JOINT ESTIMATES OF CURRENT LEVEL AND CHANGE IN THREE-STAGE SUCCESSIVE SAMPLING

Dharam S. Rana, Old Dominion University

1. Introduction - In many repeat surveys, it is required to report estimates of population mean on the current occasion and the immediately preceding occasion. These simultaneous estimates of mean and change are called joint or combined estimates. It has been seen that sometimes the required timings of the survey estimates is such that the estimate of the mean for the first occasion can wait until data for the second occasion is available. In such cases the estimate of mean on the first occasion can also be improved by using data from the second sample and the difference between these two estimated means gives the best linear estimate (i.e., unbiased estimate with minimum variance) of change that can be obtained from the data from the two samples. However, in many situations, estimates for the first occasion must be made before sample results from the second occasion are made available. In such cases, the population mean on the first occasion has to be estimated from the first sample only, and it may not be feasible to revise this initial estimate later on. It is the latter case that will be considered here to develop joint estimates of mean and change.

In three-stage successive sampling, there are many ways to alter the composition of the first sample on the second occasion. In the present paper only four important alternatives (sampling procedures) have been selected to obtain the joint estimates. It is assumed that the units in the population of interest are fixed and the sample size remains same on each occasion. The study is confined to two occasions only, but the results obtained can be extended to more than two occasions. On the first occasion, a simple random sample of n primary stage units (PSU's) is selected from the population of interest. Within each of n PSU's, a random selection of m second-stage units (SSU's) is made and in each of these nm SSU's a random sample of k third-stage units (TSU's) is taken. Selection at each stage is carried out by simple random sampling without replacement, and it is the same in case of all the four procedures.

2. Notations

Let N = Number of PSU,s in the population, M = Number of SSU's in each PSU, K = Number of TSU's in each SSU within PSU's, and $y_{hij\ell}$ = value of the ℓ-th tertiary unit in the j-th second-stage unit located in i-th first-stage unit.

$$S_{bh}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(\bar{Y}_{i..} - \bar{Y}_{...})^2 \text{ , } h = 1, 2$$

= true variance among PSU means on the h-th occasion.

$$S_{wh}^2 = \frac{1}{N(M-1)}\sum_{i=1}^{N}\sum_{j=1}^{M}(\bar{Y}_{ij.} - \bar{Y}_{i..})^2 \text{ , } h = 1, 2$$

= true variance among SSU means on the h-th occasion.

$$S_{th} = \frac{1}{NM(K-1)}\sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{\ell=1}^{K}(y_{ij\ell} - \bar{Y}_{ij.})^2 \text{ , } h = 1,2$$

= true variance among TSU's on the h-th occasion.

$$\rho_b S_{b1} S_{b2} = \frac{1}{N-1}\sum_{i=1}^{N}(\bar{Y}_{1i..} - \bar{Y}_{1...})(\bar{Y}_{2i..} - \bar{Y}_{2...})$$

$$\rho_w S_{w1} S_{w2} = \frac{1}{N(M-1)}\sum_{i=1}^{N}\sum_{j=1}^{M}(\bar{Y}_{1ij.} - \bar{Y}_{1i..})(\bar{Y}_{2ij.} - \bar{Y}_{2i..})$$

and

$$\rho_t S_{t1} S_{t2} = \frac{1}{NM(K-1)}\sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{\ell=1}^{K}(y_{1ij\ell} - \bar{Y}_{1ij.})(y_{2ij\ell} - \bar{Y}_{2ij.})$$

represent true covariances among PSU mean, SSU means and TSU's respectively between first and second occasion. In the above relations, ρ_b, ρ_w and ρ_t denote true correlations among PSU means, SSU means and TSU's respectively $\bar{Y}_{h...}$, $\bar{Y}_{hi..}$ and $Y_{hij.}$ stand for population means of TSU's, i-th PSU and j-th SSU within i-th PSU on the h-th occasion (h = 1,2) respectively. For convenience sake, dots to denote $\bar{Y}_{h...}$, $\bar{Y}_{hi..}$, etc. will be dropped.

