\sigma_{zz} + 2\bar{Y}\bar{Z}\sigma_{yz} - 2\bar{Y}\,\overline{RP}\,\sigma_{xz} - 2\sigma_{xy}\,\overline{RP}\,\bar{Z}\Big]$$
$$+ \frac{1}{n(n-1)\bar{Z}^2}\Big[\sigma_{xx}\sigma_{rprp} + (\sigma_{xrp})^2 + (\sigma_{yz})^2 + \sigma_{yy}\sigma_{zz} - 2\sigma_{xy}\sigma_{zrp} - 2\sigma_{xz}\sigma_{yrp}\Big] \qquad (3.2.10)$$

A procedure for computing an unbiased estimator analogous to

$$\frac{\bar{x}_1\,\bar{x}_2\,\ldots\,\bar{x}_{r-1}\,\bar{x}_r}{\bar{X}_1\,\bar{X}_2\ldots\,\bar{X}_{r-1}}$$ is given by

Robson in his paper (1957), where $x_1, x_2, \ldots x_{r-1}$ having negative correlation with the variable under study. He has derived the variance in the case of three variated.

Now if situations are such that more number of supplementary informations available, say (r+s) where r of them having +ve correlation and s of them having -ve correlation then we can define

$$\bar{y}_{r+s} = \frac{1}{n} \sum_{i=1}^{n} \frac{y_i \; z_{1i} \; z_{2i} \cdots z_{si}}{x_{1i} \; x_{2i} \; x_{3i} \cdots x_{ri}} \; \frac{\bar{X}_1 \bar{X}_2 \; \cdots \; \bar{X}_r}{\bar{Z}_1 \bar{Z}_2 \cdots \; \bar{Z}_s}$$

and we can find an unbiased estimate corresponding to this estimate and the variance of this estimate can also be calculated preceeding in the exact manner as it is done for $\bar{y}_{rp}$.

CHAPTER IV

MULTIVARIATE UNBIASED PRODUCT TYPE ESTIMATOR

4.1. INTRODUCTION.

In sample survey situations it is a fairly common occurence that one or more auxiliary variables (x's) are positively correlated with y, the variable in question. When positive correlation occurs, under certain conditions all the biased and unbiased ratio-type estimators have been shown to have smaller variance than the sample mean $\bar{y}$. Situations where one or more x's are negatively correlated with y are perhaps not very common. They do occur however. For most of the consumer goods, price and quantity are negatively correlated. For families with a given level of income, usually there is negative correlation between their savings and personal loans. In the field of agriculture also consumption and marketable surplus of food grains are found to be negatively correlated. For a fixed holding size, the number of family members of a cultivator and his marketable surplus are also negatively correlated.

Olkin (1958) has considered the use of multisupplementary variables in building up ratio estimator and this estimator has been found to be more efficeent in comparison to the ratio estimate using one supplementary variable. However when the supplementary variables are negatively correlated with the study variable ratio method of estimation cannot be efficiently used. In such cases for the use of single supplementary variable, Murthy (1964) has considered the complementary situation of improving upon the unbiased estimator by considering a suitable product estimator. M. P. Singh (1967b) has considered the extension of the usual biased product estimator to the multivariate case. Robson(1957) has considered an unbiased product type estimator in the univariate case. The present investigation is an extension of the unbiased product type estimator to the multivariate case. Once again the variance expressions are obtained by using multivariate symmetric means.

4.2. MULTIVARIATE UNBIASED PRODUCT METHOD OF ESTIMATION.

Let there be k supplementary variables $(x_1), (x_2), \ldots (x_k)$ information on which is available for each unit of the population. Let y and x_i be the unbiased estimators of the parameters Y and X_i corresponding to the variable under study and the ith supplementary variable $(i = 1, 2, \ldots k)$ based on any probability deample design and let X_i be known in advance.

Let $\bar{y} = (1/n)\sum_{i=1}^{n} y_i$ and $\bar{x}_j = (1/n)\sum_{i=1}^{n} x_{ji}$ $(j = 1, 2, \ldots k)$

Let $\bar{y}'_{pj}$ $(j = 1, 2, \ldots k)$ be the unbiased product type estimator using the jth supplementary variable.

Then

$$\bar{y}'_{pj} = \frac{\bar{x}_j \bar{y}}{\bar{X}_j} - \frac{(N-n)}{N(n-1)} \left[\frac{1}{n} \sum_{i=1}^{n} x_{ji} y_i - \bar{x}_j \bar{y} \right] \qquad (j = 1, 2, \ldots k) \tag{4.2.1.}$$

In terms of the symmetric means we can write

$$\bar{x}_j = \langle (\overset{1}{0}\overset{2}{0}0 \ldots \overset{j}{1}\overset{j+1}{0} \ldots \overset{k}{0}) \rangle \qquad (j = 1, 2, \ldots k)$$

$$\bar{y} = \langle (\overset{1}{1}\overset{2}{0}\overset{3}{0}0 \ldots \overset{k}{0}) \rangle$$

-2600-

Here the first element in the vector represents the y unit, the succeding k units for $x_1, x_2, \ldots x_k$ respectively.

As usual an angle bracket with a prime $\langle\ \rangle'$ is used to denote a population symmetric mean while $\langle\ \rangle$ without prime denotes a sample symmetric mean.

We can write (4.2.1.) with symmetric means notations but it will be tedius to write the expression in this form. Also to derive

$VP_{ij} = Cov(\bar{y}'_{pi}, \bar{y}'_{pj})$

keeping $\bar{y}'_{pj}$ in symmetric mean notations in this form given above will be laborious. Hence we shall derive $VP_{12} = Cov(\bar{y}'_{p1}, \bar{y}'_{p2})$ assuming that we have two supplementary information x_1, x_2 with y without loss of generality.

