COMBINATIONS OF SOME ESTIMATORS USING SUPPLEMENTARY INFORMATION

By Dhiresh Adhvaryu
*South Gujarat University, Surat**
Received in February, 1974

1. Introduction

The ratio, regression, product and difference methods of estimation play an important role whenever supplementary information is to be utilized for estimation purposes in sample surveys. Ghosh (1947), Olkin (1958), Raj (1965), Srivastava (1966) and Rao and Mudholkar (1967) have used multi-supplementary information for constructing estimates based on these methods. In the present paper some combinations of estimators are proposed when the information on two supplementary characters is available, by considering a class of estimators. Some of them are compared with Olkin's (1958) estimator, the estimator obtained in the same way as Olkin's by using product estimators in place of ratio estimators and Srivastava's (1966) estimator. The technique is extended to two-phase and stratified sampling. Although the extension of the results of the present paper when information on more than two supplementary characters is available is straightforward, the actual computation becomes somewhat more complex and hence these extensions are not considered.

2. Estimator, it's Bias and Mean Square Error

Suppose that a sample of size n is selected from N units of the population using any sampling scheme yielding unbiased estimates. The variable y under study and the supplementary variables x_1 and x_2 are measured on it. The population mean $\bar{Y}$ of y is to be estimated using information on x_1 and x_2, the population means $\bar{X}_1$ and $\bar{X}_2$ of x_1 and x_2 respectively are available. The sample means of y, x_1 and x_2, namely $\bar{y}$, $\bar{x}_1$ and $\bar{x}_2$ are unbiased estimates of $\bar{Y}$, $\bar{X}_1$ and $\bar{X}_2$ respectively.

*Work partially done during the author's employment at V. S. Patel College, Bilimora.

Firstly we estimate the population mean of y by using the information on supplementary variable x_1, with the estimator

$$\hat{\bar{Y}}_{g_1}=\bar{y}+g_1(\bar{X}_1-\bar{x}_1)$$

where random variable g_1 converges to some fixed quality G_1 as n increases. Then this estimate is used to get the estimator, using the information on second supplementary variable x_2 also, as

$$\hat{\bar{Y}}=\hat{\bar{Y}}_{g_1}+g_2(\bar{X}_2-\bar{x}_2) \quad \ldots(1)$$

where random variable g_2 converges to fixed quantity G_2 as n increases. At the outset it may be noted that the ratio, regression, difference and product estimators belong to the class of estimators $\bar{y}+h(\bar{X}-\bar{x})$ when one supplementary variable x is used, random variable h converging to fixed quantity H. h takes values $\frac{\bar{y}}{\bar{x}}$, sample regression coefficient b_{yx}; constant d and $-\frac{\bar{y}}{\bar{X}}$, for the ratio estimator

$$\hat{\bar{Y}}_R=\frac{\bar{y}}{\bar{x}}\bar{X}, \text{ regression estimator}$$

$$\hat{\bar{Y}}_{lr}=\bar{y}+b_{yx}(\bar{X}-\bar{x}), \text{ difference estimator}$$

$$\hat{\bar{Y}}_d=\bar{y}+d(\bar{X}-\bar{x}) \text{ and product estimator}$$

$$\hat{\bar{Y}}_p=\frac{\bar{y}\,\bar{x}}{\bar{X}} \text{ respectively.}$$

For deriving the expressions for bias and mean square error of the estimator we write

$$\bar{y}=\bar{Y}(1+e),\ \bar{x}_i=\bar{X}_i(1+e_i),\ (i=1,2),\ g_1=G_1(1+e_3)$$

and

$g_2=G_2(1+e_4)$ in (1). It may be noted that $E(e)=E(e_i)=0$,

$$Var(e)=\frac{Var(\bar{y})}{\bar{Y}^2},\ Var(e_i)=\frac{Var(\bar{x}_i)}{\bar{X}_i^2}$$

and

$$Cov(e,e_i)=\frac{Cov(\bar{y},\bar{x}_i)}{\bar{Y}\bar{X}_i};\ (i=1,2).$$

Also assuming the sample moderately large so that

$$E(e_j) \doteq 0 (j=3, 4)\ ;$$

$$E(e_1e_3) \doteq \frac{Cov(g_1, \bar{x}_1)}{G_1\bar{X}_1} \quad \text{and} \quad E(e_2e_4) \doteq \frac{Cov(g_2, \bar{x}_2)}{G_2\bar{X}_2}.$$

Then it would be fairly simple to establish the following theorem :

