DOUBLE SAMPLING ON SUCCESSIVE OCCASIONS USING A TWO-STAGE DESIGN

By

O. P. Kathuria

Indian Agricultural Statistics Research Institute, New Delhi

(Received : March, 1976)

1. Introduction

In a sample survey, besides the main character under study, information is generally available on a number of related characters also. If the survey is of repeat nature, the information so obtained may be utilised to improve current estimates of the character under study. When the population values of the related characters are not known, they may be estimated by resorting to double sampling. The information so obtained may then be utiltised in obtaining improved estimates of the main character under study. An illustration may be provided by the sample surveys conducted by the Indian Agricultural Statistics Research Institute (IASRI) for methodological investigations into high yielding varieties programme wherein sampling units have been partially matched between two consecutive seasons and estimates ore required to be obtained of the area under high yielding varieties of crops as well as of the crop yields.

The theory of sampling on successive occasions developed by Jessen [3], Yates [15], Patterson [6], Tikkiwal [11, 12] and others for single stage units was extended by Kathuria [4], Singh [9], Singh and Kathuria [10] and Tikkiwal [13] to multi-stage designs. Tikkiwal [14] developed the theory of multi-phase sampling on successive occasions with partial matching of units for the main and related characters. Sen [7], Singh and Srivastava [8] and Kathuria [5] made use of multi-auxiliary information in obtaining improved estimates of the population character on the current occasion.

In the present study it is assumed that the population value of the related variate is not known. Using a two-stage sampling design, by adopting a suitable replacement scheme for the primary sampling units (psu's) improved estimator of the mean of related character has first been obtained. With further sub-sampling, improved estimator of the mean of the main character under study has been obtained by utilising information available on it from the previous occasion as well on the related variates.

We shall build (*i*) a linear unbiased estimator, (*ii*) a double sample estimator, and (*iii*) a ratio-type composite estimator of the main variate under study for sampling on two occasions and examine their relative efficiencies.

2. Replacement Procedure of Samplimg Units

Consider a population Π consisting of N psu's, the i^{th} psu containing M_i second stage units (ssu's). On the first occasion, a sample of n psu's is drawn on which the related variate X is observed. Out of these n psu's a sub-sample of n_1 units is drawn on which the main variate Y is also observed. On the second occasion n_1' psu's out of n drawn on the first occasion are retained and a sample of n_2' units is drawn afresh on which the X variate is observed. Out of the n_1 units for Y on the first occasion, a sample of n_1'' units is retained on the second occasion and a sample of n_2'' units is drawn afresh : these n_2'' units could be drawn either out of n_2' units for X on the second occasion or they could be drawn afresh from the population. The selection of psu's in the sample is done with probability proportional to size with replacement (ppswr) on both the occasions, size being the number of ssu's within each psu. For the selection of ssu's we shall follow the method due to Cochran [1] which suggests that if the *i*-th psu is drawn s_i times then s_i sub-sample of size m each are drawn independently with equal probability and without replacement from the i^{th} psu, each sub-sample being replaced after it is drawn. The ppswr scheme has also been used by Des Raj [2] for sampling on two occasions using a single stage sampling design. We assume that $n'_1 \geqslant n_1$. Broadly, we envisage the following pattern of replacement of psu's in the sample :

1st occasion

X variate $\quad \underbrace{XXXXXXXXX}_{n}$

Y variate $\quad \underbrace{XXXXXXX}_{n_1}$

2nd occasion

X variate $\quad \underbrace{XXXXXXX}_{n_1'}, \quad \underbrace{XXXXXXXXX}_{n_2'},$

Y variate $\quad \underbrace{XXX}_{n_1''} \quad \underbrace{XXXX}_{n''_2} \text{ or } \underbrace{XXXX}_{n_2''}$

3. NOTATIONS

Let z_{tij} ($z=y, x$) be the value of the j-th ssu in the i-th psu and $\bar{Z}_t$ ($Z=Y, X$) the population mean on the t-th occasion. Let P_i be the probability of selecting the i-th psu such that

$$\sum_{i=1}^{N} P_i=1 \text{ and let } M_0=\sum_{i=1}^{N} M_i.$$

Let
$$S^2_{tzi}=\frac{1}{(M_i-1)}\sum_{j=1}^{M_i}(z_{tij}-\bar{Z}_{ti})^2$$

be the variance between ssu's in the i-th psu of the population. Also let

$$\sigma^2_{tzw}=\sum_{i=1}^{N}\frac{M_i^2}{M_0^2P_i}\left(\frac{1}{m}-\frac{1}{M_i}\right)S^2_{tzi},$$

$$\sigma^2_{tzl}=\sum_{i=1}^{N}P_i\left(\frac{M_i}{M_oP_i}\bar{Z}_{ti}-\bar{Z}_t\right)^2,$$

$$S^2_{tt'zui}=\frac{1}{(M_i-1)}\sum_{j=1}^{M_i}(z_{tij}-\bar{Z}_{ti})\,(u_t{}'_{ij}-\bar{U}_t{}'_i),$$

$$\sigma^2_{tt'zuw}=\sum_{i=1}^{N}\frac{M^2_i}{M^2_o\,P_i}\left(\frac{1}{m}-\frac{1}{M_i}\right)S^2_{tt'zui}.$$