2.1 Joint Estimates of Mean and Change by Procedure (1)- All the PSU's of the first sample are retained on the second occasion but only a fraction r of SSU's with their sample of TSU's in each of these PSU's is retained. The remaining fraction s of SSU's is selected afresh in a random manner so that r + s = 1. Under this sampling plan, an initial estimate of $\bar{Y}_1$ the population on the first occasion can be written as

$$\bar{y}_{1(1)} = r\,\bar{y}'_{1(1)} + s\,\bar{y}^*_{1(1)}$$

The joint linear estimates of the population mean on the second occasion and the change that occured in the characteristic of interest between first and second occasion may be expressed as

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

and

$$\Delta_{(1)} = e\,\bar{y}'_{1(1)} + f\,\bar{y}^*_{1(1)} + g\,\bar{y}'_{2(1)} + h\,\bar{y}^*_{2(1)} \qquad (2.1.1)$$

where

$$\bar{y}'_{h(1)} = \frac{1}{nrmk}\sum_{i=1}^{n}\sum_{j=1}^{rm}\sum_{\ell=1}^{k} y_{hij\ell}, \text{ } h = 1, 2$$

= mean per TSU based on nrmk matched TSU's. $\Delta_{(i)}$ = estimate of change between first and second occasion by i-th procedure (i = 1,2,3, 4) and

$$\bar{y}^*_{h(1)} = \frac{1}{nsmk}\sum_{i=1}^{n}\sum_{j=1}^{sm}\sum_{\ell=1}^{k} y_{hij\ell}, \text{ } h = 1, 2$$

= mean per TSU based on nsmk TSU's. It is to be noted here that the number within parenthesis indicates the sampling procedure.

Suppose it is desired that the estimate of change between first and second occasion equals the difference between estimated means on these occasions, that is,

$$\Delta_{(1)} = \bar{y}_{2(1)} - \bar{y}_{1(1)}$$

Using the above relation the condition of unbiasedness, the joint estimates in (2.1.1) may be rewritten as

$$y_{2(1)} = (e+r)\bar{y}'_{1(1)} - (e+r)\bar{y}^*_{1(1)} + c\bar{y}'_{2(1)} + (1-c)\bar{y}^*_{2(1)}$$

$$\Delta_{(1)} = e\bar{y}'_{1(1)} - (1+e)\bar{y}^*_{1(1)} + c\bar{y}'_{2(1)} + (1-c)\bar{y}^*_{2(1)} \qquad (2.1.2)$$

It may be useful to determine the weights c and e so as to minimize a linear function of variances of $\bar{y}_{2(1)}$ and $\Delta_{(1)}$. Let $\mathrm{Var}[\Delta_{(1)}] + \lambda\,\mathrm{Var}[\bar{y}_{2(1)}]$ represent one such linear function, where λ is a specified positive number. It will be assumed throughout this study that N, M and K are large and the true variances on the two occasions are equal, that is, $S^2_{b1} = S^2_{b2} = S^2_b$, $S^2_{w1} = S^2_{w2} = S^2_w$ and $S^2_{t1} = S^2_{t2} = S^2_{t2} = S^2_t$. Yates [5] argues that wherever the successive sampling is likely to be used the assumption of equality of variances on consecutive occasions holds. Under the simplifying assumptions of equal variances and that the finite population correction factors e.g., $\frac{n}{N}$, $\frac{m}{M}$ and $\frac{k}{K}$ are negligible, it can be shown that

$$\mathrm{Var}\,\Delta_{(1)} = e^2\left(\frac{S_b^2}{n} + \frac{S_w^2}{nrm} + \frac{S_t^2}{nrmk}\right) + (1+e)^2\left(\frac{S_b^2}{n} + \frac{S_w^2}{nsm} + \frac{S_t^2}{nsmk}\right) + c^2\left(\frac{S_b^2}{n} + \frac{S_w^2}{nrm} + \frac{S_t^2}{nrmk}\right) + (1-c)^2\left(\frac{S_b^2}{n} + \frac{S_w^2}{nsm} + \frac{S_t^2}{nsmk}\right) - 2e(1+e)\frac{S_b^2}{n} + 2ec\left(\rho_b\frac{S_b^2}{n} + \rho_w\frac{S_w^2}{nrm} + \rho_t\frac{S_t^2}{nrmk}\right) + 2e(1-c)\rho_b\frac{S_b^2}{n} - 2(1+e)c\,\rho_b\frac{S_b^2}{n} - 2(1+e)(1-c)\rho_b\frac{S_b^2}{n} + 2c(1-c)\frac{S_b^2}{n}$$

and

$$\mathrm{Var}\,y_{2(1)} = (e+r)^2\left(\frac{S_b^2}{n} + \frac{S_w^2}{nrm} + \frac{S_t^2}{nrmk}\right) + \left(\frac{S_b^2}{n} + \frac{S_w^2}{nsm} + \frac{S_t^2}{nsmk}\right) + c^2\left(\frac{S_b^2}{n} + \frac{S_w^2}{nrm} + \frac{S_t^2}{nrmk}\right) + (1-c)^2\left(\frac{S_b^2}{n} + \frac{S_w^2}{nsm} + \frac{S_t^2}{nsmk}\right) - 2(e+r)^2\frac{S_b^2}{n} + 2(e+r)c\left(\rho_b\frac{S_b^2}{n} + \rho_w\frac{S_w^2}{nrm} + \rho_t\frac{S_t^2}{nrmk}\right) - 2(e+r)c\,\rho_b\frac{S_b^2}{n} + 2c(1-c)\frac{S_b^2}{n}$$

