THREE-STAGE SAMPLING ON SUCCESSIVE OCCASIONS

Dharam S. Rana, Old Dominion University
and R. P. Chakrabarty, Jackson State University

Introduction -- To maintain an up-to-date study of a time-dependent population, technique of sampling over successive occasions is used. In the present paper, the theory of sampling on successive occasions for a three-stage sampling design has been developed. In the case of three-stage successive sampling, there are about twelve sampling procedures available to alter the composition of the current sample. The estimates of population mean on the second occasion, the change between two consecutive occasions and the overall mean for two occasions are obtained for four selected procedures namely (5), (7), (8) and (11). The limited space did not permit the inclusion of the results for the overall mean.

2.1 Consider a population Π consisting of N primary stage units (PSU's) and each PSU consisting of M second-stage units (SSU's) and each SSU containing K third-stage units (TSU's). In the present study it is assumed that the population units are fixed, the variances on the two consecutive occasions are equal, the sample sizes remain same on each occasion and N, M and K are large so that the finite population correction factors at all three stages are negligible. The discussion is confined to two occasions only although the results obtained can be extended to more than two occasions. It is further assumed that the selection of the first sample, which consists of n PSU's, m SSU's within each of the n PSU's and k TSU's within each of the nm SSU's, is carried out by the method of simple random sampling without replacement and this applies to all the four sampling plans considered in this paper.

2.2 Estimate of the Mean by Procedure (5). On the second occasion, retain all the PSU's of the first sample but retain only a fraction r of the SSU's with their samples of TSU's in each of the PSU retained and select afresh a fraction s of SSU's such that r + s = 1. Let $y_{hij\ell}$ denote the value of the response variable y for the ℓ-th tertiary unit in the j-th second-stage unit within the i-th first-stage unit on the h-th occasion. A general linear unbiased estimator of $\bar{Y}_2$ the population on the second occasion may be written as $\bar{y}_{2(5)} = a\,[\bar{y}'_{1(5)} - \bar{y}^{*}_{1(5)}] + c\,\bar{y}'_{2(5)} + (1-c)\,\bar{y}^{*}_{2(5)}$ where $\bar{y}'_{h(5)} = \frac{1}{nrmk}\sum_{i=1}^{n}\sum_{j=1}^{rm}\sum_{\ell=1}^{k} y_{hij\ell}$ $\quad h = 1,2$ $\quad \bar{y}^{*}_{h(5)} = \frac{1}{nsmk}\sum_{i=1}^{n}\sum_{j=1}^{sm}\sum_{\ell=1}^{k} y_{hij\ell}$, h = 1, 2. The variance of $\bar{y}_{2(5)}$ is given by $Var\,[\bar{y}_{2(5)}] = a^2\,[\frac{Sw^2}{nrm} + \frac{St^2}{nrmk} + \frac{Sw^2}{nsm} + \frac{St^2}{nsmk}] + c^2\,[\frac{Sb^2}{n} + \frac{Sw^2}{nrm} + \frac{St^2}{nrmk}] + (1-c)^2\,[\frac{Sb^2}{n} + \frac{Sw^2}{nsm} + \frac{St^2}{nsmk}] + 2ac\,[\rho_b\frac{Sb^2}{n} + \rho_w\frac{Sw^2}{nrm} + \rho_t\frac{St^2}{nrmk} - \rho_b\frac{Sb^2}{n}] + 2\,c\,(1-c)\,\frac{Sb^2}{n}$ where $Sb^2 = \frac{1}{N-1}\sum_{i=1}^{N}(\bar{Y}i.. - \bar{Y}...)^2$ = true variance among PSU means, $Sw^2 = \frac{1}{N(M-1)}\sum_{i=1}^{N}\sum_{j=1}^{M}(\bar{Y}ij - \bar{Y}i..)^2$ = true variance among SSU means, $St^2 = \frac{1}{NM(K-1)}\sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{\ell=1}^{K}(y_{ij\ell} - \bar{Y}ij.)^2$ = true variance among TSU's. $\rho_b\,Sb^2 = \frac{1}{N-1}\sum_{i=1}^{N}(\bar{Y}_{1i}.. - \bar{Y}_1...)(\bar{Y}_{2i}.. - \bar{Y}_2...)$ = true Covariance among PSU means, $\rho_w\,Sw^2 = \frac{1}{N(M-1)}\sum_{i=1}^{N}\sum_{j=1}^{M}(\bar{Y}_{1ij.} - \bar{Y}_{1i..})(\bar{Y}_{2ij.} - \bar{Y}_{2i..})$ = true covariance among SSU means. $\rho_t\,St^2 = \frac{1}{NM(K-1)}\sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{\ell=1}^{K}(y_{1ij\ell} - \bar{Y}_{1ij.})(y_{2ij\ell} - \bar{Y}_{2ij.})$ = true covariance among TSU's.