In terms of symmetric mean notations after simplifications we have

$$\bar{y}'_{p1} = \frac{1}{\bar{X}_1}\left[\frac{1}{N}\langle(110)\rangle + \frac{(N-1)}{N}\langle(100),(010)\rangle\right]$$

$$\bar{y}'_{p2} = \frac{1}{\bar{X}_2}\left[\frac{1}{N}\langle(101)\rangle + \frac{(N-1)}{N}\langle(100),(001)\rangle\right]$$

$$Cov(\bar{y}'_{p1}, \bar{y}'_{p2}) = E(\bar{y}'_{p1}, \bar{y}'_{p2}) - \bar{Y}^2 \qquad (4.2.2)$$

$$E(\bar{y}'_{p1}, \bar{y}'_{p2}) = E\left[\frac{1}{N^2\bar{X}_1\bar{X}_2}\langle(110)\rangle + (N-1)\langle(100),(010)\rangle\right.$$
$$\left.\langle(101)\rangle + (N-1)\langle(100),(001)\rangle\right]$$

Multiplying the symmetric means inside the bracket and then taking expectations, we have

$$E(\bar{y}'_{p1}, \bar{y}'_{p2}) = \frac{1}{N^2 n\bar{X}_1\bar{X}_2}\Big\{\langle(211)\rangle' + (n-1)\langle(110),(101)\rangle'$$

$$+ \frac{(N-1)^2}{(n-1)} \Big\{ \langle (200),(011) \rangle' + \langle (110),(101) \rangle' + (n-2) \langle (200),(010),(001) \rangle'$$

$$+ (n-2) \langle (100),(101),(010) \rangle' + (n-2) \langle (100),(110),(001) \rangle'$$

$$+ (n-2) \langle (100),(100),(011) \rangle'$$

$$+ (n-2)(n-3) \langle (100),(100),(010),(001) \rangle' \Big\}$$

$$+ (N-1) \Big\{ \langle (210),(001) \rangle' + \langle (100),(111) \rangle' + (n-2) \langle (110),(100),(001) \rangle'$$

$$+ \langle (201),(010) \rangle' + \langle (100),(111) \rangle' + (n-2) \langle (110),(100),(010) \rangle' \Big] \Big\}$$

$$\bar{y}^2 = \frac{\bar{Y}\bar{X}_1\bar{X}_2}{\bar{X}_1\bar{X}_2} = \frac{1}{N^2 \bar{X}_1\bar{X}_2} \Big[\langle (110) \rangle' + (N-1) \langle (100),(010) \rangle' \Big]$$

$$\Big[\langle (101) \rangle' + (N-1) \langle (100),(010) \rangle' \Big]$$

$$= \frac{1}{N^3 \bar{X}_1\bar{X}_2} \Big[\langle (211) \rangle' + (N-1) \langle (110),(101) \rangle'$$

$$+ (N-1) \langle (200),(011) \rangle' + \langle (110),(101) \rangle'$$

$$+ (N-2) \langle (100),(101),(010) \rangle' + (N-2) \langle (200),(010),(001)$$

$$+ (N-2) \langle (100),(110),(001) \rangle' + (N-2) \langle (100),(100),(011) \rangle'$$

$$+ (N-2)(N-3) \langle (100),(100),(010),(001) \rangle'$$

$$+ \langle (210),(001) \rangle' + \langle (100),(111) \rangle'$$

$$+ (N-2) \langle (110),(100),(001) \rangle' + \langle (201),(010) \rangle'$$

$$+ \langle (100),(111) \rangle' + (N-2) \langle (101),(100),(010) \rangle' \Big]$$

Hence (4.2.2.) becomes after rearranging the terms,

$$\text{Cov}(\bar{y}'_{p1}, \bar{y}'_{p2}) = \frac{1}{N^2 \bar{X}_1\bar{X}_2} \Big[\langle (211) \rangle' \Big\{ \frac{1}{n} - \frac{1}{N} \Big\}$$

$$+ \langle (110),(101) \rangle' \Big\{ \frac{(n-1)}{n} - \frac{(N-1)}{N} + \frac{(N-1)^2}{n(n-1)} - \frac{(N-1)}{N} \Big\}$$

$$+ \langle (200),(011) \rangle' \left\{ \frac{(N-1)^2}{n(n-1)} - \frac{(N-1)^2}{N(N-1)} \right\}$$

$$+ \left\{ \begin{array}{l} \langle (200),(010),(001) \rangle' \\ + \langle (100),(100),(011) \rangle' \\ + \langle (100),(101),(010) \rangle' \\ + \langle (100),(110),(001) \rangle' \end{array} \right\} \left\{ \frac{(N-1)^2 (n-2)}{n(n-1)} - \frac{(N-1)^2 (N-2)}{N(N-1)} \right\}$$

$$+ \left\{ \begin{array}{l} \langle (100),(101),(010) \rangle' \\ + \langle (100),(110),(001) \rangle' \end{array} \right\} \left\{ \frac{(N-1)(n-2)}{n} - \frac{(N-1)(N-2)}{N} \right\}$$

$$+ \langle (100),(100),(010),(001) \rangle' \left\{ \frac{(n-2)(n-3)}{n(n-1)} - \frac{(N-2)(N-3)}{N(N-1)} \right\} (N-1)^2$$

$$+ \langle (210),(001) \rangle' \left\{ \frac{(N-1)(n-2)}{n} - \frac{(N-1)(N-2)}{N} \right\}$$

$$+ \left\{ \begin{array}{l} 2 \langle (100),(111) \rangle' \\ + \langle (201),(001) \rangle' \end{array} \right\} \left\{ \frac{(N-1)(n-2)}{n} - \frac{(N-1)(N-2)}{N} \right\} \Big] \qquad (4.2.3.)$$

An unbiased estimator of Cov ($\bar{y}'_{p1}$. $\bar{y}'_{p2}$) is constructed simply by putting the sample symmetric means instead of the population symmetric means in the expression (4. 2. 3.).