The optimum weights c_o and e_o that will minimize the linear function $\mathrm{Var}[\Delta_{(1)}] + \lambda\,\mathrm{Var}[\bar{y}_{2(1)}]$ are obtained by solving the following equations for c_o and e_o:

$$\frac{\partial}{\partial c}\,\mathrm{Var}(\Delta_{(1)}) + \lambda\,\mathrm{Var}(\bar{y}_{2(1)}) = 0$$

$$\frac{\partial}{\partial e}\,\mathrm{Var}(\Delta_{(1)}) + \lambda\,\mathrm{Var}(y_{2(1)}) = 0$$

The optimum values of weights are

$$c_o = \frac{r\left(S_w^2 + \frac{S_t^2}{k}\right)}{\left(S_w^2 + \frac{S_t^2}{k}\right) - s\left(\rho_w S_w^2 + \rho_t\frac{S_t^2}{k}\right)} - \frac{\lambda}{(1+\lambda)}\,\frac{rs\left(S_w^2 + \frac{S_t^2}{k}\right)\left(\rho_w S_w^2 + \rho_t\frac{S_t^2}{k}\right)}{\left[\left(S_w^2 + \frac{S_t^2}{k}\right)^2 - s^2\left(\rho_w S_w^2 + \rho_t\frac{S_t^2}{k}\right)^2\right]}$$

and

$$e_o = \frac{-r\left(S_w^2 + \frac{S_t^2}{k}\right)}{\left(S_w^2 + \frac{S_t^2}{k}\right) - s\left(\rho_w S_w^2 + \rho_t\frac{S_t^2}{k}\right)} + \frac{\lambda}{(1+\lambda)}\,\frac{rs^2\left(\rho_w S_w^2 + \rho_t\frac{S_t^2}{k}\right)^2}{\left[\left(S_w^2 + \frac{S_t^2}{k}\right)^2 - s^2\left(\rho_w S_w^2 + \rho_t\frac{S_t^2}{k}\right)^2\right]}$$

The joint estimators of $\bar{Y}_2$ and $\bar{Y}_2 - \bar{Y}_1$ with optimum weights, are given by

$$\bar{y}_{2(1)} = \frac{-rs\beta_o}{(\alpha_o^2 - s^2\beta_o^2)}\left(\alpha_o + \frac{s\beta_o}{1+\lambda}\right)[\bar{y}'_{1(1)} - \bar{y}^*_{1(1)}] + \frac{r\alpha_o}{(\alpha_o^2 - s^2\beta_o^2)}\left(\alpha_o + \frac{s\beta_o}{1+\lambda}\right)[\bar{y}'_{2(1)} - \bar{y}^*_{2(1)}] + \bar{y}^*_{2(1)}$$

$$\Delta_{(1)} = \frac{-r}{(\alpha_o^2 - s\beta_o^2)}\left[\alpha_o(\alpha_o + s\beta_o) - \frac{\lambda\, s\beta_o^2}{(1+\lambda)}\right][\bar{y}'_{1(1)} - \bar{y}^*_{1(5)}] + \frac{r\,\alpha_o}{(\alpha_o^2 - s\beta_o^2)}\left(\alpha_o + \frac{s\beta_o}{(1+\lambda)}\right)[\bar{y}'_{2(1)} - \bar{y}^*_{2(1)}] + \bar{y}^*_{2(1)} - \bar{y}^*_{1(1)}$$

where

$$\alpha_o = S_w^2 + \frac{S_t^2}{k} \text{ and } \beta_o = \rho_w S_w^2 + \rho_t\frac{S_t^2}{k}$$

For the special case of $\lambda = 1$, the variances of the joint estimators with optimum weights are reduced to

$$\mathrm{Var}[\bar{y}_{2(1)}] = \frac{S_b^2}{n} + \frac{1}{nsm}(S_w^2 + \frac{S_t^2}{k})$$

$$- \frac{r(S_w^2 + \frac{S_t^2}{k})[(S_w^2 + \frac{S_t^2}{k})^2 - \frac{s^2}{4}(\rho_w S_w^2 + \rho_t \frac{S_t^2}{k})^2]}{nsm[(S_w^2 + \frac{S_t^2}{k})^2 - s^2(\rho_w S_w^2 + \rho_t \frac{S_t^2}{k})^2]}$$

and

$$\mathrm{Var}[\Delta_{(1)}] = \frac{2}{n}(1-\rho_b)S_b^2 + \frac{2}{nsm}(S_w^2 + \frac{S_t^2}{k}) + r(S_w^2 + \frac{S_t^2}{k})$$