In the above formulae, ρ_b, ρ_w and ρ_t denote the true correlation coefficients among PSU means, SSU means and TSU's respectively and $\bar{Y}_h...$, $\bar{Y}_{hi}..$ and $\bar{Y}_{hij}.$ represent true mean of the Psu means true mean of the i-th PSU and the true mean of the j-th SSU in the i-th PSU on the h-th occasion (h = 1,2) respectively. The optimum weights a_o and c_o that will minimize the variance of $\bar{y}_{2(5)}$ are

$$a_o = \frac{-rs\,\alpha_o\,\beta_o}{\alpha_o^2 - s^2\beta_o^2} \quad \text{and} \quad c_o = \frac{r\,\alpha_o^2}{\alpha_o^2 - s^2\beta_o^2} \quad \text{where}$$

$$\alpha_o = Sw^2 + \frac{St^2}{k} \text{ and } \beta_o = \rho_w\,Sw^2 + \rho_t\,\frac{St^2}{k} \qquad (2.2.1)$$

It should be noted here that both $\bar{y}_{2(5)}$ and its variance are independent of ρ_b. With optimum weights, the variance of $\bar{y}_{2(5)}$ is given by

$$Var\,[\bar{y}_{2(5)}] = \frac{1}{n}\,[Sb^2 + \frac{\alpha_o}{m}\,\frac{(\alpha_o^2 - s\,B_o^2)}{(\alpha_o^2 - s^2\beta_o^2)}] \qquad (2.2.2)$$

The optimum replacement fraction s_o that would minimize this variance is

$$s_o = 1/[1 + (1 - \frac{\beta_o^2}{\alpha_o^2})^{1/2}]$$

It may be easily shown that $s_o \geqslant 1/2$. There are two special cases of interest here e.g., (i) s = 0 and (ii) s = 1. In either case, it follows from (2.2.2) that $Var[\bar{y}_{2(5)}] = \frac{1}{n}[Sb^2 + \frac{\alpha_o}{m}]$. Thus it is clear that a complete retention or complete replacement of SSU's within the PSU's from the first sample on the second occasion does not help to improve on the estimate of the current population mean.

2.3 Estimate of the Mean by Procedure (8)

From the first sample retain only a fraction p of the PSU's along with their samples of SSU's and TSU's on the second occasion. Replace the remaining fraction q (such that p + q = 1) of the PSU's by a fresh random selection of PSU's on the second occasion.

A general linear unbiased estimator of $\bar{Y}_2$ may be

$$\bar{y}_{2(8)} = a\,[\bar{y}_{1(8)} - \bar{y}''_{1(8)}] + c\,\bar{y}'_{2(8)} + (1-c)\,\bar{y}''_{2(8)}$$

and its variance is given by

$$\mathrm{Var}\,[\bar{y}_{2(8)}] = a^2\Big[\frac{\alpha}{np} + \frac{\alpha}{nq}\Big] + c^2\,\frac{\alpha}{np} + (1-c)^2\,\frac{\alpha}{nq} + 2ac\,\frac{\delta}{np}.$$

where $\bar{y}'_{h(8)} = \frac{1}{npmk}\sum_{i=1}^{np}\sum_{j=1}^{m}\sum_{\ell=1}^{k} y_{hij\ell}$, $h = 1, 2$ $\quad \bar{y}''_{h(8)} = \frac{1}{nqmk}\sum_{i=1}^{nq}\sum_{j=1}^{m}\sum_{\ell=1}^{k} y_{hij\ell}$, $h = 1, 2$ $\quad \alpha = Sb^2 + \frac{Sw^2}{m} + \frac{St^2}{mk}$ and

$$\delta = \rho_b\,Sb^2 + \rho_w\,\frac{Sw^2}{m} + \rho_t\,\frac{St^2}{mk} \qquad (2.3.1)$$