Theorem 1. For moderately large sample the bias and the mean square error of the estimator (1) to the first order of approximation are given by

$$B(\hat{\bar{Y}}) = -[Cov\ (g_1, \bar{x}_1) + Cov\ (g_2, \bar{x}_2)] \qquad \ldots(2)$$

and

$$MSE\ (\hat{\bar{Y}}) = Var\ (\bar{y}) + G_1^2\ Var\ (\bar{x}_1) + G_2^2\ Var\ (\bar{x}_2) - 2\ G_1\ Cov\ (\bar{y}, \bar{x}_1) - 2\ G_2\ Cov\ (\bar{y}, \bar{x}_2) + 2\ G_1G_2\ Cov\ (\bar{x}_1, \bar{x}_2) \qquad \ldots(3)$$

3. Two Phase Sampling

Suppose that information on supplementary variables is not available and it is comparatively less expensive to collect information on these variables. A fairly large first phase or initial sample is taken to collect information on the supplementary variables. Then either a subsample of the large initial sample or an independent sample from the whole population, called second phase sample, is used to measure character of interest and sometimes supplementary characters also, as the case may be.

In this section the case of two phase sampling when population means of both the supplementary variables x_1 and x_2 are unknown, is considered. Let $\bar{x}_1'$ and $\bar{x}_2'$ be the means of x_1 and x_2 for the initial sample of size n'. n is the size of the second phase sample and $\bar{y}$, $\bar{x}_1$ and $\bar{x}_2$ are associated sample means for y, x_1 and x_2. We estimate $\bar{Y}$ first by $\hat{\bar{Y}}'_{g_1} = \bar{y} + g_1(\bar{x}_1' - \bar{x}_1)$ and this estimate is used to get the estimator

$$\hat{\bar{Y}}_{t_2} = \hat{\bar{Y}}'_{g_1} + g_2\ (\bar{x}_2' - \bar{x}_2) \qquad \ldots(4)$$

where, as in section 2, g_1 and g_2 converges to G_1 and G_2.

In order to find expressions for bias and mean square error of the estimator (4) we write $\bar{x}_1' = \bar{X}_1(1+e_1')$ and $\bar{x}_2' = \bar{X}_2(1+e_2')$ in addition to substituting $\bar{Y}(1+e)$, $\bar{X}_1(1+e_1)$ and $\bar{X}_2(1+e_2)$ for $\bar{y}$, $\bar{x}_1$ and $\bar{x}_2$, as in section 2; $E(e_1')=0$ and $E(e_2')=0$. As the initial sample is large and assuming second phase sample moderately large, on the similar lines as that for single phase sampling, we get the following theorems :

Theorem 2. The bias of the two phase sampling estimator (4) is given by

$$B(\hat{\bar{Y}}_{t_2}) = [Cov(g_1, \bar{x}_1') + Cov(g_2, \bar{x}_2')] - Cov(g_1, \bar{x}_1) + Cov(g_2, \bar{x}_2)] \quad \ldots(5)$$

Theorem 3. In two phase sampling if the second phase sample is a sub sample of the initial sample, mean square error of the estimator (4) to the first order, is given by

$$\begin{aligned} MSE(\hat{\bar{Y}}_{t_2}) = {} & Var(\bar{y}) + G_1^2 [Var(\bar{x}_1) + Var(\bar{x}_1') \\ & -2\, Cov(\bar{x}_1, \bar{x}_1')] + G_2^2 [Var(\bar{x}_2) + Var(\bar{x}_2') \\ & -2\, Cov(\bar{x}_2, \bar{x}_2')] + 2\, G_1[Cov(\bar{y}, \bar{x}_1') - Cov(\bar{y}, \bar{x}_1)] \\ & +2\, G_2[Cov(\bar{y}, \bar{x}_2') - Cov(\bar{y}, \bar{x}_2)] \\ & +2\, G_1G_2[Cov(\bar{x}_1, \bar{x}_2) + Cov(\bar{x}_1' \bar{x}_2') \\ & - Cov(\bar{x}_1, \bar{x}_2') - Cov(\bar{x}_1' \bar{x}_2)] \end{aligned} \quad \ldots(6)$$

Theorem 4. In two phase sampling if the initial and second phase samples are independent, mean square error of (4), to the first order, is given by

$$\begin{aligned} MSE(\hat{\bar{Y}}_{t_2}) = {} & Var(\bar{y}) + G_1^2 [Var(\bar{x}_1) + Var(\bar{x}_1')] \\ & + G_2^2 [Var(\bar{x}_2) + Var(\bar{x}_2')] - 2\, G_1 Cov(\bar{y}, \bar{x}_1) \\ & -2\, G_2\, Cov(\bar{y}, \bar{x}_2) + 2\, G_1 G_2[Cov(\bar{x}_1, \bar{x}_2) \\ & + Cov(\bar{x}_1', \bar{x}_2')] \end{aligned} \quad \ldots(7)$$

Remark. We note that in case, the population mean of one supplementary variable, say x_1, is known, the other variable x_2 is only observed on the initial sample and the population mean $\bar{X}_2$ is estimated by $\bar{x}_2'$, the initial sample mean of x_2. Then the estimator

$$\hat{\bar{Y}}_{t_1} = \hat{\bar{Y}}_{g_1} + g_2(\bar{x}_2' - \bar{x}_2) \qquad \ldots(8)$$

obtained by substituting $\bar{x}_2'$ for $\bar{X}_2$ in (1), is used to estimate the population mean $\bar{Y}$. One can find the expressions for bias and mean square error of (8) on the similar lines as that for (4). We omit them to save space.