$$[s^2(\rho_w S_w^2 + \rho_t \frac{S_t^2}{k}) - 8(S_w^2 + \frac{S_t^2}{k})\{(1+s\rho_w)S_w^2$$

$$+ (1+s\rho_t)\frac{S_t^2}{k}\}][4nsm\{(S_w^2 + \frac{S_t^2}{k})^2 - s^2(\rho_w S_w^2$$

$$+ \rho_t \frac{S_t^2}{k})^2\}]^{-1}$$

2.2 Joint Estimates of Mean and Change by Procedure (2) - Under this sampling plan, a partial replacement of units is carried out at second and third stage. Retain all the first-stage units from the first sample but retain only a fraction r of the PSU's retained and make a fresh random selection of the remaining fraction s of SSU's (such that r + s = 1) on the second occasion. Within each of the nrm SSU's retained, further retain only a fraction t of TSU's, and supplement the remaining fraction u of TSU's selected at random so that t + u = 1.

An initial estimate of $\bar{Y}_1$, based on first sample only, is given by

$$\bar{y}_{1(2)} = rt\ \bar{y}'_{1(2)} + ru\ \bar{y}^{**}_{1(2)} + s\ \bar{y}^{*}_{1(2)}$$

The joint linear and unbiased estimators of $\bar{Y}_2$ and $\bar{Y}_2 - \bar{Y}_1$ may be written as

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

and

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

$$+ (1-f-g)[\bar{y}^{*}_{2(7)} - \bar{y}_{1(2)}]$$

where

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

$$\bar{y}^{**}_{h(2)} = \frac{1}{nrmuk} \sum_{i=1}^{n} \sum_{j=1}^{rm} \sum_{\ell=1}^{k} y_{ij\ell}, \quad h = 1, 2$$

$$\bar{y}^{*}_{h(2)} = \frac{1}{nsmk} \sum_{i=1}^{n} \sum_{j=1}^{sm} \sum_{\ell=1}^{k} y_{ij\ell}, \quad h = 1, 2$$

If we impose the condition that $\Delta_{(2)} = \bar{y}_{2(2)} - \bar{y}_{1(2)}$ then by comparing coefficients on both sides it follows that a = rt - d, b = ru - e and f = d, g = e. So the joint linear unbiased estimators can be rewritten as

$$\bar{y}_{2(2)} = (rt - d)\bar{y}'_{1(2)} + (ru - e)\bar{y}^{**}_{1(2)}$$
$$- (r - d - e)\bar{y}^{*}_{1(2)} + d\bar{y}'_{2(2)} + e\bar{y}^{**}_{2(2)}$$
$$+ (1 - d - e)\bar{y}^{*}_{2(2)}$$

$$\Delta_{(2)} = d[\bar{y}'_{2(2)} - \bar{y}'_{1(2)}] + e[\bar{y}^{**}_{2(2)} - \bar{y}^{**}_{1(2)}]$$
$$+ (1 - d - e)[\bar{y}^{*}_{2(2)} - \bar{y}^{*}_{1(2)}] \quad (2.2.1)$$

and their variances are

$$\mathrm{Var}\ y_{2(2)} = [(rt - d)^2 + d^2]\alpha' + [(ru - e)^2 + e^2]\beta'$$
$$+ [(r - d - e)^2 + (1 - d - e)^2]\gamma'$$
$$+ 2[(rt - d)(ru - e) + de]\alpha^*$$
$$+ 2(rt - d)d\ \delta' + 2[(r - d - e)e$$
$$+ (ru - e)d]\delta^* + 2(1 - d - e)(d + e)$$
$$- (r - d - e)^2 \frac{S_b^2}{n} - 2(d + e)$$
$$(r - d - e)\rho_b \frac{S_b^2}{n}$$

and

$$\mathrm{Var}\ \Delta_{(2)} = 2d^2(\alpha' - \delta') + 2e^2(\beta' - \delta^*) + 2(1 - d$$
$$- e)^2(\gamma' - \rho_b \frac{S_b^2}{n}) + 4de[(1 - \rho_b)\frac{S_b^2}{n}$$
$$+ (1 - \rho_w)\frac{S_w^2}{nrm}] + 4d(1 - d - e)(1 - \rho_b)\frac{S_b^2}{n}$$
$$+ 4e(1 - d - e)(1 - \rho_b)\frac{S_b^2}{n} \quad (2.2.2)$$

where

$$\alpha' = \frac{S_b^2}{n} + \frac{S_w^2}{nrm} + \frac{S_t^2}{nrmtk}, \quad \beta' = \frac{S_b^2}{n} + \frac{S_w^2}{nrm} + \frac{S_t^2}{nrmuk}$$

