STATISTICS IN TRANSITION-new series, August 2011
Vol. 12, No. 1, pp. 9—24

SUBSAMPLING THE NONRESPONDENTS IN CLUSTER SAMPLING ON SAMPLING ON TWO SUCCESSIVE OCCASIONS

Housila P. Singh, Sunil Kumar[1]

ABSTRACT

The problem of estimation of finite population mean for current occasion in the context of cluster sampling on two successive occasions when there is non-response on both the occasions. Estimators for the current occasion are derived as the particular case when there is non-response on first occasion and second occasion only. A comparison between variances of the estimates is studied. An empirical study is made to study the performance of the proposed strategy.

Key words: Non-response, Successive sampling; Mail surveys; Study variate; Auxiliary variate.

1. Introduction

Cluster or area sampling is widely practiced in sample surveys because of its low cost and time saving device to conduct large scale and complicated surveys. In those sample surveys where there is no list of the elements in the finite population or units of the population are widely scattered, it is required to take repeated observations on the selected units. It is well known that the cluster sampling is better than the simple random sampling when the intra class correlation within the same cluster is negative and smaller than $-(M-1)^{-1}$, where M denotes the size of the cluster. It is noted that the relative efficiency of the cluster sampling with respect to simple random sampling depends upon M, the size of the cluster and the intra-class correlation coefficient, it decreases if the size of the cluster M increases considerably. In practice, the intra-class correlation coefficient is generally non-negative and decreases as M increases, but the rate of decrease is small relative to the rate of increase in M, so that ordinarily, increase in the size of a cluster leads to substantial increase in the sampling variance of the sample estimate, see Sukhatme and Sukhatme (1970).

[1] School of Studies in Statistics, Vikram University, Ujjain – 456010, M. P., India. E-mail: hpsujn@rediffmail.com; sunilbhougal06@gmail.com

Zarkovich and Krane (1965) demonstrated that the correlation between two characters in cluster sampling with clusters as sampling units is expected to be higher than correlation coefficient in element sampling.

It is well known that the use of auxiliary at the estimation stage improves the efficiency of the estimators. Out of many ratio, product and regression methods of estimation are good illustrations in this context. If the survey is repetitive in nature, past values of the variable under investigation may be used as an auxiliary to improve on the precision of the current estimate. In such repetitive surveys, a fraction of the original sample units may be retained for use at the current occasion, while the remaining fraction is selected afresh. This procedure is known as sampling on successive occasions. The papers by Jessen (1942), Ware and Cunia (1962), Frayer and Furnival (1967), Singh (1968), Kathuria (1975), Raj (1979), Arnab (1980), Singh et al (1992) and Singh and Vishwakarma (2007). Recently Pradhan (2004) considered the problem of estimating the mean on second occasion over sampling on two successive occasions when the sampling units are clusters and the observations on the first occasion are regarded as ancillary information for the observations on the second or current occasion. These authors have carried out their studies under the supposition that there is complete response from all the sample units.

In surveys covering human populations this information is usually not obtained from all the sample units even after callbacks. Hansen and Hurwitz (1946) were the first to deal with the problem of incomplete samples in mail surveys. This method has been applied by various authors including Cochran (1977), Rao (1986), Khare and Srivastava (1993, 1995, 1997), Okafor and Lee (2000), Tabasum and Khan (2004, 2006), Choudhary et al. (2004), Okafor (2001, 2005) and Singh and Kumar (2008 a, b).

In this paper the theory for use of Hansen and Hurwitz (1946) technique for estimation of population mean for current occasion in the context of cluster sampling over sampling on two successive occasions has been developed. The results obtained are depicted with the help of numerical illustration.

2. Estimation of population mean for current occasion in the presence of nonresponse on both the occasions

Suppose that the population is composed of N clusters of M elements each, and that a simple random sample of n clusters is drawn without replacement from it. Let (x_{ij}, y_{ij}), $(i = 1,2,...,N;\ j = 1,2,...,M)$ be the values of the characteristic on first and second occasions for the j^{th} unit of the i^{th} cluster, respectively. We assume that the population can be divided into two classes, those who will respond at the first attempt and those who will not. Let the sizes of these two classes be N_1 and N_2 clusters respectively. In the second occasion $m\ (= n\lambda)$ of the n clusters selected on the first occasion retained at random and the remaining

$u = n\mu = (n - m)$ of the clusters are replaced by a fresh selection. The characters x and y are supposed to be correlated when they are observed (repeatedly). We assume that in the unmatched portion of the sample on the two occasions u_1 clusters (i.e. $M\,u_1$ elements) respond and u_2 clusters (i.e. $M\,u_2$ elements) do not. Similarly, in the matched portion m_1 clusters (i.e. $M\,m_1$ elements) respond and m_2 clusters (i.e. $M\,m_2$ elements) do not. Let $m_{h_2} = m_2/k$, $k > 1$ denote the size of the sub sample drawn from the non-response class from the matched portion of the sample on the two occasions for collecting information through personal interview. Similarly, denote by $u_{h_2} = u_2/k$, $k > 1$ the size of the sub sample drawn from the non-response class from the unmatched portion of the sample on the two occasions. Further, we define:

