MULTI-STAGE SAMPLING WITH PARTIAL REPLACEMENT OF THE SAMPLE ON SUCCESSIVE OCCASIONS

R. P. Chakrabarty and D. S. Rana, University of Georgia

Summary

In this paper, we develop the theory of sampling on successive occasions when the sampling design is multi-stage. We consider four different sampling plans, with different combinations of replacement of primary and secondary sampling units for the case of the two-stage sampling on two occasions. The best linear unbiased estimates of the population mean on the current occasion, the change from one occasion to another and the overall mean over two occasions are given. The relative efficiencies of the four sampling plans are also given. It is shown that the partial replacement of the primary and secondary sampling units is generally efficient for the estimation of mean on the current occasion.

1. INTRODUCTION

In the study of any dynamic population that changes with time, it is necessary to sample the population at different time points to obtain reliable estimates of the population parameters. Generally speaking, we may be interested in estimating three different parameters: (1) The population mean on the most recent occasion, (2) the change in the mean from one occasion to the next, (3) The mean over all occasions in a given period of time. For the special case of the simple random sampling the theory of successive sampling has been developed by Jessen (1942), Yates (1960), Patterson (1950), Eckler (1955), and Rao and Graham (1964). Recently, Singh (1968) and Singh and Kathuria (1969) have obtained some results on the successive sampling for a two-stage sampling design. In this paper, we develop the theory of successive sampling in the case of two-stage sampling. The extension of the theory of successive sampling from the single stage to the two-stage sampling is of considerable practical importance since usually a survey design is two-stage or multi-stage. We consider four different sampling procedures for the case of two-stage sampling on two occasions only. In all sampling procedures, we restrict ourselves to the following two-stage design for simplicity:

N=the number of primary sampling units (psu's) in the population

M=the number of the secondary sampling units (ssu's) in each psu in the population

n=the number of psu's in the sample

m=the number of ssu's in each psu in the sample.

The psu's and ssu's are selected by simple random sampling without replacement on each occasion and n<<N&m<<M so that the finite population correction factors in the variance formulas can be ignored. In deriving the variances of the estimates we shall also ignore the covariance terms which are of order 1/N. The sample size on each occasion is nm. Some notations are introduced here that will be used in the subsequent sections.

$y_{ik\ell}$=the value of y for the ℓ-th secondary in the k-th primary on the i-th occasion (i=1,2).

$$\overline{Y}_{i..} = \sum_{k=1}^{N} \sum_{\ell=1}^{M} y_{ik\ell}/NM$$

=the population mean of y on the ith occasion (i=1,2).

$$S_{bi}^2 = \sum_{k=1}^{N} (\overline{Y}_{ik.} - \overline{Y}_{i..})^2/(N-1)$$

=mean square error between psu means in the population on the i-th occasion (1=1,2)

$$S_{wi}^2 = \frac{\sum_{k=1}^{N} \sum_{\ell=1}^{M} (y_{ik\ell} - \overline{Y}_{ik.})^2}{N(m-1)}$$

=mean square error between ssu's within psu's in the population on the i-th occasion (i= 1,2)

$$\rho_{bij} S_{bi} S_{bj} = \frac{1}{(N-1)} \sum_{k=1}^{N} (\overline{Y}_{ik.} - \overline{Y}_{i..})(\overline{Y}_{jk.} - \overline{Y}_{j..})$$

=true covariance between the psu means on the i-th and j-th occasions (i≠j=1,2).

$$\rho_{wij} S_{wi} S_{wj} = \frac{1}{N(M-1)} \sum_{k=1}^{N} \sum_{\ell=1}^{M} (y_{ik\ell} - \overline{Y}_{ik.})(y_{jk\ell} - \overline{Y}_{jk.})$$

=true covariance between ssu's within psu's on i-th and j-th occasions (i≠j=1,2).

The four sampling procedures that we consider for the two-stage sampling design are as follows:

Procedure I: Retain all primary sampling units (psu's) and from each of these psu's make a fresh selection of secondary sampling units (ssu's) on the current occasion.

The pattern of sampling may be illustrated as follows:

Sample fraction of Secondaries	nm	nm	
First Occasion	xxxx $\overline{y}^*_{1..I}$		psu's same but ssu's different
Second Occasion		xxxx $\overline{y}^*_{2..I}$	

where $\overline{y}^*_{i..I}$=mean per secondary on the i-th occasion based on nm ssu's that are present only on the i-th occasion (i=1,2)

Procedure II: Retain only a fraction p of the psu's with their samples of ssu's from the previous occasion and make a new selection of fraction q of the psu's on the current occasion, where p+q=1.

Sampling Pattern

Sample fraction of psu's	np	nq	nq
First Occasion	xxx $\overline{y}'_{1..II}$	xxxx $\overline{y}''_{1..II}$	

Second Occasion	xxx $\bar{y}'_{1..II}$	xxxx $\bar{y}''_{2..II}$
	psu's & ssu's same	psu's (& ssu's) different

where

$\bar{y}'_{i..II}$ = mean per secondary on the i-th occasion (i=1,2) based on npm units common to both occasions

$\bar{y}''_{i..II}$ = mean per secondary on the i-th occasion based on nqm units that are present on the i-th occasion only (i=1,2).