$$\sigma^2_{tt'zub}=\sum_{i=1}^{N}P_i\left(\frac{M_i}{M_oP_i}\bar{Z}_{ti}-\bar{Z}_t\right)\left(\frac{M_i}{M_oP_i}\bar{U}_t{}'_i-\bar{U}_t{}'\right)$$

$$\sigma^2_{tz}=\sigma^2_{tzb}+\sigma^2_{tzw}$$

$$R_{tt'zu}\,\sigma_{tz}\,\sigma_{t'u}=\sigma^2_{tt'zub}+\sigma^2_{tt'zuw}$$

$$=a^2_{tt'zu}$$

for all

$$t \neq t' = 1, 2 ;$$

$$z, u = y, x$$

and

$$Z, U = Y, X$$

and for

$$t, t' = 1, 2$$

when

$$z \neq u = y, x$$

and

$$Z \neq U = Y, X.$$

4. Linear Unbiased Estimator of Means of X and Y Variates

Consider first the x-variate. Let x_{1n_1}', x_{2n_1}' and x_{1n_2}', x_{2n_2}' be the sample means on first and second occasions based on $n_1'm$ and $n_2'm$ units respectively and let $\bar{x}_{1n}$ be the sample mean based on nm units on the first occasion. A minimum variance linear unbiased estimator of $\bar{X}_2$, the population mean of X on the second occasion may be written as

$$\hat{\bar{X}}_2 = \phi_x \left[\bar{x}_{2n_1}' + R_{12x} \sigma_{1x}^{-1} \sigma_{2x} (\bar{x}_{1n} - \bar{x}_{1n_1}') \right] + (1 - \phi_x) \bar{x}_{2n_2}' \quad \ldots(4.1)$$

where

$$\phi_x = nn_1' \left(n^2 - n_2'^2 R_{12x}^2 \right)^{-1} \quad \ldots(4.2)$$

$R_{tt'zu}$ is written as $R_{tt'z}$ when $z=u$ and R_{tzu} when $t=t'$. Consider now the Y-variate. Let $\bar{y}_{1n_1}''$. $\bar{y}_{2n_1}''$ and $\bar{y}_{1n_2}''$, $\bar{y}_{2n_2}''$ denote the sample means based on $n_1'' m$, $n_2''m$ units on first and second occasions respectively. We write a linear estimator of $\bar{Y}_2$, the population mean of Y on the second occasion as

$$\bar{y}_2 = a (\bar{y}_{1n_1}'' - \bar{y}_{1n_2}'') + e \bar{y}_{2n_1}'' + (1-e) \bar{y}_{2n_2}'' + g \left(\hat{\bar{X}}_2 - \bar{x}_{2n_2}' \right) \quad \ldots(4.3)$$

The conditions for $\bar{y}_2$ to have minimum variance are

$$\text{Cov} (\bar{y}_2, \bar{y}_{1n_1}'') = \text{Cov} (\bar{y}_2, \bar{y}_{1n_2}'') \quad \ldots(4.4)$$

$$\text{Cov} (\bar{y}_2, \bar{y}_{2n_1}'') = \text{Cov} (\bar{y}_2, \bar{y}_{2n_2}'') \quad \ldots(4.5)$$

$$\text{Cov} (\bar{y}_2, \hat{\bar{X}}_2) = \text{Cov} (\bar{y}_2, \bar{x}_{2n_2}') \quad \ldots(4.6)$$

Two cases arise; namely, (I) when n''_2 psu's for y on the second occasion are drawn out of n'_2 units for X and (II) when they are drawn independently from the population.

Case I : n''_2 units for y sub-sampled out of n'_2 units for x. On simplifying equations (4.4) to (4.6) we get the values of a, e and g as

$$a=-e[n''_2 R_{12y}\sigma_{2y}/n_1\sigma_{1y}) \quad \ldots(4.7)$$

$$e=\left[\frac{n_1'}{nn_2''}-\frac{1}{n_2'}\phi_x R_{2xy}\left(R_{2xy}-\frac{n_2'^2}{n^2}R_{12x}R_{12xy}\right)\right]$$

$$\left[\frac{n_1n_1'}{nn_1''n_2''}\left(1-\frac{n_2''^2}{n_1^2}R^2_{12y}\right)-\frac{n}{n_1'n_2'}\phi_x\left(R_{2xy}-\frac{n_2'^2}{n^2}R_{12x}R_{12xy}\right)^2\right]^{-1} \quad \ldots(4.8)$$

$$g=\frac{\sigma_{2y}}{n_2''.\sigma_{2x}}\left[\frac{n_1n_1'}{nn''_1}R_{2xy}\left(1-\frac{n_2''^2}{n_1^2}R^2_{12y}\right)-\left(R_{2xy}-\frac{n_2'^2}{n^2}R_{12x}R_{12xy}\right)\right]$$

$$\left[\frac{n_1n_1'}{nn_1''n_2''}\left(1-\frac{n_2'^2}{n_1^2}R^2_{12y}\right)-\frac{n}{n'_1n_2'}\phi_x\left(R_{2xy}-\frac{n_2'^2}{n^2}R_{12x}R_{12xy}\right)^2\right]^{-1} \quad \ldots(4.9)$$