- $\overline{X}_{i.} = \sum_{j=1}^{M} x_{ij}/M$ and $\overline{Y}_{i.} = \sum_{j=1}^{M} y_{ij}/M$ are means of the i^{th} cluster on the first and second occasion respectively,
- $\overline{X} = \sum_{i=1}^{N} \overline{x}_{i.}/N$ and $\overline{Y} = \sum_{i=1}^{N} \overline{y}_{i.}/N$ are cluster population means of x and y respectively.
- $\overline{X}_{NM} = \sum_{i=1}^{N}\sum_{j=1}^{M} x_{ij}/NM$ and $\overline{Y}_{NM} = \sum_{i=1}^{N}\sum_{j=1}^{M} y_{ij}/NM$ are the population means of x and y per element on the first and second occasions respectively.
- $S_x^2 = \sum_{i=1}^{N}\sum_{j=1}^{M} \left(x_{ij} - \overline{X}_{NM}\right)^2 / (NM - 1)$: the population mean square between elements on the first occasion.
- $S_y^2 = \sum_{i=1}^{N}\sum_{j=1}^{M} \left(y_{ij} - \overline{Y}_{NM}\right)^2 / (NM - 1)$: the population mean square between elements on the second occasion,
- $\rho_x = \sum_{i}^{N}\sum_{j<k}^{M} \left(x_{ij} - \overline{X}_{NM}\right)\left(x_{jk} - \overline{X}_{NM}\right) / \left\{(M-1)(NM-1)S_x^2\right\}$: the intra-class correlation coefficient between elements of a cluster on first occasion,
- $\rho_y = \sum_{i}^{N}\sum_{j<k}^{M} \left(y_{ij} - \overline{Y}_{NM}\right)\left(y_{jk} - \overline{Y}_{NM}\right) / \left\{(M-1)(NM-1)S_y^2\right\}$: the intra-class correlation coefficient between elements of a cluster on second occasion,

- $\rho_b = \sum_{i=1}^{N}\left(\overline{Y}_{i.} - \overline{Y}_{NM}\right)\left(\overline{X}_{i.} - \overline{X}_{NM}\right) \Big/ \left\{\sum_{i=1}^{N}\left(\overline{Y}_{i.} - \overline{Y}_{NM}\right)^2 \sum_{i=1}^{N}\left(\overline{X}_{i.} - \overline{X}_{NM}\right)^2\right\}^{1/2}$: the simple correlation coefficient between cluster means on both occasions.

Let an equivalent sample of nM elements be selected from the population of NM elements by simple random sampling. Then the mean per element on the first and second occasions respectively be defined by $\overline{x}_{nM} = \sum_{l=1}^{nM} x_l / nM$ and $\overline{y}_{nM} = \sum_{l=1}^{nM} y_l / nM$. Similarly, we define $\overline{x}_{uM} = \sum_{l=1}^{uM} x_l / uM$ and $\overline{y}_{uM} = \sum_{l=1}^{uM} y_l / uM$ as sample means based on a simple random sample of uM units.

Further, we denote

- $S_{x(2)}^2 = \sum_{i=1}^{N_2}\sum_{j=1}^{M}\left(x_{ij} - \overline{X}_{N_2M}\right)^2 \Big/ \left(N_2M - 1\right)$: the population mean square between the elements pertaining to the non-response class on the first occasion,
- $S_{y(2)}^2 = \sum_{i=1}^{N_2}\sum_{j=1}^{M}\left(y_{ij} - \overline{Y}_{N_2M}\right)^2 \Big/ \left(N_2M - 1\right)$: the population mean square between the elements pertaining to the non-response class on the second occasion,
- $\rho_{x(2)} = \sum_{i=1}^{N_2}\sum_{j<k}^{M}\left(x_{ij} - \overline{X}_{N_2M}\right)\left(x_{jk} - \overline{X}_{N_2M}\right) \Big/ \left\{(M-1)(N_2M-1)S_{x(2)}^2\right\}$: the intra-class correlation coefficient between elements of a cluster on first occasion pertaining to non-response class,
- $\rho_{y(2)} = \sum_{i=1}^{N_2}\sum_{j<k}^{M}\left(y_{ij} - \overline{Y}_{N_2M}\right)\left(y_{jk} - \overline{Y}_{N_2M}\right) \Big/ \left\{(M-1)(N_2M-1)S_{y(2)}^2\right\}$: the intra-class correlation coefficient between elements of a cluster on second occasion pertaining to non-response class,
- $\rho_{b(2)} = \sum_{i=1}^{N_2}\left(\overline{Y}_{i.} - \overline{Y}_{N_2M}\right)\left(\overline{X}_{i.} - \overline{X}_{N_2M}\right) \Big/ \left\{\sum_{i=1}^{N_2}\left(\overline{Y}_{i.} - \overline{Y}_{N_2M}\right)^2 \sum_{i=1}^{N_2}\left(\overline{X}_{i.} - \overline{X}_{N_2M}\right)^2\right\}^{1/2}$: the simple correlation coefficient between cluster means on both the occasions pertaining to the non-response class,

- $\overline{X}_{N_2M} = \sum_{i=1}^{N_2}\sum_{j=1}^{M} x_{ij} / N_2M$ and $\overline{Y}_{N_2M} = \sum_{i=1}^{N_2}\sum_{j=1}^{M} y_{ij} / N_2M$ are the means of x and y per element on the first and second occasions respectively pertaining to the non-response class.

Let $\overline{x}^*_{mM}$ and $\overline{x}^*_{uM}$ denote the Hansen and Hurwitz estimator for matched and unmatched portion of the sample on the first occasion. Let the corresponding estimator for the second occasion be denoted by $\overline{y}^*_{mM}$ and $\overline{y}^*_{uM}$. Thus, we have the following set-up:

I[st] Occasion	$\overline{x}^*_{uM}$	$\overline{x}^*_{mM}$	
2[nd] Occasion		$\overline{y}^*_{mM}$	$\overline{y}^*_{uM}$

where

$$\overline{y}^*_{mM} = \left(m_1 M\, \overline{y}_{m_1 M} + m_2 M\, \overline{y}_{m_{h_2} M}\right) / mM \,,\ \overline{y}^*_{uM} = \left(u_1 M\, \overline{y}_{u_1 M} + u_2 M\, \overline{y}_{u_{h_2} M}\right) / uM \,,$$

$$\overline{x}^*_{mM} = \left(m_1 M\, \overline{x}_{m_1 M} + m_2 M\, \overline{x}_{m_{h_2} M}\right) / mM \,,\ \overline{x}^*_{uM} = \left(u_1 M\, \overline{x}_{u_1 M} + u_2 M\, \overline{x}_{u_{h_2} M}\right) / uM \,,$$