Procedure III: Retain all the psu's from the preceeding occasion, but retain only a fraction p of ssu's within each of the psu's retained and make a fresh selection of the remaining fraction q of ssu's on the current occasion, (p+q=1).

Sample fraction of ssu's	npm	nqm	nqm
First Occasion	xxx $\bar{y}'_{1..III}$	xxxx $\bar{y}^*_{1..III}$	
Second Occasion	xxx $\bar{y}'_{2..III}$		xxxx $\bar{y}^*_{2..III}$
	psu's & ssu's same	psu's same but ssu's different	

where

$\bar{y}'_{i..III}$ = mean per secondary on the i-th occasion (i=1,2) based on npm units present on both occasions.

$\bar{y}^*_{i..III}$ = mean per secondary on the i-th occasion for nqm units that are present only on the i-th occasion (i=1,2).

Procedure IV: Retain a fraction p of psu's from the previous occasion and in each of these psu's retain only a fraction r of the ssu's and make a fresh selection of fraction s of ssu's (r+s=1). Also select the remaining fraction q of the psu's on the current occasion.

Sampling Pattern

	Primaries np		nq
Secondaries	nprm	npsm	nqm
First occasion	xxx $\bar{y}'_{1..IV}$	xxxx $\bar{y}^*_{1..IV}$	xxxxx $\bar{y}''_{1..IV}$
Second Occasion	xxx $\bar{y}'_{2..IV}$	xxxx $\bar{y}^*_{2..IV}$	xxxxx $\bar{y}''_{2..IV}$
	psu's,ssu's same	psu's same ssu's diff.	both psu's and ssu's different

where

$\bar{y}'_{i..IV}$ = mean per secondary on the i-th occasion (i=1,2) based on nprm units that are present on both occasions.

$\bar{y}^*_{i..IV}$ = mean per secondary on the i-th occasion (i=1,2) based on npsm units that are present only on the i-th occasion (i=1,2).

$\bar{y}''_{i..IV}$ = mean per secondary on the i-th occasion based on the nqm units that are present only on the i-th occasion (i=1,2).

We obtain the best linear unbiased estimates of the change and over all mean under procedures I, II & III. The best linear unbiased estimate of the population mean on the current occasion is given under procedure IV also. The relative efficiencies of suggested sampling procedures are given.

2. ESTIMATES OF THE CHANGE

2.1 Procedure I: A linear unbiased estimate of the change in the population mean $\bar{Y}_{2..}-\bar{Y}_{1..}$ is given by

$$\Delta_I = \bar{y}^*_{2..I} - \bar{y}^*_{1..I} \tag{2.1}$$

and its variance is

$$Var(\Delta_I) = \frac{S^2_{b1}}{n} + \frac{S^2_{w1}}{nm} + \frac{S^2_{b2}}{n} + \frac{S^2_{w2}}{nm} - 2\rho_{b12}\frac{S_{b1}S_{b2}}{n}$$

If the variance components on two occasions remain same, that is $S^2_{b1} = S^2_{b2} = S^2_b$; $S^2_{w1} = S^2_{w2} = S^2_w$ and $\rho_{b12} = \rho_b$ then the variance of Δ_I is further simplified to

$$Var(\Delta_I) = \frac{2}{n}\left[(1-\rho_b)S^2_b + \frac{S^2_w}{m}\right] \tag{2.2}$$

If independent samples are drawn on each occasion, the variance of Δ_I is

$$Var(\Delta_1) = \frac{2}{n}\left(S^2_b + \frac{S^2_w}{m}\right) \tag{2.3}$$

If the same sample is repeated on each occasion, the variance of Δ_I is

$$Var(\Delta_I) = \frac{2}{n}\left[(1-\rho_b)\, S^2_b + (1-\rho_w)\, \frac{S^2_w}{m}\right] \tag{2.4}$$

A comparison of (2.2),(2.3)&(2.4) shows that for $\rho_b > 0, \rho_w > 0$, the most precise estimate of change is obtained by observing the sample on each occasion. The partial replacement policy is, however, superior to the policy of complete replacement.

2.2. Procedure II: A general linear estimate of the change $\bar{Y}_{2..} - \bar{Y}_{1..}$ may be written in the following form:

$$\Delta_{II} = a\bar{y}''_{1..II} + b\bar{y}'_{1..II} + c\bar{y}'_{2..II} + d\bar{y}''_{2..II} \tag{2.5}$$

Now $E(\Delta_{II}) = (a+b)\,\bar{Y}_{1..} + (c+d)\,\bar{Y}_{2..}$

If Δ_{II} is to be an unbiased estimate of the change $\bar{Y}_{2..} - \bar{Y}_{1..}$, then a+b = -1 and c+d = 1, and

$$\Delta_{II} = a\bar{y}''_{1..II} - (1+a)\bar{y}'_{1..II} + c\bar{y}'_{2..II} + (1-c)\bar{y}''_{2..II} \tag{2.6}$$

Consequently, the variance of Δ_{II} is

$$Var(\Delta_{II}) = a^2\frac{\alpha_1}{nq} + (1+a)^2\frac{\alpha_1}{np} + c^2\frac{\alpha_2}{np} + (1-c)^2\frac{\alpha_2}{nq} - 2c(1+a)\frac{\beta_{12}}{np} \tag{2.7}$$

where

$$S^2_{bi} + \frac{S^2_{wi}}{m} = \alpha_i, \quad i=1,2$$

$$\rho_{bij}\frac{S_{bi}S_{bj}}{} + \rho_{wij}\frac{S_{wi}S_{wj}}{m} = \beta_{ij}, \ i\neq j = 1,2$$