The optimum weights are; $a_o = -\frac{pq\,\alpha\,\delta}{\alpha^2 - q^2\delta^2}$ and $c_o = \frac{p\,\alpha^2}{\alpha^2 - q^2\delta^2}$. The variance of $\bar{y}_{2(8)}$ with optimum weights is

$$\mathrm{Var}\,[\bar{y}_{2(8)}] = \frac{\alpha}{n}\,\frac{(\alpha^2 - q\delta^2)}{(\alpha^2 - q^2\delta^2)} \qquad (2.3.2)$$

The optimum replacement fraction q_o that will minimize the variance in (2.3.2) is

$$q_o = \frac{\alpha^2 - \alpha\sqrt{\alpha^2 - \delta^2}}{\delta^2} \quad \text{and} \quad \mathrm{Var}\,[\bar{y}_{2(8)}]_{opt} = \frac{\delta^2}{2n[\alpha - \sqrt{\alpha^2 - \delta^2}]}$$

2.4 Estimate of the Mean by Procedure (7):

Retain all the PSU's from the first sample. In each of the PSU's retained, further retain only a fraction r of the SSU's and select a fraction s of the new SSU's such that r + s = 1. In each of the matched SSU's, retain only a fraction t of TSU's and select fresh a fraction u of TSU's such that t + u = 1.

A general linear unbiased estimator of $\bar{Y}_2$, may be

$$\bar{y}_{2(7)} = a\,\bar{y}'_{1(7)} + b\,\bar{y}^{**}_{1(7)} - (a+b)\,\bar{y}^{*}_{1(7)} + d\,\bar{y}'_{2(7)} + e\,\bar{y}^{**}_{2(7)} + (1-d-e)\,\bar{y}^{*}_{2(7)}$$

where

$$\bar{y}'_{h(7)} = \frac{1}{nrmtk}\sum_{i=1}^{n}\sum_{j=1}^{rm}\sum_{\ell=1}^{tk} y_{hij\ell}, \quad h = 1,2$$

$\bar{y}^{**}_{h(7)} = \frac{1}{nrmuk}\sum_{i=1}^{n}\sum_{j=1}^{rm}\sum_{\ell=1}^{uk} y_{hij\ell}$ and $\bar{y}^{*}_{h(7)} = \frac{1}{nsmk}\sum_{i=1}^{n}\sum_{j=1}^{sm}\sum_{\ell=1}^{k} y_{hij\ell}$. The variance of $\bar{y}_{2(7)}$ is given by

$$\mathrm{Var}\,[\bar{y}_{2(7)}] = (a^2 + d^2)\,\alpha' + (b^2 + e^2)\,\beta' + [(a+b)^2 + (1-d-e)^2]\gamma' + 2(ab + de)\alpha^* + 2(ae + bd + be)\delta^* + 2ad\delta' - 2(a+b)(d+e)\,\rho_b\,\frac{\delta b^2}{n} + 2[(d+e)(1-d-e) - (a+b)^2]\frac{Sb^2}{n}$$

where

$\alpha' = \frac{Sb^2}{n} + \frac{Sw^2}{nrm} + \frac{St^2}{nrmtk}$, $\beta' = \frac{Sb^2}{n} + \frac{Sw^2}{nrm} + \frac{St^2}{nrmuk}$

$\gamma' = \frac{Sb^2}{n} + \frac{Sw^2}{nsm} + \frac{St^2}{nsmk}$, $\delta' = \rho_b\,\frac{Sb^2}{n} + \rho_w\,\frac{Sw^2}{nrm} + \rho_t\,\frac{St^2}{nrmtk}$, $\alpha^* = \frac{Sb^2}{n} + \frac{Sw^2}{nrm}$ and $\delta^* + \rho_b\,\frac{Sb^2}{n} + \rho_w\,\frac{Sw^2}{nrm}$.