3.1. Special Cases

Simple random sampling and systematic sampling are among frequently used sampling designs sample surveys, so these two specified cases may be of interest.

Simple Random Sampling

Suppose that simple random sampling with replacement is used. Then for the estimator (1) one can easily write down the expressions for mean square error, hence we omit it.

We give below in (9) and (10) the expressions for mean square error of the estimator (4) for the cases (*a*) when the second phase sample is a sub sample of the first phase sample and (*b*) when they are independent, with simple random sampling with replacement at both the phases.

$$MSE(\hat{\bar{Y}}_{t_{2srs}}) = \frac{\sigma_y^2}{n} + \left(\frac{1}{n} - \frac{1}{n'}\right)(G_1^2\sigma_{x_1}^2 + G_2^2\sigma_{x_2}^2$$
$$-2G_1\rho_{yx_1}\sigma_y\sigma_{x_1} - 2G_2\rho_{yx_2}\sigma_y\sigma_{x_2}$$
$$+2G_1G_2\rho_{x_1x_2}\sigma_{x_1}\sigma_{x_2}) \qquad \ldots(9)$$

$$MSE(\hat{\bar{Y}}_{t_{2srs}}) = \frac{1}{n}(\sigma_y^2 - 2G_1\rho_{yx_1}\sigma_y\sigma_{x_1} - 2G_2\rho_{yx_2}\sigma_y\sigma_{x_2})$$
$$+\left(\frac{1}{n} - \frac{1}{n'}\right)\Big(G_1^2\sigma_{x_1}^2 + G_2^2\sigma_{x_2}^2$$
$$+2G_1G_2\rho_{x_1x_2}\sigma_{x_1}\sigma_{x_2}) \qquad \ldots(10)$$

where σ^2's have their usual meaning and ρ_{yx_1}, ρ_{yx_2} and $\rho_{x_1 x_2}$ are the correlations between y and x_1, y and x_2 and x_1 and x_2 respectively. Let us consider the cost function of the form

$$c=c_1 n'+c_2 n+c_3 \qquad ...(11)$$

where c_1 and c_2 are cost per unit for collecting information at second and first phase respectively and c_3 is overhead cost. To get the optimum value of n and n' we note that (9) and (10) are of the form

$$M=\frac{M_1}{n'}+\frac{M_2}{n} \qquad ...(12)$$

where M_1 and M_2 are 'variance type' functions. Then the optimum values of n' and n which minimize mean square error (12) for a fixed cost c_0 are

$$n'_{opt}=\frac{(c_0-c_3)\sqrt{M_1 c_2}}{c_1\sqrt{M_1 c_1}+c_2\sqrt{M_2 c_1}} \qquad ...(13)$$

and

$$n_{opt}=\frac{(c_0-c_3)\sqrt{M_2 c_1}}{c_1\sqrt{M_1 c_2}+c_2\sqrt{M_2 c_1}} \qquad ...(14)$$

Systematic Sampling

In sytematic sampling with sampling interval μ, the population can be divided into μ clusters, the ith cluster containing units numbered i, $i+\mu$, ..., ..., $i+(n-1)\mu$, assuming that the $N(=n\mu)$ units of the population are numbered from 1 to N. For two phase sampling we select λ clusters in the initial sample and in the second sample a cluster is selected randomly either (a) from these λ clusters or (b) from the μ clusters of the population. For simplicity we assume that intra cluster correlations of y, x_1 and x_2 are equal, that is $\rho_y=\rho_{x_1}=\rho_{x_2}=\rho$(say). Using well known results of cluster sampling the mean squar error of (4) under (a) and (b) are obtained as under:

$$MES(\hat{\bar{Y}}_{t_2}{}^{sy})=\frac{1}{n}[1+\rho(n-1)][\sigma_y^2+\left(1-\frac{1}{\lambda}\right)(G_1^2\sigma_{x_1}^2$$
$$+G_2^2\sigma_{x_2}^2-2G_1\rho_{yx_1}\sigma_y\sigma_{x_1}-2G_2\rho_{yx_2}\sigma_y\sigma_{x_2}$$
$$+2G_1G_2\rho_{x_1x_2}\sigma_{x_1}\sigma_{x_2})] \qquad ...(15)$$

$$MSE(\hat{\bar{Y}}_{t_2}{}^{sy})=\frac{1}{n}[1+\rho(n-1)][\sigma_y^2-2G_1\rho_{yx_1}\sigma_y\sigma_{x_1}$$
$$-2G_2\rho_{yx_2}\sigma_y\sigma_{x_2}+\left(1+\frac{1}{\lambda}\right)(G_1^2\sigma_{x_1}^2+G_2^2\sigma_{x_2}^2$$
$$+2G_1G_2\rho_{x_1x_2}\sigma_{x_1}\sigma_{x_2})] \qquad ...(16)$$

