Aligarh Journal of Statistics
Vol. 27 (2007), 43-59

ON EFFICIENT ESTIMATION IN *PSNR* SAMPLING SCHEME USING AUXILIARY INFORMATION

Meenakshi Srivastava and Priyanka Jain

ABSTRACT

The work presented in this paper is based on the work of Shukla and Dubey (2001), relating to mean estimation in mail surveys under *PSNR* (Post Stratified non response) sampling scheme. An alternative estimator has been proposed for estimating the population mean, under the similar set up, using auxiliary information. The bias and approximate variance of the estimator have been derived and compared with the estimator of Shukla and Dubey (2001). It is found that the bias of proposed estimator is very small and it is more efficient. The cost analysis is incorporated to estimate the optimal sample size. Theoretical derivations are numerically justified.

1. INTRODUCTION

The non-response in the mail surveys is a kind of incompleteness that occurs when respondents do not reply through their mails and the sample remains incomplete. In the set-up of stratified sampling, when stratum sizes are unknown, the post stratification is a useful strategy (see, Cochran (1999), Mukhopadhyay (1999) etc.).The non-response in surveys could be handled by various techniques and methodologies [see Lessler and Kalsbeek (1992), Grover and Couper (1998), Khare (1987) and Khot (1994)] and one of them is the imputation of available data as described by Hinde and Chambers (1991). The post-stratification, discussed by Hansen *et al.* (1953). Jager *et al.* (1985), advocated that with respect to relevant criteria, it may improve upon the estimation strategy subsequently over the sample mean or ratio estimator. Silva and Skinner (1995) used the technique of post-stratification for estimating distribution function with auxiliary information. Shukla and Trivedi (2001, 2006), Shukla *et al*. (2002) derived methodologies for parameter estimation in sampling under post-stratification. Some other useful contributions are by Wywial (2001), Smith (1991), Holt and Smith (1979) etc.

The mixture of post-stratification and non-response is due to Zhang (1999) who has obtained some results regarding the effect of post-stratification while handling binary survey data subject to non-response. Shukla and Dubey (2001) proposed *PSNR* sampling scheme for dealing with non-response. Shukla and Dubey (2004, 2006) have taken into consideration some new aspects related to similar problem.

This paper takes into account the use of auxiliary information, in the approach of Shukla and Dubey (2001), and suggests a new estimator under the *PSNR* sampling scheme. This one is more efficient than Shukla and Dubey`s estimator under the same set-up.

2. NOTATIONS

Let a population of size N be divided into k strata. The size of the $i-$th strata be N_i, $(i=1,\cdots,k)$, such that $\sum_{i=1}^{k} W_i = \sum_{i=1}^{k} \frac{N_i}{N} = 1$. In what follows, symbols are established below:

N_i' : Number of respondents (RS) in $i-$th strata.

N_i'' : Number of non respondents (NRS) in $i-$th strata.

$\overline{Y}_i'$: Population mean of $i-$th strata for response group (R) for Y.

$\overline{Y}_i''$: Population mean of $i-$th strata for non-response group (NR) for Y.

$\overline{Y}$: Population mean for Y, i.e. $\overline{Y} = \sum_{i=1}^{k} W_i \overline{Y}_i$.

S_{1iY}^2 : Population mean square in $i-$th strata for $R-$group (R) for Y.

S_{2iY}^2 : Population mean square in $i-$th strata for $NR-$group (NR) for Y.

C_{1iY} : Coefficient of variation in $i-$th strata for $R-$group (R) for Y.

C_{2iY} : Coefficient of variation in $i-$th strata for $NR-$group (NR) for Y.

We assume X an auxiliary variable, correlated with Y.

$\overline{X}_i'$: Population mean in $i-$th strata for response group (R) of X.

$\overline{X}_i''$: Population mean in $i-$th strata for non-response group (NR) of X.

S_{1iX}^2 : Population mean square in $i-$th strata for $R-$group for X.

S_{2iX}^2 : Population mean square in $i-$th strata for $NR-$group for X.

C_{1iX} : Coefficient of variation in $i-$th strata for $R-$group for X.

C_{2iX} : Coefficient of variation in $i-$th strata for $NR-$group for X.

ρ_i' : Correlation coefficient for $R-$group (between X_i' and Y_i').

ρ_i'' : Correlation coefficient for NR-group (between X_i'' and Y_i'').

R_i' : Ratio in $i-$th strata for $R-$group, i.e. $R_i' = \dfrac{\overline{Y}_i'}{\overline{X}_i'}$.

R_i'' : Ratio in $i-$th strata for $NR-$group, i.e. $R_i'' = \dfrac{\overline{Y}_i''}{\overline{X}_i''}$.

This is to note that $N_i = N_i' + N_i''$ and $n_i = n_i' + n_i''$ hold throughout what follows.

3. POST STRATIFIED NON-RESPONSE (*PSNR*) SAMPLING SCHEME WITH AUXILIARY VARIABLE

Following lines of Shukla and Dubey (2001) and using auxiliary information X, the $PSNR$ sampling scheme is described into following steps:

Step I: Select a sample of size n by $SRSWOR$ from the population N and post-stratified into k strata, such that n_i units represent to $N_i \left(\sum_{i=1}^{k} n_i = n \right)$. An auxiliary source of information (other than X) or prior guess may be used for this purpose.

Step II: Mail questionnaires to all the random n_i units for response over the variable Y under study and wait until a deadline. If possible complete response occurs, $\overline{y}_i$ is sample mean from $i-$th strata and $\overline{y} = (n)^{-1} \sum_{i=1}^{k} n_i \overline{y}_i$.

