Jour. Ind. Soc. Ag. Statistics
54(2), 2001 : 221–235

Mean Estimation in Deeply Stratified Population Under Post-Stratification

D. Shukla and Manish Trivedi
Dr. Hari Singh Gour University, Sagar, (M.P.), 470003
(Received : August, 1999)

SUMMARY

Suppose a population is stratified according to two attributes, each having three levels, in particular, then it constitutes a 3×3 deep stratification and interesting for survey practitioners being close to reality. This paper, presents the problem of mean estimation under above population, when frames of each 3×3 stratum are assumed to be unknown. An estimation strategy has been proposed using the post-stratified sampling scheme. The optimum properties are examined and relative efficiencies are compared. Mathematical finding is numerically supported.

Key words : Post-stratification, SRSWOR, Optimal, Deep stratification.

1. Introduction

We assume existence of a 3×3 deeply stratified population of size N in particular. Let Y_{ijk} be the k^{th} value of $(i, j)^{th}$ strata having size N_{ij} of a variable Y under study ($i = 1, 2, 3$; $j = 1, 2, 3$ and $k = 1, 2, \ldots, N_{ij}$). A random sample of size n is drawn by SRSWOR and post-stratified into n_{ij} units such that

$$n_{ij}\left(\sum_{i=1}^{3}\sum_{j=1}^{3} n_{ij} = n\right) \text{comes from } N_{ij}\left(\sum_{i=1}^{3}\sum_{j=1}^{3} N_{ij} = N\right)$$

Let $\overline{Y}_{ij}$ be the mean and S_{ij}^2 be the population mean square of $(i, j)^{th}$ strata. Also, $\overline{Y}$ and S^2 represent entire population mean and population mean square along-with $\overline{y}_{ij}$ and $\overline{y}$ as sample means based on n_{ij} and n units respectively.

Moreover, $N_{i.}\left(=\sum_{j=1}^{3} N_{ij}\right), N_{.j}\left(=\sum_{i=1}^{3} N_{ij}\right), n_{i.}\left(=\sum_{j=1}^{3} n_{ij}\right)$ and $n_{.j}\left(=\sum_{i=1}^{3} n_{ij}\right)$ are row and column totals and $\overline{Y}_{i.}, \overline{Y}_{.j}, \overline{y}_{i.}, \overline{y}_{.j}$ are population and sample means based on them respectively.

1.1 An Example

In an educational survey, students are classified as per their Academic-merit and Economic-background. Let an educational institution has P_1 proportion of meritorious students, P_2 average and P_3 below average students ($P_1 + P_2 + P_3 = 1$). Whereas same has P_4 proportion of economically poor students, P_5 from middle-class income and P_6 from above middle-income level ($P_4 + P_5 + P_6 = 1$). This constitutes 3 × 3 classification where P_m (m=1,2,....6) are known alongwith total strength of students in the institution but, each cell frequency and cell-frames are unknown. The survey practitioner wants to estimate the average monthly expenditure of students by an effective utilization of prior information on proportions P_m^s.

2. Derivation of Some Useful Theorems

With usual notations,

$$W_{ij} = \left(\frac{N_{ij}}{N}\right), W_{i.} = \left(\frac{N_{i.}}{N}\right), W_{.j} = \left(\frac{N_{.j}}{N}\right), p_{ij} = \left(\frac{n_{ij}}{n}\right), p_{i.} = \left(\frac{n_{i.}}{n}\right) \text{ and } p_{.j} = \left(\frac{n_{.j}}{n}\right)$$

assume sample n is large enough to support following

$$p_{ij} = W_{ij}\left(1+\varepsilon_{ij}\right), p_{i'j} = W_{i'j}\left(1+\varepsilon_{i'j}\right), p_{ij'} = W_{ij'}\left(1+\varepsilon_{ij'}\right) \tag{2.1}$$

where, $E\left[\varepsilon_{ij}\right] = E\left[\varepsilon_{i'j}\right] = E\left[\varepsilon_{ij'}\right] = 0;\ i \neq i' = 1,2,3;\ j \neq j' = 1,2,3$

$$E\left[\varepsilon_{ij}^2\right] = \left(\frac{1}{W_{ij}^2}\right)\left[\frac{(N-n)W_{ij}\left(1-W_{ij}\right)}{(N-1)n}\right]; E\left[\varepsilon_{i'j}^2\right] = \left(\frac{1}{W_{i'j}^2}\right)\left[\frac{(N-n)W_{i'j}\left(1-W_{i'j}\right)}{(N-1)n}\right]$$

$$E\left[\varepsilon_{ij'}^2\right] = \left(\frac{1}{W_{ij'}^2}\right)\left[\frac{(N-n)W_{ij'}\left(1-W_{ij'}\right)}{(N-1)n}\right]$$

$$E\left[\varepsilon_{i'j}\,\varepsilon_{ij'}\right]=\left\{\frac{-1}{W_{i'j}\,W_{ij'}}\right\}\left[\frac{(N-n)W_{ij'}\,W_{i'j}}{(N-1)n}\right]$$

$$E\left[\varepsilon_{i'j}\,\varepsilon_{ij}\right]=\left\{\frac{-1}{W_{ij}\,W_{i'j}}\right\}\left[\frac{(N-n)W_{ij}\,W_{i'j}}{(N-1)n}\right]$$

$$E\left[\varepsilon_{ij}\,\varepsilon_{ij'}\right]=\left\{\frac{-1}{W_{ij}\,W_{ij'}}\right\}\left[\frac{(N-n)W_{ij'}\,W_{ij}}{(N-1)n}\right]$$

2.1 Justification

For sample mean $\bar{y}$ based on n units and $E(\bar{y})=\bar{Y}$, Sukhatme *et al.* [2] have used one of approximations as $\bar{y}=\bar{Y}\,(1+\varepsilon)$, $E(\varepsilon)=0$, assuming sample size large and derived expressions of m.s.e. for ratio, product and regression estimators upto first and second order of approximations. If, in particular, for an attribute A in the same population, suppose

$Y_i = 1$ if i^{th} population unit possess A

$= 0$ otherwise

then $\bar{y}=w$, $E(w)=W$ holds where w and W are sample and population proportions respectively with respect to A. Without loss of generality, one can write $w=W\,(1+\varepsilon'),E(\varepsilon')=0$ for a large n.