When population size N is very large (4. 2. 3.) becomes

$$\text{Cov}(\bar{y}'_{p1}, \bar{y}'_{p2}) = \frac{1}{n \bar{X}_1 \bar{X}_2} \Big[\langle (100),(101),(010) \rangle' + \langle (100),(110),(001) \rangle' + \langle (200),(010),(001) \rangle' + \langle (100),(100),(011) \rangle' - 4 \langle (100),(100),(010),(001) \rangle' \Big]$$

$$+ \frac{1}{\bar{X}_1 \bar{X}_2 n(n-1)} \Big[\langle (110),(101) \rangle' + \langle (200),(011) \rangle'$$

$$- \langle (200),(010),(001) \rangle' - \langle (100),(100),(011) \rangle'$$
$$- \langle (100),(101),(010) \rangle' - \langle (100),(110),(001) \rangle'$$
$$+ 2 \langle (100),(100),(001),(010) \rangle' \Big] \qquad (4.2.4)$$

Further we can use the approximation

$$\langle (a_1), (a_2), \ldots (a_n) \rangle' = \langle (a_1) \rangle' \langle (a_2) \rangle' \ldots \langle (a_n) \rangle'$$

Expressing the above expression (4.2.5.) in terms of usual notations we have

$$\mathrm{Cov}(\bar{y}'_{p1}, \bar{y}'_{p2}) = \frac{1}{\bar{X}_1 \bar{X}_2 n} \Big[\bar{Y} \bar{X}_1 \sigma_{x_2 y} + \bar{Y} \bar{X}_2 \sigma_{x_1 y} + \bar{X}_1 \bar{X}_2 \sigma_{yy} + \bar{Y}^2 \sigma_{x_1 x_2} \Big] + \frac{1}{\bar{X}_1 \bar{X}_2 n(n-1)} \Big[\sigma_{x_1 y} \sigma_{x_2 y} + \sigma_{yy} \sigma_{x_1 x_2} \Big] \qquad (4.2.5)$$

In general,

$$\mathrm{Cov}(\bar{y}'_{pi}, \bar{y}'_{pj}) = \frac{1}{\bar{X}_i \bar{X}_j n} \Big[\bar{X}_i \bar{Y} \sigma_{x_j y} + \bar{X}_j \bar{Y} \sigma_{x_i y} + \bar{X}_i \bar{X}_j \sigma_{yy} + \bar{Y}^2 \sigma_{x_i x_j} \Big] + \frac{1}{\bar{X}_i \bar{X}_j n(n-1)} \Big[\sigma_{x_i y} \sigma_{x_j y} + \sigma_{yy} \sigma_{x_i x_j} \Big] \qquad (4.2.6)$$

$$\mathrm{Cov}(\bar{y}'_{pi}, \bar{y}'_{pi}) = V(\bar{y}'_{pi})$$
$$= \frac{1}{n \bar{X}_i^2} \Big[\bar{X}_i^2 \sigma_{yy} + 2 \bar{X}_i \bar{Y} \sigma_{x_i y} + \bar{Y}^2 \sigma_{x_i x_i} \Big] + \frac{1}{\bar{X}_i^2 n(n-1)} \Big[(\sigma_{x_i y})^2 + \sigma_{x_i x_i} \sigma_{yy} \Big] \qquad (4.2.7.)$$

An unbiased estimate of (4.2.7) can be estimated as in the same lines done by Goodman and Hartley (1958), using Fisher's Bivariate 'k' statistics, and is omitted here.

Again using Olkin's method, we can define the multivariate unbiased product type estimate of the population mean as

$$\bar{y}'_{p(M.V.)} = \sum_{j=1}^{k} w_j \bar{y}'_{pj} \qquad (4.2.8.)$$

where $\bar{y}'_{pj}$ is given by (4.2.1) and weights w_j are chosen such that their sum

$$\sum_{j=1}^{k} w_j = 1$$

(4.2.8) can be written in the matrix form as

$$\bar{y}'_{p(M.V.)} = W' \bar{y}'_p$$

$$\text{where} \quad W = \begin{bmatrix} w_1 \\ w_2 \\ \cdot\cdot \\ \cdot\cdot \\ w_k \end{bmatrix} \quad \text{and} \quad \bar{y}'_p = \begin{bmatrix} \bar{y}'_{p1} \\ \bar{y}'_{p2} \\ \cdot\cdot \\ \cdot\cdot \\ \bar{y}'_{pk} \end{bmatrix}$$

and W' is the transpose of W.

$$\text{If} \quad V = \begin{bmatrix} VP_{11} & VP_{12} & \cdot & \cdot & VP_{1k} \\ VP_{21} & VP_{22} & \cdot & \cdot & VP_{2k} \\ \cdot\cdot & \cdot\cdot & \cdot & \cdot & \cdot\cdot \\ \cdot\cdot & \cdot\cdot & \cdot & \cdot & \cdot\cdot \\ VP_{k1} & VP_{k2} & \cdot & \cdot & VP_{kk} \end{bmatrix}$$

Where VP_{ij} is given by (4.2.6.) and VP_{ii} by (4.2.7).

The weights w_j ($j = 1, 2, \ldots k$) are chosen so as to minimise the variance of $\bar{y}'_{p(M.V.)}$.

Again following the algebra of Olkin we have $w_j = R_j / D$ where R_j is the sum of all the elements in the jth row of the matrix V^{-1}, and D is the sum of all the elemnts in the matrix V^{-1}, and V^{-1} is the inverse o f V. The minimum variance of $\bar{y}'_{p(M.V.)}$ is 1/D .

For estimating the weights and the variance of the estimated mean using all the auxiliary variables in the multivariate unbiased estimator, the estimators of VP_{ij} at (4.2.6.) and (4.2.7.) are substituted for VP_{ij} in the matrix V.

4.3. COMPARISON BETWEEN MULTIVARIATE UNBIASED PRODUCT TYPE ESTIMATORS WITH DIFFERENT NUMBER OF SUPPLEMENTARY VARIABLES.

THEOREM. Let $\bar{y}'_{pq}$ and $\bar{y}'_{pk}$ denote respectively the multivariate unbiased product type estimators of $\bar{Y}$ with optimum weights based on the set of supplementary variables $(x_1), (x_2), \ldots (x_k)$ and $(x_1), (x_2) \ldots \ldots (x_k)$ where k is greater than q.

Then $V(\bar{y}'_{pq}) \geq V(\bar{y}'_{pk})$

Proof of the theorem is similar to Olkin (1958) and is omitted here.

Application of the above results can be studied for any particulra sampling scheme. The results are illustrated with some emperical data in Chapter VI.