2.5 Estimate of the Mean by Procedure (11)

Retain a fraction p of the PSU's from the first sample and select a fraction q (such that p + q = 1) of new PSU's on the second occasion. Further, retain a fraction r of the SSU's with their samples of TSU's in each of the PSU's retained and select anew the remaining fraction s of SSU's such that r + s = 1.

A general linear unbiased estimator of $\bar{Y}_2$ is

$$\bar{y}_{2(11)} = a\,\bar{y}'_{1(11)} + b\,\bar{y}^{*}_{1(11)} + (1-d-e)\,\bar{y}''_{1(11)} + d\,\bar{y}'_{2(11)} + e\,\bar{y}^{*}_{2(11)} + (1-d-e)\,\bar{y}''_{2(11)}$$

$\bar{y}'_{h(11)} = \frac{1}{nprmk}\sum_{i=1}^{np}\sum_{j=1}^{rm}\sum_{\ell=1}^{k} y_{hij\ell}$, $\bar{y}^{*}_{h(11)} = \frac{1}{npsmk}\sum_{i=1}^{np}\sum_{j=1}^{sm}\sum_{\ell=1}^{k} y_{hij\ell}$

and

$$\bar{y}''_{h(11)} = \frac{1}{nqmk}\sum_{i=1}^{nq}\sum_{j=1}^{m}\sum_{\ell=1}^{k} y_{hij\ell}, \quad h = 1, 2$$

The variance of $\bar{y}_{2(11)}$ is

$$\mathrm{Var}\,[\bar{y}_{2(11)}] = \frac{a^2}{np}\Big[Sb^2 + \frac{\alpha_o}{rm}\Big] + \frac{b^2}{np}\Big[Sb^2 + \frac{\alpha_o}{sm}\Big] + \frac{(a+b)^2}{nq}\,\alpha + \frac{d^2}{np}\Big[Sb^2 + \frac{\alpha_o}{rm}\Big] + \frac{e^2}{np}\Big[Sb^2 + \frac{\alpha_o}{sm}\Big] + \frac{(1-d-e)^2}{nq}\,\alpha + 2ab\,\frac{Sb^2}{np} + 2\,\frac{ad}{np}\Big[Sb^2 + \frac{\beta_o}{rm}\Big] + \frac{2\,Sb^2}{np}[a\,e\,\rho_b + bd\,\rho_b + be\,\rho_b + de].$$

The optimum weights are given by

$$a_o = -\,\Omega\,pr\,\alpha_o\Big[s\,Sb^2\,\beta_o + q\,\rho_b\,Sb^2\,\alpha_o + \frac{(q + ps)}{m}\,\alpha_o\,\beta_o\Big]$$

$$b_o = \Omega\,ps\,\Big[Sb^2(qs\,\rho_b\,\beta_o^2 - q\,\rho_b\,\alpha_o^2 + r\,\alpha_o\,\beta_o) + \frac{pr}{m}\,\alpha_o\,\beta_o\Big]$$

$$d_o = \Omega\ pr\ \alpha_o\ [Sb^2\ (\alpha_o + qs\ \rho_b\ \beta_o) + \frac{\alpha_o^2}{m}]$$

and

$$e_o = \Omega\ ps\ [Sb^2(\alpha_o^2 - qr\ \rho_b\ \alpha_o\ \beta_o - s\ \beta_o^2) + \frac{\alpha_o}{m}$$

$$(\alpha_o^2 - (q + ps)\ \beta_o^2)]$$

where

$$\Omega = \alpha/[(qs\ \rho_b\ \beta_o + Sb^2\ \alpha_o + \frac{\alpha_o^2}{m})^2 - (q\ \rho_b\ Sb\ \alpha_o +$$

$$s\ Sb^2\ \beta_o + \frac{(q + ps)}{m}\ \beta_o\ \alpha_o)^2]$$ and α_o, β_o and α are the same as defined in section 2.2 and 2.3.