Step III: Assume that non-response observed when the deadline of returning questionnaire is over, and there are n_i' respondents, n_i'' non-respondents in the $i-$th strata $(n_i' + n_i'' = n_i)$. The $\overline{y}_i'$ is mean of responding n_i' units for the Y and $\overline{x}_i'$ is mean of corresponding n_i' units over auxiliary variable X. Moreover, $R_i = \dfrac{Y_i}{X_i}$ is the true ratio in the stratum i and it is assumed that respondents for Y must have responded for X also among n_i'.

Step IV: From non-responding n_i'', select sub-samples of size n_i''' by $SRSWOR$, maintaining a prefixed fraction $f_i = \left(\dfrac{n_i'}{n_i'''} \right)$ over all the k strata.

Step V: Conduct a personal interview for n_i''' units and assume all these responded well during that period over Y and X both. The $\overline{y}_i'''$ is the mean based on n_i''' and $\overline{x}_i'''$ is the corresponding mean of X .

4. THE PROPOSED ESTIMATOR

The proposed estimation strategy is

$$\bar{y}_{rPSNR} = \sum_{i=1}^{k} W_i \left[\left(\frac{n'_i}{n_i} \right) (\bar{y}'_{ir}) + \left(\frac{n''_i}{n_i} \right) (\bar{y}'''_{ir}) \right],$$

where $rPSNR$ stands for "ratio post-stratified non-response", $\bar{y}'_{ir}$ is the ratio estimate for mean of responding units of $i-$th strata in stratified random sampling [i.e. $\bar{y}'_{ir} = \frac{\bar{y}'_i}{\bar{x}'_i} \bar{X}'_i$] and $\bar{y}'''_{ir}$ is the ratio estimate for mean based on n'''_i units in stratified random sampling [i.e. $\bar{y}'''_{ir} = \frac{\bar{y}'''_i}{\bar{x}'''_i} \bar{X}''_i$], where $\bar{X}'_i$ and $\bar{X}''_i$ is assumed to be known.

Theorem 4.1: $\bar{y}_{rPSNR}$ is biased for $\bar{Y}$ and the approximate amount of bias is

$$\text{Bias}\,(\bar{y}_{rPSNR}) = \sum_i W_i \left(\frac{N'_i}{N_i} \right) \bar{Y}'_i \left(\frac{N-n}{Nn} \right) (C^2_{1iX} - \rho'_i C_{1iX} C_{1iY})$$

$$+ \sum_i W_i \left\{ \frac{1}{nW_i} + \frac{(N-n)(1-W_i)}{n^2 (N-1) W_i^2} \right\} \bar{Y}''_i (f_i - 1)(C^2_{2iX} - \rho''_i C_{2iX} C_{2iY})$$

Proof: The conditional expectation of $\bar{y}_{rPSNR}$ given n_i and (n'_i, n''_i) is denoted by

$$E(\bar{y}_{rPSNR}) = E\left[E\left[E\left\{ \sum_{i=1}^{k} W_i \left(\frac{n'_i}{n_i} \right) \bar{y}'_{ir} + \sum_{i=1}^{k} W_i \left(\frac{n''_i}{n_i} \right) \bar{y}'''_{ir} \right\} \middle| \, n_i \, (n'_i, n''_i) \right] \right]$$

$$= E\left[E\left[\sum_{i=1}^{k} W_i \left(\frac{n'_i}{n_i} \right) \bar{y}'_{ir} + \sum_{i=1}^{k} W_i \left(\frac{n''_i}{n_i} \right) \bar{y}''_i \right. \right.$$

$$\left. \left. \left\{ 1 + \frac{n''_i - n'''_i}{n''_i n'''_i} (C^2_{2iX} - \rho''_i C_{2iX} C_{2iY}) \right\} \middle| \, n_i \, (n'_i, n''_i) \right] \right]$$

$$= E\left[\sum_{i=1}^{k} W_i \left(\frac{N'_i}{N_i} \right) \bar{Y}'_i + \sum_{i=1}^{k} W_i \left(\frac{N''_i}{N_i} \right) \bar{Y}''_i + \sum_{i=1}^{k} W_i \left(\frac{N'_i}{N_i} \right) \bar{Y}'_i \left(\frac{N'_i - n'_i}{N'_i n'_i} \right) \right.$$

$$\left. (C^2_{1iX} - \rho'_i C_{1iX} C_{1iY}) + \sum_{i=1}^{k} W_i \left(\frac{1}{n_i} \right) \bar{Y}''_i (f_i - 1)(C^2_{2iX} - \rho''_i C_{2iX} C_{2iY}) / n_i \right]$$

$$= \bar{Y} + \sum_{i=1}^{k} W_i \left(\frac{N'_i}{N_i}\right) \bar{Y}'_i \left(\frac{N'_i - n'_i}{N'_i\, n'_i}\right) (C_{1iX}^2 - \rho'_i\, C_{1iX}\, C_{1iY})$$

$$+ \sum_{i=1}^{k} W_i \left\{ \frac{1}{n W_i} + \frac{(N-n)(1-W_i)}{n^2 (N-1) W_i^2} \right\} \bar{Y}''_i (f_i - 1)(C_{2iX}^2 - \rho''_i\, C_{2iX}\, C_{2iY}) .$$

For mathematical simplicity, we shall use approximation

$\frac{N'_i - n'_i}{N'_i\, n'_i} \cong \frac{N-n}{N\, n}$ for all $i = 1, 2, \cdots, k$, then we get

$$E(\bar{y}_{rPSNR}) = \sum_{i=1}^{k} W_i \left(\frac{N'_i}{N_i}\right) \bar{Y}'_i \left(\frac{N-n}{N\, n}\right) (C_{1iX}^2 - \rho'_i\, C_{1iX}\, C_{1iY})$$

$$+ \sum_{i=1}^{k} W_i \left\{ \frac{1}{n W_i} + \frac{(N-n)(1-W_i)}{n^2 (N-1) W_i^2} \right\} \bar{Y}''_i (f_i - 1)(C_{2iX}^2 - \rho''_i\, C_{2iX}\, C_{2iY})$$

Hence the theorem.