$$\overline{y}_{m_1 M} = \sum_{i=1}^{m_1}\sum_{j=1}^{M} y_{ij} / m_1 M \,,\ \overline{y}_{m_{h_2} M} = \sum_{i=1}^{m_{h_2}}\sum_{j=1}^{M} y_{ij} / m_{h_2} M \,,\ \overline{x}_{m_1 M} = \sum_{i=1}^{m_1}\sum_{j=1}^{M} x_{ij} / m_1 M \,,$$

$$\overline{x}_{m_{h_2} M} = \sum_{i=1}^{m_{h_2}}\sum_{j=1}^{M} x_{ij} / m_{h_2} M \,, \overline{y}_{u_1 M} = \sum_{i=1}^{u_1}\sum_{j=1}^{M} y_{ij} / u_1 M \,, \overline{y}_{u_{h_2} M} = \sum_{i=1}^{u_{h_2}}\sum_{j=1}^{M} y_{ij} / u_{h_2} M \,,$$

$$\overline{x}_{u_1 M} = \sum_{i=1}^{u_1}\sum_{j=1}^{M} x_{ij} / u_1 M \text{ and } \overline{x}_{u_{h_2} M} = \sum_{i=1}^{u_{h_2}}\sum_{j=1}^{M} x_{ij} / u_{h_2} M \,.$$

With these notations and background we define a generalized estimator of the population mean $\overline{Y}$ on second (current) occasion as

$$\hat{D}_{12} = a\overline{x}^*_{uM} + b\overline{x}^*_{mM} + c\overline{y}^*_{mM} + d\overline{y}^*_{uM} \,, \tag{2.1}$$

where *a, b, c* and *d* are suitably chosen constants. We have

$$E\left(\hat{D}_{12}\right) = (a+b)\overline{Y}_{NM} + (c+d)\overline{X}_{NM} \tag{2.2}$$

$$= (a+b)\overline{Y} + (c+d)\overline{X} \text{ (since clusters are of equal size).}$$

In order that $\hat{D}_{12}$ is an unbiased estimator of $\overline{Y}_{NM} \left(or\ \overline{Y}\right)$, we have

$(a+b) = 0$ and $(c+d) = 1$.

Substituting the values of *a* and *d* we obtain

$$\hat{D}_{12} = a(\overline{x}^*_{uM} - \overline{x}^*_{mM}) + c\overline{y}^*_{mM} + (1-c)\overline{y}^*_{uM} \,. \tag{2.3}$$

The variance of $\hat{D}_{12}$ is given by

$$Var\left(\hat{D}_{12}\right)=\left[a^2 Var(\bar{x}^*_{uM})+a^2 Var(\bar{x}^*_{mM})+c^2 Var(\bar{y}^*_{mM})+(1-c)^2 Var(\bar{y}^*_{uM})\right.$$
$$\left.-2acCov(\bar{x}^*_{mM},\bar{y}^*_{mM})\right] \tag{2.4}$$

Assuming that N is sufficiently large, other covariance terms being zero, the variances and covariance involved in (2.4) are given by

$$Var\left(\bar{x}^*_{uM}\right)=\frac{1}{uM}\left\{\rho^*_x S^2_x+W_2(k-1)\rho^*_{x(2)}S^2_{x(2)}\right\},$$

$$Var\left(\bar{y}^*_{uM}\right)=\frac{1}{uM}\left\{\rho^*_y S^2_y+W_2(k-1)\rho^*_{y(2)}S^2_{y(2)}\right\},$$

$$Var\left(\bar{x}^*_{mM}\right)=\frac{1}{mM}\left\{\rho^*_x S^2_x+W_2(k-1)\rho^*_{x(2)}S^2_{x(2)}\right\},$$

$$Var\left(\bar{y}^*_{mM}\right)=\frac{1}{mM}\left\{\rho^*_y S^2_y+W_2(k-1)\rho^*_{y(2)}S^2_{y(2)}\right\},$$

$$Cov\left(\bar{x}^*_{mM},\bar{y}^*_{mM}\right)=\frac{1}{mM}\left\{\rho_b\sqrt{\rho^*_x\rho^*_y}S_yS_x+W_2(k-1)\rho_{b(2)}\sqrt{\rho^*_{x(2)}\rho^*_{y(2)}}S_{y(2)}S_{x(2)}\right\},$$

where $\rho^*_x=\{1+\rho_x(M-1)\}$, $\rho^*_{x(2)}=\{1+\rho_{x(2)}(M-1)\}$, $\rho^*_y=\{1+\rho_y(M-1)\}$ and $\rho^*_{y(2)}=\{1+\rho_{y(2)}(M-1)\}$.

Putting the above variances and covariance expressions in (2.4) we get the variance of $\hat{D}_{12}$ as

$$Var\left(\hat{D}_{12}\right)=\left[\frac{1}{M}\left(\frac{1}{u}+\frac{1}{m}\right)\left\{\left(\rho^*_x S^2_x+W_2(k-1)\rho^*_{x(2)}S^2_{x(2)}\right)a^2+\left(\rho^*_y S^2_y+W_2(k-1)\rho^*_{y(2)}S^2_{y(2)}\right)c^2\right\}\right.$$
$$+\frac{1}{uM}(1-2c)\left(\rho^*_y S^2_y+W_2(k-1)\rho^*_{y(2)}S^2_{y(2)}\right)$$
$$\left.-\frac{2ac}{mM}\left(\rho_b\sqrt{\rho^*_x\rho^*_y}S_yS_x+W_2(k-1)\rho_{b(2)}\sqrt{\rho^*_{x(2)}\rho^*_{y(2)}}S_{y(2)}S_{x(2)}\right)\right]. \tag{2.5}$$

which is minimized for

$$\left.\begin{aligned} a&=\frac{\lambda\mu V_yV_{xy}}{\left(V_xV_y-\mu^2V^2_{xy}\right)}\\ c&=\frac{\lambda V_yV_x}{\left(V_xV_y-\mu^2V^2_{xy}\right)}\end{aligned}\right\} \tag{2.6}$$

where

$$V_y=\left(\rho^*_y S^2_y+W_2(k-1)\rho^*_{y(2)}S^2_{y(2)}\right),$$
$$V_x=\left(\rho^*_x S^2_x+W_2(k-1)\rho^*_{x(2)}S^2_{x(2)}\right),$$

$$V_{xy} = \left(\rho_b \sqrt{\rho_x^* \rho_y^*} S_y S_x + W_2 (k-1) \rho_{b(2)} \sqrt{\rho_{x(2)}^* \rho_{y(2)}^*} S_{y(2)} S_{x(2)} \right).$$