2.6 Relative Efficiency

In a design problem, it is important to study the relative performance of the different estimators. To study the relative efficiency of the four sampling procedures, it is necessary to have the same overall replacement fraction. Denoting the overall replacement fraction by q^*, it may be shown that for the Procedure (7), $q^* = s + u - s \times u$ and for the Procedure (11), $q^* = q + s - q \times s$. Let RM85 = Var $[\bar{y}_{2(5)}]$/ Var $[\bar{y}_{2(8)}]$, RM87 = Var $[\bar{y}_{2(7)}]$/Var $[\bar{y}_{2(8)}]$ and RM811 = Var $[\bar{y}_{2(11)}]$/Var $[\bar{y}_{2(8)}]$.

The relative efficiencies are computed for some selected values of the parameters and the design quantities. Limited space permits to present only a small fraction of the results in table 2.1. In most cases, procedure (8) is more efficient than the other procedures. The effects of various quantities on the relative efficiencies are discussed in details in the technical report. In tables 2.1 and 3.1, the symbols $\phi = Sw^2/Sb^2$, $\psi = St^2/Sb^2$, m = 16, k = 8, q = u = .5, s = .1 and $q^* = .55$.

3.1 Estimate of the Change by Procedure (5)

A general linear unbiased estimator of $(\bar{Y}_2 - \bar{Y}_1)$ the change, may be written as

$$\Delta_{(5)} = a\ \bar{y}'_{1(5)} - (1 + a)\ \bar{y}^*_{1(5)} + c\ \bar{y}'_{2(5)} + (1 - c)\ \bar{y}^*_{2(5)}$$

and its variance is given by

$$\text{Var}\ [\Delta_{(5)}] = a^2[\frac{Sb^2}{n} + \frac{\alpha_o}{nrm}] + (1 + a)^2[\frac{Sb^2}{n} + \frac{\alpha_o}{nsm}]$$

$$+ c^2\ [\frac{Sb^2}{n} + \frac{\alpha_o}{nrm}] + (1 - c)^2\ [\frac{Sb^2}{n} + \frac{\alpha_o}{nsm}] + 2ac$$

$$[\rho_b\ \frac{Sb^2}{n} + \frac{\beta_o}{nrm}] + 2[a(1 - c) - c(1 + a) - (1 + a)$$

$$(1 - c)] \times \rho_b\ \frac{Sb^2}{n} - 2[a(1 + a) - c(1 - c)]\ \frac{Sb^2}{n}$$

For $\bar{y}'_{h(5)}$, $\bar{y}^*_{h(5)}$, α_o and β_o see Section 2.2.

The optimum weights are

$$a_o = \frac{-r\ \alpha_o}{(\alpha_o - s\ \beta_o)} \quad \text{and} \quad c_o = \frac{r\ \alpha_o}{(\alpha_o - s\ \beta_o)}$$

The variance of $\Delta_{(5)}$ with optimum weights is

$$\text{Var}\ [\Delta_{(5)}] = \frac{2}{n}\ [(1 - \rho_b)Sb^2 + \frac{\alpha_o}{m}\frac{(\alpha_o - \beta_o)}{(\alpha_o - s\beta_o)}]$$

(3.1.1)

<u>Special Cases</u>: It is easily seen from (3.1.1) that if s = 0,

$$\text{Var}\ [\Delta_{(5)}] = \frac{2}{n}\ [(1 - \rho_b)Sb^2 + (1 - \rho_w)\frac{Sw^2}{m} + (1 - \rho_t)\frac{St^2}{mk}] \qquad (3.1.2)$$

and if s = 1, we obtain

$$\text{Var}\ [\Delta_{(5)}] = \frac{2}{n}\ [(1 - \rho_b)\ Sb^2 + \frac{Sw^2}{m} + \frac{St^2}{mk}] \qquad (3.1.3)$$

From (3.1.1) and (3.1.3) it is clear that for positive correlations, it is advantageous to retain a fraction of SSU's from the first sample to estimate the change.