Theorem 2.1 : Using (2.1) and avoiding terms of higher order, an approximate result, for $j\neq j'$, is

$$A_{ij(j')}=E\left[\frac{p_{ij'}}{p_{ij}}\right]=\frac{W_{ij'}}{W_{ij}}\left[1+\frac{Var(p_{ij})}{W_{ij}^2}-\frac{Cov(p_{ij}\,p_{ij'})}{W_{ij}\,W_{ij'}}\right]$$

Proof:

$$E\left[\frac{p_{ij'}}{p_{ij}}\right]=E\left[\frac{W_{ij'}\,(1+\varepsilon_{ij'})}{W_{ij}\,(1+\varepsilon_{ij})}\right]=\frac{W_{ij'}}{W_{ij}}E\left[1+\varepsilon_{ij'}-\varepsilon_{ij}-\varepsilon_{ij}\,\varepsilon_{ij'}+\varepsilon_{ij}^2+\varepsilon_{ij}^2\,\varepsilon_{ij'}\ldots\right]$$

Avoiding all higher order terms $\left[(\varepsilon_{ij})^r\,(\varepsilon_{ij'})^s\right]$ for $(r+s)>2$, theorem holds.

Theorem 2.2 : Using (2.1), an approximate result, for $j \neq j'$ is

$$B_{ij(j')} = E\left[\frac{p_{ij'}^2}{p_{ij}}\right] = \frac{W_{ij'}^2}{W_{ij}}\left[1 + \frac{Var(p_{ij})}{W_{ij}^2} + \frac{Var(p_{ij'})}{W_{ij'}^2} - \frac{2\,Cov(p_{ij}\,p_{ij'})}{W_{ij}\,W_{ij'}}\right]$$

Proof :

$$E\left[\frac{p_{ij'}^2}{p_{ij}}\right] = E\left[\frac{W_{ij'}^2\left(1+\varepsilon_{ij'}\right)^2}{W_{ij}\left(1+\varepsilon_{ij}\right)}\right] = \frac{W_{ij'}^2}{W_{ij}} E\left[1+\varepsilon_{ij'}^2 + 2\varepsilon_{ij'} - \varepsilon_{ij} - 2\varepsilon_{ij}\,\varepsilon_{ij'} + \varepsilon_{ij}^2 + \ldots\right]$$

Avoiding $\left[\left(\varepsilon_{ij}\right)^r\left(\varepsilon_{ij'}\right)^s\right]$ for $(r+s)>2$, theorem holds.

Theorem 2.3 : Using (2.1), an approximate result, for $i \neq i', j \neq j'$, is

$$C_{ij(i'j')} = E\left[\frac{p_{ij'}\,p_{i'j}}{p_{ij}}\right] = \frac{W_{ij'}\,W_{i'j}}{W_{ij}}\left[1 + \frac{Var(p_{ij})}{W_{ij}^2} + \frac{Cov(p_{i'j}\,p_{ij'})}{W_{i'j}\,W_{ij'}} - \frac{Cov(p_{ij}\,p_{i'j})}{W_{ij}\,W_{i'j}} - \frac{Cov(p_{ij}\,p_{ij'})}{W_{ij}\,W_{ij'}}\right]$$

Proof: $E\left[\frac{P_{i'j}\,P_{ij'}}{P_{ij}}\right] = E\left[\frac{W_{i'j}\left(1+\varepsilon_{i'j}\right)W_{ij'}\left(1+\varepsilon_{ij'}\right)}{W_{ij}\left(1+\varepsilon_{ij}\right)}\right]$

$$= \frac{W_{i'j}\,W_{ij'}}{W_{ij}} E\left[1+\varepsilon_{i'j} + \varepsilon_{ij'} + \varepsilon_{i'j}\,\varepsilon_{ij'} - \varepsilon_{ij} - \varepsilon_{ij}\,\varepsilon_{i'j} - \varepsilon_{ij}\,\varepsilon_{ij'} + \varepsilon_{ij}^2 \ldots\right]$$

On avoiding terms $\left[\left(\varepsilon_{ij}\right)^r\left(\varepsilon_{ij'}\right)^s\left(\varepsilon_{i'j}\right)^t\right]$ for $(r+s+t)>2$, we get result.