$$\delta' = \rho_b \frac{S_b^2}{n} + \rho_w \frac{S_w^2}{nrm} + \rho_t \frac{S_t^2}{nrmtk}, \quad \alpha^* = \frac{S_b^2}{n} + \frac{S_w^2}{nrm}$$

and $\delta^* = \rho_b \frac{S_b^2}{n} + \rho_w \frac{S_w^2}{nrm}$. The optimum weights that will minimize the linear function $\mathrm{Var}[y_{2(2)}] + \mathrm{Var}[\Delta_{(2)}]$ are

$$d_o = \frac{rt[(3+(1-s\rho_w)(1-u\rho_t))S_w^2 + (4-(s+ru)\rho_t)\frac{S_t^2}{k}]}{[(1-u\rho_t)(1-s\rho_w)S_w^2 + (1-(s+ru)\rho_t)\frac{S_t^2}{k}]}$$

and e_o =

$$\frac{ru[(3(1-\rho_t)+(1-s\rho_w)(1-u\rho_t))S_w^2 + (4-3\rho_t-(s+ru)\rho_t)\frac{S_t^2}{k}]}{[(1-u\rho_t)(1-s\rho_w)S_w^2 + (1-(s+ru)\rho_t)\frac{S_t^2}{k}]}$$

By substituting the optimum values of the weights in equations (2.2.1) and (2.2.2) the joint linear unbiased estimators mean and change and their variances with optimum weights can be obtained.

2.3 Joint Estimates of Mean and Change by Procedure (3) - In this plan, only a fraction p of the PSU's along with their samples of SSU's and TSU's from the first sample is retained and a fresh random selection of the remaining fraction q of PSU's is made on the second occasion. Note that p + q = 1 so that the sample size remains same on the two occasions.

Using data from first sample only, a linear unbiased estimate of $\bar{Y}_1$ is

$$\bar{y}_{1(3)} = p\,\bar{y}'_{1(3)} + q\,\bar{y}''_{1(3)}$$

The joint linear unbiased estimators of $\bar{Y}_2$ and $\bar{Y}_2 - \bar{Y}_1$ subject to the constraint $\Delta_{(3)} = \bar{y}_{2(3)} - \bar{y}_{1(3)}$ may be expressed as

$$\bar{y}_{2(3)} = (e+p)\bar{y}'_{1(3)} - (e+p)\bar{y}''_{1(3)} + c\bar{y}'_{2(3)} + (1-c)\bar{y}''_{2(3)}$$

and

$$\Delta_{(3)} = e\bar{y}'_{1(3)} - (1+e)\bar{y}''_{1(3)} + c\bar{y}'_{2(3)} + (1-c)\bar{y}''_{2(3)} \qquad (2.3.1)$$

where

$$\bar{y}'_{h(3)} = \frac{1}{npmk}\sum_{i=1}^{np}\sum_{j=1}^{m}\sum_{\ell=1}^{k} y_{ij\ell}\,,\ h = 1,\ 2$$

$$\bar{y}''_{h(3)} = \frac{1}{nqmk}\sum_{i=1}^{nq}\sum_{j=1}^{m}\sum_{\ell=1}^{k} y_{ij\ell}\,,\ h = 1,\ 2$$

The variances of joint estimators are

$$\mathrm{Var}[\bar{y}_{2(3)}] = \frac{(e+p)^2}{np}[S_b^2+\frac{S_w^2}{m}+\frac{S_t^2}{mk}] + \frac{(e+p)^2}{nq}[S_b^2+\frac{S_w^2}{m}+\frac{S_t^2}{mk}]$$
$$+ \frac{c^2}{np}[S_b^2+\frac{S_w^2}{m}+\frac{S_t^2}{mk}] + \frac{(1-c)^2}{nq}[S_b^2+\frac{S_w^2}{m}+\frac{S_t^2}{mk}]$$
$$+ \frac{2(e+p)c}{np}[\rho_b S_b^2 + \rho_w\frac{S_w^2}{m} + \rho_t\frac{S_t^2}{mk}]$$

and

$$\mathrm{Var}[\Delta_{(3)}] = \frac{e^2}{np}[S_b^2 + \frac{S_w^2}{m} + \frac{S_t^2}{mk}] + \frac{(1+e)^2}{nq}[S_b^2 + \frac{S_w^2}{m}$$
$$+ \frac{S_t^2}{mk}] + \frac{c^2}{np}[S_b^2 + \frac{S_w^2}{m} + \frac{S_t^2}{mk}] + \frac{(1-c)^2}{nq}[S_b^2$$
$$+ \frac{S_t^2}{mk}] + \frac{2ec}{np}[\rho_b S_b^2 + \rho_w\frac{S_w^2}{m} + \rho_t\frac{S_t^2}{mk}] \qquad (2.3.2)$$