The values of a and c that minimize the variance of Δ_{II} are obtained by solving the following equations

$$\frac{\partial}{\partial a}\mathrm{Var}(\Delta_{II}) = 0 \text{ and } \frac{\partial}{\partial c}\mathrm{Var}(\Delta_{II}) = 0$$

as

$$a_0 = \frac{\beta_{12}pq\,\alpha_2+\beta_{12}^2q^2-\alpha_1\alpha_2q}{\alpha_1\alpha_2-\beta_{12}^2q^2}; \quad c_0 = \frac{\alpha_1p(\alpha_2+\beta_{12}q)}{\alpha_1\alpha_2-\beta_{12}^2q^2} \tag{2.8}$$

If we assume that

$$S_{b1}^2 = S_{b2}^2 = S_b^2 \text{ and } S_{w1}^2 = S_{w2}^2 = S_w^2 \quad \text{(say)}$$

$$\text{then } \alpha_1 = \alpha_2 = \alpha \text{ and } \beta_{12} = \beta \tag{2.9}$$

Now, using (2.6), through (2.9) we get

$$\Delta_{II} = \left[\frac{(\alpha-\beta)q}{\alpha-\beta q}\right](\bar{y}''_{2..II}-\bar{y}''_{1..II}) + \frac{\alpha p}{\alpha-\beta q}(\bar{y}'_{2..II}-\bar{y}'_{1..II}) \tag{2.10}$$

and

$$\mathrm{Var}(\Delta_{II}) = 2\frac{\alpha}{n}\left(\frac{\alpha-\beta}{\alpha-\beta q}\right); \ 0<q<1 \text{ and } \alpha>\beta \tag{2.11}$$

Thus, $\mathrm{Var}(\Delta_{II})$ is minimum when q=0 and is maximum when q=1, for $\beta>0$. Hence to estimate change it is best to repeat the same sample on the second occasion if $\beta>0$.

Relative Efficiency: Let REC21 denote the relative efficiency of procedure II w.r.t. procedure I for the estimation of the change. From the equations (2.2) and (2.11) it can be shown that

$$\mathrm{REC21} = \frac{(1-\rho_b+\frac{\phi}{m})[(1-\rho_bq)+(1-\rho_wq)\frac{\phi}{m}]}{(1+\frac{\phi}{m})[(1-\rho_b)+(1-\rho_w)\frac{\phi}{m}]} \tag{2.12}$$

where $\phi = S_w^2/S_b^2$.

It is interesting to note that for q=0 (complete Matching), $\rho_b>0$ and $\rho_w>0$, REC21<1, and REC21<1 if q = 0, $\rho_b<0$ and $\rho_w<0$. But for 0<q<1, it is difficult to make analytic comparisons. The numerical evaluation of REC21 given by us in a technical report (1974) shows that for small values of m and q, and large values of ϕ, ρ_b and ρ_w, procedure II is generally superior to procedure I for estimating the change in the population mean.

2.3 Procedure III: Following the methods of the previous section and under the assumption (2.9) it can be shown that the best estimate of the change is given by

$$\Delta_{III} = \frac{(1-q)}{1-\rho_wq}(\bar{y}'_{2..III}-\bar{y}'_{1..III}) + \frac{q(1-\rho_w)}{1-\rho_wq}(\bar{y}^*_{2..III}-\bar{y}^*_{1..III}) \tag{2.13}$$

and its variance after some algebraic manipulation can be obtained as

$$\mathrm{Var}(\Delta_{III}) = \frac{2}{n}\left[(1-\rho_b)S_b^2+\left(\frac{1-\rho_w}{1-\rho_wq}\right)\frac{S_w^2}{m}\right] \tag{2.14}$$

Special Cases

$$\text{(i) For q=0, } \mathrm{Var}(\Delta_{III}) = \frac{2}{n}\left[(1-\rho_b)S_b^2+(1-\rho_w)\frac{S_w^2}{m}\right] \tag{2.15}$$

$$\text{(ii) For q=1, } \mathrm{Var}(\Delta_{III}) = \frac{2}{n}\left[(1-\rho_b)S_b^2+\frac{S_w^2}{m}\right] \tag{2.16}$$

It may be noted that for q=1, the procedure III becomes identical to procedure I, and the variance in (2.16) checks with that in (2.2).

Now, from (2.14), (2.15) and (2.16) it can be seen that for q=0, the variance of Δ_{III} is minimized. Therefore, the most precise estimate of change is obtained by repeating the same sample on each occasion. It follows from (2.14) and (2.16) that a policy of partial retention is superior to the policy of selecting independent samples on each occasion if $\rho_w>0$.

Relative Efficiency: (i) Denoting the relative efficiency of procedure III with respect to procedure I by REC31, it can be seen from (2.2) and (2.15) that

$$\mathrm{REC31} = \frac{1-\rho_b+\frac{\phi}{m}}{1-\rho_b+\left(\frac{1-\rho_w}{1-\rho_wq}\right)\frac{\phi}{m}} > 1, \text{ for } \rho_b,\rho_w>0. \tag{2.17}$$

Thus, procedure III is superior to procedure I for estimating change when $\rho_b, \rho_w > 0$.