Remark 4.1: We use some standard results in the proof of above expression

i) $E\left[\left(\frac{n'_i}{n_i}\, \bar{y}'_{ir}\right) | n_i\right] = \frac{N'_i}{N_i} \bar{Y}'_i \left[1 + \frac{N'_i - n'_i}{N'_i\, n'} (C_{1iX}^2 - \rho'_i\, C_{1iX}\, C_{1iY})\right]$

ii) $E\left[\left(\frac{n''_i}{n_i}\, \bar{y}''_i\right) | n_i\right] = \frac{N''_i}{N_i} \bar{Y}''_i$

iii) $E\left(\frac{1}{n_i}\right) = \left[\frac{1}{nW_i} + \frac{(N-n)(1-W_i)}{n^2 (N-1) W_i^2}\right].$

Theorem 4.2: The approximate expression of variance of $\bar{y}_{rPSNR}$ is

$$V(\bar{y}_{rPSNR}) = \sum_{i=1}^{k} W_i^2 \left[\left(E\left(\frac{1}{n_i}\right)\right) \left(\frac{N''_i}{N_i}\right) (f_i - 1) A'_i \right]$$

$$+ \sum_{i=1}^{k} W_i^2 \left\{ E\left(\frac{1}{n_i}\right) W_i (1 - W_i) \right\} \left[\left(\frac{N-n}{N\, n}\right) B'_i \right.$$

$$\left. + \left\{ 1 + \left(\frac{N-n}{N\, n}\right) C'_i \right\}^2 \left\{ \left(\frac{1}{n} - \frac{1}{N}\right) S_{2iy}^2 \right\} \right],$$

where

$$A_i' = [S_{2iy}^2 + R_i''^2\, S_{2ix}^2 - 2\,\rho_i''\, R_i''\, S_{2iy}\, S_{2ix}],$$

$$B_i' = [S_{1iy}^2 + R_i'^2\, S_{1ix}^2 - 2\,\rho_i'\, R_i'\, S_{1iy}\, S_{1ix}]$$

$$C_i' = (C_{2ix}^2 - \rho_i'' C_{2iy}\, C_{2ix})\,.$$

Proof: We use standard results given below

i) $V\,(\bar{y}_i'' \mid n_i) = \left(\frac{1}{n_i''} - \frac{1}{N_i''}\right) S_{2iy}^2$

ii) $MSE\,[\bar{y}_{ir}''' \mid n_i, (n_i', n_i'')] = \left\{ \frac{n_i'' - n_i'''}{n_i''\, n_i'''} (s_{2iy}^2 + R_i''^2\, s_{2ix}^2 - 2\,\rho_i''\, R_i''\, s_{2iy}\, s_{2ix}) \right\}$

In order to avoid derivational complexity, we use approximation

$MSE\,(\bar{y}_{rPSNR}) \cong V\,(\bar{y}_{rPSNR})$ and $MSE\,(\bar{y}_{ir}''') \cong V\,(\bar{y}_{ir}''')$

The conditional variance is a sum of three components

$$\begin{aligned} V\,(\bar{y}_{rPSNR}) &= E\,[E\,\{V\,(\bar{y}_{rPSNR}) \mid n_i, (n_i',\, n_i'')\}] \\ &\quad + E\,[V\,\{E\,(\bar{y}_{rPSNR}) \mid n_i, (n_i',\, n_i'')\}] \\ &\quad + V\,[E\,\{V\,(\bar{y}_{rPSNR}) \mid n_i, (n_i',\, n_i'')\}]\,. \end{aligned}$$

Consider the first one

$$= E\left[E\left[V\left\{ \sum_{i=1}^{k} \left(\frac{W_i}{n_i}\right) (n_i'\, \bar{y}_{ir}') + \sum_{i=1}^{k} \left(\frac{W_i}{n_i}\right) (n_i''\, \bar{y}_{ir}''') \right\} \Big|\, n_i, (n_i',\, n_i'') \right]\right]$$

$$= E\left[E\left[0 + \sum_{i=1}^{k} \left(\frac{W_i\, n_i''}{n_i}\right)^2 \left\{ \frac{n_i'' - n_i'''}{n_i''\, n_i'''} (s_{2iy}^2 + R_i''^2 s_{2ix}^2 - 2\,\rho_i''\, R_i''\, s_{2iy}\, s_{2ix}) \right\} \Big|\, n_i, (n_i',\, n_i'') \right]\right]$$

$$= \sum_{i=1}^{k} W_i^2 \left\{ E\left(\frac{1}{n_i}\right) \left(\frac{N_i''}{N_i}\right) (f_i - 1)\, (s_{2iy}^2 + R_i''^2 s_{2ix}^2 - 2\,\rho_i''\, R_i''\, s_{2iy}\, s_{2ix}) \right\}$$

$$= \sum_{i=1}^{k} W_i^2 \left\{ \left(\frac{1}{n W_i} + \frac{(N-n)\,(1-W_i)}{n^2\,(N-1)\, W_i^2} \right) \left(\frac{N_i''}{N_i}\right) (f_i - 1)\, A_i' \right\}.$$

The second component provides

$$= E\left[V\left\{ \sum_{i=1}^{k} W_i \left(\frac{n_i'}{n_i}\right) \bar{y}_{ir}' + \sum_{i=1}^{k} W_i \left(\frac{n_i''}{n_i}\right) E(\bar{y}_{ir}''') \middle| n_i, (n_i', n_i'') \right\} \right]$$

$$= E\left[V\left[\sum_{i=1}^{k} W_i \left(\frac{n_i'}{n_i}\right) \bar{y}_{ir}' + \sum_{i=1}^{k} W_i \left(\frac{n_i''}{n_i}\right) \bar{y}_i'' \left\{ 1 + \frac{n_i'' - n_i'''}{n_i'' n_i'''} \right.\right.\right.$$

$$\left.\left.\left. (C_{2iX}^2 - \rho_i'' C_{2iX} C_{2iY}) \middle| n_i, (n_i', n_i'') \right\} \right] \right] \tag{4.1}$$