The optimum weights that will minimize a linear function $\mathrm{Var}[\Delta_{(3)}] + \mathrm{Var}[y_{2(3)}]$ are obtained in the usual manner.

$$c_o = \frac{p\alpha}{\alpha - q\delta} - \frac{pq\alpha\delta}{2(\alpha^2-q^2\delta^2)}$$

and

$$e_o = \frac{-p\alpha}{\alpha - q\delta} + \frac{pq^2\delta^2}{2(\alpha^2-q^2\delta^2)} \qquad (2.3.3)$$

where

$$\alpha = S_b^2 + \frac{S_w^2}{m} + \frac{S_t^2}{mk} \text{ and } \delta = \rho_b S_b^2 + \rho_w\frac{S_w^2}{m} + \rho_t\frac{S_t^2}{mk}$$

From equations (2.3.2) and (2.3.3) the optimum variances of the joint estimators are given by

$$\mathrm{Var}[y_{2(3)}] = \frac{\alpha}{4n}\,\frac{(4\alpha^2-3q\delta^2-q^2\delta^2)}{(\alpha^2-q^2\delta^2)}$$

and

$$\mathrm{Var}[\Delta_{(3)}] = \frac{\alpha}{4n}\,\frac{(8\alpha^2-8q\delta^2+pq\delta^2-8p\alpha\delta)}{(\alpha^2-q^2\delta^2)}$$

From equations (2.3.1) and (2.3.3) the joint estimators of mean and change with optimum weights can be obtained.

2.4 Joint Estimates of Mean and Change by Procedure (4) - Partial replacement of units at primary as well as secondary stage is considered in this procedure. Only a fraction p of the PSU's from the first sample is retained on the second occasion and the rest of the fraction q of PSU's is selected anew. Within each of the np PSU's retained, only a fraction r of SSU's with their samples of TSU's is retained and the remaining fraction s of SSU's is selected afresh in a random manner. Again, p + q = 1 and r + s = 1 so as to keep the same sample size on the two occasions.

An initial estimate of $\bar{Y}_2$ based on first sample only is given by

$$\bar{y}_{1(4)} = pr\,\bar{y}'_{1(4)} + ps\,\bar{y}^*_{1(4)} + q\,\bar{y}''_{1(4)}$$

Subject to the constraint $\Delta_{(4)} = \bar{y}_{2(4)} - \bar{y}_{1(4)}$ the joint linear unbiased estimators of $\bar{Y}_2$ and $\bar{Y}_2 - \bar{Y}_1$ may be written as

$$\bar{y}_{2(4)} = (pr-d)\bar{y}'_{1(4)} + (ps-e)\bar{y}^*_{1(4)} - (p-d-e)\bar{y}''_{1(4)}$$
$$+ d\,\bar{y}'_{2(4)} + e\,\bar{y}^*_{2(4)} + (1-d-e)\,\bar{y}''_{2(4)}$$

and

$$\Delta_{(4)} = d[\bar{y}'_{2(4)} - \bar{y}'_{1(4)}] + e[\bar{y}^*_{2(4)} - \bar{y}^*_{1(4)}]$$
$$+ (1-d-e)[\bar{y}''_{2(4)} - \bar{y}''_{1(4)}] \qquad (2.4.1)$$

The variances of the joint estimators in (2.4.1) are given by

$$\mathrm{Var}[y_{2(4)}] = [(pr-d)^2+d^2]\alpha'' + [(ps-e)^2+e^2]\beta''$$
$$+ [(p-d-e)^2+(1-d-e)^2]\alpha'' + 2(pr-d)d\,\delta''$$
$$+ 2\,\frac{S_b^2}{np}[(pr-d)(ps-e) + (pr-d)e\,\rho_b$$
$$+ (ps-e)d\,\rho_b + \rho_b + de]$$

and

$$\mathrm{Var}[\Delta_{(4)}] = 2d^2(\alpha'' - \delta'') + 2e^2(\beta'' - \rho_b\frac{S_b}{np})$$

$$+ 2(1 - d - e)^2\gamma'' + 4\, de(1 - \rho_b)\frac{S_b^2}{np}$$

(2.4.2)

where

$$\alpha'' = \frac{S_b^2}{np} + \frac{S_w^2}{nprm} + \frac{S_t^2}{nprmk}, \quad \beta'' = \frac{S_b^2}{np} + \frac{S_w^2}{npsm} + \frac{S_t^2}{npsmk}$$

$$\gamma'' = \frac{S_b^2}{nq} + \frac{S_w^2}{nqm} + \frac{S_t^2}{nqmk} \quad \text{and} \quad \delta'' = \frac{\rho_b S_b^2}{np} + \frac{\rho_w S_w^2}{nprm} + \frac{\rho_t S_t^2}{nprmk}$$