Substituting the values of a and c the minimum linear unbiased estimator for the population mean on the second occasion is given by

$$\hat{D}_{12}^* = \left[\frac{\lambda V_y V_x}{\left(V_x V_y - \mu^2 V_{xy}^2\right)} \left\{ \left(\frac{\mu V_{xy}}{V_x} \right) \left(\overline{x}_{uM}^* - \overline{x}_{mM}^* \right) + \overline{y}_{mM}^* \right\} + \left(\frac{\mu V_x V_y - \mu^2 V_{xy}^2}{V_x V_y - \mu^2 V_{xy}^2} \right) \overline{y}_{uM}^* \right]. \tag{2.7}$$

The variance of $\hat{D}_{12}^*$ is obtained as

$$Var\left(\hat{D}_{12}^*\right) = \frac{\left(\mu V_x V_y - \mu^2 V_{xy}^2\right) V_y}{nM\mu\left(V_x V_y - \mu^2 V_{xy}^2\right)} = \frac{\left(V_x V_y - \mu V_{xy}^2\right) V_y}{nM\left(V_x V_y - \mu^2 V_{xy}^2\right)}. \tag{2.8}$$

In case there is no-response then $Var\left(\hat{D}_{12}^*\right)$ reduces to

$$Var\left(\hat{D}_0^*\right) = \frac{\left(1 - \mu\rho_b^2\right) V_y}{nM\left(1 - \mu^2 \rho_b^2\right)} \tag{2.9}$$

which is the same as obtained by Pradhan (2004) for the optimum estimator of the population mean on the current occasion in the context of sampling on two occasion when there is non-response.

$$\hat{D}_0^* = \left[\frac{S_y}{S_x} \left\{ \frac{1 + (M-1)\rho_y}{1 + (M-1)\rho_x} \right\}^{1/2} \frac{\mu(1-\mu)\rho_b}{(1-\mu^2\rho_b^2)} \left(\overline{X}_{uM} - \overline{X}_{mM} \right) + \frac{(1-\mu)}{(1-\mu^2\rho_b^2)} \overline{Y}_{uM} + \left\{ 1 - \frac{(1-\mu)}{(1-\mu^2\rho_b^2)} \right\} \overline{Y}_{mM} \right]. \tag{2.10}$$

Minimizing the variance of $\hat{D}_{12}^*$ given at (2.8) yields the optimum fraction of unmatched and matched portion of the sample as

$$\mu_{opt} = \frac{V_x V_y}{V_x V_y + \sqrt{V_x^2 V_y^2 - V_x V_y V_{xy}^2}}, \tag{2.11}$$

$$\lambda_{opt} = \frac{\sqrt{V_x^2 V_y^2 - V_x V_y V_{xy}^2}}{V_x V_y + \sqrt{V_x^2 V_y^2 - V_x V_y V_{xy}^2}}. \tag{2.12}$$

Thus, the resulting minimum value of the variance $\hat{D}_{12}^*$ is given by

$$\min.Var\left(\hat{D}_{12}^*\right) = \frac{V_{xy}^2 V_y}{2nM\left(V_x V_y - \sqrt{V_x^2 V_y^2 - V_x V_y V_{xy}^2} \right)}. \tag{2.13}$$

Remark 2.1

When there is non-response only on first occasion, then minimum variance unbiased linear estimator (MVULE) for the estimator for the population mean on current occasion can be obtained as follows:

$$\hat{D}_1 = \left\{a\left(\bar{x}^*_{uM} - \bar{x}^*_{mM}\right) + c\bar{y}_{mM} + (1-c)\bar{y}_{uM}\right\}. \tag{2.14}$$

The variance of $\hat{D}_1$ is given by

$$Var\left(\hat{D}_1\right) = \frac{1}{M}\left[\left\{\frac{1}{u}(1-2c) + \left(\frac{1}{m} + \frac{1}{u}\right)c^2\right\}V_{y(1)} - 2ac\frac{V_{xy(1)}}{m} + a^2\left(\frac{1}{m} + \frac{1}{u}\right)V_x\right], \tag{2.15}$$

where $V_{y(1)} = \rho^*_y S^2_y$ and $V_{xy(1)} = \rho_b S_y S_x \sqrt{\rho^*_y S^2_x}$.

The variance of $\hat{D}_1$ is minimized for

$$\left.\begin{aligned} a &= \frac{\lambda\mu V_{y(1)} V_{xy(1)}}{V_x V_{y(1)} - \mu^2 V^2_{xy(1)}} \\ c &= \frac{\lambda V_x V_{y(1)}}{V_x V_{y(1)} - \mu^2 V^2_{xy(1)}} \end{aligned}\right\}. \tag{2.16}$$

Thus, the MVULE for the current occasion in this case is given by

$$\hat{D}^*_1 = \left[\left(\frac{\lambda V_x V_{y(1)}}{V_x V_{y(1)} - \mu^2 V^2_{xy(1)}}\right)\left\{\left(\frac{\mu V_{xy(1)}}{V_x}\right)\left(\bar{x}^*_{uM} - \bar{x}^*_{mM}\right) + \bar{y}^*_{mM}\right\} + \left(\frac{\mu V_x V_{y(1)} - \mu^2 V^2_{xy(1)}}{V_x V_{y(1)} - \mu^2 V^2_{xy(1)}}\right)\bar{y}^*_{uM}\right] \tag{2.17}$$