(ii) let REC32 denote the relative efficiency of procedure III w.r.t. II for estimating the change. It follows from (2.11) and (2.15) that

$$\mathrm{REC32} = \frac{(1-\rho_wq)(1+\frac{\phi}{m})[1-\rho_b+(1-\rho_w)\frac{\phi}{m}]}{[(1-\rho_b)(1-\rho_wq)+(1-\rho_w)\frac{\phi}{m}][1-\rho_bq+(1-\rho_wq)\frac{\phi}{m}]} \tag{2.18}$$

It may be noted that REC32=1 for q=0 as it is expected and for q=1 REC32>1 if $\rho_b>0$. For 0<q<1, it is difficult to make analytical comparisons of the efficiencies. From the numerical evaluation of REC32, we find that the procedure III generally gives a better estimate of change than procedure II, and the gains are appreciable when q, ρ_b, ρ_w and m are large and ϕ is small.

3. ESTIMATES OF THE OVER-ALL MEAN FOR TWO OCCASIONS

3.1 Procedure I: A general linear estimate of the over-all mean on two occasions, $\bar{Y}=\frac{1}{2}(\bar{Y}_{1..}+\bar{Y}_{2..})$ is given by

$$\bar{y}_I = \theta_1\bar{y}^*_{1..I}+\theta_2\bar{y}^*_{2..I} \tag{3.1}$$

where θ_1 and θ_2 are suitable weights depending on the relative importance of the two occasions and $\theta_1+\theta_2=1$. Suppose in the study of a disease it is intended to estimate the average number of persons affected by it, then θ_1 and θ_2 may be taken in proportion to the number of affected persons on each occasion. Under the assumption (2.9) the variance of $\bar{y}_I$ is obtained as

$$Var(\bar{y}_I)=[\theta_1^2+\theta_2^2](\frac{S_b^2}{n}+\frac{S_w^2}{nm})+2\theta_1\theta_2\rho_b\frac{S_b^2}{n} \quad (3.2)$$

If $\theta_1=\theta_2=\frac{1}{2}$ then

$$Var(\bar{y}_I)=\frac{1}{2n}[(1+\rho_b)S_b^2+\frac{S_w^2}{m}] \quad (3.3)$$

If new samples are drawn on each occasion, then

$$Var(\bar{y}_I)=\frac{1}{2n}[S_b^2+\frac{S_w^2}{m}]$$

Thus, for $\rho_b>0$ procedure I does not provide an efficient estimate of the over-all mean. However, if $\rho_b<0$ procedure I is perferable to drawing different samples on each occasion.

3.2 Procedure II: One possible linear form of the estimate of the over-all population mean may be

$$\bar{y}_{II}=\theta_1[a\bar{y}'_{1..II}+(1-a)\bar{y}''_{1..II}]+\theta_2[b\bar{y}'_{2..II}+(1-b)\bar{y}''_{2..II}] \quad (3.4)$$

Where θ_1 and θ_2 are suitable weights that may be chosen according to the relative importance of the two occasions so that $\theta_1+\theta_2=1$; a and b are constants to be chosen in such a way that Var $(\bar{y}_{II})$ is minimum. From Singh (1968) who developed this procedure we have the variance of $\bar{y}_{II}$ under assumption (2.9) as

$$Var(\bar{y}_{II})=\frac{1}{2n}\frac{(S_b^2+\frac{S_w^2}{m})[(1+\rho_b)S_b^2+(1+\rho_w)\frac{S_w^2}{m}]}{(1+\rho_b q)S_b^2+(1+\rho_w q)\frac{S_w^2}{m}} \quad (3.5)$$

$Var(\bar{y}_{II})$ is minimum when q=1 and maximum when q=0 for $\rho_b>0$ and $\rho_w>0$. Hence, it follows, that to estimate the over-all mean for two occasions it is best to draw independent samples. However, if $\rho_b<0$, $\rho_w<0$, a partial replacement procedure will be more efficient.

Relative Efficiency: The relative efficiency of procedure II w.r.t. procedure I for the estimation of the over-all mean from (3.3) and (3.5) is given by

$$REOM21=\frac{(1+\rho_b+\frac{\phi}{m})[(1+\rho_b q)+(1+\rho_w q)\frac{\phi}{m}]}{(1+\frac{\phi}{m})[(1+\rho_b)+(1+\rho_w)\frac{\phi}{m}]} \quad (3.6)$$

Thus, REOM21>1 for q=1, $\rho_b>0$ and REOM21<1 for q=0 & ρ_b, $\rho_w>0$. Further, the numerical evaluation of REOM21 shows that procedure II is superior to procedure I for ρ_b, ρ_w >.5, $1\le\phi\le5$ and q> .5.