The estimator (4.3) takes the form

$$\bar{y}_2=e\,[\bar{y}_{2n_1''}+(R_{12y}\sigma_{2y}/\sigma_{1y})(\bar{y}_{1n_1}-\bar{y}_{1n_1''})]+(1-e)$$

$$\bar{y}_{2n''_2}+g\,(\hat{\bar{X}}_2+\bar{x}_{2n_2'}) \quad \ldots(4.10)$$

and

$$V(\bar{y}_2)=\text{Cov}(\bar{y}_2,\bar{y}_{2n''_2})$$

$$=(1-e)\,\sigma^2_{2y}/n''_2-g\,\phi_x R_{2xy}\sigma_{2x}\sigma_{2y}/n'_2 \quad \ldots(4.11)$$

When x and y are observed on the same set of units on both the occasions *i.e.*

when

$$n_1=n, \qquad n''_1=n'_1$$

and

$$n''_2=n'_2$$

and if further it is assumed that

$$\sigma^2_{tzb} = \sigma^2_{zb} \;,\; \sigma^2_{tzw} = \sigma^2_{zw}$$

such that

$$\sigma^2_{tz} = \sigma^2_{zb} + \sigma^2_{zw} = \sigma^2_z$$

for all

$$t=1, 2 ; \qquad z=y, x$$

and

$$\sigma^2_{tt'zu} = R_z \sigma^2_z \qquad \ldots(4.12)$$

for

$$t \neq t'=1, 2 ; \qquad 2=u=y, x$$

$$= R_{zu} \sigma^2_{zu}$$

for

$$z \neq u=x, y$$

then a, e and g simplify to the following form :

$$a=-en'_2 R_y/n \qquad \ldots(4.13)$$

$$e=\left[\frac{n_1'}{n} - \phi_x R^2_{xy}\left(1-\frac{n'^2_2}{n^2} R_x\right)\right]$$

$$\left[1-\frac{n'^2_2}{n^2} R^2_y - \frac{n}{n_1'} \phi_x R^2_{xy}\left(1-\frac{n'^2_2}{n^2} R_x\right)^2\right]^{-1} \ldots(4.14)$$

$$g=\frac{\sigma_y}{\sigma_x}\left[\frac{n'^2_2}{n^2} R_{xy}\left(R_x-R^2_y\right)\right]$$

$$\left[1-\frac{n'^2_2}{n^2} R^2_y - \frac{n}{n'_1} \phi_x R^2_{xy}\left(1-\frac{n'^2_2}{n^2} R_x\right)^2\right]^{-1} \ldots(4.15)$$

where

$$\phi_x=nn'_1\left(n_2-n'^2_2 R^2_x\right)^{-1}$$

The estimator $\bar{y}_2$ becomes

$$\bar{y}_2=e\left[\bar{y}_{2n'_1}+R_y\left(\bar{y}_{1n_1}-\bar{y}_{1n'_1}\right)\right]$$

$$+(1-e)\,\bar{y}_{2n'_2}+g\left(\hat{\bar{X}}_2-\bar{x}_{2n_2'}\right) \qquad \ldots(4.16)$$

and

$$V(\bar{y}_2)=(1-e-g'\,\phi_x R_{xy})\,\sigma^2_y\Big/n'_2 \qquad \ldots(4.17)$$

where

$$g' = g\sigma_x/\sigma_y.$$

Remark 1 : When $\bar{X}_2$ is known, we get

$$e = \frac{n_1'}{n}\left(1 - R^2_{xy}\right)\left(1 - R^2_{xy} - \frac{n'^2_2}{n^2} R^2_y\right)^{-1}$$

and

$$g = -\frac{n'^2_2}{n^2} R_{xy} R^2_y \left(1 - R^2_{xy} - \frac{n'^2_2}{n^2} R^2_y\right)^{-1}.$$

Writing

$$\mu = n'_2/n, \qquad \lambda = n'_1/n$$

such that

$$\mu + \lambda = 1,$$

the optimum proportion of units to be replaced from first occasion to the second is given by

$$\mu_0 = \left(R^2_y\right)^{-1}\left[\left(1 - R^2_{xy}\right) - \sqrt{\left(1 - R^2_{xy}\right)\left(1 - R^2_{xy} - R^2_y\right)}\right] \qquad \ldots(4.18)$$

and the minimum variance is

$$V_{min}(\bar{y}_2) = \left(1 - R^2_{xy} - \mu_o R^2_y\right)\left(1 - R^2_{xy} - \mu^2_o R^2_y\right)^{-1} \sigma^2_y / n$$

$$+ \lambda_o \mu_o R^2_{xy} R^2_y \left(1 - R^2_{xy} - \mu^2_o R^2_y\right)^{-1} \sigma_x \sigma_y / n \qquad \ldots(4.19)$$

If there had been no auxiliary information, $R_{xy} = 0$ and the results of the univariate case follow. It may be verified that with auxiliary information, the optimum replacement fraction and the precision of the estimator are higher than the corresponding values if there were no auxiliary information.