3.2 Estimate of Change by Procedure (8)

A general linear unbiased estimator of $(\bar{Y}_2 - \bar{Y}_1)$ is

$$\Delta_{(8)} = a\ \bar{y}'_{1(8)} - (1 + a)\ \bar{y}''_{1(8)} + c\ \bar{y}'_{2(8)} + (1 - c)\ \bar{y}''_{2(8)}$$

and its variance is given by

$$\text{Var}\ [\Delta_{(8)}] = a^2\ \frac{\alpha}{np} + (1 + a)^2\ \frac{\alpha}{nq} + c^2\ \frac{\alpha}{np} + (1 - c)^2\ \frac{\alpha}{nq} + 2ac\ \frac{\delta}{np}$$

where $\bar{y}'_{h(8)}$, $\bar{y}''_{h(8)}$, α and δ are the same as defined in Section 2.3. The optimum weights are

$$a_o = \frac{-p\ \alpha}{\alpha - q\delta} \quad \text{and} \quad c_o = \frac{p\ \alpha}{\alpha - q\delta}$$

The variance of $\Delta_{(8)}$ with optimum weights is

$$\text{Var}\ [\Delta_{(8)}] = \frac{2}{n}\ \alpha\ \frac{(\alpha - \delta)}{(\alpha - q\delta)}$$

3.3 Estimate of the Change by Procedure (7)

One possible linear unbiased estimator of the change may be of the form

$$\Delta_{(7)} = a[\bar{y}'_{2(7)} - \bar{y}'_{1(7)}] + b[\bar{y}^{**}_{2(7)} - \bar{y}^{**}_{1(7)}] + (1 - a - b)[\bar{y}^{**}_{2(7)} - \bar{y}^*_{1(7)}]$$

and its variance is

$$\text{Var}\ [\Delta_{(7)}] = \frac{2}{n}[a^2(\lambda_1 - \lambda_4 + \lambda_3) + b^2(\lambda_2 - \rho_w\ \frac{Sw^2}{rm}$$

$+ \lambda_3) + (1 - \rho_b)Sb^2 + \lambda_3 + 2ab \{(1 - \rho_w) \frac{Sw^2}{rm} + \lambda_3\} - 2 \lambda_3(a + b)]$

where

$$\lambda_1 = \frac{Sw^2}{rm} + \frac{St^2}{rmtk}, \quad \lambda_3 = \frac{Sw^2}{sm} + \frac{St^2}{smk},$$

$$\lambda_2 = \frac{Sw^2}{rm} + \frac{St^2}{rmuk}, \quad \lambda_4 = \rho_w \frac{Sw^2}{rm} + \rho_t \frac{St^2}{rmtk}$$

The optimum weights are given by

$$a_o = rt\alpha_o [(1 - u\rho_t)(1 - s\rho_w)Sw^2 + (1 - s\rho_t - ru\rho_t)\frac{St^2}{k}]^{-1}$$

$$b_o = ru(1 - \rho_t)\alpha_o [(1 - u\rho_t)(1 - s\rho_w) Sw^2 + (1 - s\rho_t - ru\rho_t) \frac{St^2}{k}]^{-1}$$

3.4 Estimate of the Change by Procedure (11)

One possible linear unbiased estimator of the change is

$$\Delta_{(11)} = a[\bar{y}'_{2(11)} - \bar{y}'_{1(11)}] + b[\bar{y}^*_{2(11)} - \bar{y}^*_{1(11)}] + (1 - a - b)[\bar{y}''_{2(11)} - \bar{y}''_{1(11)}]$$

and its variance is

$$\mathrm{Var}\,[\Delta_{(11)}] = 2a^2[\frac{Sb^2}{np} + \frac{\alpha_o}{nprm} - (\rho_b \frac{Sb^2}{np} + \frac{\beta_o}{nprm})] + 2b^2[\frac{Sb^2}{np} + \frac{\alpha_o}{npsm} - \rho_b \frac{Sb^2}{np}] + 2(1 - a - b)^2 \frac{\alpha}{nq} + 4ab(1 - \rho_b)\frac{Sb^2}{np}$$

The optimum weights are

$a_o = pr\ \alpha_o\ \alpha/D'$ and $b_o = ps(\alpha_o - \beta_o)\alpha/D'$

where

$$D' = Sb^2(1 - q\rho_b)[\alpha_o - s\beta_o] + \frac{\alpha_o}{m} [\alpha_o - (q + ps)\beta_o]$$