3.3 Procedure III: A general linear estimate of the over-all mean, under procedure III, may be written as

$$\bar{y}_{III}=\theta_1[a\bar{y}'_{1..III}+(1-a)\bar{y}^*_{1..III}]+\theta_2[c\bar{y}'_{2..III}+(1-c)\bar{y}^*_{2..III}] \quad (3.7)$$

where θ_1 and θ_2 are suitable weights $(\theta_1+\theta_2=1)$. It can be shown that under the assumption (2.9) the values of a and c that minimize $Var(\bar{y}_{III})$ are given by

$$a_0=\frac{p(\theta_1-\theta_2\rho_w q)}{\theta_1(1-\rho_w^2q^2)}\ ;\ c_0=\frac{p(\theta_2-\theta_1\rho_w q)}{\theta_2(1-\rho_w^2q^2)} \quad (3.8)$$

Therefore, with optimum values of a and c, $\bar{y}_{III}$ becomes

$$\bar{y}_{III}=\theta_1[\frac{p(\theta_1-\theta_2\rho_w q)}{\theta_1(1-\rho_w^2q^2)}(\bar{y}'_{1..III}-\bar{y}^*_{1..III})+\bar{y}^*_{1..III}]+\theta_2[\frac{p(\theta_2-\theta_1\rho_w q)}{\theta_2(1-\rho_w^2q^2)}(\bar{y}'_{2..III}-\bar{y}^*_{2..III})+\bar{y}^*_{2..III}] \quad (3.9)$$

and its variance is,

$$Var(\bar{y}_{III})=[(2\theta_1\theta_2\rho_w q-\theta_1^2-\theta_2^2)p+(\theta_1^2+\theta_2^2)(1-\rho_w^2q^2)]\frac{S_w^2}{nqm(1-\rho_w^2q^2)}+(\theta_1^2+\theta_2^2+2\theta_1\theta_2\rho_b)\frac{S_b^2}{n} \quad (3.10)$$

If $\theta_1=\theta_2=\frac{1}{2}$, then

$$Var(\bar{y}_{III})=\frac{1}{2n}[(1+\rho_b)S_b^2+(\frac{1+\rho_w}{1+\rho_w q})\frac{S_w^2}{m}] \quad (3.11)$$

Relative Efficiency: Note that for q=1, $Var(\bar{y}_{III})$ $=Var(\bar{y}_I)$ given in (3.3); but for $0\le q<1$ and $\rho_b>0$, $\rho_w>0$ procedure I is more efficient than procedure III. The relative efficiency of procedure II w.r.t III for the estimation of the over-all mean is given by

$$REOM23=\frac{[1+\rho_b+(\frac{1+\rho_w}{1+\rho_w q})\frac{\phi}{m}][(1+\rho_b q)+(1+\rho_w q)\frac{\phi}{m}]}{(1+\frac{\phi}{m})[1+\rho_b+(1+\rho_w)\frac{\phi}{m}]} \quad (3.12)$$

We note that REOM23=1 for q=0 and for q=1, REOM23 >1 if $\rho_b>o$ & REOM23<1 if $\rho_b<0$. For 0<q<1 the numerical evaluation of REOM23 shows that procedure II is more efficient than procedure III and the gains in efficiency are appreciable for ρ_b,ρ_w >.5 and q>.5.

We find from the above results that for the estimation of over-all mean it will be statistically more efficient to draw independent samples on each occasion when the correlation between the values of the sampling units on successive occasions is positive. However, practical considerations such as construction of frame, operational problems and costs may often weigh in favor of retaining a part of the sample from one occasion to the next. This policy of partial replacement is, of course, better if the correlation is negative.

4. ESTIMATES OF THE MEAN ON THE CURRENT OCCASION

4.1 Procedure I: A general linear estimate of $\bar{Y}_{2..}$, the population mean on the second occasion may be written as

$$\bar{y}_{2..I}=a\bar{y}^{*}_{1..I}+b\bar{y}^{*}_{2..I}$$

If $\bar{y}_{2..I}$ is to be an unbiased estimate of $\bar{Y}_{2..}$, we must have a=0 and b=1. Thus procedure I becomes a trivial case.

4.2 Procedure II: A linear unbiased estimate of $\bar{Y}_{2..}$ may be expressed as

$$\bar{y}_{2..II}=a(\bar{y}'_{1..II}-\bar{y}''_{1..II})+c\bar{y}'_{2..II}+(1-c)\bar{y}''_{2..II} \tag{4.1}$$

From Singh(1968) we have the optimum values of a and c under the assumption (2.9) as

$$a_0=pq\alpha\beta/(\alpha^2-q'^2\beta^2) \qquad c_0=p\alpha^2/(\alpha^2-q'^2\beta^2) \tag{4.2}$$

and
$$\mathrm{Var}(\bar{y}_{2..II})=\alpha(\alpha^2-q'\beta^2)/n(\alpha^2-q'^2\beta^2) \tag{4.3}$$

Where q' denotes the replacement fraction for psu's. Now $\mathrm{Var}(\bar{y}_{..II})=\alpha/n$ for q'=0 or q'=1 and for all other values of q'(0<q'<1) the partial replacement will be more efficient if $\beta\neq 0$. Singh also shows that q' should always exceed $\frac{1}{2}$.