Case II : n''_2 units for y drawn independently from the population.

In this case

$$\text{Cov}\,(\bar{y}_{2n_2''}, \bar{x}_{2n_2'}) = 0$$

and consequently

$$\text{Cov}\,(\bar{y}_{2n_2''}, \hat{\bar{X}}_2) = 0.$$

The values of e and g (herein called e_1 and g_1) become

$$e_1 = \left[\frac{n_1}{n''_1}\left(1 - \frac{n''^2_2}{n_1^2} R^2_{12_y}\right) \frac{n^{2''} n_2'}{n'^2_1} \phi_a \left(R_{2.xy} - \frac{n_2'}{n} R_{12xy} R_{12x}\right)^2\right]^{-1} \qquad \ldots(4.20)$$

and

$$g_1 = -\frac{n_2'}{n_1'} \frac{\sigma_{2y}}{\sigma_{2x}} \left(R_{2xy} - \frac{n'_2}{n} R_{12x} R_{12xy} \right)$$

$$\left[\frac{n_1}{n''_1} \left(1 - \frac{n''^2_2}{n_1^2} R^2_{12y} \right) - \frac{n''_2 \, n'_2}{n'^2_1} \phi_x \right.$$

$$\left. \left(R_{2xy} - \frac{n'_2}{n} R_{12x} R_{12xy} \right)^2 \right]^{-1} \quad \text{.. (4.21)}$$

It may be observed from (4.21) that so long as x and y are positively correlated, g_1 will generally be negative.

Assuming equality of sample sizes for x and y and (4.12) we have

$$e_1 = \left[\frac{n}{n'_1} \left(1 - \frac{n''^2_2}{n^2} R^2_y \right) - \frac{n''^2_2}{n'^2_1} \phi_x R^2_{xy} \right.$$

$$\left. \left(1 - \frac{n'_2}{n} R_x \right)^2 \right]^{-1} \quad \text{...(4.22)}$$

$$g_1 = -\frac{n'_2}{n'_1} \frac{\sigma_y}{\sigma_x} R_{xy} \left(1 - \frac{n'_2}{n} R_x \right) \left[\frac{n}{n'_1} \left(1 - \frac{n'^2_2}{n^2} R^2_y \right) \right.$$

$$\left. - \frac{n'^2_2}{n'^2_1} \phi_x R^2_{xy} \left(1 - \frac{n'_2}{n} R_x \right)^2 \right]^{-1} \quad \text{...(4.23)}$$

and

$$V(\bar{y}_2) = (1 - e_1) \, \sigma^2_y \Big/ n_2'. \quad \text{...(4.24)}$$

Remark 2 : When $\bar{X}_2$ is known and writing $\mu = n_2'/n$, $\lambda = n_1'/n$ we get

$$e_1 = \lambda \left[1 - \mu^2 \left(R^2_{xy} + R^2_y \right) \right]^{-1}$$

and

$$g_1 = -\mu \left(R_{xv} \frac{\sigma_y}{\sigma_v} \right) \left[1 - \mu^2 \left(R^2_{xy} + R^2_y \right) \right]^{-1}$$

$$\mu_{opt} = \left[1 - \sqrt{1 - \left(R^2_{xy} + R^2_y \right)} \right] \left[R^2_{ay} + R^2_y \right]^{-1}$$

and

$$V_{opt}(\bar{y}_2) = \left[1 + \sqrt{1 - \left(R^2_{xy} + R^2_y \right)} \right] \sigma^2_y \Big/ 2n.$$

For the univariate case and for single stage sampling when Pi's are all equal R_{xy} and R_y become the population correlation coefficients ρ_{xy} and ρ_y and therefore

$$\mu_{opt} = \left[1 - \sqrt{1 - \left(\rho^2_{xy} + \rho^2_y \right)} \right] \left[\rho^2_{xy} + \rho^2_y \right]^{-1} \quad \text{.. (4.25)}$$

and

$$V_{opt}(\bar{y}_2)=\left[1+\sqrt{1-\left(\rho_{xy}^2+\rho_y^2\right)}\right]\sigma_y^2\Big/2n \qquad \ldots(4.26)$$

This corresponds to the well known case of sampling on two occasions except that ρ_y^2 is replaced by $\left(\rho_{xy}^2+\rho_y^2\right)$. Therefore higher correlations between X and Y will result in greater precision of the estimator.

5. Double Sample Estimator of y on Both Occasions

In Section 4 we built the estimator y_2 without giving regard to the fact that the n_1 units for y on the first occasion were taken as a sub-sample of the n units for x on that occasion, so that any correlation between x and y on the first occasion was not taken care of. In this section, we first propose to build a double sample estimator of y on the first occasion and then a double sample estimator for y on the second occasion by taking into account the matched and unmatched units for x and y on both the occasions.