CHAPTER V

GENERALIZED MULTIVARIATE UNBIASED ESTIMATION FOR THE MEAN OF FINITE POPULATIONS.

5.1. INTRODUCTION.

Just as ratio estimators are often indicated when the variable y is positively correlated with an auxiliary variable x so the product estimators are indicated when y is negatively correlated with x. The method of estimation suggested by Olkin (1958) is extended to the case where some or all of the auxiliary variables are positively correlated and some or all are negatively correlated to the variable under study. S.R.S. Rao and Govind S. Mudholkor (1967) have suggested a multivariate estimate with p+q supplementary variables where the first p of them having positive correlation with the variable under study and the remaining q having negative correlation with the variable under study. Similar studies have been done by S.K. Srivastava (1965).

In the present investigation a multivariate unbiased estimation of the population mean $\bar{Y}$ is considered and the expressions are obtained by using the multivariate symmetric means.

5.2. GENERALIZED MULTIVARIATE UNBIASED ESTIMATION.

Let there be (p+q) supplementary variables (x_1), (x_2), . . (x_p), (x_{p+1}), . . (x_{p+q}) information on which is available for each unit of the population. Let y and x_i be the unbiased estimators of the parameters Y and X_i corresponding to the variable under study and the ith supplementary variable (i = 1, 2, . . p+k) based on any probability sample design and let X_i be known in advance.

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

and $\bar{x}_j = \frac{1}{n} \sum_{i=1}^{n} x_{ji}$ $(j = 1, 2, \ldots p+k)$

Let a random sample of n observations $(y_i, x_{1i}, x_{2i}, \ldots x_{(p+k)i},$
$r_{1i}, r_{2i}, \ldots r_{pi})$ $(i = 1, 2, \ldots n)$
be obtained from a population of size N where $r_{ji} = y_{ji} / x_{ji}$,
$(i = 1, 2, \ldots n, \; j = 1, 2, \ldots p)$

Let $\bar{r}_j = \frac{1}{n} \sum_{i=1}^{n} y_{ji}/x_{ji}$ $(j = 1, 2, \ldots p)$

$$\text{Let } \bar{y}_j' = \bar{r}_j \bar{X}_j - \frac{n(N-1)}{N(n-1)} (\bar{x}_j \bar{r}_j - \bar{y}) \qquad (5.2.1.)$$
$$(j = 1, 2, \ldots p)$$

$$\text{Let } \bar{y}_{pj}' = (\bar{x}_j \bar{y} / \bar{X}_j) - \frac{(N-n)}{Nn(n-1)} \sum_{i=1}^{n} (x_{ji} - \bar{x}_j) \bar{y} \qquad (5.2.2.)$$
$$(j = p+1, p+2, \ldots p+k)$$

(5.2.1.) is the Hartley and Ross unbiased ratio type estimate and

(5.2.2.) is the Robson's unbiased product type estimator.

Interms of symmetric mean we can write

$$\bar{r}_j = \langle (\overset{1}{0} \overset{2}{0} 0 \ldots \overset{p}{0} \ldots \overset{p+k}{0} \; \overset{1}{0} \overset{2}{0} \ldots \overset{j}{1} \overset{j+1}{0} \ldots \overset{p}{0}) \rangle$$
$$(j = 1, 2, \ldots p)$$

$$\bar{x}_j = \langle (\overset{1}{0} \; \overset{2}{0} \; 0 \ldots \overset{j}{1} \; \overset{j+1}{0} \ldots \overset{p+k}{0} \; \overset{1}{0} \ldots \overset{p}{0}) \rangle \quad (j = 1, 2, \ldots p+k)$$

$$\bar{y} = \langle (1 \; \overset{1}{0} \; \overset{2}{0} \; \ldots \; \overset{p+k}{0} \; \overset{1}{0} \ldots \overset{p}{0}) \rangle$$

Here the first element in the vector represents the y unit, the succedding (p+k) units for $(x_1, x_2, \ldots x_{p+k})$ and the remaining p units for $r_1, r_2, \ldots r_p$ respectively. An angle bracket with a prime $\langle\ \rangle'$ is used to denote a population symmetric mean while $\langle\ \rangle$ without prime denotes a sample symmetric mean. We can write (5.2.1.) and (5.2.2.) in symmetric mean notations but it will be tedius to write it in this form. Also to derive

$V_{rp_{ij}} = \text{cov}\ (\bar{y}'_{ri}, \bar{y}'_{pj})$ keeping $\bar{y}_j$, $\bar{y}'_{pj}$ in symmetric mean notations will be laborious. Hence we shall derive

$V_{rp_{12}} = \text{Cov}\ (\bar{y}'_{r1}, \bar{y}'_{p2})$ assuming that we have two supplementary information x_1, x_2 (x_1 having +ve correlation with y and x_2 having negative correlation with y) without loss of generality.

In terms of symmetric mean notations we have

$$\bar{y}'_{r1} = \langle(0100)\rangle\langle(0001)\rangle + \frac{(N-1)}{N}\left[\langle(1000)\rangle - \langle(0100),(0001)\rangle\right] \qquad (5.2.3.)$$

$$\bar{y}'_{p2} = \frac{1}{\bar{X}_2}\left[(1/N)\ \langle(1010)\rangle + \frac{(N-1)}{N}\ \langle(1000),(0010)\rangle\right] \qquad (5.2.4.)$$

$$\text{Cov}\ (\bar{y}'_{r1}, \bar{y}'_{p2}) = E\ (\bar{y}'_1, \bar{y}'_{p2}) - \bar{Y}^2 \qquad (5.2.5.)$$