We note that in this case the cost function (11) will be $c=(c_1\lambda+c_2)n+c_3$ and expressions (15) and (16) are of the form :

$$M=[1+\rho(n-1)]\left[\frac{M_1}{n\lambda}+\frac{M_2}{n}\right] \qquad ...(17)$$

where M_1 and M_2 are 'variance type' functions. The optimum values of λ and n which minimize (17) for a fixed cost c_0 are

$$\lambda_{opt}=\sqrt{\frac{M_1}{M_2}\left[\frac{c_2}{c_1}+\left(\frac{\rho}{1-\rho}\right)\frac{c_0-c_3}{c_1}\right]} \qquad ...(18)$$

and

$$n_{opt}=\frac{c_0-c_3}{c_1\lambda_{opt}+c_2} \qquad ...(19)$$

Remark. One can obtain the optimum values which minimize cost for a fixed mean square error in simple random sampling and systematic samling.

4. Particular Cases

Keeping in view that the ratio, regression and product estimators belong to the class $\bar{y}+h(\bar{X}-\bar{x})$, for one supplementary variable, the following combinations of estimators are constructed as particular cases of (1) :

(*i*) Ratio—Ratio estimator

$$\hat{\bar{Y}}_{R,\,R,}\left(g_1=\frac{\bar{y}}{\bar{x}_1}\;;\;g_2=\frac{\hat{\bar{Y}}_R}{\bar{x}_2}\right)$$

(*ii*) Ratio—Regression estimator

$$\hat{\bar{Y}}_{R,\,Reg,}\left(g_1=\frac{\bar{y}}{\bar{x}_1}\;;\;g_2=b_{yx_2}\right)$$

(*iii*) Ratio—Product estimator

$$\hat{\bar{Y}}_{R,\,P,}\left(g_1=\frac{\bar{y}}{\bar{x}_1}\;;\;g_2=-\frac{\hat{\bar{Y}}_R}{\bar{x}_2}\right)$$

(*iv*) Regression—Regression estimator

$$\hat{\bar{Y}}_{Reg,\,Reg,}\;(g_1=b_{yx_1}\;;\;g_2=b_{yx_2})$$

(*v*) Regression—Ratio estimator

$$\hat{\bar{Y}}_{Reg,\,R,}\left(g_1=b_{yx_1}\;;\;g_2=\frac{\hat{\bar{Y}}_{1r}}{\bar{x}_2}\right)$$

(*vi*) Regression—Product estimator

$$\hat{\bar{Y}}_{Reg,\ P}\left(g_1 = b_{yx1}\ ;\ g_2 = -\frac{\hat{\bar{Y}}_{1r}}{\bar{X}_2}\right)$$

(*vii*) Productor—Productor estimator

$$\hat{\bar{Y}}_{p,\ p,}\left(g_1 = -\frac{\bar{y}}{\bar{X}_1}\ ;\ g_2 = -\frac{\hat{\bar{Y}}_P}{\bar{X}_2}\right)$$

(*viii*) Product—Ratio estimator

$$\hat{\bar{Y}}_{P,\ R,}\left(g_1 = -\frac{\bar{y}}{\bar{X}_1}\ ;\ g_2 = \frac{\hat{\bar{Y}}_P}{\bar{x}_2}\right)$$

(*ix*) Product—Regression estimator

$$\hat{\bar{Y}}_{P,\ Reg,}\left(g_1 = -\frac{\bar{y}}{\bar{X}_1}\ ;\ g_2 = b_{yx_2}\right)$$

where $\hat{\bar{Y}}_R$, $\hat{\bar{Y}}_P$ and $\hat{\bar{Y}}_{1r}$ are the ratio, product and regression estimators of $\bar{Y}$ based on the first supplementary variable x_1 and b_{yx_1} and b_{yx_2} are the sample regression coefficients of y on x_1 and y on x_2 respectively.