To determine the suitable values of weights d and e that will minimize the linear function $\mathrm{Var}[y_{2(4)}] + \mathrm{Var}[\Delta_{(4)}]$, it can be shown that the solution of the simultaneous equations

$$\frac{\partial}{\partial d}\{\mathrm{Var}\ y_{2(4)}] + \mathrm{Var}[\Delta_{(4)}]\} = 0$$

and

$$\frac{\partial}{\partial e}\{\mathrm{Var}\ y_{2(4)}] + \mathrm{Var}[\Delta_{(4)}]\} = 0$$

provides the following optimum values: $e_o =$

$$\frac{ps[S_b^2\{4(\alpha_o-\beta_o)-q\rho_b(\alpha_o-s\beta_o)+r\beta_o\}+\frac{\alpha_o}{m}\{4(\alpha_o-\beta_o)+pr\beta_o\}]}{4[S_b^2(1-q\rho_b)(\alpha_o-s\beta_o)+\frac{\alpha_o}{m}\{\alpha_o-\beta_o(q+ps)\}]}$$

and

$$d_o = \frac{pr[S_b^2\{4\alpha_o-q\rho_b(\alpha_o-s\beta_o)-s\beta_o\}+\frac{\alpha_o}{m}\{4\alpha_o-(q+ps)\beta_o\}]}{4[S_b^2(1-q\rho_b)(\alpha_o-s\beta_o)+\frac{\alpha_o}{m}\{\alpha_o-\beta_o(q+ps)\}]}$$

(2.4.3)

where α_o and β_o are defined in section 2.1. The joint estimators of mean and change with optimum weights can be obtained from equations (2.4.1) and (2.4.3) and their optimum variance can be obtained from equations (2.4.2) and (2.4.3).

2.5 Comparison - Combined estimates of mean and change have been obtained by four different sampling plans. It is important to find out which of the plans is more efficient. Relative performance of these plans is studied here for the same overall replacement fraction, say q^*. By equating total number of units replaced, it follows that $q^* = s + u - su$ for procedure (2) and $q^* = q + s - qs$ for procedure (4). From these preceding relations, it is seen that $q = u$. In case of procedure (1) and (3), it is obvious that $q^* = s$ and $q^* = q$ respectively. For convenience, the ratios S_w^2/S_b^2 and S_t^2/S_b^2 are denoted by ϕ and ψ respectively. It can be shown that for a three-stage sampling design to be useful, ϕ and ψ must satisfy the following conditions:

$$0 < \phi < M \quad \text{and} \quad 0 < \psi < K\,\phi$$

Let $\mathrm{RJM34} = \mathrm{Var}[\bar{y}_{2(4)}]/\mathrm{Var}[\bar{y}_{2(3)}]$ represent the relative efficiency of the jointly estimated mean by procedure (3) with respect to the joint estimate of mean by procedure (4). The symbols RJM31 and RJM32 have similar meanings. Similarly the symbols RJC14, RJC13 and RJC14 represent the relative efficiency of sampling procedure (1) with respect to procedures (4), (3) and (2) respectively in the combined estimation of change. The relative efficiencies of the four sampling plans are studied numerically for an arbitrarily selected range of values of the parameters and design quantities. Some of the results are arranged in Tables 1 through 4. Some important observations made from these tables are as follows:

(i) As ρ_b increases from 0.5 to 0.9, RJM34 increases moderately and gains in RJM32 and RJM31 are relatively more significant.

(ii) As ρ_w and ρ_t increase from 0.5 to 0.9, the changes produced in the values of RJM34, RJM31 and RJM32 are negligibly small.

(iii) When ϕ changes from 0.5 to 10, RJM34 remains practically unaltered, but RJM31 decreases moderately and RJM32 shows a slight decline with ϕ.

(iv) All three relative efficiencies register a slight decline as overall replacement fraction q^* changes from 0.65 to 0.85.

Some of the results from the numerical investigation of the joint estimates of change are presented in Tables 3 and 4. The following observations are made from these tables:

(i) RJC14 and RJC13 increase as ρ_b increases from 0.5 to 0.9, and RJC12 decreases slightly with ρ_b in most cases.

(ii) As ρ_w and ρ_t increase from 0.5 to 0.9, RJC13 increases slowly, however RJC12 and RJC14 remain almost unaltered.

(iii) All the three relative efficiencies decrease as ϕ increases from 0.5 to 10.

(iv) When q^* changes from 0.65 to 0.85, RJC13 shows significant gains but RJC12 and RJC14 remain practically unchanged.