Multiplying (5.2.1.) with (5.2.4.), taking expectations, (5.2.5.) becomes after grouping the terms ,

$$\text{Cov}\ (\bar{y}'_{r1}, \bar{y}'_{p2}) =$$

$$\frac{1}{N^2 n \bar{X}_2}\Big[\langle(1111)\rangle' + (N-1)\ \langle(0100),(1011)\rangle'$$
$$+ (n-1)\ \langle(1110),(0001)\rangle' + (n-1)\ \langle(1010),(0101)\rangle'$$
$$+ (n-1)(N-2)\ \langle(0100),(1010),(0001)\rangle'\Big]$$

$$+ \frac{(N-1)}{N^2 n \bar{X}_2} \Big[\langle (1100),(0011) \rangle' + \langle (1000),(0111) \rangle' + (N-2) \langle (1000),(0100),(0011) \rangle' \Big]$$

$$+ \frac{(N-1)(n-2)}{N^2 n \bar{X}_2} \Big[\langle (1100),(0010),(0001) \rangle' + \langle (0110),(1000),(0001) \rangle' + \langle (0101),(1000),(0010) \rangle' + (N-3) \langle (1000),(0100),(0010),(0001) \rangle' \Big]$$

$$+ \frac{(N-1)}{N^2 n \bar{X}_2} \Big[\langle (2000),(0010) \rangle' + \langle (1000),(1010) \rangle' + (n-2) \langle (1000),(1000),(0010) \rangle' \Big]$$

$$- \frac{(N-1)}{N^2 n (n-1) \bar{X}_2} \Big[\langle (1100),(0011) \rangle' + \langle (1001),(0110) \rangle' + (n-2) \langle (1100),(0010),(0001) \rangle' + \langle (0110),(0001),(1000) \rangle' + \langle (0011),(1000),(0100) \rangle' + \langle (1001),(0100),(0010) \rangle' + (n-3) \langle (1000),(0010),(0100),(0001) \rangle' \Big]$$

$$- \frac{1}{N^2 \bar{X}_2} \Big[\langle (2010) \rangle' + (N-1) \langle (1000),(1010) \rangle' \Big]$$

$$- \frac{(N-1)}{N^2 \bar{X}_2} \Big[\langle (2000),(0010) \rangle' + \langle (1000),(1010) \rangle' + (N-2) \langle (1000),(1000),(0010) \rangle' \Big] \qquad (5.2.6.)$$

An unbiased estimate of Cov ($\bar{y}'_{r1}$, $\bar{y}'_{p2}$) is constructed simply by substituting the sample symmetric mean for population symmetric means in (5.2.6.)

When the population size N is very large, (5.2.6.) can be written after rearranging the terms as

$$- \frac{1}{n(n-1)\,\bar{X}_j} \left[\sigma_{x_i y} \; \sigma_{x_j r_i} + \sigma_{y r_i} \; \sigma_{x_i x_j} \right] \tag{5.2.9.}$$

$(i = 1, 2, \ldots p,\; j = p+1, p+2, \ldots p+k)$

$Cov(\bar{y}'_{ri}, \bar{y}'_{rj})$ $(i = 1, 2, \ldots p,\; j = 1, 2, \ldots p)$ and

$Cov(\bar{y}'_{pi}, \bar{y}'_{pj})$ $(i = p+1, p+2, \ldots p+k,\; j = p+1, p+2, \ldots p+k)$

are already defined in 2.2.7. and 2.2.8. and 4.2.7. and 4.2.8. respectively.

Following Olkin's method, we can define the generalised multivariate unbiased ratio type estimate of the population mean as

$$\bar{y}'_G = \sum_{i=1}^{p} w_i \bar{y}'_{ri} + \sum_{i=p+1}^{p+k} w_i \bar{y}'_{pi} \tag{5.2.10.}$$

where $\bar{y}'_{ri}$, $\bar{y}'_{pi}$ are given by 5.2.1. and 5.2.2. respectively, and weights w_i are chosen such that their sum $\sum_{i=1}^{p+k} w_i = 1$.

$$\text{If } V_G = \begin{bmatrix} V_{G11} & V_{G12} & \cdot\cdot & V_{G1p} & V_{G1(p+1)} & \cdot\cdot & V_{G1(p+k)} \\ V_{G21} & \cdot\cdot & \cdot & \cdot\cdot & \cdot\cdot & \cdot\cdot & V_{G2(p+k)} \\ \cdot\cdot & \cdot\cdot & \cdot & \cdot\cdot & \cdot\cdot & \cdot\cdot & \cdot\cdot \\ \cdot\cdot & \cdot\cdot & \cdot & \cdot\cdot & \cdot\cdot & \cdot\cdot & \cdot\cdot \\ V_{G(p+k)1} & \cdot\cdot & & \cdot\cdot & \cdots & & V_{G(p+k)(p+k)} \end{bmatrix}$$

where V_{Gij} $(i = 1, 2, \ldots p,\; j = 1, 2, \ldots p)$ are given by (2.2.7) and (2.2.8.)

V_{Gij} ($i = p+1, p+2, \ldots p+k$, $j = p+1, p+2, \ldots p+k$)

are given by (4.2.7 and 4.2.8.).

V_{Gij} ($i = 1, 2, \ldots p$, $j = p+1, p+2, \ldots p+k$)

are given by (5.2.9.)

The weights w_i are chosen so as to minimise the variance of $\bar{y}'_G$. Again following the algebra of Olkin, we have $w_i = R_i/D$ where R_i is the sum of all elements in the ith row of the matrix V_G^{-1} and D is the sum of all the elements in the matrix V_G^{-1} . V_G^{-1} is the inverse of V_G. The minimum variance of $\bar{y}'_G$ is 1/D.

For estimating the weights and the variance of the estimated mean using all the auxiliary variable in the multivariate unbiased estimator, the estimates of V_{Gij} at (5.2.9.), (2.2.7.), (2.2.8.), (4.2.7.) and (4.2.8.) are substituted for V_{Gij} in the matrix V_G. When we have supplementary variables with only positive correlation or with only negative correlation then $\bar{y}'_G$ will be reduced to a multivariate unbiased ratio type estimate or a multivariate unbiased product type estimate.

5.3. COMPARISON BETWEEN GENERALIZED UNBIASED ESTIMATE WITH DIFFERENT NUMBER OF SUPPLEMENTARY VARIABLES.

THEOREM

Let $\bar{y}'_{G(p+k)}$ and $\bar{y}'_{G(r+s)}$ denote respectively the generalized multivariate unbiased estimators of $\bar{Y}$ with optimum weights based on the set of supplementary variables

$(x_1), (x_2) \ldots (x_p), (x_{p+1}), \ldots (x_{p+k})$ (where the first p variables having positive correlation with y and the remaining k having negative correlation with y) and $(x_1), (x_2), \ldots (x_r), (x_{r+1}), \ldots (x_{r+s})$ (where the first r variables having positive correlation with y and the remaining s having negative correlation with y) where $r > p$ and $s > k$ then

$$V(\bar{y}'_{G(r+s)}) \leq V(\bar{y}'_{G(p+k)})$$

Proff Proof of the theorem is similar to Olkin (1958) and is omitted here.