For simplicity, we have taken $\beta_i^* = \left(\frac{n_i'}{n_i}\right)$ and $\gamma_i^* = \left(\frac{n_i''}{n_i}\right)$. To simplify further, we shall use approximation like

i) $\frac{n_i'' - n_i'''}{n_i'' n_i'''} \cong \frac{N-n}{Nn}$ for all $i = 1, 2, \cdots, k$

ii) $V[\{(\beta_i^*)(\bar{y}_{ir}')\} | n_i] \cong V[(\beta_i^*) | n_i] V[(\bar{y}_{ir}') | n_i]$

iii) $V[\{(\gamma_i^*)(\bar{y}_i'')\} | n_i] \cong V[(\gamma_i^*) | n_i] V[(\bar{y}_i'') | n_i]$

iv) $V(\beta_i^* | n_i) = \left(\frac{N_i - n_i}{N_i - 1}\right)\left(\frac{W_i(1-W_i)}{n_i}\right) \cong \left(\frac{W_i(1-W_i)}{n_i}\right)$

v) $V(\gamma_i^* | n_i) = \left(\frac{N_i - n_i}{N_i - 1}\right)\left(\frac{W_i(1-W_i)}{n_i}\right) \cong \left(\frac{W_i(1-W_i)}{n_i}\right)$

using above results, the expression results into

$$= E\left[\sum_{i=1}^{k} W_i^2 V\{(\beta_i^*)(\bar{y}_{ir}')\} + \sum_{i=1}^{k} W_i^2 \left\{ 1 + \frac{N-n}{Nn} (C_{2iX}^2 - \rho_i'' C_{2iY} C_{2iX}) \right\}^2 \right.$$

$$\left. V\{(\gamma^*)(\bar{y}_i'')\} \middle| n_i \right]$$

$$= E\left[\sum_{i=1}^{k} W_i^2 \left\{ \frac{W_i(1-W_i)}{n_i} \right\} \left\{ \frac{N-n}{Nn} (S_{1iy}^2 + R_i'^2 S_{1ix}^2 - 2\rho_i' R_i' S_{1iy} S_{1ix}) \right\} \right.$$

$$+ \sum_{i=1}^{k} W_i^2 \left\{ 1 + \frac{N-n}{Nn} (C_{2ix}^2 - \rho_i'' C_{2iy} C_{2ix}) \right\}^2 \left\{ \frac{W_i(1-W_i)}{n_i} \right\}$$

$$\left. \left\{ \left(\frac{1}{n} - \frac{1}{N}\right) S_{2iy}^2 \right\} \middle| n_i \right]$$

$$= \sum_{i=1}^{k} W_i^2 \left\{ E\left(\frac{1}{n_i}\right) W_i (1 - W_i) \right\} \left[\left(\frac{N-n}{N\,n} \right) B_i' \right.$$

$$\left. + \left\{ 1 + \left(\frac{N-n}{N\,n} \right) C_i' \right\}^2 \left\{ \left(\frac{1}{n} - \frac{1}{N} \right) S_{2iy}^2 \right\} \right]$$

The third component vanishes and by adding the first and second component, we get the required result.

5. COST ANALYSIS

Cost analysis is incorporated to estimate the optimal sample size, consider $i-$th stratum-based approach by assuming cost C_{0i}, C_{1i}, C_{2i} varying over all k strata.

C_{0i} : cost of n_i units of the $i-$th stratum.

C_{1i} : cost of collecting, editing and processing per n_i' units in the $i-$th strata of response class.

C_{2i} : cost of personal interview and processing per n_i''' units in the $i-$th strata for non-response class.

The $i-$th stratum has total cost

$$C_i = [C_{0_i}\, n_i + C_{1i}\, n_i' + C_{2i}\, n_i'''] \tag{5.1}$$

Total cost over k strata

$$T_c = \sum_{i=1}^{k} C_i = \sum_{i=1}^{k} [C_0\, n_i + C_1\, n_i' + C_2\, n_i''']$$

$$E(T_c) = \left(\frac{n}{N} \right) \sum_{i=1}^{k} \left\{ C_0 N_i + C_1 N_i' + C_2 \frac{N_i''}{f_i} \right\}$$

To get optimum f_i, λ_i and n, define a function L_i, with lagrange multiplier λ_i and pre-fixed level of variance V_{0i}

$$L_i = [\text{expected cost of } i-\text{th strata}] + \lambda_i \, [V(\bar{y}_{rPSNR})_i - V_{0i}]$$

$$L_i = \left(\frac{n}{N} \right) \left[C_{0_i} N_i + C_{1i} N_i' + C_{2i} \left(\frac{N_i''}{f_i} \right) \right] + \lambda_i \left[\left\{ \left(E\left(\frac{1}{n_i} \right) \right) \right\} \left(\frac{N_i''}{N_i} \right) (f_i - 1) A_i' \right]$$

$$+\lambda_i\left\{E\left(\frac{1}{n_i}\right)W_i\,(1-W_i)\right\}\left[\left(\frac{N-n}{N\,n}\right)B_i'+\left\{1+\left(\frac{N-n}{N\,n}\right)C_i'\right\}^2\right.$$

$$\left.\left\{\left(\frac{1}{n}-\frac{1}{N}\right)S_{2iy}^2\right\}\right]-\lambda_i\,V_{0i} \tag{5.2}$$

On differentiating (5.2) with respect to f_i, λ_i and n, we get three equations.