The variance of $\Delta_{(11)}$ with optimum weights is

$$\mathrm{Var}\,[\Delta_{(11)}] = \frac{2}{n} (Sb^2 + \frac{\alpha_o}{m})[Sb^2(1 - \rho_b)(\alpha_o - s\beta_o) + \frac{\alpha_o}{m} (\alpha_o - \beta_o)] \times [Sb^2(1 - q\rho_b)(\alpha_o - s\beta_o) + \frac{\alpha_o}{m} \{\alpha_o - (q + ps)\beta_o\}]^{-1}$$

3.5 Relative Efficiency

Let RC75 = Var $[\Delta_{(5)}]$/Var $[\Delta_{(7)}]$, denote the relative efficiency of Procedure (7) with respect to Procedure (5). Symbols RC78 and RC711 have similar meanings. These relative efficiencies are studied numerically and some of the results are presented in Table 3.1. The Procedure (7) provides the most efficient estimate of the change.

4.1 Sample Allocation

In most applications of sampling, cost is an important factor since the resources for sample surveys are always limited. Therefore, it is important to study the optimum allocation of the sample subject to a given cost. The optimum distribution of the sample to estimate the current population mean by Procedures (5) and (8) in two-stage successive sampling is considered here. It is assumed that the travel cost between units is unimportant.

4.2 Allocation of the Sample in Procedure (5)

A simple cost function for two occasions may be

$$c(1) = c_1n + c_2nm; \quad c(2) = c_2'nrm + c_2'' nsm$$

where c_1 is the cost of preparing frame and c_2 the cost of enumeration on the first occasion. c_2' and c_2'' are the costs of enumeration on the matched and unmatched parts of the sample on the second occasion. The total cost for two occasion is

$$c = c_1n + (c_2 + c_2'r + c_2'' s)nm \qquad (4.2.1)$$

where

$c = c(1) + c(2)$. The variance of $\bar{y}_{2(5)}$ in two-stage, successive sampling is

$$\mathrm{Var}\,[\bar{y}_{2(5)}] = \frac{Sb^2}{n} + (\frac{1 - s\rho_w^2}{1 - s^2\rho_w^2}) \frac{Sw^2}{nm} \qquad (4.2.2)$$

From (4.2.1) and (4.2.2) it may be shown that the optimum values of m and n are

$$m_o = \{ c_1\phi(1-s\rho_w^2)/[(1-s^2\rho_w^2)(c_2+c_2'r+c_2''s)] \}^{1/2}$$

$$n_o = \frac{[c(1-s^2\rho_w^2)]^{1/2}}{c_1(1-s^2\rho_w^2)^{1/2}+ \{ c_1\phi(c_2+c_2'r+c_2''s)(1-s\rho_w^2)\}^{1/2}}$$

4.3 Allocation of the Sample in Procedure (8)

The total cost for two occasions in procedure (8) is

$$c = (c_1+c_1'q)n + (c_2+c_2p+c_2''q)nm \qquad (4.3.1)$$

It is noted here that $c_1'q$ is the additional cost of frame due to new selection of a fraction q of PSU's on the second occasion. The variance of $\bar{y}_{2(8)}$ for two-stage successive sampling is

$$\mathrm{Var}\,[\bar{y}_{2(8)}] = \frac{1}{n}(Sb^2 + \frac{Sw^2}{m}) \times$$

$$\frac{[(Sb^2 + \frac{Sw^2}{m})^2 - q(\rho_b Sb^2 + \rho_w \frac{Sw^2}{m})^2]}{[(Sb^2 + \frac{Sw^2}{m})^2 - q^2(\rho_b Sb + \rho_w \frac{Sw^2}{m})^2]} \qquad (4.3.2)$$

Eliminating n from (4.3.1) and (4.3.2) we obtain

$$\mathrm{Var}[\bar{y}_{2(8)}] = \frac{1}{c}[c_1 + \tfrac{1}{2}c_1' + \frac{m}{2}(2c_2 + c_2' + c_2'')](Sb^2 + \frac{Sw^2}{m}) \quad x$$

$$\frac{[(Sb^2 + \frac{Sw^2}{m})^2 - \frac{1}{2}(\rho_b Sb^2 + \rho_w \frac{Sw^2}{m})^2]}{[(Sb^2 + \frac{Sw^2}{m})^2 - \frac{1}{4}(\rho_b Sb^2 + \rho_w \frac{Sw^2}{m})^2]} \qquad (4.3.3)$$

Now $\frac{\partial}{\partial m}$ [Var $(\bar{y}_{2(8)}]$ = 0 provides a sixth degree equation in m which is solved for the optimum m by the method of successive approximations. From (4.3.1), optimum n is obtained.