4.3 Procedure III: The estimation of the mean on the current occasion under this procedure was investigated by Singh and Kathuria (1969). Under the assumption (2.9) they obtained the estimate and its variance as

$$\bar{y}_{2..III}=\frac{p}{1-\rho_w q'^2}[\bar{y}'_{2..III}+\rho_w q'(\bar{y}^{*}_{1..III}-\bar{y}'_{1..III})] + \frac{q'(1-\rho_w^2 q')}{(1-\rho_w^2 q'^2)}\bar{y}^{*}_{2..III} \tag{4.4}$$

and

$$\mathrm{Var}(\bar{y}_{2..III})=\frac{S_b^2}{n}+[\frac{(1-\rho_w^2 q')}{1-\rho_w^2 q'^2}]\frac{S_w^2}{nm}, \quad 0<q'<1$$

$$=\frac{S_b^2}{n}+\frac{S_w^2}{nm} \quad \text{for } q'=0,1 \tag{4.5}$$

where q' denotes replacement fraction for ssu's. From the results in (4.5) it is easily seen that it is advantageous to retain a part of the sample for estimation of the population mean on the second occasion.

4.4 Procedure IV: A general linear estimate of $\bar{Y}_{2..}$, the population mean on the second occasion may be written as

$$\bar{y}_{2..IV}=a\bar{y}'_{1..IV}+b\bar{y}^{*}_{1..IV}+e\bar{y}''_{1..IV}+c\bar{y}'_{2..IV}+d\bar{y}^{*}_{2..IV}+f\bar{y}''_{2..IV} \tag{4.6}$$

Now $E(\bar{y}_{2..IV})=(a+b+e)\bar{Y}_{1..}+(e+d+f)\bar{Y}_{2..}$

If $\bar{Y}_{1..}\neq\bar{Y}_{2..}$, and $\bar{y}_{2..IV}$ is to be an unbiased estimate of $\bar{Y}_{2..}$, we must have a+b+e=0 and c+d+f=1 . Therefore,

$$\bar{y}_{2..IV}=a\bar{y}'_{1..IV}+b\bar{y}^{*}_{1..IV}-(a+b)\bar{y}''_{1..IV}+c\bar{y}'_{2..IV}+d\bar{y}^{*}_{2..IV}+(1-c-d)\bar{y}''_{2..IV} \tag{4.7}$$

The variance of $\bar{y}_{2..IV}$, ignoring the finite population correction terms and covariance terms which are of order 1/N, is given by

$$\begin{aligned}V=\mathrm{Var}(\bar{y}_{2..IV})=&a^2(\frac{S_{b1}^2}{np}+\frac{S_{w1}^2}{nprm})+b^2(\frac{S_{w1}^2}{np}+\frac{S_{w1}^2}{npsm})\\&+(a+b)^2(\frac{S_{b1}^2}{nq}+\frac{S_{w1}^2}{nqm})+c^2(\frac{S_{b2}^2}{np}+\frac{S_{w2}^2}{nprm})\\&+d^2(\frac{S_{b2}^2}{np}+\frac{S_{w2}^2}{npsm})+(1-c-d)^2(\frac{S_{b2}^2}{nq}+\frac{S_{w2}^2}{nqm})\\&+2ab\frac{S_{b1}^2}{np}+2ac(\rho_{b12}\frac{S_{b1}S_{b2}}{np}+\rho_{w1}\frac{S_{w1}S_{w2}}{nprm})\\&+2ad\rho_{b12}\frac{S_{b1}S_{b2}}{np}+2bc\rho_{b12}\frac{S_{b1}S_{b2}}{np}\\&+2bd\rho_{b12}\frac{S_{b1}S_{b2}}{np}+2cd\frac{S_{b2}^2}{np}\end{aligned} \tag{4.8}$$

The optimum values of the weights a,b,c and d that will minimize $\mathrm{Var}(\bar{y}_{2..IV})$ are obtained by solving the following system of linear equations:

$$\frac{\partial V}{\partial a}=0,\ \frac{\partial V}{\partial b}=0,\ \frac{\partial V}{\partial c}=0 \text{ and } \frac{\partial V}{\partial d}=0.$$

Under the assumption (2.9) these lead to the following equations.

$$\begin{aligned}&a[m(1-s)+(q+(1-q)(1-s))\phi]\\&+b[m(1-s)+(1-q)(1-s)\phi]+c[\rho_b mq(1-s)+\rho_w q\phi]\\&+d[mq(1-s)\rho_b]=0\end{aligned} \tag{4.9}$$

$$a[ms+(1-q)s\phi]+b[ms+(q+(1-q)s)\phi]+cmqs\rho_b+dmqs\rho_b=0 \tag{4.10}$$

$$\begin{aligned}&a[mq(1-s)\rho_b+q\rho_w\phi]+b[mq(1-s)\rho_b]+c[m(1-s)\\&+(q+(1-q)(1-s))\phi]+d[m(1-s)+(1-q)(1-s)\phi]\\&=m(1-q)(1-s)+(1-q)(1-s]\phi\end{aligned} \tag{4.11}$$

$$\begin{aligned}&amqs\rho_b+bmqs\rho_b+c[ms+(1-q)s\phi]+d[ms+(q+(1-q)s)\phi]\\&=m(1-q)s+(1-q)s\phi\end{aligned} \tag{4.12}$$