Substituting (2.16) in (2.15) we get the variance of $\hat{D}^*_1$ as

$$Var\left(\hat{D}^*_1\right) = \frac{V_x V_{y(1)} - \mu V^2_{xy(1)}}{nM\left(V_x V_{y(1)} - \mu^2 V^2_{xy(1)}\right)}. \tag{2.18}$$

The optimum fraction to be unmatched is given by

$$\left.\begin{aligned} \mu_{opt} &= \frac{V_x V_{y(1)}}{V_x V_{y(1)} + \sqrt{V^2_x V^2_{y(1)} - V_x V_{y(1)} V^2_{xy(1)}}} \\ \lambda_{opt} &= \frac{\sqrt{V^2_x V^2_{y(1)} - V_x V_{y(1)} V^2_{xy(1)}}}{V_x V_{y(1)} + \sqrt{V^2_x V^2_{y(1)} - V_x V_{y(1)} V^2_{xy(1)}}} \end{aligned}\right\}. \tag{2.19}$$

Thus, the resulting minimum value of $Var\left(\hat{D}_1^*\right)$ at (2.18) is given by

$$\min.Var\left(\hat{D}_1^*\right)=\frac{V_{xy(1)}^2 V_{y(1)}}{2nM\left(V_x V_{y(1)}-\sqrt{V_x^2 V_{y(1)}^2-V_x V_{y(1)} V_{xy(1)}^2}\right)}. \tag{2.20}$$

Remark 2.2

When there is non-response only on second occasion, the MVULE for population mean on the current occasion can be obtained as follows:

$$\hat{D}_2=\left\{a\left(\bar{x}_{uM}-\bar{x}_{mM}\right)+c\,\bar{y}_{mM}^*+(1-c)\bar{y}_{uM}^*\right\}. \tag{2.21}$$

The variance of $\hat{D}_2$ is given by

$$Var\left(\hat{D}_2\right)=\frac{1}{M}\left[\frac{1}{mu}\left\{c^2+(1-2c)\lambda\right\}V_y-2ac\frac{V_{xy(1)}}{n\lambda}+\frac{a^2}{mu}V_{x(1)}\right], \tag{2.22}$$

where $V_{x(1)}=\rho_x^* S_x^2$.

The variance of $\hat{D}_2$ is minimized for

$$\left.\begin{aligned} a&=\frac{\lambda\mu V_y V_{xy(1)}}{V_{x(1)}V_y-\mu^2 V_{xy(1)}^2}\\ c&=\frac{\lambda V_{x(1)}V_y}{V_{x(1)}V_y-\mu^2 V_{xy(1)}^2}\end{aligned}\right\}.$$

(2.23)

Thus, the MVULE in this case is given by

$$\hat{D}_2^*=\left[\left(\frac{\lambda V_{x(1)}V_y}{V_{x(1)}V_y-\mu^2 V_{xy(1)}^2}\right)\left\{\left(\frac{\mu V_{xy(1)}}{V_{x(1)}}\right)\left(\bar{x}_{uM}-\bar{x}_{mM}\right)+\bar{y}_{mM}^*\right\}+\left(\frac{\mu V_{x(1)}V_y-\mu^2 V_{xy(1)}^2}{V_{x(1)}V_y-\mu^2 V_{xy(1)}^2}\right)\bar{y}_{uM}^*\right] \tag{2.24}$$

Substituting (2.23) in (2.22) we get the variance of $\hat{D}_2^*$ as

$$Var\left(\hat{D}_2^*\right)=\frac{V_{x(1)}V_y-\mu V_{xy(1)}^2}{nM\left(V_{x(1)}V_y-\mu^2 V_{xy(1)}^2\right)}. \tag{2.25}$$

The optimum fraction to be unmatched is given by

$$\left.\begin{aligned}\mu_{opt}&=\frac{V_{x(1)}V_y}{V_{x(1)}V_y+\sqrt{V_{x(1)}^2V_y^2-V_{x(1)}V_yV_{xy(1)}^2}}\\ \lambda_{opt}&=\frac{\sqrt{V_{x(1)}^2V_y^2-V_{x(1)}V_yV_{xy(1)}^2}}{V_{x(1)}V_y-\sqrt{V_{x(1)}^2V_y^2-V_{x(1)}V_yV_{xy(1)}^2}}\end{aligned}\right\}. \tag{2.26}$$

Thus, the resulting minimum value of $Var\left(\hat{D}_2^*\right)$ at (2.25) is given by

$$\min.Var\left(\hat{D}_2^*\right)=\frac{V_{xy(1)}^2 V_y}{2nM\left(V_{x(1)}V_y-\sqrt{V_{x(1)}^2V_y^2-V_{x(1)}V_yV_{xy(1)}^2}\right)}. \tag{2.27}$$

3. Comparison between variances of the estimator

3.1. Comparison between variance of $\hat{D}_{12}^*$ and $\hat{D}_2^*$.

We note from (2.13) and (2.27) that
$\min.Var\left(\hat{D}_{12}^*\right)>\min.Var\left(\hat{D}_2^*\right)$ if

$$\frac{V_{xy}^2}{\left(V_xV_y-\sqrt{V_x^2V_y^2-V_xV_yV_{xy}^2}\right)}>\frac{V_{xy(1)}^2}{\left(V_{x(1)}V_y-\sqrt{V_{x(1)}^2V_y^2-V_{x(1)}V_yV_{xy(1)}^2}\right)}. \tag{3.1}$$

3.2. Comparison between variance of $\hat{D}_{12}^*$ and $\hat{D}_1^*$.

We note from (2.13) and (2.20) that
$\min.Var\left(\hat{D}_{12}^*\right)>\min.Var\left(\hat{D}_1^*\right)$ if

$$\frac{V_y}{V_{y(1)}}\left(\frac{V_{xy}}{V_{xy(2)}}\right)^2>\frac{\left(V_{x^2}V_y-\sqrt{V_x^2V_y^2-V_xV_yV_{xy}^2}\right)}{\left(V_xV_{y(1)}-\sqrt{V_x^2V_{y(1)}^2-V_xV_{y(1)}V_{xy(1)}^2}\right)}. \tag{3.2}$$

For the sake of convenience we assume that

$\rho_x=\rho_y=\rho\,(say)\Rightarrow\rho_x^*=\rho_y^*=\{1+\rho(M-1)\}=\rho^*\,(say)$,

$\rho_{x(2)}=\rho_{y(2)}=\rho_{(2)}\,(say)\Rightarrow\rho_{x(2)}^*=\rho_{y(2)}^*=\{1+\rho_{(2)}(M-1)\}=\rho_{(2)}^*\,(say)$,

$S_x^2=S_y^2=S^2\,(say)$, $S_{x(2)}^2=S_{y(2)}^2=S_{(2)}^2\,(say)$.