4.4 Relative Efficiency

Let REMC58 = Var $[\bar{y}_{2(8)}]$/Var $[\bar{y}_{2(5)}]$ represent the relative efficiency of Procedure (5) with respect to Procedure (8). On the basis of numerical study (some of the results presented in Table 4.1) made, it is observed that the Procedure (5) is more efficient than the Procedure (8) to estimate the current population mean.

Conclusions: It is observed from the extensive numerical study of the relative efficiencies that if the sampling is inexpensive and the precision of the estimates is of major interest, the sampling Procedure (8) is more efficient than the other procedures in most cases. However, the gains of Procedure (8) over Procedure (5) are modest in most cases. If cost is taken into consideration, Procedure (5) is more efficient than Procedure (8) to estimate the current population mean.

References

[1] Chakrabarty, R. P. and D. S. Rana (1974). "Multi-Stage Sampling with Partial Replacement of the Sample on Successive Occasions". Proceedings of the ASA (Social Statistics Section) 1974. pp. 262-268.

[2] Eckler, A. R. (1955). "Rotation Sampling". Ann. Math. Stat., 26, pp. 665-685.

[3] Singh, D. (1968). "Estimates in Successive Sampling Using a Multi-Stage Design". Jour. Amer. Stat. Assoc., 63, pp. 99-112.

[4] Singh, D. and Kathuria, O. P. (1969). "On Two-Stage Successive Sampling". The Australian Journal of Statistics, 11, No. 2, pp. 59-66.

[5] Chakrabarty, R. P. and D. S. Rana, "Three-Stage Sampling on Successive Occasions". Technical Report. Department of Stat. and Comp. Sci., University of Georgia. August, 1976.

Table 2.1 Values of RM85, RM87, RM811

ρ_b	ρ_w	ρ_t	$\phi=.5,\ \psi=2$ x	y	z	$\phi=5,\ \psi=2$ x	y	z
	.5	.5	100	107	100	100	107	100
.5	.7	.7	92	107	100	93	108	100
	.9	.9	79	107	100	82	109	100
	.5	.5	108	116	100	106	113	100
.7	.7	.7	100	116	100	100	115	100
	.9	.9	86	116	100	88	118	100
	.5	.5	125	133	101	116	124	101
.9	.7	.7	116	134	102	111	128	101
	.9	.9	100	135	102	100	133	101

x = RM85 y = RM87 z = RM811

Table 3.1 Values of RC75, RC711, RC78

ρ_b	ρ_w	ρ_t	$\phi=.5,\ \psi=2$ x	y	z	$\phi=5,\ \psi=2$ x	y	z
	.5	.5	101	133	137	108	132	136
.5	.7	.7	101	133	138	108	137	141
	.9	.9	101	135	139	105	142	148
	.5	.5	102	151	159	111	147	152
.7	.7	.7	102	152	161	113	152	159
	.9	.9	101	154	163	108	159	168
	.5	.5	104	175	188	119	165	171
.9	.7	.7	104	176	190	125	171	179
	.9	.9	103	179	195	120	180	191

x = RC75 y = RC711 z = RC78

Table 4.1 Values of REMC 58

ρ_b	ϕ	$q^* = .5$ $\rho_w=.5$	.7	.9	$q^* = .7$ $\rho_w=.5$	.7	.9
	.5	129	129	133	141	142	147
.5	.5	117	118	121	123	124	130
	10	114	115	118	119	120	126
	.5	122	122	124	133	133	136
.7	5	113	113	115	119	119	122
	10	111	111	113	116	116	119
	.5	113	111	111	121	118	117
.9	.5	109	107	107	114	112	110
	10	108	107	106	113	110	109

$c = 10550 \quad c_1' = 70.50 \quad c_2' = 12.50$

$c_1 = 64.40 \quad c_2 = 14.75 \quad c_2'' = 16.25$