Using theorems 1 and 2 the expressions for bias of these estimators in single phase sampling and two phase sampling when both $\bar{X}_1$ and $\bar{X}_2$ are unknown are as as given below :

$$B(\hat{\bar{Y}}_{R,\ R}) = \frac{1}{\bar{X}_1}[R_1\ Var\ (\bar{x}_1) + Cov\ (\bar{y},\ \bar{x}_1)] + \frac{1}{\bar{X}_2}[R_2\ Var\ (\bar{x}_2) + Cov\ (\bar{y},\ \bar{x}_2) - R_1\ Cov\ (\bar{x}_1,\ \bar{x}_2)] \quad ...(20)$$

$$B(\hat{\bar{Y}}_{R,\ P}) = \frac{1}{\bar{X}_1}[R_1\ Var\ (\bar{x}_1) - Cov\ (\bar{y},\ \bar{x}_1)] + \frac{1}{\bar{X}_2}[Cov\ (\bar{y},\ \bar{x}_2) - R_1\ Cov\ (\bar{x}_1,\ \bar{x}_2)] \quad ...(21)$$

$$B(\hat{\bar{Y}}_{R,\ Reg}) = \frac{1}{\bar{X}_1}[R_1\ Var\ (\bar{x}_1) - Cov\ (\bar{y},\ \bar{x}_1)] - Cov\ (b_{yx_2},\ \bar{x}_2) \quad ...(22)$$

$$B(\hat{\bar{Y}}_{P,\ P}) = \frac{1}{\bar{X}_1}Cov\ (\bar{y},\ \bar{x}_1) + \frac{1}{\bar{X}_2}[Cov\ (\bar{y},\ \bar{x}_2) + R_1 Cov\ (\bar{x}_1,\ \bar{x}_2)] \quad ...(23)$$

$$B(\hat{\bar{Y}}_{P,\,R})=\frac{1}{\bar{X}_1}\,Cov\,(\bar{y},\,\bar{x}_1)+\frac{1}{\bar{X}_2}\,[R_2\,Var\,(\bar{x}_2)-Cov\,(\bar{y},\,\bar{x}_2)-R_1\,Cov\,(\bar{x}_1,\,\bar{x}_2)] \quad \ldots(24)$$

$$B(\hat{\bar{Y}}_{P,\,Reg})=\frac{1}{\bar{X}_1}\,Cov\,(\bar{y},\,\bar{x}_1)-Cov(b_{yx_2}\,,\,\bar{x}_2) \quad \ldots(25)$$

$$B(\hat{\bar{Y}}_{Reg,\,Reg})=-[Cov\,(b_{yx_1}\,,\,\bar{x}_1)+Cov\,(b_{yx_2},\,\bar{x}_2)] \quad \ldots(26)$$

$$B(\hat{\bar{Y}}_{Reg,\,R})=-Cov\,(b_{yx_1},\,\bar{x}_1)+\frac{1}{\bar{X}_2}\,[R_2\,Var\,(\bar{x}_2)-Cov(\bar{y},\,\bar{x}_2)+B_{yx_1}\,Cov(\bar{x}_1,\,\bar{x}_2)] \quad \ldots(27)$$

$$B(\hat{\bar{Y}}_{Reg,\,P})=-Cov\,(b_{yx_1},\,\bar{x}_1)+\frac{1}{\bar{X}_2}\,[Cov(\bar{y},\,\bar{x}_2)-B_{yx_1}\,Cov\,(\bar{x}_1,\,\bar{x}_2)] \quad \ldots(28)$$

For two phase sampling (both $\bar{X}_1$ and $\bar{X}_2$ unknown)

$$B(\hat{\bar{Y}}_{R,\,R_{(2)}})=\frac{1}{\bar{X}_1}\,[A'-R_1D+R_2(B+B')]+\frac{1}{\bar{X}_2}\,[A-R_2C] \quad \ldots(29)$$

$$B(\hat{\bar{Y}}_{R,\,P_{(2)}})=\frac{1}{\bar{X}_1}\,[A'-R_1D-R_2(B+B')]-\frac{1}{\bar{X}_2}\,[A+R_2C'] \quad \ldots(30)$$

$$B(\hat{\bar{Y}}_{R,\,Reg_{(2)}})=\frac{1}{\bar{X}_1}[A'-R_1D]+Cov(b_{yx_2}\,,\,\bar{x}'_2)-Cov(b_{yx_2},\,\bar{x}_2) \quad \ldots(31)$$

$$B(\hat{\bar{Y}}_{P,\,P_{(2)}})=-\frac{1}{\bar{X}_1}[A'+R_1D'-R_2(B+B')]-\frac{1}{\bar{X}_2}[A+R_2C'] \quad \ldots(32)$$

$$B(\hat{\bar{Y}}_{P,\,R_{(2)}})=-\frac{1}{\bar{X}_1}\,[A'+R_1D'+R_2(B+B')]+\frac{1}{\bar{X}_2}\,[A-R_2C] \quad \ldots(33)$$

$$B(\hat{\bar{Y}}_{P,\,Reg_{(2)}})=-\frac{1}{\bar{X}_1}\,[A'+R_1D']+Cov(b_{yx_2}\,,\,\bar{x}'_2)-Cov(b_{yx_2}\,,\,\bar{x}_2) \quad \ldots(34)$$