2.2 Some Symbols

$$D_{ij} = E\left[\frac{1}{n_{ij}}\right] = \frac{1}{nW_{ij}} + \frac{(N-n)\left(1-W_{ij}\right)}{(N-1)\,n^2 W_{ij}^2}$$

$$F_{i.} = E\left[\frac{p_{i.}^2}{N_{ij}}\right] = \left(\frac{1}{N_{ij}}\right)\left\{\frac{(N-n)}{(N-1)}\frac{W_{i.}(1-W_{i.})}{n} + W_{i.}^2\right\}$$

$$F_{.j} = E\left[\frac{p_{.j}^2}{N_{ij}}\right] = \left(\frac{1}{N_{ij}}\right)\left\{\frac{(N-n)}{(N-1)}\frac{W_{.j}(1-W_{.j})}{n} + W_{.j}^2\right\}; M_i = \sum_{j=1}^{3}\overline{Y}_{ij}$$

$$F_{ij} = E\left[\frac{p_{i.}p_{.j}}{N_{ij}}\right] = \left(\frac{1}{N_{ij}}\right)\left\{Cov\left(p_{i.}\,p_{.j}\right) + E(p_{i.})E(p_{.j})\right\}; M_j = \sum_{i=1}^{3}\overline{Y}_{ij}$$

3. Proposed Estimation Strategy

To recall assumptions are **(a)** a setup of 3 × 3 deeply stratified population N **(b)** frame of N units available for non-stratifying variable **(c)** sample size n is large (d) stratum sizes N_{ij} are unknown but information about $N_{i.}$ and $N_{.j}$ are known by some other sources.

To estimate $\overline{Y}$ a Deeply stratified Post-stratified estimator is

$$\overline{y}_{dps} = \sum_{i=1}^{3}\sum_{j=1}^{3} W_{\alpha ij}\,\overline{y}_{ij} \qquad (3.1)$$

where $W_{\alpha ij} = \left[\left(\frac{\alpha}{2}\right)\left\{\left(\frac{n_{i.}}{n}\right)+\left(\frac{N_{i.}}{N}\right)\right\}+\left(\frac{1-\alpha}{2}\right)\left\{\left(\frac{n_{.j}}{n}\right)+\left(\frac{N_{.j}}{N}\right)\right\}\right]$

The constant α be suitably chosen such that $0 \le \alpha \le 1$.

3.1 Motivation

I. The usual post-stratified estimator for a 3× 3 set-up is

$$\overline{y}_{ps} = \sum_{i=1}^{3}\sum_{j=1}^{3} W_{ij}\,\overline{y}_{ij} \qquad (3.2)$$

with $W_{ij} = \left(\frac{N_{ij}}{N}\right)$ which essentially requires a knowledge of N_{ij}

II. When only information of $N_{i.}$ and $N_{.j}$ available but not N_{ij}, the usual estimator (3.2) fails to perform estimation.

III. The information $N_{i.}$ and $N_{.j}$ are more common to be priorly known.

IV. An effective utilization of known $N_{i.}$ and $N_{.j}$ for estimation of $\overline{Y}$, is required.

V. A contribution by Agrawal and Panda [1] supports for choosing $W_{\alpha ij}$ in the present form.

3.2 Properties of Strategy

(I) At $\alpha = 1$, estimator $(\bar{y}_{dps})_1$ with $W_{1ij} = \left(\frac{1}{2}\right)\left[\frac{n_{i.}}{n} + \frac{N_{i.}}{N}\right]$

(II) At $\alpha = 0$, estimator $(\bar{y}_{dps})_0$ with $W_{0ij} = \left(\frac{1}{2}\right)\left[\frac{n_{.j}}{n} + \frac{N_{.j}}{N}\right]$

(III) At $\alpha = \frac{1}{2}$, estimator $(\bar{y}_{dps})_{1/2}$ with

$$W_{1/2ij} = \left(\frac{1}{4}\right)\left[\left\{\frac{n_{i.}}{n} + \frac{N_{i.}}{N}\right\} + \left\{\frac{n_{.j}}{n} + \frac{N_{.j}}{N}\right\}\right]$$

We have $(\bar{y}_{dps})_1$ purely based on row totals, $(\bar{y}_{dps})_0$ on column totals and $(\bar{y}_{dps})_{1/2}$ on an average of these two.

Theorem 3.1 : The estimator $\bar{y}_{dps}$ is biased for $\overline{Y}$.

Proof: Denote $E[(.)/n_{ij}]$ as a conditional expectation given n_{ij}

$$E\left(\bar{y}_{dps}\right) = E\left[E\left\{\frac{(\bar{y}_{dps})}{n_{ij}}\right\}\right] = E\left[\frac{\left\{\sum_{i=1}^{3}\sum_{j=1}^{3} W_{\alpha ij}\, E(\bar{y}_{ij})\right\}}{n_{ij}}\right]$$

$$= \sum_{i=1}^{3}\sum_{j=1}^{3} E(W_{\alpha ij})\overline{Y}_{ij} = \overline{Y} + \alpha V_1 + (1-\alpha)V_2$$

where, $V_1 = \sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{j'=1}^{3} W_{ij}\,\overline{Y}_{ij'}$; $V_2 = \sum_{i=1}^{3}\sum_{i'=1}^{3}\sum_{j=1}^{3} W_{ij}\,\overline{Y}_{i'j}$

Theorem 3.2: The Mean square error of $\overline{y}_{dps}$ is

$$M\left(\overline{y}_{dps}\right)=\left(\frac{1}{4}\right)\Big[(U_1+U_2)+\alpha^2\left(R_1+R_2+4V_1^2\right)+(1-\alpha)^2\left(S_1+S_2+4V_2^2\right)$$
$$+2\alpha(1-\alpha)(T_1+T_2+4V_1V_2)\Big]$$

where,

$$U_1=\left[\left(\frac{11}{n}\right)-\left(\frac{15}{N}\right)\right]\sum_{i=1}^{3}\sum_{j=1}^{3}W_{ij}S_{ij}^2,\ U_2=0 \text{ (considered for symmetry)}$$