Solving the equations (4.9) through (4.12) the optimum values of a,b,c, and d can be expressed in terms of m,q and s, and parameters ρ_b,ρ_w and $\phi=S_w^2/S_b^2$ but they are too cumbersome to present here. The numerical values of the weights for selected values of the parameters are presented in tables 1 and 2. Substitution of the optimum values of the weights a,b,c and d in (4.8) provide $Var(\bar{y}_{2..IV})$ which is too complicated for analytical investigation of the replacement policy. To compare procedure IV with procedures II and III the over-all replacement fraction should be same in each case. By equating the number of unmatched units we obtain

$$nq'm = npsm + nqm$$

$$\text{or } q' = q+s-qs \tag{4.13}$$

where q' is the replacement fraction of psu's in procedure II and that of ssu's in procedure III. The relative efficiencies, REM42 and REM43 of procedure IV w.r.t procedures II and III respectively have been evaluated numerically for selected values of m,q,s and parameters ρ_b, ρ_w and ϕ. The results are given in tables 3 and 4. It can be seen from tables 3 and 4 that (1) procedures II and IV are superior to procedure III, and (2) for $q \geq .5$ and $s > .5$, REM42>1, but the gains achieved are not high. Thus the choice between procedure II and IV for estimating mean on the current occasion may depend on other practical considerations such as cost and operational problems in a given survey situation.

We may mention that the efficiencies of the suggested sampling procedures have been illustrated using 'live data' from the Georgia Agricultural Facts (1964) for the estimation of average corn-yield per acre on the current occasion. The procedure IV was generally superior to procedures II & III but the gains achieved were modest. Further, the assumption (2.9)was found to be nearly valid in this case. These results are given in a technical report (1974).

REFERENCES

[1] Eckler, A. R. (1955). "Rotation Sampling". Ann. Math. Stat., 26, pp. 664-685.

[2] Georgia Agricultural Facts (August, 1964) Crop Reporting Service, Athens, Georgia.

[3] Jessen, R. J. (1942. "Statistical Investigation of a sample survey for obtaining farm facts." Iowa Agri. Exp. Stn. Res. Bul. 304.

[4] Patterson, H.D.(1950). "Sampling on successive occasions with partial replacement of units." Jour. Roy. Stat. Soc., Series B, pp. 241-255.

[5] Rana, D. S. and Chakrabarty, R. P. (1974). Two-stage sampling on successive occasions with partial replacement of units. Technical Report No. 104. Statistics & Computer Science Dept. University of Georgia.

[6] Rao, J.N.K. and Graham, J.E.(1964). "Rotation designs for sampling on repeated occasions." Jour. Amer. Stat. Asso., 59, PP. 492-509.

[7] Singh. D. (1968). "Estimates in successive sampling using a multistage design." Jour. Amer. Stat. Assoc., 63, pp. 99-112.

[8] Singh, D. and Kathuria, O.P.(1969). "On two-stage successive sampling." The Australian Journal of Statistics, 11, No. 2, pp. 59-66.

[9] Yates, F. (1960). Sampling methods for censuses and surveys. Third edition, Charles Griffin and Co., Ltd., London.

Table 1. Values of the weights a, b, c and d for ρ_b = .7, ϕ = .5, m = 4 and a set of values of ρ_w q and s.

	q = .3				q = .5				q = .7				
s	a	b	c	d	a	b	c	d	a	b	c	d	
.1	-.1613	.0190	.6639	.0648	-.1901	.0074	.5141	.0466	-.1738	.0002	.3509	.0293	
.3	-.1774	.0389	.5366	.1920	-.1872	.0097	.4194	.1396	-.1608	-.0072	.2881	.0890	
.5	-.1687	.0343	.4060	.3216	-.1658	-.0060	.3196	.2367	-.1359	-.0257	.2204	.1525	ρ_b=.7
.7	-.1312	.0014	.2637	.4621	-.1225	-.0426	.2086	.3438	-.0966	-.0574	.1439	.2232	ρ_b=.5
.9	-.0564	-.0680	.0976	.6250	-.0505	-.1064	.0774	.4691	-.0384	-.1062	.0532	.3062	
.1	-.1834	.0319	.6727	.0605	-.2139	.0173	.5280	.0402	-.1973	.0067	.3684	.0256	
.3	-.2204	.0736	.5600	.1739	-.2271	.0365	.4462	.1231	-.1951	.0109	.3149	.0764	
.5	-.2271	.0852	.4462	.2872	-.2178	.0339	.3596	.2071	-.1777	.0011	.2555	.1311	ρ_b=.7
.7	-.1951	.0591	.3149	.4161	-.1777	.0026	.2555	.3059	-.1389	-.0274	.1817	.1971	ρ_b=.7
.9	-.0969	-.0306	.1336	.5919	-.0844	-.0775	.1083	.4429	-.0631	-.0875	.0763	.2889	
.1	-.2063	.0454	.6829	.0553	-.2394	.0278	.5443	.0365	-.2241	.0141	.3897	.0209	
.3	-.2691	.1127	.5912	.1496	-.2747	.0682	.4819	.1006	-.2387	.0334	.3518	.0587	
.5	-.3047	.1526	.5081	.2339	-.2904	.0893	.4219	.1605	-.2403	.0408	.3124	.0961	ρ_b=.7
.7	-.3053	.1584	.4147	.3265	-.2778	.0842	.3488	.2304	-.2209	.0304	.2599	.1426	ρ_b=.9
.9	-.2127	.0761	.2485	.4887	-.1849	.0078	.2064	.3599	-.1398	.0299	.1517	.2325	