Application of the above results can be studied for any particular sampling scheme. The results are illustrated with the help of some emperical data in chapter VI.

CHAPTER VI

ILLUSTRATION

6.1. INTRODUCTION.

Marketable surplus represents the surplus available for disposal with the producer, after the requirements of family consumption, feed, seed and wastage have been met. By consumption of the cultivator we mean the following.

(a) Quantity set apart for seed.

(b) Quantity consumed or set apart for later consumption.

For marketable surplus we consider the following types of disposals.

(a) Quantity sold during the season for cash.

(b) Quantity stored for later disposal.

(c) Quantity paid to land owner as rent.

(d) Loans returned.

(e) Quantity of other disposals.

It has been observed that marketable surplus of a particular crop is highly correlated with the volume of out-put, and holding size of the cultivator. When we fix the holding sizd, then marketable surplus is negatively correlated with the number of family member. Also marketable surplus is observed to be having negative correlation with consumption.

The results which we have obtained in Chapters III, IV, V will now be illustrated with the help of the data collected in the Bench-Mark and Assesment Survey under Intensive Agricultural Development Programme (IADP) conducted by Institute of Agricultural Research Statistics, (I.C.A.R.) in Tanjore district in the year 1967-68.

The sampling plan for the survey was one of the stratified multistage, the blocks being the stratas, villages and cultivators as first and second stage units repectively. From each stratum about 5 villages were selected at random and from each village a sample of 8 cultivators was selected for collecting information on production, consumption and disposal of crops produced of each of the ultimate stage units of selection, vis the cultivators. The following items of information from the survey were utilised for the analysis.

i.) Number of blocks.

ii.) Number of villages selected from each block.

iii.) Number of cultivators selected from each village.

iv.) Number of family members in the household of each of the selected cultivators.

v.) Holding size of each of the sampled cultivator.

vi.) Production of paddy obtained by enquiry method of each of the selected cultivator.

vii.) Surplus reported to be availabe for marketing by each of the cultivator.

The complete information on items mentioned above were availble for 529 cultivators in " Samba" season and 693 cultivators in Rabi " Kuruvai" season, in the year 1967-68. The 529 cultivators in "Samba" season and 693 cultivators in "Kuruvai" season were considered as forming two populations for purposes of the study. A random sample (without replacement) of 100 cultivators from Samba season and another 100 cultivators from Kuruvai season were taken and the following observations were made.

1. Marketable surplus of paddy per cultivator in kilograms (y).

2. Holding size of the cultivator (x_1) (in acres)

3. Number of family members (x_2) in the cultivators family.

4. Consumption of the cultivator (x_3)

6.2. Notations and definitions.

N - m number of cultivators in the population.

n - number of cultivators in the sample.

$\bar{Y}$ - Average marketable surplus of paddy per cultivator (in kgs.)

$\bar{X}_1$ - Average holding sese of cultivator (in acres) per cultivator.

$\bar{X}_2$ - Average number of family members per cultivator.

$\bar{X}_3$ - Average quantity of paddy consumption per cultivator (in kg.)

$\bar{y}$, $\bar{x}_1$, $\bar{x}_2$, $\bar{x}_3$ are the unbiased estimates of $\bar{Y}$, $\bar{X}_1$, $\bar{X}_2$, $\bar{X}_3$ based on a random sample of 100 observations.

$\bar{y}_{r1}$ - the estimated average marketable surplus based on the biased ratio

$$= \frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{x_{1i}}\ \bar{X}_1$$

$\bar{y}'_{r1}$ - the estimated average marketable surplus based on the unbiased Hartley and Ross unbiased ratio using the supplementary variable x_1.

$\bar{y}_{p2}$ - the estimated average marketable surplus based on the biased product estimate.

$$= \frac{\bar{y}\,\bar{x}_2}{\bar{X}_2}$$

$\bar{y}'_{p2}$ - the estimated average marketable surplus based on the unbiased product type estimate using the supplementary variable x_2.

$\bar{y}_{p3}$ - the estimated average marketable surplus based on the biased product.

$$= \quad \bar{y}\,\bar{x}_3 / \bar{X}_3$$

$\bar{y}'_{p3}$ - the estimated average marketable surplus based on the unbiased product type estimate using the supplementary variable x_3.

$\bar{y}_{r1p2}$ - the estimated average marketable surplus based on the biased ratio-cum-product estimate.

$$= \frac{1}{n} \sum_{i=1}^{n} \frac{y_i x_{2i}}{x_{1i}} \quad \frac{\bar{X}_1}{\bar{X}_2}$$

$\bar{y}'_{r1p2}$ - the estimated average marketable surplus based on the unbiased ratio-cum-product estimate suggested in chapter 3, using the supplementary variable x_1, and x_2.

$\bar{y}_{p(M.V.)}$ - the estimated average marketable surplus based on the multivariate biased product estimate using the supplementary variables x_2, and x_3.

$\bar{y}'_{p(M.V.)}$ - the estimated average markotable surpus based on the unbiased multivariate product type estimate using the same supplementary variables x_2, and x_3.

$\bar{y}_{G(M.V.)}$ - the estimated average marketable surpus based on the generalised biased multivariate estimate using the supplementary variables x_1, and x_2.

$\bar{y}'_{G(M.V.)}$ - the estimated average marketable surpus based on the generalised unbiased multivariate estimate using the supplementary variables x_1, and x_2.