$$R_1=\sum_{i=1}^{3}\sum_{j\neq j'=1}^{3}\left(\frac{1}{n}\right)B_{ij(j')}S_{ij}^2+\sum_{i=1}^{3}\sum_{j=1}^{3}W_{i.}^2D_{ij}S_{ij}^2+\sum_{i\neq i'=1}^{3}\sum_{j\neq j'=1}^{3}\left(\frac{2}{n}\right)C_{ij(i',j')}S_{ij}^2$$
$$+\left(\frac{2}{n}\right)\sum_{i=1}^{3}\sum_{j\neq j'=1}^{3}W_{i.}A_{ij(j')}S_{ij}^2-\sum_{i=1}^{3}\sum_{j=1}^{3}F_{i.}S_{ij}^2$$
$$-\left(\frac{3}{N}\right)\sum_{i=1}^{3}\sum_{j\neq j'=1}^{3}\left\{\frac{\left(W_{ij'}^2S_{ij}^2\right)}{W_{ij}}\right\}-\left(\frac{6}{N}\right)\sum_{i\neq i'=1}^{3}\sum_{j\neq j'=1}^{3}\left(\frac{W_{i'j}W_{ij'}}{W_{ij}}\right)S_{ij}^2$$

$$R_2=\left\{\frac{(N-n)}{(N-1)n}\right\}\left[\sum_{i=1}^{3}W_{i.}(1-W_{i.})M_i^2-\sum_{i=1}^{3}\sum_{i'=1}^{3}M_iM_{i'}W_{i.}W_{i'.}\right]$$

$$S_1=\sum_{i=1}^{3}\sum_{j\neq j'=1}^{3}\left(\frac{1}{n}\right)B_{ij(j')}S_{ij}^2+\sum_{i=1}^{3}\sum_{j=1}^{3}W_{.j}^2D_{ij}S_{ij}^2+\sum_{i\neq i'=1}^{3}\sum_{j\neq j'=1}^{3}\left(\frac{2}{n}\right)C_{ij(i',j')}S_{ij}^2$$
$$+\left(\frac{2}{n}\right)\sum_{i=1}^{3}\sum_{j\neq j'=1}^{3}W_{.j}A_{ij(j')}S_{ij}^2-\sum_{i=1}^{3}\sum_{j=1}^{3}F_{.j}S_{ij}^2-\left(\frac{3}{N}\right)\sum_{i=1}^{3}\sum_{j\neq j'=1}^{3}\left\{\frac{\left(W_{ij'}^2S_{ij}^2\right)}{W_{ij}}\right\}$$
$$-\left(\frac{6}{N}\right)\sum_{i\neq i'=1}^{3}\sum_{j\neq j'=1}^{3}\left(\frac{W_{i'j}W_{ij'}}{W_{ij}}\right)S_{ij}^2$$

$$S_2=\left\{\frac{(N-n)}{(N-1)n}\right\}\left[\sum_{j=1}^{3}W_{.j}(1-W_{.j})M_j^2-\sum_{j=1}^{3}\sum_{j'=1}^{3}M_jM_{j'}W_{.j}W_{.j'}\right]$$

$$T_1 = \sum_{i=1}^{3} \sum_{j=1}^{3} W_{i.} W_{.j} D_{ij} S_{ij}^2 + \sum_{i \neq i'=1}^{3} \sum_{j \neq j'=1}^{3} \left(\frac{1}{n}\right) C_{ij(i',j')} S_{ij}^2 - \sum_{i=1}^{3} \sum_{j=1}^{3} F_{ij} S_{ij}^2$$

$$+\left(\frac{1}{n}\right) \sum_{i=1}^{3} \sum_{j \neq j'=1}^{3} (W_{i.} + W_{.j}) A_{ij(j')} S_{ij}^2 - \left(\frac{3}{N}\right) \sum_{i \neq i'=1}^{3} \sum_{j \neq j'=1}^{3} \left(\frac{W_{i'j} W_{ij'}}{W_{ij}}\right) S_{ij}^2$$

$$T_2 = -\left\{\frac{(N-n)}{(N-1)n}\right\} \sum_{i=1}^{3} \sum_{j=1}^{3} W_{i.} W_{.j} M_i M_j$$

Proof: $M(\bar{y}_{dps}) = E\left[\frac{V(\bar{y}_{dps})}{n_{ij}}\right] + V\left[\frac{E(\bar{y}_{dps})}{n_{ij}}\right] + \left[\text{Bias}(\bar{y}_{dps})\right]^2$

$$E\left[\frac{V(\bar{y}_{dps})}{n_{ij}}\right] = E\left[\sum_{i=1}^{3} \sum_{j=1}^{3} \left(\frac{1}{n_{ij}}\right) W_{\alpha ij}^2 S_{ij}^2\right] - E\left[\sum_{i=1}^{3} \sum_{j=1}^{3} \left(\frac{1}{N_{ij}}\right) W_{\alpha ij}^2 S_{ij}^2\right] \quad (3.3)$$