An improved estimator of $\bar{Y}_1$, the population mean on the first occasion may be written as

$$\hat{\bar{y}}_1=\bar{y}_{1n_1}+b'\left(\bar{x}_{1n}-\bar{x}_{1n_1}\right) \qquad \ldots(5.1)$$

where y_{1n_1}, $\bar{x}_{1n_1}$ are the sample means based on n_1 psu's on the first occasion on which both x and y were observed and $\bar{x}_{1n}$ is the sample mean based on n psu's on which x alone was observed and b' is an unknown quantity. It may be seen that $\hat{\bar{y}}_1$ is a biased estimator of $\bar{Y}_1$, the bias in $\hat{\bar{y}}_1$ being given by [Cov $(b', \bar{x}_{1n})$ − Cov $(b', \bar{x}_{1n_1})$] which will be neglegible if n is sufficiently large. The value of b' obtained by minimising the variance of $\hat{\bar{y}}_1$ is given by

$$\hat{b}'=[\text{Cov}(\bar{y}_{1n},\bar{x}_{1n_1})-\text{Cov}(\bar{y}_{1n_1},\bar{x}_{1n})][V(\bar{x}_{1n_1})-V(\bar{x}_{1n})]^{-1}$$

$$=R_{1xy}\,\sigma_{1y}/\sigma_{1x}$$

and

$$V(\hat{\bar{y}}_1)=\frac{1}{n_1}\,\sigma_{1y}^2-\left(\frac{1}{n_1}-\frac{1}{n}\right)R_{1xy}^2\,\sigma_{1y}^2 \qquad \ldots(5.2)$$

Based on matched units for y, an estimator of $\bar{Y}_2$ may be written as

$$\bar{z}_2=\bar{y}_{2n''_1}+B\left(\hat{\bar{y}}_1-\bar{y}_{1n''_1}\right) \qquad \ldots(5.3)$$

Assuming B to be an unknown constant, its value obtained by minimising $V(\bar{z}_2)$ is given by.

$$\hat{B}=\left[\left(\frac{1}{n''_1}-\frac{1}{n_1}\right) R_{12y}+\left(\frac{1}{n_1}-\frac{1}{n}\right) R_{1xy}\ R_{12xy}\right]$$

$$\left[\left(\frac{1}{n''_1}-\frac{1}{n_1}\right)+\left(\frac{1}{n_1}-\frac{1}{n}\right) R^2_{1xy}\right]^{-1} \sigma_{2y}/\sigma_{1y} \qquad \ldots(5.4)$$

By taking into account the correlation of y with x variate, an overall estimator of $\bar{Y}_2$ may be wtitten as

$$y_{2d}=\alpha_1\bar{z}_2+(1-\alpha_1)\ y_{2n_2''}+\beta_1\left(\hat{\bar{X}}_2-\bar{x}_{2n_2'}\right) \qquad \ldots(5.5)$$

We shall assume that the n''_2 units for y on the second occasion are drawn independently from the population, corresponding to case (II) of section 4. The values of α_1 and β_1 obtained by minimising $V(\bar{y}_{2d})$ are given by

$$\hat{\alpha}_1=\frac{1}{n''_2}\ \sigma^2_{2y}\left[\frac{n_1}{n''_1\ n''_2}\ \sigma^2_{2y}-\hat{\beta}^2\left\{\left(\frac{1}{n''_1}-\frac{1}{n_1}\right)\right.\right.$$

$$+\left(\frac{1}{n_1}-\frac{1}{n}\right) R^2_{1xy}\Big\}\ \sigma^2_{1y}-n''_2\ \phi_x\left\{\frac{1}{n'_1}\ R_{2xy}\ \sigma_{2y}\right.$$

$$-\left(\frac{1}{n_1}-\frac{1}{n}\right)\hat{\beta}\ R_{1xy}\ R_{12x}\ \sigma_{1y}-\left(\frac{1}{n'_1}-\frac{1}{n}\right)$$

$$\left.\left. R_{12x}\ R_{12xy}\ \sigma_{2y}\right\}^2\right]^{-1} \qquad \ldots(5.6)$$

$$\hat{\beta}_1=-\frac{n''_2}{\sigma_{2x}}\ \sigma\left[\frac{1}{n'_1}\ R_{2xy}\ \sigma_{2y}-\left(\frac{1}{n_1}-\frac{1}{n}\right)\right.$$

$$\left.\hat{\beta}\ R_{1xy}R_{12y}\ R_{12x}\ \sigma_{1y}-\left(\frac{1}{n_1'}-\frac{1}{n}\right) R_{12x}\ R_{12xy}\ \sigma_{2y}\right] \ldots(5.7)$$

Under the assumption of (4.12), the values of $\hat{\alpha}_1$, $\hat{\beta}_1$ simplify to

$$\hat{\alpha}_1=\frac{1}{n''_2}\left[\frac{n_1}{n''_1\ n''_2}-\hat{\beta}^2\left\{\left(\frac{1}{n''_1}-\frac{1}{n_1}\right)+\left(\frac{1}{n_1}-\frac{1}{n}\right)\right.\right.$$

$$R^2_{xy}-n''_2\ \phi_x\ R^2_{xy}\left\{\frac{1}{n'_1}-\left(\frac{1}{n_1}-\frac{1}{n}\right)\hat{\beta}\ R_x\right.$$

$$\left.\left.-\left(\frac{1}{n'_1}-\frac{1}{n}\right) R_x\right\}^2\right]^{-1} \qquad \ldots(5.8)$$