$$B(\hat{\bar{Y}}_{Reg,\,Reg_{(2)}})=-Cov\,(b_{yx_1}\,,\,\bar{x}'_1)-Cov(b_{yx_1}\,,\,\bar{x}_1)+Cov(b_{yx_2},\,\bar{x}'_2)-Cov(b_{yx_2}\,,\,\bar{x}_2) \quad \ldots(35)$$

$$B(\hat{\bar{Y}}_{Reg,\ R_{(2)}})=Cov(b_{yx_1},\ \bar{x}_1')-Cov(b_{yx_1},\ \bar{x}_1)$$
$$+\frac{1}{\bar{X}_2}[A+B_{yx_1}(B+B')-R_2C] \qquad ...(36)$$

$$B(\hat{\bar{Y}}_{Reg,\ P_{(2)}})=Cov(b_{yx_1},\ \bar{x}_1')-Cov(b_{yx_1},\ \bar{x}_1)$$
$$-\frac{1}{\bar{X}_2}[A-B_{yx_1}(B+B')+R_2C'] \qquad ...(37)$$

where

$$R_1=\frac{\bar{Y}}{\bar{X}_1},\ R_2=\frac{\bar{Y}}{\bar{X}_2}$$

$$A=Cov(\bar{y},\ \bar{x}_2')-Cov(\bar{y},\ \bar{x}_2),\ A'=Cov(\bar{y},\ \bar{x}_1')-Cov(\bar{y},\ \bar{x}_1)$$

$$B=Cov(\bar{x}_1,\ \bar{x}_2)-Cov(\bar{x}_1,\ \bar{x}_2'),\ B'=Cov(\bar{x}_1'),\bar{x}_2'-Cov(\bar{x}_1^1,\ \bar{x}_2)$$

$$C=Cov(\bar{x}_2,\ \bar{x}_2')-Var(\bar{x}_2),\ C'=Cov(\bar{x}_2',\ \bar{x}_2)-Var(\bar{x}_2')$$

$$D=Cov(\bar{x}_1,\ \bar{x}_1')-Var(\bar{x}_1),\ D'=Cov(\bar{x}_1,\ \bar{x}_1')-Var(\bar{x}_1')$$

and B_{yx_1} and B_{yx_2} are the population regression coefficients of y on x_1 and y on x_2. Expression of the bias will be different in two phase sampling if the first and second phase samples are independent. Also one can easily obtain the expressions for mean square error of the combinations by using (3), (6) and (7). We omit them.

4.1. Some Comparisons

Consider the following multivariate estimators

$$\hat{\bar{Y}}'_{R,\ R}=W_1\hat{\bar{Y}}_{R_1}+W_2\hat{\bar{Y}}_{R_2} \qquad ...(38)$$

$$\hat{\bar{Y}}'_{P,\ P}=W_1\hat{\bar{Y}}_{P_1}+W_2\hat{\bar{Y}}_{P_2} \qquad ...(39)$$

and

$$\hat{\bar{Y}}'_{Reg,\ Reg}=W_1\hat{\bar{Y}}_{Reg_1}+W_2\hat{\bar{Y}}_{Reg_2} \qquad ...(40)$$

where $\hat{\bar{Y}}_{R_1}$, $\hat{\bar{Y}}_{P_1}$, $\hat{\bar{Y}}_{Reg_1}$ and $\hat{\bar{Y}}_{R_2}$, $\hat{\bar{Y}}_{P_2}$, $\hat{\bar{Y}}_{Reg_2}$ are ratio, product and regression estimators based on x_1 and x_2 respectively. For the sake of simplicity we assume that the coefficients of variation of $\bar{x}_1$ and $\bar{x}_2$ are equal, that is $C_{\bar{x}_1} = C_{\bar{x}_2} = C_{\bar{x}}$ (say) and there is same correlation ρ_{yx} between y and x_i $(i=1, 2)$.

Then the mean square errors of these estimators are approximately given by :

$$MSE\,(\hat{\bar{Y}}'_{R,\,R}) = \bar{Y}^2[C^2_{\bar{y}} - 2\,\rho_{yx} C_{\bar{y}} C_{\bar{x}} + \frac{C^2_{\bar{x}}}{2}\,(1+\rho_{x_1 x_2})] \qquad \ldots(41)$$

$$MSE(\hat{\bar{Y}}'_{P,\,P}) = \bar{Y}^2[C^2_{\bar{y}} + 2\,\rho_{yx} C_{\bar{y}} C_{\bar{x}} + \frac{C^2_{\bar{x}}}{2}(1+\rho_{x_1 x_2})] \qquad \ldots(42)$$

$$MSE(\hat{\bar{Y}}'_{Reg,\,Reg}) = \frac{\sigma_y{}^2}{n}\left[1 - \rho^2_{yx}\frac{(3-\rho_{x_1 x_2})}{2}\right] \text{(Srivastava -- 1966)} \qquad (43)$$