For further derivation of (3.3) following are used

$$a_1 : E\left[\sum_{i=1}^{3} \sum_{j=1}^{3} \left(\frac{1}{n_{ij}}\right) \left\{\left(\frac{n_{i.}}{n}\right) + \left(\frac{N_{i.}}{N}\right)\right\}^2 S_{ij}^2\right] = \left(\frac{11}{n}\right) \sum_{i=1}^{3} \sum_{j=1}^{3} W_{ij} S_{ij}^2$$

$$+\left(\frac{1}{n}\right) \sum_{i=1}^{3} \sum_{j \neq j'=1}^{3} B_{ij(j')} S_{ij}^2 + \left(\frac{2}{n}\right) \sum_{i \neq i'=1}^{3} \sum_{j \neq j'=1}^{3} C_{ij(i',j')} S_{ij}^2$$

$$+\sum_{i=1}^{3} \sum_{j=1}^{3} W_{i.}^2 D_{ij} S_{ij}^2 + \left(\frac{2}{n}\right) \sum_{i=1}^{3} \sum_{j \neq j'=1}^{3} W_{i.} A_{ij(j')} S_{ij}^2$$

$$a_2 : E\left[\sum_{i=1}^{3} \sum_{j=1}^{3} \left(\frac{1}{n_{ij}}\right) \left\{\left(\frac{n_{.j}}{n}\right) + \left(\frac{N_{.j}}{N}\right)\right\}^2 S_{ij}^2\right] = \left(\frac{11}{n}\right) \sum_{i=1}^{3} \sum_{j=1}^{3} W_{ij} S_{ij}^2$$

$$+\left(\frac{1}{n}\right) \sum_{i=1}^{3} \sum_{j \neq j'=1}^{3} B_{ij(j')} S_{ij}^2 + \left(\frac{2}{n}\right) \sum_{i \neq i'=1}^{3} \sum_{j \neq j'=1}^{3} C_{ij(i',j')} S_{ij}^2$$

$$+\sum_{i=1}^{3} \sum_{j=1}^{3} W_{.j}^2 D_{ij} S_{ij}^2 + \left(\frac{2}{n}\right) \sum_{i=1}^{3} \sum_{j \neq j'=1}^{3} W_{.j} A_{ij(j')} S_{ij}^2$$

$$a_3 : E\left[\sum_{i=1}^{3}\sum_{j=1}^{3}\left(\frac{1}{n_{ij}}\right)\left\{\left(\frac{n_{i.}}{n}+\frac{N_{i.}}{N}\right)+\left(\frac{n_{.j}}{n}+\frac{N_{.j}}{N}\right)\right\}S_{ij}^2\right]=\left(\frac{11}{n}\right)\sum_{i=1}^{3}\sum_{j=1}^{3} W_{ij}S_{ij}^2$$

$$+\sum_{i=1}^{3}\sum_{j=1}^{3} W_{i.}W_{.j}D_{ij}S_{ij}^2+\left(\frac{1}{n}\right)\sum_{i\neq i'=1}^{3}\sum_{j\neq j'=1}^{3} C_{ij(i',j')}S_{ij}^2$$

$$+\left(\frac{1}{n}\right)\sum_{i=1}^{3}\sum_{j\neq j'=1}^{3}\left(W_{i.}+W_{.j}\right)A_{ij(j')}S_{ij}^2$$

To obtain results in a_1, a_2, a_3 theorems 2.1, 2.2 and 2.3 are used wherever required. With a_1, a_2, a_3, α and other terms the resultant expression is

$$E\left[\sum_{i=1}^{3}\sum_{j=1}^{3}\left(\frac{1}{n_{ij}}\right)W_{\alpha ij}^2 S_{ij}^2\right]$$

$$=\left(\frac{11}{4}\right)\sum_{i=1}^{3}\sum_{j=1}^{3} W_{ij}S_{ij}^2+\left(\frac{\alpha^2}{4}\right)\left[\left(\frac{1}{n}\right)\sum_{i=1}^{3}\sum_{j\neq j'=1}^{3} B_{ij(j')}S_{ij}^2\right.$$

$$+\left(\frac{2}{n}\right)\sum_{i\neq i'=1}^{3}\sum_{j\neq j'=1}^{3} C_{ij(i',j')}S_{ij}^2+\sum_{i=1}^{3}\sum_{j=1}^{3} W_{i.}^2 D_{ij}S_{ij}^2$$

$$\left.+\left(\frac{2}{n}\right)\sum_{i=1}^{3}\sum_{j\neq j'=1}^{3} W_{i.}A_{ij(j')}S_{ij}^2\right]+\left(\frac{(1-\alpha)^2}{4}\right)\left[\left(\frac{1}{n}\right)\sum_{i=1}^{3}\sum_{j\neq j'=1}^{3} B_{ij(j')}S_{ij}^2\right.$$

$$+\left(\frac{2}{n}\right)\sum_{i\neq i'=1}^{3}\sum_{j\neq j'=1}^{3} C_{ij(i',j')}S_{ij}^2+\sum_{i=1}^{3}\sum_{j=1}^{3} W_{.j}^2 D_{ij}S_{ij}^2$$

$$\left.+\left(\frac{2}{n}\right)\sum_{i=1}^{3}\sum_{j\neq j'=1}^{3} W_{.j}A_{ij(j')}S_{ij}^2\right]+\left(\frac{\alpha(1-\alpha)}{2}\right)\left[\left(\frac{1}{n}\right)\sum_{i\neq i'=1}^{3}\sum_{j\neq j'=1}^{3} C_{ij(i',j')}S_{ij}^2\right.$$

$$\left.+\sum_{i=1}^{3}\sum_{j=1}^{3} W_{i.}W_{.j}D_{ij}S_{ij}^2+\left(\frac{1}{n}\right)\sum_{i=1}^{3}\sum_{j\neq j'=1}^{3}\left(W_{i.}+W_{.j}\right)A_{ij(j')}S_{ij}^2\right] \quad (3.3.1)$$