Differentiating with respect to f_i we get,

$$-\left(\frac{n}{N}\right)\frac{C_{2i}}{f_i^2}N_i''+\lambda_i\left[E\left(\frac{1}{n_i}\right)\left(\frac{N_i''}{N_i}\right)A_i'\right]=0 \tag{5.3}$$

To simplify the expression further, we shall use approximation

$$\frac{N_i'}{N_i}=\frac{N_i''}{N_i}=\frac{N_i}{N}\cong W_i$$

or $$\lambda_i=(n\,C_{2i}\,W_i)\left\{f_i^2\,E\left(\frac{1}{n_i}\right)A_i'\right\}^{-1} \tag{5.4}$$

Secondly, differentiating (5.2) with respect to λ_i, we get

$$\left[\left\{E\left(\frac{1}{n_i}\right)\right\}\left(\frac{N_i''}{N_i}\right)(f_i-1)\,A_i'\right]+\left\{E\left(\frac{1}{n_i}\right)W_i\,(1-W_i)\right\}$$

$$\left[\left(\frac{N-n}{N\,n}\right)B_i'+\left\{1+\left(\frac{N-n}{N\,n}\right)C_i'\right\}^2\left\{\left(\frac{1}{n}-\frac{1}{N}\right)S_{2iy}^2\right\}\right]-V_{0i}=0 \tag{5.5}$$

Thirdly, differentiating (5.2) with respect to n, and neglecting the terms of order n^{-3}, we get

$$[C_{0i}\,W_i+C_{1i}\,W_i]+C_{2i}\,\frac{W_i}{f_i}-\left(\frac{\lambda_i}{n}\right)\left[\left\{E\left(\frac{1}{n_i}\right)+\frac{N\,(1-W_i)}{n^2\,(N-1)\,W_i^2}\right\}W_i\,(f_i-1)\,A_i'\right]$$

$$-\left(\frac{\lambda_i}{n^2}\right)\left[\left\{2\,E\left(\frac{1}{n_i}\right)+\frac{N\,(1-W_i)}{n^2\,(N-1)\,W_i^2}\right\}W_i\,(1-W_i)\,(S_{2iy}^2+B_i'^2)\right]=0 \tag{5.6}$$

putting the value of λ_i from (5.4) in (5.6), we get

$$l_{1i}\,f_i^2-l_{2i}\,f_i+l_{3i}=0 \tag{5.7}$$

or $f_i = \left(\frac{1}{2 l_{1i}}\right)\left(l_{2i} \pm \sqrt{l_{2i}^2 - 4 l_{1i}\, l_{3i}}\right)$, (5.8)

where

$$l_{1i} = (C_{0i}\, W_i + C_{1i}\, W_i), \qquad l_{2i} = \left(C_{2i}\, W_i - \frac{C_{2i}\, W_i^2\, \alpha_i}{E(1/n_i)}\right)$$

$$l_{3i} = \left(\frac{C_{2i}\, \alpha_i\, W_i^2}{E(1/n_i)} - \frac{C_{2i}\, W_i\, G_i}{E(1/n_i)\, A_i'}\right), \quad \alpha_i = E\left(\frac{1}{n_i}\right) + \frac{N(1-W_i)}{n^2 (N-1) W_i^2},$$

$$G_i = \left\{2E\left(\frac{1}{n_i}\right) + \frac{N(1-W_i)}{n^2 (N-1) W_i^2}\right\} W_i (1-W_i) \left(\frac{S_{2iy}^2 + B_i'}{n}\right).$$

The positive value of f_i will occur if the condition $l_{2i}^2 > (l_{1i}\, l_{3i})$ holds. On substituting the standard values of $E\left(\frac{1}{n_i}\right)$ in (5.5), n optimum can be obtained from the following equation:

$$n^2 V_{0i} - n E_i - F_i = 0, \tag{5.9}$$

where

$$E_i = \frac{P_i}{W_i} - \frac{(1-W_i) P_i}{(N-1) W_i^2}, \; F_i = \frac{(1-W_i) P_i}{W_i^2} + \frac{Q_i}{W_i},$$

$$Q_i = W_i (1-W_i)(B_i' + S_{2iy}^2), \qquad P_i = W_i (f_i - 1) A_i'.$$

This equation gives the optimum $(n_{opt})_i$ using f_i from (5.9) along with a suitable choice of V_{0i}.

Remark 5.1: The equation (5.9), varies over i, therefore it will give k value of optimal sample size $(n_{opt})_i$, for $i = 1, 2, \cdots, k$, one can choose any of these for practical purpose or one may choose an average integer value of $(n_{opt})_i$ or, minimum / maximum of $(n_{opt})_i$, over all k strata. It depends upon the situation faced by survey practitioner.

6. NUMERICAL ILLUSTRATIONS

The data set is same as taken by Dubey and Shukla (2001). Except the value of X is generated by taking into account the linear regression between Y on X. The values of X have been rounded off. Suppose a population of 400 units, has four strata I, II, III, IV, each one is divided into two groups as response (R) and non-response (NR). The values of the population parameters computed from the data are given below:

$$N = 400, \quad \bar{Y} = 81.65, \quad \bar{X} = 21.4251, \quad S^2 = 1857.025.$$

Table 6.1:

Type	Division	Strata			
		I	II	III	IV
Size of Strata	Response class (R)	$N_1' = 30$	$N_2' = 40$	$N_3' = 60$	$N_4' = 70$
	Non-Response class (NR)	$N_1'' = 30$	$N_2'' = 40$	$N_3'' = 60$	$N_4'' = 70$
	Total N_i	$N_1 = 60$	$N_2 = 80$	$N_3 = 120$	$N_4 = 140$
	$\bar{Y}_i'$	$\bar{Y}_1' = 12.800$	$\bar{Y}_2' = 46.475$	$\bar{Y}_3' = 81.133$	$\bar{Y}_4' = 126.828$
	$\bar{Y}_i''$	$\bar{Y}_1'' = 14.167$	$\bar{Y}_2'' = 49.275$	$\bar{Y}_3'' = 89.866$	$\bar{Y}_4'' = 126.671$
	$\bar{X}_i'$	$\bar{X}_1' = 10.866$	$\bar{X}_2' = 38.800$	$\bar{X}_3' = 70.950$	$\bar{X}_4' = 109.786$
	$\bar{X}_i''$	$\bar{X}_1'' = 13.133$	$\bar{X}_2'' = 41.650$	$\bar{X}_3'' = 79.816$	$\bar{X}_4'' = 112.000$
Mean Squares	Response class (R)	$S_{11y}^2 = 67.062$	$S_{12y}^2 = 156.667$	$S_{13y}^2 = 228.015$	$S_{14y}^2 = 237.97$
		$S_{11x}^2 = 64.395$	$S_{12x}^2 = 140.677$	$S_{13x}^2 = 178.726$	$S_{14x}^2 = 235.967$
	Non-Response class (NR)	$S_{21y}^2 = 62.971$	$S_{22y}^2 = 204.870$	$S_{23y}^2 = 248.79$	$S_{24y}^2 = 254.745$
		$S_{21x}^2 = 51.567$	$S_{22x}^2 = 226.797$	$S_{23x}^2 = 201.847$	$S_{24x}^2 = 353.884$
Weights	W_i	$W_1 = 0.15$	$W_2 = 0.20$	$W_3 = 0.30$	$W_4 = 0.35$
Coefficient of variation	C_{1iX}	$C_{11X} = 0.7385$	$C_{12X} = 0.3057$	$C_{13X} = 0.1884$	$C_{14X} = 0.1399$
	C_{1iY}	$C_{11Y} = 0.6398$	$C_{12Y} = 0.2693$	$C_{13Y} = 0.1861$	$C_{14Y} = 0.1216$
	C_{2iX}	$C_{21X} = 0.5468$	$C_{22X} = 0.3616$	$C_{23X} = 0.1779$	$C_{24X} = 0.1680$
	C_{2iY}	$C_{21Y} = 0.5601$	$C_{22Y} = 0.2904$	$C_{23Y} = 0.1755$	$C_{24Y} = 0.1260$
Correlation coefficient	ρ_i'	$\rho_1' = 0.89949$	$\rho_2' = 0.84058$	$\rho_3' = 0.8125$	$\rho_4' = 0.80223$
	ρ_i''	$\rho_1'' = 0.87278$	$\rho_2'' = 0.8997$	$\rho_3'' = 0.8896$	$\rho_4'' = 0.85074$

A sample of size $n = 80$ is drawn by *SRSWOR* and post-stratified into n_i units. When a deadline of receiving mailed questionnaire is over, n_i' and n_i'' units are as below along with prefixed n_i''' and f_i.

Table 6.2: Sample description

Type	Strata				Total
Sample size	I	II	III	IV	
n_i	16	18	22	24	80
n_i'	6	8	9	10	33
n_i''	10	10	13	14	47
n_i'''	3	4	4	05	16
$f_i = n_i'' / n_i'''$	3.33	2.50	2.50	2.00	10.33

i) For calculation purpose, the cost function is considered with cost C_{0i}, C_{1i} and C_{2i} and cost optimum f_i has been calculated from (5.8) which is shown below:

.

Table 6.3:

Strata I	Strata II	Strata III	Strata IV
$C_{01} = 0.5$	$C_{02} = 0.5$	$C_{03} = 0.5$	$C_{04} = 0.5$
$C_{11} = 0.5$	$C_{12} = 0.5$	$C_{13} = 1.0$	$C_{14} = 0.5$
$C_{21} = 5.0$	$C_{22} = 6.0$	$C_{23} = 6.0$	$C_{24} = 5.0$
$f_1 = 4$	$f_2 = 5$	$f_3 = 3$	$f_4 = 3$

Table 6.3 reveals that C_{0i} and C_{1i} approximately same but C_{2i} is higher to these two.

ii) The computed optimal sample size $(n_{opt})_{irPSNR}$ and value of V_{0i} are given below:

Table 6.4a:

	Strata I	Strata II	Strata III	Strata IV
V_{0i}	90	125	235	105
$(n_{opt})_{irPSNR}$	3	4	2	4

Again, in order to compare the cost optimal sample size of our procedure to that of Shukla and Dubey procedure, we have taken the same V_{0i} and computed $(n_{opt})_i$. The, following table 6.4b gives the comparison of the optimal sample size computed by the two procedures.

Table 6.4b: Comparison of the optimal sample size calculated by Shukla and Dubey and the suggested procedure

	Strata I	Strata II	Strata III	Strata IV
V_{0i}	100	150	275	125
$(n_{opt})_i$	23	59	27	40
$(n_{opt})_{irPSNR}$	2	4	2	3

For convenience, the entire data used for the estimation of cost optimal sample size is given below in table 6.5:

Table 6.5:

Strata		Sample 1	Sample 2	Sample 3	Sample 4
I	n_1'	6	6	6	6
	y_1'	4, 13, 14, 30, 7, 10	9, 12, 14, 8, 1, 29	2, 17, 22, 19, 7, 20	9, 6, 11, 15, 19, 7
	x_1'	1, 5, 5, 10, 3, 3	3, 5, 5, 3, 1, 8	1, 6, 8, 6, 3, 8	3, 3, 4, 5, 8, 3
	n_1''	10	10	10	10
	y_1''	4, 14, 6, 4, 22, 25, 7, 15, 9, 1	4, 22, 14, 26, 8, 30, 20, 15, 6, 8	2, 4, 16, 8, 10, 22, 14, 6, 18, 20	1, 13, 25, 7, 19, 21, 15, 15, 27, 9
	x_1''	1, 5, 4, 1, 8, 8, 4, 5, 1, 1	1, 8, 5, 9, 4, 10, 7, 5, 1, 4	1, 1, 6, 4, 4, 7, 5, 1, 6, 7	1, 4, 8, 5, 7, 7, 5, 5, 9, 1
	n_1'''	3	3	3	3
	y_1'''	22, 4, 15	26, 30, 15	22, 2, 20	21, 9, 15
	x_1'''	8, 1, 5	9, 10, 5	7, 1, 7	7, 4, 4
	n_1	16	16	16	16
	$\bar{y}_1'$	13	12.167	14.50	11.167
	$\bar{x}_1'$	4.50	4.167	5.33	4.00
	$\bar{y}_1'''$	13.667	23.667	14.667	15.00
	$\bar{x}_1'''$	4.667	8.00	5.00	5.00