The mean square errors of $\hat{\bar{Y}}_{R,\,R}$, $\hat{\bar{Y}}_{P,\,P}$, and $\hat{\bar{Y}}_{Reg,\,Reg}$ are approximately given by

$$MSE(\hat{\bar{Y}}_{R,\,R}) = \bar{Y}^2\left[C^2_{\bar{y}} - 4\rho_{yx} C_{\bar{y}} C_{\bar{x}} + 2\,C_{\bar{x}}(1+\rho_{x_1 x_2})\right] \qquad \ldots(44)$$

$$MSE(\bar{Y}_{P,\,P}) = \bar{Y}^2\left[C^2_{\bar{y}} + 4\,\rho_{yx} C_{\bar{y}}\, C_{\bar{x}} + 2\,C^2_{\bar{x}}\,(1+\rho_{x_1 x_2})\right] \ldots(45)$$

$$MSE(\hat{\bar{Y}}_{Reg,\,Reg}) = \frac{\sigma_y{}^2}{n}\left[1 - 2\,\rho^2_{xy}(1-\rho_{x_1 x_2})\right] \qquad \ldots(46)$$

Comparing the mean square errors of $\hat{\bar{Y}}_{R,\,R}$, $\hat{\bar{Y}}_{P,\,P}$ and $\hat{\bar{Y}}_{Reg,\,Reg}$ with that of $\hat{\bar{Y}}'_{R,\,R}$ $\hat{\bar{Y}}'_{P,\,P}$ and $\hat{\bar{Y}}'_{Reg,\,Reg}$ respectively it will be observed that the former estimators will be more efficient than the later estimators if the following conditions (47), (48) and (49) respectively are satisfied :

$$\frac{\rho_{yx}}{1+\rho_{x_1 x_2}} \; \frac{C_{\bar{y}}}{C_{\bar{x}}} > \frac{3}{4} \qquad \ldots(47)$$

$$\frac{\rho_{yx}}{1+\rho_{x1x2}} \; \frac{C_{\bar{y}}}{C_{\bar{x}}} < -\frac{3}{4} \qquad \ldots(48)$$

and

$$\rho_{x_1 x_2} < \frac{1}{3} \qquad \ldots(49)$$

It may be noted that the conditions depend on the signs of $\bar{Y}$, $\bar{X}_1$ and $\bar{X}_2$. The above conditions are derived under the assumption that all $\bar{Y}$, $\bar{X}_1$ and $\bar{X}_2$ are either positive or negative simultaneously.

5. Stratified Sampling

Let there be strata of sizes $N_1, N_2 \ldots, N_L$ from which samples of sizes $n_1, n_2, \ldots, n_L$ respectively are taken, sampling being independent in each stratum :

$(\sum_{h=1}^{L} N_h = N, \sum_{h=1}^{L} n_h = n.)$ Let $\bar{y}_h, \bar{x}_{1h}$ and $\bar{x}_{2h}$ $(h=1, 2, \ldots, L)$ be the sample means of the variables y, x_1 and x_2 for the h^{th} stratum. $\bar{y}_{st} = \sum_{h=1}^{L} W_h \bar{y}_h$, $\bar{x}_{1st} = \sum_{h=1}^{L} W_h \bar{x}_{1h}$ and $\bar{x}_{2st} = \sum_{x=1}^{L} W_h \bar{x}_{2h}$, where $W_h = \frac{N_h}{N}$, are stratified sample means of y, x_1 and x_2. Then the following two types of estimators of $\bar{Y}$ can be formed

(1) Separate estimator $\hat{\bar{Y}}_s = \sum_{h=1}^{L} W_h \hat{\bar{Y}}_h$ (50)

where $\hat{\bar{Y}}_h = \bar{y}_{g_1 h} + g_{2h}(\bar{X}_{2h} - \bar{x}_{2h})$; $\bar{y}_{g1h} = \bar{y}_h + g_{1h}(\bar{X}_{1h} - \bar{x}_{1h})$ and g_{1h} and g_{2h} are the quntities as g_1 and g_2 for h^{th} stratum.