We also have,

$$a_4 : E\left[\sum_{i=1}^{3}\sum_{j=1}^{3}\left(\frac{1}{N_{ij}}\right)\left\{\left(\frac{n_{i.}}{n}\right)+\left(\frac{N_{i.}}{N}\right)\right\}^2 S_{ij}^2\right]=\left(\frac{15}{N}\right)\sum_{i=1}^{3}\sum_{j=1}^{3} W_{ij}S_{ij}^2$$

$$+\sum_{i=1}^{3}\sum_{j=1}^{3} F_{i.}S_{ij}^2+\left(\frac{3}{N}\right)\sum_{i=1}^{3}\sum_{j\neq j'=1}^{3}\frac{W_{ij'}^2}{W_{ij}}S_{ij}^2+\left(\frac{6}{N}\right)\sum_{i\neq i'=1}^{3}\sum_{j\neq j'=1}^{3}\frac{W_{i'j}W_{ij'}}{W_{ij}}S_{ij}^2$$

$$a_5 : E\left[\sum_{i=1}^{3}\sum_{j=1}^{3}\left(\frac{1}{N_{ij}}\right)\left\{\left(\frac{n_{.j}}{n}\right)+\left(\frac{N_{.j}}{N}\right)\right\}^2 S_{ij}^2\right]=\left(\frac{15}{N}\right)\sum_{i=1}^{3}\sum_{j=1}^{3} W_{ij}S_{ij}^2$$

$$+\sum_{i=1}^{3}\sum_{j=1}^{3} F_{.j}S_{ij}^2+\left(\frac{3}{N}\right)\sum_{i=1}^{3}\sum_{j\neq j'=1}^{3}\frac{W_{ij'}^2}{W_{ij}}S_{ij}^2+\left(\frac{6}{N}\right)\sum_{i\neq i'=1}^{3}\sum_{j\neq j'=1}^{3}\frac{W_{i'j}W_{ij'}}{W_{ij}}S_{ij}^2$$

$$a_6 : E\left[\sum_{i=1}^{3}\sum_{j=1}^{3}\left(\frac{1}{N_{ij}}\right)\left\{\left(\frac{n_{i.}}{n}+\frac{N_{i.}}{N}\right)+\left(\frac{n_{.j}}{n}+\frac{N_{.j}}{N}\right)\right\}S_{ij}^2\right]=\left(\frac{15}{N}\right)\sum_{i=1}^{3}\sum_{j=1}^{3} W_{ij}S_{ij}^2$$

$$+\sum_{i=1}^{3}\sum_{j=1}^{3} F_{ij}S_{ij}^2+\left(\frac{3}{N}\right)\sum_{i\neq i'=1}^{3}\sum_{j\neq j'=1}^{3}\frac{W_{i'j}W_{ij'}}{W_{ij}}S_{ij}^2$$

Theorem 2.1 , 2.2 and 2.3 are also used to derive a_4, a_5, and a_6 and, we get

$$E\left[\sum_{i=1}^{3}\sum_{j=1}^{3}\left(\frac{1}{N_{ij}}\right)W_{\alpha ij}^2 S_{ij}^2\right]=$$

$$\left(\frac{15}{4N}\right)\sum_{i=1}^{3}\sum_{j=1}^{3} W_{ij}S_{ij}^2+\left(\frac{\alpha^2}{4}\right)\left[\sum_{i=1}^{3}\sum_{j=1}^{3} F_{i.}S_{ij}^2+\left(\frac{3}{N}\right)\sum_{i=1}^{3}\sum_{j\neq j'=1}^{3}\frac{W_{ij'}^2}{W_{ij}}S_{ij}^2\right.$$

$$\left.+\left(\frac{6}{N}\right)\sum_{i\neq i'=1}^{3}\sum_{j\neq j'=1}^{3}\frac{W_{i'j}W_{ij'}}{W_{ij}}S_{ij}^2\right]+\left(\frac{(1-\alpha)^2}{4}\right)\left[\sum_{i=1}^{3}\sum_{j=1}^{3} F_{.j}S_{ij}^2\right.$$

$$\left.+\left(\frac{3}{N}\right)\sum_{i=1}^{3}\sum_{j\neq j'=1}^{3}\frac{W_{ij'}^2}{W_{ij}}S_{ij}^2+\left(\frac{6}{N}\right)\sum_{i\neq i'=1}^{3}\sum_{j\neq j'=1}^{3}\frac{W_{i'j}W_{ij'}}{W_{ij}}S_{ij}^2\right]$$

$$+\left(\frac{\alpha(1-\alpha)}{2}\right)\left[\sum_{i=1}^{3}\sum_{j=1}^{3} F_{ij}S_{ij}^2+\left(\frac{3}{N}\right)\sum_{i\neq i'=1}^{3}\sum_{j\neq j'=1}^{3}\frac{W_{i'j}W_{ij'}}{W_{ij}}S_{ij}^2\right] \qquad (3.3.2)$$

$$V\left[\frac{E(\bar{y}_{dps})}{n_{ij}}\right] = V\left[\sum_{i=1}^{3}\sum_{j=1}^{3} W_{\alpha ij}\,\bar{Y}_{ij}\right]$$

$$= \left(\frac{\alpha^2}{4}\right)\sum_{i=1}^{3} V(p_{i.})\bar{Y}_{i.}^2 + \left(\frac{(1-\alpha)^2}{4}\right)\sum_{j=1}^{3} V(p_{.j})\bar{Y}_{.j}^2$$

$$+ \left(\frac{\alpha(1-\alpha)}{2}\right)\sum_{i=1}^{3}\sum_{j=1}^{3} Cov(p_{i.}, p_{.j})\bar{Y}_{i.}\bar{Y}_{.j}$$

$$= \left(\frac{\alpha^2}{4}\right)R_2 + \left\{\frac{(1-\alpha)^2}{4}\right\}S_2 + \left\{\frac{\alpha(1-\alpha)}{2}\right\}T_2 \qquad (3.4)$$

$$\left[Bias(\bar{y}_{dps})\right]^2 = \left[\alpha^2 V_1^2 + (1-\alpha)^2 V_2^2 + 2\alpha(1-\alpha)V_1 V_2\right] \qquad (3.5)$$

Use of (3.3.1), (3.3..2), (3.4) and (3.5) provides the proof of theorem.