2.6 Sample Allocation - In a design problem, it is important to study optimum allocation of sample. In the present section, optimum distribution of sample will be considered for a special case that is two-stage successive sampling. Assuming that travel cost among units is unimportant, one possible cost function for procedure (1) may be of the form

$$c = c_1 n + (c_2 + c_2' r + c_2'' s)nm$$

where c is total cost for two occasions, c_1 is the cost of preparing frame and c_2, c_2' and c_2'' are enumeration cost. The optimum values of m and n that will minimize the linear function $\mathrm{Var}[\bar{y}_{2(1)}] + \mathrm{Var}[\Delta_{(1)}]$ subject to the above cost function are

$$m_o = \frac{[12c_1S_w^2 - rc_1(12+8s\rho_w-2s^2\rho_w^2)S_w^2/(1-s^2\rho_w^2)]^{\frac{1}{2}}}{[4(c_2+c_2'r+c_2''s)sS_b^2(3-2\rho_b)]^{\frac{1}{2}}}$$

and

$$n_o = c[c_1+m_o(c_2+c_2'r+c_2''s)]^{-1}$$

A possible cost function for procedure (3) may be written as

$$c = (c_1+c_2'q)n + (c_2+c_2'p+c_2''q)nm$$

The optimum values of n and m that will minimize $Var[y_{2(3)}] + Var[\Delta_{(3)}]$ are obtained by method of successive approximation. The relative performance of sampling plans (1) and (3) are studied numerically and it is noted that procedure (1) is more efficient than procedure (3).

Table 1. $q = .5,\ s = .3,\ q^* = .65$

ρ_b	ρ_w	ρ_t	$\phi = .5$ x	y	z	$\phi = 10$ x	y	z
	.5	.5	102	105	105	102	103	104
.5	.7	.7	102	105	105	102	103	105
	.9	.9	102	105	105	102	101	106
	.5	.5	104	111	111	103	107	108
.7	.7	.7	104	111	112	104	107	110
	.9	.9	104	111	112	104	107	112
	.5	.5	109	125	125	105	112	113
.9	.7	.7	109	125	126	107	114	117
	.9	.9	109	126	126	108	116	122

Table 2. $q = .5,\ s = .7,\ q^* = 85$

ρ_b	ρ_w	ρ_t	$\phi = .5$ x	y	z	$\phi = 10$ x	y	z
	.5	.5	100	103	103	101	102	102
.5	.7	.7	100	103	103	100	101	101
	.9	.9	100	103	103	98	99	99
	.5	.5	100	107	107	101	104	104
.7	.7	.7	100	108	108	101	105	105
	.9	.9	100	107	107	100	103	104
	.5	.5	105	121	121	103	109	108
.9	.7	.7	106	121	122	104	111	111
	.9	.9	106	122	122	105	114	115

In Tables 1 and 2, x = RJM34, y = RJM31, z = RJM32, m = 16, k = 8, $\psi = 0.5$.

Table 3. $q = .5,\ s = .3,\ q^* = .65$

ρ_b	ρ_w	ρ_t	$\phi = .5$ x	y	z	$\phi = 10$ x	y	z
	.5	.5	133	147	99	118	126	90
.5	.7	.7	133	148	99	120	131	87
	.9	.9	134	150	94	131	149	88
	.5	.5	154	183	99	125	137	87
.7	.7	.7	154	184	98	126	141	83
	.9	.9	156	187	99	138	163	84
	.5	.5	189	247	96	131	147	83
.9	.7	.7	190	249	95	129	145	75
	.9	.9	195	260	96	137	164	72

Table 4. $q = .5,\ s = .7,\ q^* = .85$

ρ_b	ρ_w	ρ_t	$\phi = .5$ x	y	z	$\phi = 10$ x	y	z
	.5	.5	133	171	99	117	136	94
.5	.7	.7	133	171	99	118	140	90
	.9	.9	134	174	99	123	163	87
	.5	.5	154	237	99	126	155	93
.7	.7	.7	154	239	99	125	159	88
	.9	.9	155	245	98	129	187	83
	.5	.5	190	388	98	135	178	90
.9	.7	.7	189	388	96	131	173	83
	.9	.9	191	407	96	130	192	75

In Tables 3 and 4, x = RJC14, y = RJC13, z = RJC12, m = 16, k = 8, $\psi = 0.5$.

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 Amer. Stat. Assoc. Social Statistics Section, 1974, pp. 262-268.

[2] Rana, Dharam S. and R. P. Chakrabarty (1977). Three-stage Sampling on Successive Occasion. Submitted to Communications in Statistics.

[3] Hansen, M. H., Hurwitz, W. N. and Madow, W. Y. (1953). Sample Survey Methods and Theory. Vol. II. John Wiley and Sons, Inc., New York.

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

[5] Yates, F. (1960). Sampling Methods for Censuses and Surveys. Third edition, Charles Griffin and Co., Ltd., London.