(2) Combined estimator $\hat{\bar{Y}}_C = \bar{y}_{g_1 st} + g_2(\bar{X}_2 - \bar{x}_{2st})$ ...(51)

where

$$\bar{y}_{gst} = \bar{y}_{st} + g_1(\bar{X}_1 - \bar{x}_{1st})$$

Bias and mean square error of (50) and (51) are given by the following theorems :

Theorem 5. Bias and mean square error of the separate estimator to the first order, are

$$B(\hat{\bar{Y}}_s) = -\sum_{h=1}^{L} W_h \left[Cov(g_{1h}, \bar{x}_{1h}) + Cov(g_{2h}, \bar{x}_{2h})\right] \quad \ldots(52)$$

and

$$MSE(\hat{\bar{Y}}_s) = \sum_{h=1}^{L} W_h^2 \, MSE\,(\hat{\bar{Y}}_h) \quad \ldots(53)$$

Theorem 6. Bias and mean square error of the combined estimator to the first order, are

$$B(\hat{\bar{Y}}_C) = -[Cov(g_1, \bar{x}_{1st}) + Cov(g_2, \bar{x}_{2st})] \quad \ldots(54)$$

and

$$MSE(\hat{\bar{Y}}_C) = Var(\bar{y}_{st}) + G_1^2\, Var(\bar{x}_{1st}) + G_2^2\, Var(\bar{x}_{2st})$$
$$-2\, G_1\, Cov(\bar{y}_{st}, \bar{x}_{1st}) - 2\, G_2\, Cov(\bar{y}_{st}, \bar{x}_{2st})$$
$$+2\, G_1 G_2\, Cov(\bar{x}_{1st}, \bar{x}_{2st}) \quad ...(55)$$

6. Concluding Remarks

With Ratios, Products and Regressions, and with two supplementary variates, one can formulate $3^2=9$ estimators as given in Section 4. If we add the difference method to the list there will be 16 different estimators. With these four methods and three supplementary variates there will be 64 different estimators and so on, many more estimators can be generated by this method.

Finally we note that the ratio cum product estimators proposed by Singh (1965, 1967) are the Ratio-Ratio and Ratio-Product estimators of the present paper. So the present paper gives different approach to Singh's estimators. Mohanty (1967) has discussed Regression-Ratio estimator of the present paper, to which the author's attention was drawn after the first draft of the paper was ready.

Summary

We consider estimation of the mean $\bar{Y}$ of a finite population with the help of information on two auxiliary characters x_1 and x_2, from sample of size n selected from the population of N units using any sampling scheme for which sample means $\bar{y}$, $\bar{x}_1$ and $\bar{x}_2$ are the unbiased estimators of population means $\bar{Y}$, $\bar{X}_1$ and $\bar{X}_2$ of the characters y, x_1 and x_2 respectively. The population mean $\bar{Y}$ of y is firstly estimated by using the estimator $\hat{\bar{Y}}_{g_1} = \bar{y} + g_1(\bar{X}_1 - \bar{x}_1)$ and this estimate is used to get the estimator $\hat{\bar{Y}} = \hat{\bar{Y}}_{g_1} + g_2(\bar{X}_2 - \bar{x}_2)$ where g_1 and g_2 converges to G_1 and G_2 respectively as n increases. The bias and mean square error of this estimator for moderately large samples are obtained. Case of two phase sampling when either $\bar{X}_1$ or $\bar{X}_2$ or both are unknown is considered. By giving different values to g_1 and g_2 various estimators are constucted. Some of them are compared with Olkin's (1958) estimator, the estimator obtained in the same way as Olkin's by using product estimators in place of ratio estimators and Srivastava's (1966) estimator.

ACKNOWLEDGEMENTS

I have great pleasure in expressing my thanks to Dr. S.W. Dharmadhikari for going through some sections of the manuscript and making some helpful suggestions. I am also grateful to Dr. S. M. Shah for drawing my attention to the work of Mohanty (1967).

REFERENCES

GHOSH, B. (1947) : Double sampling with many auxiliary variates, Cal. Stat. Assoc. Bull., 1, 91–93.

MOHANTY, S. (1967) : Combination of regression and ratio estimate, J. Ind. Stat. Assoc., 5, 16—29.

MURTHY, M N. (1964) : Product method of estimation, Sankhya Series A, 26, 69—72.

MURTHY, M.N. (1967) : *Sampling Theory and Methods*, Statistical Publishing Society, Calcutta.

OLKIN, I. (1958) : Multivariate ratio estimation for finite population Biometrika, 45, 154—165.

RAJ, DES (1965) : On a method of using multi-auxiliary information, in sample surveys. J.—Amer. Stat. Assoc., 60, 270—277.

RAJ, DES (1968) : *Sampling Theory*, McGraw Hill Inc.

RAO, P.S.R.S. AND MUDHOLKAR, G.S. (1967) : Generalized multivariate estimator for the mean of finite population. J. Amer. Stat. Assoc, 62, 1009—1012.

SINGH, M.P. (1965) : On the estimation of ratio and product of the population parameters, Sankhya Series B, 27 321—328.

SINGH, M.P. (1967) : Ratio cum Product method of estimation. Metrika, 12, 34—42.

SRIVASTAVA, S.K. (1966) : On ratio and linear regression methods of estimation with several auxiliary variables. J. Ind. Stat. Assoc., 4, 66–72.