4. Optimum Choice

$$\alpha_{opt} = \left[\frac{\left(S_1 + S_2 + 4V_2^2\right) - \left(T_1 + T_2 + 4V_1 V_2\right)}{\left(R_1 + R_2 + 4V_2^2\right) + \left(S_1 + S_2 + 4V_2^2\right) - 2\left(T_1 + T_2 + 4V_1 V_2\right)}\right]$$

$$M\left[\left(\bar{y}_{dps}\right)_{opt}\right] = \left(\frac{1}{4}\right)\left[\left(U_1 + U_2\right) + \left\{\frac{\left(R_1 + R_2 + 4V_1^2\right)\left(S_1 + S_2 + 4V_2^2\right) - \left(T_1 + T_2 + 4V_1 V_2\right)^2}{\left(R_1 + R_2 + 4V_1^2\right) + \left(S_1 + S_2 + 4V_2^2\right) - 2\left(T_1 + T_2 + 4V_1 V_2\right)}\right\}\right]$$

5. Efficiency Comparison

I. $\left(\bar{y}_{dps}\right)_1$ is efficient over $\left(\bar{y}_{dps}\right)_0$ if $\left(R_1 + R_2 + 4V_1^2\right) \le \left(S_1 + S_2 + 4V_2^2\right)$

II. $\left(\bar{y}_{dps}\right)_1$ is efficient over $\left(\bar{y}_{dps}\right)_{1/2}$ if

$$\left(R_1+R_2+4V_1^2\right)\leq\frac{1}{3}\left[\left(S_1+S_2+4V_2^2\right)+2\left(T_1+T_2+4V_1V_2\right)\right]$$

III. $\left(\bar{y}_{dps}\right)_0$ is efficient over $\left(\bar{y}_{dps}\right)_{1/2}$ if

$$\left(S_1+S_2+4V_2^2\right)\leq\frac{1}{3}\left[\left(R_1+R_2+4V_1^2\right)+2\left(T_1+T_2+4V_1V_2\right)\right]$$

6. *Numerical Illustrations*

Consider two populations of size N=650 and N=490 and samples of size 260 and 196 by SRSWOR respectively and post-stratified according to 3 × 3 classification.

Parameters of population are given in Table 6.1 and Table 6.2.

(a) For data set -I

$M\left[\left(\bar{y}_{dps}\right)_1\right]$ = 132.5416 $B\left[\left(\bar{y}_{dps}\right)_1\right]$ = 9.0127

$M\left[\left(\bar{y}_{dps}\right)_0\right]$ = 118.6221 $B\left[\left(\bar{y}_{dps}\right)_0\right]$ = 8.8626

$M\left[\left(\bar{y}_{dps}\right)_{1/2}\right]$ = 14.8312 $B\left[\left(\bar{y}_{dps}\right)_{1/2}\right]$ = 8.6274

$M\left[\left(\bar{y}_{dps}\right)_{opt}\right]$ = 7.0314 $B\left[\left(\bar{y}_{dps}\right)_{opt}\right]$ = 8.1721

with α_{opt} = 0.4613

For data set -II

$M\left[\left(\bar{y}_{dps}\right)_1\right]$ = 98.1312 $B\left[\left(\bar{y}_{dps}\right)_1\right]$ = 6.7285

$M\left[\left(\bar{y}_{dps}\right)_0\right]$ = 87.6234 $B\left[\left(\bar{y}_{dps}\right)_0\right]$ = 6.5578

$M\left[\left(\bar{y}_{dps}\right)_{1/2}\right]$ = 13.1394 $B\left[\left(\bar{y}_{dps}\right)_{1/2}\right]$ = 6.4432

$M\left[\left(\bar{y}_{dps}\right)_{opt}\right]$ = 4.3122 $B\left[\left(\bar{y}_{dps}\right)_{opt}\right]$ = 5.7055

with α_{opt} = 0.4913

(b) This is to recall that it is not possible to get estimate of $\bar{Y}$ from usual sample mean estimator since N_{ij}'s are assumed unknown.

(c) It seems that estimator $\left(\bar{Y}_{dps}\right)_{1/2}$ is more efficient than $\left(\bar{Y}_{dps}\right)_0$ and $\left(\bar{Y}_{dps}\right)_1$

Table 6.1 (for data set 1)