Thus $Var\left(\hat{D}_0^*\right)$, $Var\left(\hat{D}_1^*\right)$, $Var\left(\hat{D}_2^*\right)$ and $Var\left(\hat{D}_{12}^*\right)$ respectively in (2.9), (2.16), (2.23) and (2.8) reduce to:

$$Var\left(\hat{D}_0^*\right)=\frac{\left(1-\mu\rho_b^2\right)V}{nM\left(1-\mu^2\rho_b^2\right)}, \tag{3.3}$$

$$Var\left(\hat{D}_1^*\right) = Var\left(\hat{D}_2^*\right) = \frac{VV_{(1)} - \mu V_{(1)}^*}{nM\left(VV_{(1)} - \mu^2 V_{(1)}^{*2}\right)}, \tag{3.4}$$

$$Var\left(\hat{D}_{12}^*\right) = \frac{\left(V^2 - \mu V^{*^2}\right)V}{nM\left(V^2 - \mu^2 V^{*^2}\right)}, \tag{3.5}$$

where $V = V_y = V_x = \left\{\rho^* S^2 + W_2(k-1)\rho_{(2)}^* S_{(2)}^2\right\}$,

$V^* = V_{xy} = \left\{\rho_b \rho^* S^2 + W_2(k-1)\rho_{b(2)}\rho_{(2)}^* S_{(2)}^2\right\}$,

$V_{(1)} = V_{y(1)} = V_{x(1)} = \rho^* S^2$ and $V_{(1)}^* = V_{xy(1)} = \rho_b S^2 \sqrt{\rho^* S^2}$.

Hence, the expression for the percentage loss in precision of $\hat{D}_{12}^*$ over $\hat{D}_0^*$ is given by

$$L_{12} = \left\{\frac{Var\left(\hat{D}_{12}^*\right)}{Var\left(\hat{D}_0^*\right)} - 1\right\} \times 100, \tag{3.6}$$

where $Var\left(\hat{D}_{12}^*\right)$ and $Var\left(\hat{D}_0^*\right)$ are respectively given by (3.6) and (3.3).

The percentage loss in precision of $\hat{D}_2^*$ (or $\hat{D}_1^*$) over $\hat{D}_0^*$ can be achieved by replacing $Var\left(\hat{D}_{12}^*\right)$ with $Var\left(\hat{D}_2^*\right)$(or $Var\left(\hat{D}_1^*\right)$) given by (3.4). The percentage loss in precision of $\hat{D}_{12}^*, \hat{D}_2^*$ (or $\hat{D}_1^*$) over $\hat{D}_0^*$ for different values of W_2, $(k-1)$, M, μ, ρ, $\rho_{(2)}, \rho^*, \rho_{(2)}^*, \rho_b, \rho_{b(2)}, S^2$ and $S_{(2)}^2$ shown in Table 1 and 2. It is assumed that $N = 300$ and $n = 50$.

Table 1. Percentage loss in precision of $\hat{D}_{12}^{*}, \hat{D}_{2}^{*}$ (or $\hat{D}_{1}^{*}$) over $\hat{D}_{0}^{*}$ for different values of W_2, $(k-1)$, μ, ρ, $\rho_{(2)}$, ρ_b, $\rho_{b(2)}$, S^2 and $S_{(2)}^2$ when $M=2$.

W_2	$(k-1)$	μ	ρ	$\rho_{(2)}$	ρ^*	$\rho_{(2)}^*$	ρ_b	$\rho_{b(2)}$	S^2	$S_{(2)}^2$	L_{12}	L_1
$\rho<\rho_{(2)}$												
0.7	2.5	0.5	0.1	0.8	1.1	1.8	0.8	0.2	0.4	0.2	17.04	12.62
0.7	2.5	0.5	0.2	0.7	1.2	1.7	0.8	0.2	0.4	0.2	16.37	12.12
0.7	2.5	0.5	0.3	0.6	1.3	1.6	0.8	0.2	0.4	0.2	15.68	11.61
0.7	2.5	0.5	0.4	0.5	1.4	1.5	0.8	0.2	0.4	0.2	14.95	11.10
$\rho>\rho_{(2)}$												
0.7	2	0.5	0.8	0.5	1.8	1.5	0.7	0.2	0.4	0.2	7.94	0.17
0.7	2	0.5	0.8	0.4	1.8	1.4	0.7	0.2	0.4	0.2	7.67	2.64
0.7	2	0.5	0.8	0.3	1.8	1.3	0.7	0.2	0.4	0.2	7.37	5.24
0.7	2	0.5	0.8	0.2	1.8	1.2	0.7	0.2	0.4	0.2	7.05	7.98
$\rho=\rho_{(2)}$												
0.4	2	0.4	0.7	0.7	1.7	1.7	0.8	0.2	0.5	0.2	7.33	3.76
0.4	2	0.4	0.7	0.7	1.7	1.7	0.8	0.2	0.5	0.2	7.33	3.76
0.4	2	0.4	0.7	0.7	1.7	1.7	0.8	0.2	0.5	0.2	7.33	3.76
0.4	2	0.4	0.7	0.7	1.7	1.7	0.8	0.2	0.5	0.2	7.33	3.76
$\rho_b<\rho_{b(2)}$												
0.7	1.5	0.3	0.5	0.8	1.5	1.8	0.4	0.5	0.8	0.2	*	*
0.7	1.5	0.3	0.5	0.8	1.5	1.8	0.3	0.5	0.8	0.2	*	*
0.7	1.5	0.3	0.5	0.8	1.5	1.8	0.2	0.5	0.8	0.2	*	*
0.7	1.5	0.3	0.5	0.8	1.5	1.8	0.1	0.5	0.8	0.2	*	*
$\rho_b>\rho_{b(2)}$												
0.6	2.5	0.7	0.5	0.7	1.5	1.7	0.7	0.4	0.4	0.2	6.58	2.59
0.6	2.5	0.7	0.5	0.7	1.5	1.7	0.7	0.3	0.4	0.2	8.22	2.59
0.6	2.5	0.7	0.5	0.7	1.5	1.7	0.7	0.2	0.4	0.2	9.64	2.59
0.6	2.5	0.7	0.5	0.7	1.5	1.7	0.7	0.1	0.4	0.2	10.87	2.59
$\rho_b=\rho_{b(2)}$												
0.6	1.5	0.5	0.4	0.5	1.4	1.5	0.2	0.2	0.5	0.3	2.22	*
0.6	1.5	0.5	0.4	0.5	1.4	1.5	0.2	0.2	0.5	0.3	2.22	*
0.6	1.5	0.5	0.4	0.5	1.4	1.5	0.2	0.2	0.5	0.3	2.22	*
0.6	1.5	0.5	0.4	0.5	1.4	1.5	0.2	0.2	0.5	0.3	2.22	*
$S^2<S_{(2)}^2$												
0.6	1.5	0.4	0.5	0.7	1.5	1.7	0.7	0.2	0.1	0.6	12.59	7.21
0.6	1.5	0.4	0.5	0.7	1.5	1.7	0.7	0.2	0.2	0.5	11.37	7.15
0.6	1.5	0.4	0.5	0.7	1.5	1.7	0.7	0.2	0.3	0.4	9.83	6.85
0.6	1.5	0.4	0.5	0.7	1.5	1.7	0.7	0.2	0.4	0.3	7.95	6.29