$$\hat{\beta}_1=-n''_2\ \alpha\ (R_{xy}\ \sigma_y/\sigma_x)\left[\frac{1}{n'_1}-\left(\frac{1}{n_1}-\frac{1}{n}\right)\hat{\beta}\ R_x\right.$$

$$\left.-\left(\frac{1}{n'_1}-\frac{1}{n}\right) R_x\right] \qquad \ldots(5.9)$$

The estimator $\bar{y}_{2d}$ takes the form

$$\bar{y}_{2d}=\hat{\alpha}_1\left[\bar{z}_2-n''_2\,(R_{xy}\,\sigma_y/\sigma_x)\left\{\frac{1}{n'_1}-\left(\frac{1}{n_1}-\frac{1}{n}\right)\hat{\beta}\,R_x\right.\right.$$

$$\left.\left.-\left(\frac{1}{n'_1}-\frac{1}{n}\right)R_x\right\}\right]\left(\hat{X}_2-\bar{x}_{2n'_2}\right)+\left(1-\hat{\alpha}_1\right)y_{2n''_2} \quad \ldots(5.10)$$

and its variance is given by

$$V(\bar{y}_{2d})=(1-\hat{\alpha}_1)\,\sigma^2_{2y}\Big/n''_2 \quad \ldots(5.11)$$

We work out the relative efficiency of $\bar{y}_{2d}$ in relation to $\bar{y}_2$ for different values of n''_2/n_1, n'_2/n, R_x, R_y and R_{xy} and is given in Table 1. In order not to involve the ratios of the type n''_2/n'_1 we degress from the assumption of equality of sample sizes for x and y made earlier and only assume that $n'_1=n_1$. As would be seen from Table 1 that $\bar{y}_{2d}$ is more efficient than $\bar{y}_2$ for different values of the parameters involved. The relative efficiency increases rapidly as R_y increases and with increase in R_{xy} for higher values of R_y.

Remark 3 :

When $\bar{X}_1$ and $\bar{X}_2$ are known, we get $\bar{y}_{2d}$, μ_{opt} and $V_{min}\,(\bar{y}_{2d})$ equal to corresponding values of section 4, case II.

6. Ratio-type Composite Estimators of Means of x and y Variates

We shall first obtain a ratio-type composite estimator for $\bar{X}_2$ and utilise this estimator to obtain a ratio-type composite estimator for $\bar{Y}_2$. We assume that n''_2 units for y on second occasion are drawn independently from the population, corresponding to case II of section 4.

A ratio-type estimator of $\bar{X}_2$ may be written as

$$\bar{X}_{2R}=P\bar{Z}^*+(1-P)\,\bar{x}_{2n'_2}$$

where

$$\bar{Z}^*=(\bar{x}_{2n'_1}/\bar{x}_{1n'_1})\,\bar{x}_{1n}$$

It may easily be seen that $\bar{X}_{2R}$ is a biased estimator of $\bar{X}_2$, its relative bias being

$$RB\,(\bar{X}_{2R})=P\,\frac{n'_2}{n'_1\,n}\left(\sigma^2_{1x}\Big/\bar{X}^2_1-R_{12x}\,\sigma_{1x}\sigma_{2x}/\bar{X}_1\bar{X}_2\right)$$

where P obtained by minimising $MSE\,(\bar{X}_{2R})$ is given by

$$P=\frac{n'_1}{n}\,\sigma^2_{2x}\left[\sigma^2_{2x}+\frac{n'^2_2}{n^2}\,U^2_x\,\sigma^2_{1x}-\frac{2n'^2_2}{n^2}\,U^x R_{12x}\,\sigma_{1x}\,\sigma_{2x}\right]^{-1}$$

$$U_x=\bar{X}_2/\bar{X}_1$$

TABLE 1

Relative efficiency of the estimator y_{2d} w.r.t. the estimator y_2 for different values of R_x, R_{xy}, R_y, n'_2/n and n'_2/n_1