B \ A		Attribute A: Low	Attribute A: Medium	Attribute A: High	Total
Attribute B	Low	$N_{11} = 71, n_{11} = 28$ $\overline{Y}_{11} = 48.7464$ $W_{11} = 0.10923$ $S_{11}^2 = 857.187$	$N_{12} = 65, n_{12} = 26$ $\overline{Y}_{12} = 147.6923$ $W_{12} = 0.1$ $S_{12}^2 = 791.2476$	$N_{13} = 68, n_{13} = 27$ $\overline{Y}_{13} = 247.4264$ $W_{13} = 0.10461$ $S_{13}^2 = 876.9092$	$N_{1.} = 205$ $n_{1.} = 81$ $\overline{Y}_{1.} = 146.49$ $W_{1.} = 0.3138$
	Medium	$N_{21} = 77, n_{21} = 31$ $\overline{Y}_{21} = 346.8831$ $W_{21} = 0.11846$ $S_{21}^2 = 866.6208$	$N_{22} = 74, n_{22} = 30$ $\overline{Y}_{22} = 431.9054$ $W_{22} = 0.11384$ $S_{22}^2 = 964.5512$	$N_{23} = 80, n_{23} = 32$ $\overline{Y}_{23} = 549.525$ $W_{23} = 0.12307$ $S_{23}^2 = 829.1359$	$N_{2.} = 231$ $n_{2.} = 93$ $\overline{Y}_{2.} = 444.29$ $W_{2.} = 0.355$
	High	$N_{31} = 73, n_{31} = 29$ $\overline{Y}_{31} = 654.315$ $W_{31} = 0.1123$ $S_{31}^2 = 787.885$	$N_{32} = 70, n_{32} = 28$ $\overline{Y}_{32} = 737.957$ $W_{32} = 0.10769$ $S_{32}^2 = 1044.759$	$N_{33} = 72, n_{33} = 29$ $\overline{Y}_{33} = 846.597$ $W_{33} = 0.11076$ $S_{33}^2 = 780.469$	$N_{3.} = 215$ $n_{3.} = 86$ $\overline{Y}_{3.} = 745.93$ $W_{3.} = 0.3307$
Total		$N_{.1} = 221, n_{.1} = 88$ $\overline{Y}_{.1} = 352.6515$ $W_{.1} = 0.3399$	$N_{.2} = 209, n_{.2} = 84$ $\overline{Y}_{.2} = 446.01912$ $W_{.2} = 0.32153$	$N_{.3} = 220, n_{.3} = 88$ $\overline{Y}_{.3} = 553.3727$ $W_{.3} = 0.3384$	$N = 650$ $n = 260$ $\overline{Y} = 450.82$

Table 6.2 (for data set 2)

B \ A		Attribute A: Low	Attribute A: Medium	Attribute A: High	Total
Attribute B	Low	$N_{11} = 56, n_{11} = 22$ $\overline{Y}_{11} = 37.232$ $W_{11} = 0.1143$ $S^2_{11} = 504.1068$	$N_{12} = 50, n_{12} = 20$ $\overline{Y}_{12} = 113.02$ $W_{12} = 0.102$ $S^2_{12} = 434.947$	$N_{13} = 52, n_{13} = 21$ $\overline{Y}_{13} = 188.8846$ $W_{13} = 0.1061$ $S^2_{13} = 505.163$	$N_{1.} = 185$ $n_{1.} = 63$ $\overline{Y}_{1.} = 111.1265$ $W_{1.} = 0.3224$
	Medium	$N_{21} = 48, n_{21} = 19$ $\overline{Y}_{21} = 267.3333$ $W_{21} = 0.09796$ $S^2_{21} = 500.926$	$N_{22} = 62, n_{22} = 25$ $\overline{Y}_{22} = 321.258$ $W_{22} = 0.1265$ $S^2_{22} = 768.06$	$N_{23} = 58, n_{23} = 23$ $\overline{Y}_{23} = 413.776$ $W_{23} = 0.11836$ $S^2_{23} = 466.716$	$N_{2.} = 168$ $n_{2.} = 67$ $\overline{Y}_{2.} = 337.792$ $W_{2.} = 0.34282$
	High	$N_{31} = 54, n_{31} = 22$ $\overline{Y}_{31} = 483.037$ $W_{31} = 0.1102$ $S^2_{31} = 564.56$	$N_{32} = 60, n_{32} = 24$ $\overline{Y}_{32} = 553.666$ $W_{32} = 0.12245$ $S^2_{32} = 529.17$	$N_{33} = 50, n_{33} = 20$ $\overline{Y}_{33} = 625.000$ $W_{33} = 0.102$ $S^2_{33} = 562.53$	$N_{3.} = 164$ $n_{3.} = 66$ $\overline{Y}_{3.} = 552.1583$ $W_{3.} = 0.33465$
Total		$N_{.1} = 158, n_{.1} = 63$ $\overline{Y}_{.1} = 259.4999$ $W_{.1} = 0.32246$	$N_{.2} = 172, n_{.2} = 69$ $\overline{Y}_{.2} = 341.796$ $W_{.2} = 0.35095$	$N_{.3} = 160, n_{.3} = 64$ $\overline{Y}_{.3} = 406.694$ $W_{.3} = 0.32646$	$N = 490$ $n = 196$ $\overline{Y} = 336.451$

(d) The estimator $\left(\bar{y}_{dps}\right)$ has made possible to estimate $\bar{Y}$ in a 3×3 set-up even without the prior knowledge of N_{ij} and frames. It has an effective utilization of row and column totals $N_{i.}$ and $N_{.j}$.

(e) The estimator is found most efficient at optimal selection of $\alpha = 0.4613$ for set-I and $\alpha = 0.4913$ for set -II.

(f) On the basis of data considered herein, one can think of choosing α to a value near to 0.5 which reveals that almost a fifty percent fraction of row sum of size- proportions $\left[\left(\frac{n_{i.}}{n}\right)+\left(\frac{N_{i.}}{N}\right)\right]$ and rest fifty percent same from column generates an ideal, quick and easy choice of α. Thus, the proposed estimator provides an easy optimum choice $\alpha = \frac{1}{2}$ or very close to it.

REFERENCES

[1] Agarwal, M.C. and Panda, K.B. (1993). An efficient estimator in post-stratification. *Metron*, **5**, **3.4**, 179-187.

[2] Sukhatme, P.V., Sukhatme, B.V., Sukhatme, S. and Asok, C. (1984). *Sampling Theory of Surveys with Applications*. Iowa State University Press, Indian Society of Agricultural Statistics Publication, New Delhi.