Table 2. Values of the weights a, b, c and d for ρ_b = .9, ϕ = .5, m = 4 and a set of values of ρ_w, q and s.

	q = .3				q = .5				q = .7				
s	a	b	c	d	a	b	c	d	a	b	c	d	
.1	-.1972	.0159	.6804	.0646	-.2469	.0033	.5524	.0477	-.2485	-.0047	.4129	.0321	
.3	-.2070	.0296	.5522	.1923	-.2344	-.0033	.4533	.1440	-.2228	-.0224	.3408	.0983	
.5	-.1914	.0183	.4194	.3237	-.2023	-.0287	.3472	.2461	-.1834	-.0525	.2617	.1700	ρ_b=.9
.7	-.1461	-.0221	.2734	.4674	-.1465	-.0764	.2276	.3600	-.1274	-.0974	.1712	.2510	ρ_b=.5
.9	-.0619	-.1003	.1016	.6353	-.0594	-.1535	.0847	.4949	-.0496	-.1614	.0633	.3465	
.1	-.2204	.0293	.6909	.0596	-.2746	.0139	.5710	.0421	-.2823	.0030	.4414	.0271	
.3	-.2520	.0655	.5785	.1723	-.2800	.0260	.4865	.1245	-.2692	-.0013	.3804	.0823	
.5	-.2529	.0715	.4634	.2864	-.2613	.0152	.3949	.2120	-.2383	-.0210	.3104	.1436	ρ_b=.9
.7	-.2136	.0385	.3287	.4181	-.2089	-.0266	.2822	.3172	-.1815	-.0623	.2211	.2195	ρ_b=.7
.9	-.1048	-.0610	.1400	.6002	-.0975	-.1217	.1200	.4657	-.0803	-.1399	.0924	.3264	
.1	-.2445	.0431	.7030	.0537	-.3048	.0254	.5927	.0354	-.3224	.0118	.4765	.0207	
.3	-.3035	.1065	.6132	.1457	-.3352	.0610	.5307	.0982	-.3308	.0264	.4359	.0592	
.5	-.3352	.1424	.5307	.2290	-.3452	.0770	.4691	.1585	-.3241	.0278	.3909	.0993	ρ_b=.9
.7	-.3308	.1435	.4359	.3224	-.3241	.0648	.3909	.2316	-.2907	.0081	.3269	.1522	ρ_b=.9
.9	-.2279	.0518	.2597	.4907	-.2119	-.0268	.2321	.3729	-.1782	-.0716	.1892	.2591	

Table 3. REM42 and REM43 for ρ_b = .7, ϕ = .5, m = 4 and a set of values of ρ_w, q and s.

	q = .3			q = .5			q = .7			
s	q'	REM42	REM43	q'	REM42	REM43	q'	REM42	REM43	
.1	.37	99	110	.55	100	114	.73	100	113	
.3	.51	97	110	.65	100	114	.79	102	113	
.5	.65	97	110	.75	101	113	.85	103	113	ρ_b=.7
.7	.79	99	110	.85	104	113	.91	106	112	ρ_b=.5
.9	.93	104	109	.95	108	112	.97	109	112	
.1	.37	99	111	.55	100	114	.73	100	114	
.3	.51	97	111	.65	100	114	.79	102	114	
.5	.65	97	111	.75	101	114	.85	104	114	ρ_b=.7
.7	.79	99	110	.85	104	114	.91	106	113	ρ_b=.7
.9	.93	104	110	.95	109	113	.97	109	112	
.1	.37	99	111	.55	100	114	.73	101	114	
.3	.51	97	111	.65	100	114	.79	102	114	
.5	.65	97	111	.75	102	114	.85	105	114	ρ_b=.7
.7	.79	99	111	.85	105	114	.91	108	114	ρ_b=.9
.9	.93	106	110	.95	110	114	.97	111	113	

Table 4. REM42 and REM43 for ρ_b = .9, ϕ = .5, m = 4 and a set of values of ρ_w, q and s.

	q = .3			q = .5			q = .7			
s	q'	REM42	REM43	q'	REM42	REM43	q'	REM42	REM43	
.1	.37	97	119	.55	99	128	.73	100	130	
.3	.51	93	119	.65	97	127	.79	101	129	
.5	.65	90	118	.75	98	126	.85	103	128	ρ_b=.9
.7	.79	92	118	.85	101	126	.91	108	127	ρ_b=.5
.9	.93	102	117	.95	112	124	.97	117	125	
.1	.37	97	120	.55	99	129	.73	100	133	
.3	.51	92	119	.65	97	129	.79	101	132	
.5	.65	90	119	.75	97	128	.85	103	131	ρ_b=.9
.7	.79	91	119	.85	100	127	.91	108	129	ρ_b=.7
.9	.93	100	118	.95	111	125	.97	117	127	
.1	.37	97	120	.55	99	130	.73	100	135	
.3	.51	92	120	.65	97	130	.79	101	135	
.5	.65	89	120	.75	97	130	.85	104	134	ρ_b=.9
.7	.79	90	120	.85	100	129	.91	109	133	ρ_b=.9
.9	.93	100	119	.95	111	127	.97	119	130	