6.3. Discussion of the results.

Tables 6.1 and 6.2 given at the end of the chapter gives the various estimates suggested above with their variances. The results show that the unbiased estimates are closer to the true value than the biased estimates. The results indicate that the unbiased ratio and ratio-cum-product estimates are superior to the biased estimates in all cases because the former estimates are not accompanied by significantly larger errors in comparison to the latter, while being unbiased. The product estimate and the ratio-cum-product estimates (both biased and unbiased) are less efficient than the simple mean estimate and this is due to very small negative correlations between the variables y and x_2 and also between y and x_3.

In the data considered for this illustration, the coefficients of correlation between the character studied and various supplementary variables were of low magnitude. Better results should be expected for cases where stronger correlations (positive and negative) exist. For such cases the estimators discussed in this dissertation could be more advantageously used.

. . . .

Table - 2.1.

Estimate of average number of pepper standards per village in Kerala State during 1966-67 obtained by various estimation procedures along with their variances and biases.

serial number.	estimator.	estimate of average number of pepper standards per village.	variance.	percentage gain in precision over simple mean.	percentage bias.
(1)	(2)	(3)	(4)	(5)	(6)
1.	$\hat{\bar{Y}}_1$ (S.M.)	25199	19514801	--	--
2.	$\hat{\bar{Y}}_{1R_{11}}$	28737	10638999	83.4	0.7301
3.	$\hat{\bar{Y}}_{1R_{12}}$	21787	11060294	76.4	1.2938
4.	$\bar{y}'_{1r_{11}}$	29242	11973724	62.9	--
5.	$\bar{y}'_{1r_{12}}$	21002	14873373	31.2	--
6.	$\hat{\bar{Y}}_{1R(M.V.)}$	28096	9534271	104.6	(0.7804)
7.	$\bar{y}'_{1r(M.V.)}$	30195	11942231	63.4	--

Table - 2.2.

Estimate of average number of pepper standards per village in Kerala State during 1967-68 obtained by various estimation procedures along with their variances and biases.

serial number.	estimator.	estimate of average number of pepper standards per village.	variance.	percentage gain in precision over simple mean.	percentage bias.
(1)	(2)	(3)	(4)	(5)	(6)
1.	$\hat{Y}_1$ (S.M.)	16384	6975448	--	--
2.	$\hat{Y}_{1R_{11}}$	26868	4595802	51.8	.4023
3.	$\hat{Y}_{1R_{12}}$	23194	2161201	227.7	-.4013
4.	$\bar{y}'_{1r_{11}}$	27978	4827874	44.5	--
5.	$\bar{y}'_{1r_{12}}$	23324	2150652	224.3	--
6.	$\hat{Y}_{1R(M.V.)}$	23194	2122555	228.2	-.4122
7.	$\bar{y}'_{1r(M.V.)}$	22586	2099536	232.23	--

Table - 2.9.

Estimates of average area under pepper per village in Kerala State during 1966-67 obtained by various estimation procedures along with thori variances and biases.

serial number.	estimator.	estimate of average area (area in acres) per village.	variance.	percentage gain in precision over simple mean.	percentage bias.
(1)	(2)	(3)	(4)	(5)	(6)
1.	$\hat{\bar{Y}}_2$(S.M.)	632.01	14595	- -	- -
2.	$\hat{\bar{Y}}_{2R_{21}}$	720.73	3980	266.7	-.1767
3.	$\hat{\bar{Y}}_{2R_{22}}$	546.44	3507	316.1	.2538
4.	$\bar{y}'_{2r_{21}}$	719.94	4013	263.8	- -
5.	$\bar{y}'_{2r_{22}}$	543.31	3598	305.6	- -
6.	$\hat{\bar{Y}}_{2R(M.V.)}$	524.11	3502	316.7	.3089
7.	$\bar{y}'_{2r(M.V.)}$	574.72	3578	307.9	- -

Table - 2.4.

Estimates of average area under pepper per village in Kerala State during 1967-68 obtained by various estimation procedures along with their variances and biases.

serial number.	estimator.	estimated average area (areain acres) per village.	variance.	percenatage gain in precision over simple mean.	percentage bias.
(1)	(2)	(3)	(4)	(5)	(6)
1.	$\hat{\bar{Y}}_2$ (S.M.)	376.81	2862	- -	- -
2.	$\hat{\bar{Y}}_{2R_{21}}$	617.94	2122	34.8	.6655
3.	$\hat{\bar{Y}}_{2R_{22}}$	533.44	687	316.5	-.2127
4.	$\bar{y}'_{2r_{21}}$	628.01	2187	30.8	- -
5.	$\bar{y}'_{2r_{22}}$	525.54	716	299.7	- -
6.	$\hat{\bar{Y}}_{2R(M.V.)}$	515.55	647	342.3	-.3986
7.	$\bar{y}'_{2r(M.V.)}$	499.18	652	338.9	- -

Table 6.1.

Various estimates of the average marketable surplus of Paddy in kilograms per cultivator in Tanjore District (Tamil Nadu) in the year 1967-68 under the IAD Programme during 'Samba' Season.

serial number.	estimator.	estimate of average marketable surplus of paddy per cultivator.	variance.
(1)	(2)	(3)	(4)
1.	Simple mean. ($\bar{y}$)	1594	48800
2.	Biased ratio type ($\bar{y}_{r1}$)	1496	42023
3.	Unbiased ratio type ($\bar{y}'_{r1}$)	1458	42095
4.	Biased product ($\bar{y}_{p2}$)	1791	55416
5.	Unbiased product type ($\bar{y}'_{p2}$)	1785	55507
6.	Biased product ($\bar{y}'_{p3}$)	1136	103625
7.	Unbiased product type ($\bar{y}'_{p3}$)	1138	105188
8.	Biased ratio-cum-product ($\bar{y}'_{r1p1}$)	1660	54019
9.	Unbiased ratio-cum-product ($\bar{y}'_{r1p1}$)	1649	54135
10.	Biased multivariate product ($\bar{y}_{p(M.V.)}$)	1767	55346
11.	Unbiased multivariate product ($\bar{y}'_{p(M.V.)}$)	1762	55435
12.	Biased multivariate generalised ($\bar{y}_{G(M.V.)}$)	1548	41488
13.	Unbiased multivariate genralised ($\bar{y}'_{G(M.V.)}$)	1512	41536

True value($\bar{Y}$) = 1287 Correlation between y and x_1 - +.3975
Sample size = 100 Correlation between y and x_2 - -.0523
Population size = 529 Correlation between y and x_3 - -.0223

Table - 6.2.