II	n_2'	8	8	8	8
	y_2'	59, 35, 47, 26, 34, 42, 60, 58	35, 50, 62, 56, 43, 64, 47, 60	49, 36, 42, 31, 64, 26, 58, 34	38, 62, 42, 55, 43, 48, 39, 63
	x_2'	20, 55, 50, 25, 45, 30, 40, 48	45, 25, 20, 31, 31, 32, 52, 60	35, 50, 40, 25, 31, 40, 61, 51	20, 35, 30, 40, 56, 39, 31, 51
	n_2''	10	10	10	10
	y_2''	41, 26, 36, 64, 66, 66, 70, 48, 49, 30	50, 62, 25, 46, 28, 40, 62, 34, 63, 62	25, 28, 47, 24, 55, 40, 69, 47, 50, 53	55, 58, 61, 66, 53, 49, 67, 44, 42, 70
	x_2''	35, 20, 32, 62, 60, 55, 68, 40, 41, 26	25, 31, 20, 32, 21, 36, 54, 30, 61, 60	20, 21, 45, 18, 52, 35, 64, 40, 45, 36	40, 46, 59, 65, 40, 68, 38, 30, 50, 20
	n_2'''	4	4	4	4
	y_2'''	26, 70, 49, 66	34, 62, 63, 28	47, 55, 69, 50	58, 66, 49, 44
	x_2'''	35, 26, 32, 60	31, 54, 36, 61	45, 18, 36, 64	40, 68, 59, 20
	n_2	18	18	18	18
	$\bar{y}_2'$	45.125	52.125	42.5	48.75
	$\bar{x}_2'$	39.125	37	41.625	37.75
	$\bar{y}_2'''$	52.75	46.75	55.25	54.25
	$\bar{x}_2'''$	38.25	45.5	40.75	46.75
III	n_3'	9	9	9	9
	y_3'	93, 68, 62, 80, 84, 59, 66, 100, 70	88, 71, 91, 86, 100, 106, 74, 68, 59	84, 66, 92, 84, 59, 102, 70, 80, 67	62, 91, 78, 66, 84, 96, 102, 106, 88
	x_3'	96, 70, 67, 65, 75, 71, 96, 90, 80	85, 80, 75, 72, 70, 75, 74, 60, 71	80, 71, 69, 70, 71, 50, 69, 61, 75	58, 67, 65, 90, 80, 70, 42, 71, 100
	n_3''	13	13	13	13
	y_3''	61, 90, 84, 89, 99, 108, 120, 103, 65, 91, 89, 83, 105	61, 103, 105, 107, 109, 70, 63, 84, 98, 95, 90, 99, 98	96, 74, 72, 84, 67, 70, 98, 105, 89, 87, 72, 97, 81	97, 88, 66, 84, 83, 79, 80, 95, 98, 98, 72, 105, 81
	x_3''	60, 85, 67, 80, 85, 65, 60, 74, 78, 90, 80, 78, 80	55, 75, 80, 72, 80, 70, 60, 85, 90, 7, 8, 60, 71, 70	85, 78, 71, 90, 70, 102, 90, 112, 89, 95, 71, 86, 78	110, 80, 55, 95, 65, 90, 42, 85, 55, 100, 110, 80, 90
	n_3'''	4	4	4	4
	y_3'''	61, 89, 103, 91	103, 107, 70, 99	96, 84, 67, 70	97, 83, 80, 72
	x_3'''	80, 85, 90, 90	90, 75, 85, 90	85, 90, 70, 102	110, 55, 90, 95
	n_3	22	22	22	22
	$\bar{y}_3'$	75.778	82.556	78.22	85.889
	$\bar{x}_3'$	78.889	73.55	68.444	71.444
	$\bar{y}_3'''$	86	94.75	79.25	83
	$\bar{x}_3'''$	86.25	85	86.75	87.5