$S^2 > S^2_{(2)}$												
0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
$S^2 = S^2_{(2)}$												
0.5	0.5	0.8	0.2	0.5	1.2	1.5	0.8	0.4	0.7	0.7	6.88	3.63
0.5	0.5	0.8	0.2	0.5	1.2	1.5	0.8	0.4	0.7	0.7	6.88	3.63
0.5	0.5	0.8	0.2	0.5	1.2	1.5	0.8	0.4	0.7	0.7	6.88	3.63
0.5	0.5	0.8	0.2	0.5	1.2	1.5	0.8	0.4	0.7	0.7	6.88	3.63
W_2												
0.7	0.5	0.6	0.5	0.8	1.5	1.8	0.8	0.1	0.7	0.2	5.49	*
0.6	0.5	0.6	0.5	0.8	1.5	1.8	0.8	0.1	0.7	0.2	4.85	0.12
0.5	0.5	0.6	0.5	0.8	1.5	1.8	0.8	0.1	0.7	0.2	4.16	1.56
0.4	0.5	0.6	0.5	0.8	1.5	1.8	0.8	0.1	0.7	0.2	3.42	3.05
μ												
0.8	0.5	0.8	0.6	0.7	1.6	1.7	0.8	0.2	0.5	0.5	11.07	3.51
0.8	0.5	0.7	0.6	0.7	1.6	1.7	0.8	0.2	0.5	0.5	11.93	5.62
0.8	0.5	0.6	0.6	0.7	1.6	1.7	0.8	0.2	0.5	0.5	11.51	5.70
0.8	0.5	0.5	0.6	0.7	1.6	1.7	0.8	0.2	0.5	0.5	10.35	4.42
$(k-1)$												
0.8	2	0.4	0.5	0.8	1.5	1.8	0.8	0.1	0.2	0.3	18.34	3.09
0.8	1.5	0.4	0.5	0.8	1.5	1.8	0.8	0.1	0.3	0.3	16.20	8.48
0.8	1	0.4	0.5	0.8	1.5	1.8	0.8	0.1	0.4	0.4	14.38	0.46
0.8	0.5	0.4	0.5	0.8	1.5	1.8	0.8	0.1	0.5	0.5	10.61	4.84

Table 1 exhibit that for all cases, i.e. $\rho(<,=,>)\rho_{(2)}$, $\rho_b(<,=,>)\rho_{b(2)}$, $S^2(<,=,>)S^2_{(2)}$, W_2, μ and $(k-1)$, when $M=2$, the percentage loss in precision ver $\hat{D}_0^*$ is maximum in $\hat{D}_{12}^*$ while it is least in $\hat{D}_1^*$ (or $\hat{D}_2^*$). It is noted that in some cases the percentage loss in precision of $\hat{D}_{12}^*$ over $\hat{D}_0^*$ is less than that of $\hat{D}_1^*$ (or $\hat{D}_2^*$) which contravene the condition (3.1). For the cases $\rho < \rho_{(2)}$, $S^2 < S^2_{(2)}$ and $S^2 > S^2_{(2)}$, the percentage loss in precision of all the estimators over $\hat{D}_0^*$ decreases. For the case $\rho > \rho_{(2)}$, the percentage loss in precision decreases in $\hat{D}_{12}^*$ while it increases in $\hat{D}_1^*$ (or $\hat{D}_2^*$). However, for the cases $\rho = \rho_{(2)}$, $\rho_b = \rho_{b(2)}$ and $S^2 = S^2_{(2)}$, the percentage loss in precision remains constant for all the estimators. For $\rho_b > \rho_{b(2)}$, the percentage loss in precision increases in $\hat{D}_{12}^*$ and remains constant in $\hat{D}_1^*$ (or $\hat{D}_2^*$). For different values of W_2, the percentage loss in precision of all the estimators over $\hat{D}_0^*$ increases in $\hat{D}_{12}^*$ while it decreases in $\hat{D}_1^*$ (or $\hat{D}_2^*$). For the case of μ, the percentage loss in precision of all the estimators over $\hat{D}_0^*$ first increases and then decreases. For different values of $(k-1)$, the percentage loss in precision decreases in case of $\hat{D}_{12}^*$ and in $\hat{D}_1^*$ (or $\hat{D}_2^*$) it first increases and then decreases.