R_x	R_{xy} \ R_y	$n'_2/n=50$								$n'_2/n=.75$							
		$n_2''/n_1=.50$				$n_2''/n_1=.75$				$n_2''/n_1=.50$				$n_2''/n_1=.75$			
		.50	.70	.90	.95	.50	.70	.90	.95	.50	.70	.90	.95	.50	.70	.90	.95
	.50	1.02	1.04	1.05	1.06	1.01	1.02	1.04	1.05	1.02	1.04	1.07	1.08	1.01	1.02	1.04	1.06
	.70	1.04	1.07	1.11	1.12	1.01	1.03	1.07	1.09	1.06	1.10	1.17	1.20	1.01	1.03	1.08	1.12
.50	.90	1.07	1.12	1.21	1.21	1.02	1.04	1.12	1.18	1.12	1.21	1.40	1.50	1.03	1.06	1.18	1.29
	.95	1.08	1.14	1.25	1.30	1.02	1.05	1.14	1.21	1.14	1.25	1.51	1.66	1.03	1.07	1.22	1.36
	.50	1.02	1.03	1.05	1.01	1.01	1.01	1.03	1.04	1.03	1.05	1.08	1.04	1.01	1.02	1.04	1.06
	.70	1.04	1.06	1.00	1.12	1.01	1.02	1.06	1.08	1.07	1.12	1.20	1.23	1.02	1.04	1.10	1.14
.70	.90	1.08	1.12	1.20	1.24	1.02	1.04	1.11	1.16	1.15	1.25	1.47	1.58	1.03	1.07	1.21	1.33
	.95	1.09	1.14	1.24	1.29	1.02	1.05	1.13	1.19	1.18	1.18	1.06	1.76	1.04	1.08	1.25	1.42
	.50	1.02	1.03	1.05	1.05	1.01	1.01	1.03	1.04	1.04	1.06	1.10	1.12	1.01	1.02	1 06	1.08
	.70	1.04	1.07	1.10	1.12	1.01	1.02	1.06	1.08	1.08	1.14	1.24	1.29	1.02	1.04	1.12	1.18
.90	.90	1.08	1.12	1.21	1.25	1.02	1.04	1.11	1.16	1.18	1.30	1.58	1.73	1.04	1.08	1.26	1·42
	.95	1.09	1,15	1.25	1.30	1 02	1.04	1.12	1.19	1.21	1.36	1.74	1.26	1.04	1.10	1.31	1.53
	.50	1.02	1.03	1.05	1.05	1.01	1.01	1.03	1.04	1.04	1.06	1.11	1.13	1.01	1.02	1.06	1.09
	.70	1.04	1.07	1.11	1.12	1.01	1.02	1.06	1.08	1.09	1.15	1.26	1.31	1.02	1.05	1.14	1.20
.95	.90	1.08	1.13	1.22	1.26	1.02	1.04	1.11	1.16	1.19	1.31	1.63	1.80	1.04	1.09	1.28	1.46
	.95	1.09	1.15	1.26	1.31	1.02	1.05	1.13	1.19	1.22	1.38	1.80	2.05	1.05	1.10	1.34	1.58

and

$$MSE\ (\bar{X}_{2R})=(1-P)\ \sigma^2_{2x}\Big/n_2' \quad ..\ (6.3)$$

Consider now the Y-variate. A ratio estimator of $\bar{Y}_2$ based on matched units is given by

$$\bar{y}'_{2R}=y_{2n1}''/y''_{1n_1})\ y_{1n_1}$$

Also since we have an estimator $\bar{X}_{2R}$ for $\bar{X}_2$, we get another estimator as

$$\bar{y}''_{2R}=(y''_{2n_1}/\bar{x}'_{2n_1})\ \bar{X}_{2R}.$$

Therefore an estimator combining $\bar{y}_{2R}'$ and $\bar{y}_{2R}''$ may be written as

$$\bar{y}_{2R}=w_1\bar{y}'_{2R}+w_2\bar{y}''_{2R}$$

where w_1, w_2 are weights which add to unity. Since $\bar{y}_{2R}'$ and $\bar{y}_{2R}''$ are biased, therefore $\bar{y}_{2R}$ is also biased, its relative bias being

$$\begin{aligned} RB\ (\bar{y}_{2R})=\frac{1}{n_1'}\left(C^2_{2x}-R_{2xy}\,C_{2x}\,C_{2y}\right)+P\Bigg[\frac{n'_2}{n'_1 n}\left(C^2_{1x}-R_{12xy}\,C_{1x}C_{2y}\right) \\ -\frac{1}{n'_1}\left(C^2_{2x}-R_{2xy}\,C_{2x}\,C_{2y}\right)\Bigg] \\ +w_1\Bigg[\frac{n''_2}{n''_1 n_1}\left(C^2_{1y}-R_{12y}\,C_{1y}\,C_{2y}\right) \\ -\frac{1}{n'_1}\left(C^2_{2x}-R_{2xy}\,C_{2x}C_{2y}\right)\Bigg] \\ -P\Bigg\{\frac{n_2'}{n'_1 n}\left(C^2_{1x}-R_{12xy}\,C_{1x}\,C_{2y}\right) \\ -\frac{1}{n'_1}\left(C^2_{2x}-R_{2xy}\,C_{2x}\,C_{2y}\right)\Bigg\}\Bigg] \quad ...(6.4) \end{aligned}$$

where

$$C^2_{tz}=\sigma^2_{tz}\Big/\bar{Z}^2_t$$

for

$$t=1,2\ ;\qquad Z=x,y\ ;\qquad Z=X,Y.$$

If we write

$$V_1=MSE\ (\bar{y}'_{2R}),$$

$$V_2=MSE\ (\bar{y}''_{2R})$$

and

$$V_{12} = \text{Cov}\,(\bar{y}'_{2R},\ \bar{y}''_{2R}),$$

the values of w_1 which minimises $MSE\,(\bar{y}_{2R})$ is given by

$$w_1 = (V_2 - V_{12})\,(V_1 + V_2 - 2V_{12})^{-1}$$

and

$$MSE\,(\bar{y}_{2R}) = (1 - w_1)\,V_2 + w_1 V_{12} \qquad \text{...(6.5)}$$

We now also consider the n''_2 units drawn on the second occasion for which y alone was observed. A combined ratio-type composite estimator of $\bar{Y}_2$ may be written as

$$y_{2CR} = Q\bar{y}_{2R} + (1 - Q)\,y_{2n''_2} \qquad \text{...(6.6)}$$

Writing

$$V_3 = V\,(\bar{y}_{2n_2''})$$

and since

$$\text{Cov}\,(\bar{y}_{2R},\ \bar{y}_{2n_2''}) = 0,$$

we get

$$Q = V_3\,[V_3 + (1 - w_1)\,V_2 + w_1\,V_{12}]^{-1}$$

and

$$MSE\,(y_{2CR}) = (1 - Q)\,\sigma^2_{2y}\big/n''_2 \qquad \text{...(6.7)}$$

Remark 4 :