IV	n'_4	10	10	10	10
	y'_4	112, 104, 149, 133, 120, 102, 131, 137, 121, 107	101, 148, 133, 117, 124, 149, 105, 150, 137, 129	146, 108, 126, 133, 136, 127, 144, 135, 106, 150	127, 112, 124, 148, 136, 105, 121, 150, 128, 140
	x'_4	140, 70, 100, 102, 102, 75, 111, 110, 120, 145	140, 145, 111, 98, 120, 121, 135, 90, 120, 140	120, 111, 102, 136, 135, 108, 110, 120, 98, 135	96, 112, 96, 100, 115, 150, 120, 114, 96, 105
	n''_4	14	14	14	14
	y''_4	139, 109, 100, 121, 130, 119, 132, 148, 150, 104, 103, 149, 107, 134	137, 147, 106, 105, 108, 145, 126, 130, 106, 108, 127, 147, 104, 148	149, 145, 140, 149, 106, 143, 110, 138, 144, 132, 111, 125, 121, 112	127, 113, 124, 118, 124, 132, 134, 116, 112, 146, 125, 125, 108, 146
	x''_4	120, 125, 135, 75, 126, 140, 88, 98, 75, 88, 75, 111, 102, 136	101, 98, 75, 88, 80, 120, 85, 105, 120, 135, 100, 99, 85, 111	120, 120, 85, 99, 136, 138, 108, 111, 112, 114, 135, 138, 80, 75	112, 120, 114, 110, 108, 104, 75, 120, 85, 138, 99, 102, 104, 75
	n'''_4	5	5	5	5
	y'''_4	139, 100, 119, 148, 134	137, 108, 126, 147, 148	149, 106, 110, 125, 112	127, 118, 132, 116, 108
	x'''_4	120, 75, 88, 140, 136	120, 130, 120, 111, 106	120, 99, 114, 75, 80	138, 104, 120, 85, 75
	n_4	24	24	24	24
	$\bar{y}'_4$	121.6	129.3	131.1	129.1
	$\bar{x}'_4$	107.5	122	119.2	110.4
	$\bar{y}'''_4$	128.0	133.2	120.4	120.2
	$\bar{x}'''_4$	111.8	117.4	97.6	104.4
$\bar{y}_{PSNR}$		80.3608	83.3530	80.0590	79.3960
$\bar{y}_{rPSNR}$		82.4945	83.4996	83.7183	84.2076

Table 6.6:

$\lvert$Bias $\bar{y}_{rPSNR}\rvert$	0.076158
$V(\bar{y}_{rPSNR})$	0.613844
$V(\bar{y}_{PSNR})$	2.228

7. DISCUSSION AND CONCLUSIONS

The proposed estimator using auxiliary information has been compared with $\bar{y}_{PSNR}$ of Shukla and Dubey in which auxiliary information has not been used. It is seen that the proposed estimator is more efficient for mean estimation in mail surveys.

i) Though it is biased estimator but the amount of bias is almost negligible (see Table 6.6).

ii) Table 6.6 shows the variance $V(\bar{y}_{PSNR}) = 2.228$, which is substantially higher than the variance $V(\bar{y}_{rPSNR}) = 0.61384$ of the proposed estimator.

iii) Table 6.4a indicates, the average integer value of the cost optimal sample size in the proposed procedure is $n = 13$, taken over all four strata.

iv) Comparison of the cost optimal sample size [See Table 6.4b], reveals that the average integer value of n_{opt} reduces to 11 as compared to 37 of Shukla and Dubey (2001) procedure.

This clearly depicts the efficiency of the proposed estimator over $\bar{y}_{PSNR}$ in the estimation of mean in mail surveys under all similar setup.

REFERENCES

Cochran, W.G. (1999): *Sampling Techniques*. John Wiley, New York.

Grover, R.M. and Couper, M.P. (1998): *Non-response in Household Surveys*. John Wiley, New York.

Hansen, M.H., Hurwitz, W.N. and Madow, W.G. (1953): *Sample Survey Methods and Theory*. John Wiley, New York.

Hinde, R.L. and Chambers, R.L. (1991): Non response imputation with multiple source of non-response. *J. Official Statist*., **7**, 169-179.

Holt, D. and Smith, T.M.F. (1979): Post-stratification. *J. Roy. Statist. Soc., Ser A,* **142**, 33-36.

Jagers, P., Oden, A. and Trulsson, L. (1985): Post-stratification and ratio estimation: usages of auxiliary information in survey sampling and opinion polls. *Internat. Statist. Rev*., **53**, 221-238.

Khare, B.B. (1987): Allocation in stratified sampling in presence of non-response. *Metron*, **XLV**, 213-221.

Khot, P.S. (1994): A note on handling non-response in sample surveys. *J. Amer. Statist. Assoc*., **89**, 693-696.

Lessler, J.T. and Kalsbeek, W.D. (1992): *Non-response Error in Surveys*. John Wiley, New York.

Mukhopadhyay, P. (1999): *Theory and Methods of Survey Sampling*, Prentice-Hall of India, New Delhi.

Shukla, D.and Dubey, J. (2001). Estimation in mail surveys under *PSNR* sampling scheme. *J. Indian Soc. Agricultural Statist.,* **54**, 288-302.

Shukla, D. and Trivedi, M. (2001): Mean estimation in deeply stratified population under post-stratification. *J. Indian Soc. Agricultural Statist.,* **54**, 221-235.

Shukla, D. and Trivedi, M. (2006): Examining stability of variability ratio with application in post-stratification. *Int. J. Manag. Syst*., **22**, 59-70.

Shukla, D., Bankey, A. and Trivedi, M. (2002): Estimation in post-stratification using prior information and grouping strategy. *J. Indian Soc. Agricultural Statist.,* **55**, 158-173.

Shukla, D. and Dubey, J. (2006): On earmarked strata in post-stratification. *Statist. Transition*, **7**, 1067-1085.

Shukla, D. and Dubey, J. (2004): On estimation under post-stratified two phase non-response sampling scheme in mail surveys. *Int. J. Manag. Syst*., **20**, 269-278.

Silva, P.L.D.N. and Skinner, C.J. (1995): Estimation of distribution functions with auxiliary information using post stratification. *J. Official Statist*., **3**, 277-294.

Smith, T.M.F. (1991): Post-stratification. *The Statisticians*, **40**, 323-330.

Sukhatme, P.V., Sukhatme, B.V., Sukhatme, S. and Ashok, C. (1984): *Sampling Theory of Surveys with Applications*. Iowa State University Press, U.S.A.

Wywial, J. (2001): Stratification of the population after sample selection. *Statistics in Transition*, **5**, 327-348.

Zhang, L.C. (1999): A note on post-stratification when analyzing binary survey data subject to non response *J. Official Statist*., **15**, 329-334.

Received : 05-07-2006

Department of Statistics,
Institute of Social Sciences,
Dr. B.R. Ambedkar University, Agra-282 002. India.
e-mail: msrivastava_iss@hotmail.com