4. Conclusions

In sampling on two occasions we have considered the estimation of population mean on second occasion when the sampling units are clusters and the observations on the first occasion are regarded as ancillary information for the observations on the second or current occasion. In many surveys, information is usually not obtained from all the sample units even after callbacks. The technique of Hansen and Hurwitz (1946) for estimating population mean for current occasion in the context of cluster sampling over sampling on two successive occasions has been developed. Three different possible cases when there is non-response (i) on both the occasions, (ii) only on the first occasion, and (iii) only on the second occasion, have been discussed. The loss in precision by using the estimators $\hat{D}_{12}^*$ (optimum estimate in case (i)), $\hat{D}_1^*$ (optimum estimate in case (ii)) and $\hat{D}_2^*$ (optimum estimate in case (iii)) over direct estimator $\bar{y}^*$ of the population mean $\bar{Y}$ on the second occasion using no information gathered on the first occasion, have been computed. It is envisaged numerically that the loss in precision due to $\hat{D}_{12}^*$ over $\bar{y}^*$ is maximum as compared to $\hat{D}_1^*$ and $\hat{D}_2^*$ for $m =$ 2 and 4. Thus, the present study is recommended when there is need to correct for non-response in cluster sampling over two occasions.

Acknowledgement

The Authors express sincere thanks to the editor and the referee for their useful suggestions which have helped in no small way to the improvement of the quality of this paper.

REFERENCES

ARNAB, R. (1980): Two-stage sampling over two occasions. *Aust. J. Statist., 22,* 349-357.

COCHRAN, W. G. (1977): *Sampling Techniques*, 3rd edition, John Wiley and Sons. New York.

CHOUDHARY, R. K., BATHLA, H. V. L. and SUD, U. C. (2004): On non-response in sampling on two occasions. *J. Indian Soc. Agricultural Statist., 58,* 331-343.

FRAYER, W. E. and FURNIVAL, G. M. (1967): Area change estimates from sampling with partial replacement. *Forest Science, 13,* 72-77.

HANSEN, M. H. and HURWITZ, W. N. (1994): The problem of the non-response in sample surveys. *J. Amer. Statist. Assoc., 41,* 517-529.

JESSEN, R. K. (1942): Statistical investigation of a sample survey for obtaining farm Facts. *Iowa Agr. Exp. Stat. Res. Bull.,* 304.

KATHURIA, O. P. (1975): Some estimators in two-stage sampling on successive occasions with partial matching at both stages. *Sankhya, 37,* 147-162.

KHARE, B. B. and SRIVASTAVA, S. (1993): Estimation of population mean using auxiliary character in presence of nonresponse. *Nat. Acad. Sc. Letters, India, 16(3),* 111-114.

KHARE, B. B. and SRIVASTAVA, S. (1995): Study of conventional and alternative two -phase sampling ratio, product and regression estimators in presence of nonresponse. *Proc. Nat. Acad. Sci. India, 65(A), II,* 195-203.

KHARE, B. B. and SRIVASTAVA, S. (1997): Transformed ratio type estimators for the population mean in the presence of nonresponse. *Comm. Statist. - Theory Methods, 26(7),* 1779-1791.

OKAFOR, F. C. (2001): Treatment of nonresponse in successive sampling, *Statistica, LXI, (2),* 195-204.

OKAFOR, F. C. (2005): Sub-sampling the nonrespondents in two-stage sampling over two successive occasions. *Jour. Ind. Statist. Assoc., 43,* 33-49.

OKAFOR, F. C. and LEE, H. (2000): Double sampling for ratio and regression estimation with sub-sampling the nonrespondent. *Survey Methodology, 26,* 183-188.

PRADHAN, B. K. (2004): On efficient of cluster sampling on sampling on two Occasions. *Statistica, anno LXIV, 1,* 183-191.

RAJ, D. (1979): *Sampling Theory*, Tata Mc Graw Hill, New Delhi, 152-162.

RAO, P. S. R. S. (1986): Ratio estimation with sub sampling the non-respondents. *Survey methodology, 12(2),* 217-230.

SINGH, D. (1968): Estimates in successive sampling using a multistage design,. *J. Amer. Statist. Assoc., 63,* 99-112.

SINGH, H. P. and KUMAR, S. (2008 a): Estimation of population product in presence of non-response in successive sampling. *Statistical Papers, 51(4),* 975-996.

SINGH, H. P. and KUMAR, S. (2008 b): Effect of non-response in sampling over two successive occasions using auxiliary information. Statistics in Transition – new series, 9, (2), 273-296.

SINGH, H. P., SINGH, HARI P. and SINGH, V. P. (1992): A generalized efficient class of estimators of population mean in two-phase and successive sampling. *Int. Jour. Management and Systems, 8, (2),* 173-183.

SINGH, H. P. and VISHWAKARMA, G. K. (2007): A general class of estimators in successive sampling. *METRON, LXV, (2),* 201-227.

SUKHATME, P. V. and SUKHATME, B. V. (1970): *Sampling theory of surveys with Applications.* Food and Agriculture Organization, Rome, Second Edition.

TABASUM, R.. and KHAN, I. A. (2004): Double sampling for ratio estimation with non Response. *Jour. Ind. Soc. Agril. Statist., 58(3),* 300-306.

TABASUM, R.. and KHAN, I. A. (2006): Double Sampling Ratio Estimator for the population mean in Presence of Non-Response. *Assam Statist. Review, 20(1),* 73-83.

WARE, K. D. and CUNIA, T. (1962): Continuous forest inventory with partial replacement of samples. *Forest Science Monograph No. 3, Society of American Foresters, Bethesda.*

ZARKOVICH, S. S. and KRANE, J. (1965): Some efficient ways of cluster sampling. *Proceedings of 35th session of International Statistical Institute, Belgrade*.