When $\bar{X}_2$ is known and assuming that

$$n_1 = n, \qquad n_1'' = n'_1, \qquad n''_2 = n'_2$$

and (4.12) and writing

$$C_{tz} = C_z$$

for

$$t = 1, 2\ ; \qquad z = x, y$$

we get

$$RB\,(\bar{y}_{2R}) = \frac{1}{n'_1}(1 - w_1)\left(C^2_x - R_{xy}\,C_x\,C_y\right) + \frac{n_2'}{n_1'\,n}\,w_1\,C^2_y\,(1 - R_y)$$

$$V_1 = \frac{1}{n'_1}\,\bar{y}^2_2\left[C^2_y + \frac{n'_2}{n}\,(1 - 2R_y)\,C^2_y\right]$$

$$V_2 = \frac{1}{n'_1}\,\bar{y}^2_2\left[C^2_y + \frac{n}{n_2'}\,C^2_x - 2R_{xy}\,C_x\,C_y\right]$$

$$V_{12} = \frac{1}{n'_1}\,\bar{y}^2_2\left[C^2_y - R_{xy}\,C_x\,C_y - \frac{n'_2}{n}\,R_y\,C^2_y\right]$$

The values of w_1 and Q can be readily obtained. It may be observed that for moderately large values of R_y and R_{xy}, the relative bias in

$\bar{y}_{2R}$ and consequently in $\bar{y}_{2CR}$ may not be very large and therefore y_{2CR} can provide a reasonably good estimate of $\bar{Y}_2$. The mean square error of $\bar{y}_{2CR}$ can be readily obtained from (6.7). The performance of $\bar{y}_{2CR}$ in relation with the estimators obtained in previous sections is possible only with the help of actual data.

SUMMARY

Use of auxiliary information has been considered in sampling on successive occasions using a two-stage sampling design. When population value of the auxiliary variate is not known, it is first estimated by drawing a larger sample of units from the population and the information so obtained is used to build estimators of the main character under study. Three different estimators called (*i*) a linear unbiased estimator, (*ii*) a double sample estimator and (*iii*) a ratio-type composite estimator were built and their relative efficiencies examined.

REFERENCES

[1] Cochran, W.G. (1963) : Sampling Techniques, 2nd Ed. ; John Wiley & Sons, New York.

[2] Des Raj (1965) : On sampling on two occasions with probability proportional to size. A M S, 36, 327-330.

[3] Jessen, R.J. (1942) : Statistical investigations of a simple survey for obtaining farm facts. Iowa Agri. Expt. St. Res. Bul. No. 304, Ames., Ia,

[4] Kathuria, O.P. (1959) : Some aspects of successive sampling in multistage designs. Thesis submitted for award of ICAR Diploma (unpublished).

[5] Kathuria, O.P. (1975) : Use of multi-auxiliary information in sampling from dynamic populations—Paper read in the symposium on sampling from dynamic populations at the 29th annual conference of Ind. Soc. Agri., Jaipur.

[6] Patterson, H.D. (1950) : Sampling on successive occasions with partial replacement of units. JRSS Series B, 12, 241-55.

[7] Sen, A.R. (1972) : Successive sampling with $p(p \geqslant 1)$ auxiliary variables. A.M.S. 43, 2031-34.

[8] Shivtar Singh and Srivastava, A.K. (1973) : Use of Auxiliary information in two stage successive sampling, Jl. Ind. Soc. Agri. Stat., 25, 101-114.

[9] Singh, D. (1968) : Estimation in successive sampling using a multistage design, JASA, 63, 99-112.

[10] Singh, D. and Kathuria, O.P. (1969) : On two-stage successive sampling *Aust. Jl. Stat.* 11, 59-66.

[11] Tikkiwal, B.D. (1953) : Optimum allocation in successive sampling. *Jl. Ind. Soc. Agri. Stat.*, 5, 100-102,

[12] Tikkiwal, B.D. (1956) : A further contribution to the theory of univariate sampling on successive occasions. *Jl. Ind. Soc. Agri. Stat.*, **8**, 84-90.

[13] Tikkiwal, B.D. (1964) : A note on two-stage sampling on successive occasions, *Sankhya,* Series A, **26**, 97-100.

[14] Tikkiwal, B.D. (1967) : Theory of multi-phase sampling from a finite or an infinite population in successive occasions. *Rev. Int. Stat. Instt.*, **35**, 247-263.

[15] Yates, F. (1960) : Sampling methods for censuses and surveys, Charles Griffin and Co. London.