Various estimates of the average marketable surplus of Paddy in kilograms per cultivator in Tanjore District(Tamil Nadu) in the year 1967-68 under IAD Programme during ' Kuruvai ' Season.

serial number.	estimator.	estimate of average markatable surplus of paddy per cultivator.	variance.
(1)	(2)	(3)	(4)
1.	Simple mean. ($\bar{y}$)	467	85742
2.	Biased ratio type ($\bar{y}_{r1}$)	763	52630
3.	Unbiased ratio type ($\bar{y}'_{r1}$)	639	56106
4.	Biased product. ($\bar{y}_{p2}$)	412	103326
5.	Unbiased product type. ($\bar{y}'_{p2}$)	414	106420
6.	Biased product. ($\bar{y}'_{p3}$)	404	106911
7.	Unbiased product type ($\bar{y}'_{p3}$)	409	111177
8.	Biased ratio-cum-product ($\bar{y}_{r1p2}$)	632	102836
9.	Unbiased ratio-cum-product type. ($\bar{y}'_{r1p2}$)	705	117206
10.	Biased multivariate product type ($\bar{y}_{p(M.V.)}$	418	95478
11.	Unbiased multivariate product type ($\bar{y}'_{p(M.V.)}$)	421	97284
12.	Biased multivariate generalised ($\bar{y}_{G(M.V.)}$)	802	47870
13.	Unbiased multivariate generalised ($\bar{y}'_{G(M.V.)}$)	734	54218

True value ($\bar{Y}$) = 672
Sample size = 100
Population size = 693

Correlation between y and x_1 = + .4173
Correlation between y and x_2 = - .0125
Correlatian between y and x_3 = - .0097

SUMMARY.

In the present study the question of developing the unbiased ratio type, product type, and ratio cum product type estimators has been taken up and appropriate estimators along with their variances have been worked out. The following estimates, the variances of the estimates and the estimates of these variances are obtained by using the concept of multivariate symmetric means.

1. Multivariate unbiased ratio type estimator of the population mean using data on two or more supplementary variables.
2. Multivariate unbiased product type estimator.
3. Generalised multivariate unbiased estimate where p of the supplementary variables having positive correlation and q of them having negative correlation with the variable under study.
4. Unbiased ratio cum product type estimator.

An emperical study has been made by using these estimators and results are included as illustration.

BIBLIOGRAPHY

1. Bowly, (1926) " Measurement of the precision attained in sampling ", Bulletin of International Statistics Institute, 22, (1), 1 - 62.

2. Cochran, W.G., (1942) " Sampling theory when the sampling units are of unequal sizes", Journal of American Statistical Association, 37, 199 - 212.

3. Cochran, W.G., (1953) " Sampling techniques", (Second edition, 1963), John Wiley and Sons Ltd., New York.

4. Deming, W.E. (1950) " Some theory of sampling", John Wiley and Sons Ltd., New York.

5. Goodman, L.A., and Hartley, H.O., (1958) " The precision of unbiased ratio type estimators", Jornal of Americah Statistical Association, 53, 491 - 508.

6. Hansen and M.H. Hurwitz, (1943) " On the theory of sampling from finite populations", Annals of Mathematical Statistics, 14, 333 - 362.

7. Hansen, M.H. and Hurwitz, (1946) " The problem of non response in sample surveys", Journal of American Statistical Association, 41, 517 - 529.

8. Hansen, M.H., Hurwitz and Madow, W.G., (1953) " sample survey methods and theory", Vol. I and II, John Wiley and Sons Ltd., New York.

9. Hartley, H.O., and Ross, A, (1954) " Unbiased ratio estimators", Nature, 174, 270 - 271.

10. Madow, W.G., and Madow, C.H., (1944) " On the theory of systematic sampling" , Annals of Mathematical Statistics, 15, 1 - 24.

11. Mahalanobis, (1940) " A sample survey of Acreage under jute in Bengal", Sankhya, 4, 511 - 530.

12. Mahalanobis, (1944) " On large scale sample surveys", Philosophical Transactions of Royal Society, London, 231(B), 329-451.

13. Mahalonobis, (1946) " Recent experiments in statistical sampling in the Indian Statistical Institute", Journal of Royal Statistical Society (A), ~~26, 69 - 74.~~ 109, 320 - 378

14. Murthy, M.N. (1964) " Product method of estimation" Sankhya Series (A), 26, 69- 74.

15. Murthy, M.N. (1967). " Sampling theory and methods", Statistical Publishing Society, Calcutta.

16. Neyman (1934) " On the two different aspects of the representative method of purposive selection", Journal of Royal Statistical Society, 97, 558 - 625.

17. Olkin, I. (1958) " Multivariate ratio estimation for finite populations" Biometrika, 45, 154-165.

18. Rao Podurai, S.R.S., and Mudholkar Govidd, S. (1967). " Generalised multivariate estimator for the mean of finite populations", Journal of American Statistical Association, 62, 511 - 512.

19. Robson, D.S., (1957) " Application of multivariate polykays to the theory of unbiased ratio type estimation, Journal of American Statistical Association, 52, 511 - 522.

20. Singh, M.P., (1965) " On the estimation of ratio and product of the population parameters", Sankhya, 27 (B), 321 - 328.

21. Singh, M.P., (1967a) " Ratio cum Product method of estimation", Metrika, 12.

22. Singh, M.P., (1967b) " Multivariate product method of estimation", Journal of the Indian Society of Agricultural Statistics, Dec., 1967, 1 - 10.

23. S.K. Srivastava, (1965) " An estimate of the mean of a finite population using several auxiliary variables", Journal of Indian Statistical Association, Vol. 3, 189 - 194.

24. Sukhatme, P.V., (1953) " Sampling theory of surveys with applications", Indian Society of Agricultural Statistics, New Delhi and Iowa State College Press, Ames, Iowa.

25. Tukey, J.W. (1956) " Keeping like moments in computation simple" Annals of Mathematical Statistics., 27, 37 - 54.