Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF AGRICULTURAL STATISTICS 68(1) 2014 39-54

Estimation of Domain Mean using Two Stage Sampling with Sub-Sampling of Non-respondents

Kaustav Aditya, U.C. Sud and Hukum Chandra
Indian Agricultural Statistics Research Institute, New Delhi

Received 30 January 2012; Revised 24 September 2013; Accepted 14 October 2013

SUMMARY

The problem of estimation of domain mean in the presence of nonresponse is considered when the sampling design is two-stage and the response mechanism is assumed to follow deterministic model. In particular, three different cases of occurrence of non-response are considered and accordingly suitable estimators based on sub-sampling of non-respondents are proposed. Expressions for the variances of the proposed estimators are also suggested. Optimum values of sample sizes are also obtained by considering a suitable cost function. In our empirical evaluation, the percentage reduction in the expected cost are computed to examine the efficiency of the proposed estimators.

Keywords: Cost function, Nonresponse, Population mean, Sub-sampling, Two-stage sampling, Percentage reduction in the expected cost.

1. INTRODUCTION

For large or medium scale surveys we are often faced with the scenario that the sampling frame of ultimate stage response units is not available and the cost of construction of the frame is very high. Sometimes the population elements are scattered over a wide area resulting in a widely scattered sample. Therefore, not only the cost of enumeration of units in such a sample may be very high, the supervision of field work may also be very difficult. For such situations, two-stage or multi-stage sampling designs are very effective. It is also the case that, in many human surveys, information is not available from all the units in surveys. The problem of nonresponse persists even after call backs. The estimates obtained from incomplete data may be biased, particularly when the respondents differ from the non-respondents. Hansen and Hurwitz (1946) proposed a technique for adjusting for nonresponse to address the problem of bias. The technique consists of selecting a sub-sample of non-respondents. Through specialized efforts data are collected from the non-respondents so as to obtain an estimate for non-responding units in the population. Foradori (1961) studied the sub-sampling of the non-respondents technique to estimate the population total in two stages using unequal probability sampling. Srinath (1971) used a different procedure for selecting the sub-sample of respondents where the sub-sampling procedure varied according to the nonresponse rates.

Oh and Scheuren (1983) attempted to compensate for nonresponse by weighing adjustment. Kalton and Karsprzyk (1986) tried the imputation technique. Tripathi and Khare (1997) extended the sub-sampling of non-respondents approach to multivariate case. Okafor and Lee (2000) extended the approach to double sampling for ratio and regression estimation. Okafor (2001, 2005) further extended the approach in the context of element sampling and two-phase sampling respectively on two successive occasions. It may be mentioned that the weighting and imputation procedures aim at elimination of bias caused by nonresponse. However, these procedures are based on

Corresponding author: Kaustav Aditya
E-mail address : kaustav@iasri.res.in

certain assumptions on the response mechanism. When these assumptions do not hold good the resulting estimate may be seriously biased. Further, when the nonresponse is confounded, i.e. the response probability is dependent on the survey character, it becomes difficult to eliminate the bias entirely. Rancourt *et al.* (1994) provided a partial correction for the situation. Hansen and Hurwitz's sub-sampling approach although costly, is free from any assumptions. When the bias caused by nonresponse is serious this technique is very effective, *i.e*. one does not have to go for 100 percent response, which can be substantially more expensive.

Besides the population estimates, the estimates for different subgroups/domains of population are often required (Sarndal *et al.* 1992). In the context of estimation of the domain parameters (mean/total), Agrawal and Midha (2007) proposed a two phase sampling design when the size of the domain was not known. Sud *et al.* (2010) considered the problem of estimation of finite population mean of a domain in the presence of nonresponse when the response mechanism follows deterministic model. Chhikara and Sud (2009) used the sub-sampling of non-respondents approach for estimation of population and domain totals in the context of item nonresponse. However, the results in both of these studies were limited to single-stage sampling design. Again, Aditya *et al.* (2013) used the sub-sampling of non-respondents approach for estimation of domain totals under two stage sampling design when the size of the domain is assumed unknown. Sarndal *et al*. (1992) also described the case of domain estimation when domain sizes are known. In real life there are situations when domain sizes are known. For example, let us consider case of crop estimation surveys for mixed crops, where there are more than one crops cultivated simultaneously and interest lies in estimation of crop area. Here, domains can be formed with respect to crops, that is, crop domains. In this case population sizes for domains are known. In what follows, three different cases of estimation of the domain mean (or total) have been considered based on three different case of occurrence of nonresponse in two-stage sampling design. We took an example of socio-economic survey to explain these three cases of occurrence of nonresponse. Here, different socio-economic classes can be considered as a domain and surveying villages can be taken as first-stage units and the households within the villages as the second-stage units. There may be situation, where some households within the selected villages do not respond at the first attempt of data collection through mail/postal enquiry creating a nonresponse situation in the ssus. This situation is termed as Case 1 of nonresponse. Under Case 2 of nonresponse, we have considered the situation where all the persons belonging to some of the selected villages respond at the first attempt of data collection whereas in the remaining selected villages some households do not responds at the first attempt. In Case 3 we have considered the situation where there is full response in some of the selected villages, partial non-response in some other selected villages and complete non-response in the remaining selected villages. Accordingly, for these cases, estimators of domain mean (or total) are proposed under two-stage sampling design using sub sampling of non-respondents technique in Section 2. Here the response mechanism is assumed to follow deterministic model. Also given in this section are expressions for variance of the proposed estimators under three cases of occurrence of nonresponse. Besides, optimum values of sample sizes are obtained by minimizing the expected cost for a fixed variance. The results are empirically illustrated in Section 3.

2. THEORETICAL DEVELOPMENTS

Let the finite population U under consideration consists of N known primary stage units (psus) labeled 1 through N. Let the i-th psu comprise M second stage units (ssus). Let us consider a population $U = (1, ..., k, ..., N)$ of size N partitioned into D sub-sets $U_1, ..., U_d, ..., U_D$ (hereafter referred as domains) and let N_d be the size of U_d $(d = 1, ..., D)$ such that $U = \bigcup_{d=1}^{D} U_d$ and $N = \sum_{d=1}^{D} N_d$. Here, N_d is assumed to be known and large. We assume that a sample s of n psus is drawn from the population under srswor sampling design. Let s_d denote the part of sample s that happens to fall in U_d, that is, $s_d = s \cap U_d$. Let us denote by n_d the size of s_d such that $s = \bigcup_{d=1}^{D} s_d$ and $n = \sum_{d=1}^{D} n_d$. Note that throughout this article the sample size n is fixed and known; however, the domain sample size n_d is random with $E(n_d) = n(N_d / N)$. Here, N_d is assumed known, hence the expected value of n_d is also known. Further, when the domain sizes are small, n_d may turn out to be very small or it may be equal to '0' in some cases. In such cases small area estimation techniques are needed for reliable estimation at the domain level. However, we do not consider this case of small area estimation here. Let M_d be the size of the units in each psu belonging to the d-th domain, and from each selected psu m_d ssus are selected by srswor, and letters/mails containing questionnaires are sent to each unit in the sample. With the random sample of observations, the statistician's task is to make the best possible estimate for the domain. Let y_{dkj} be the value of study character pertaining to j-th ssu in the k-th psu in d-th domain, $k = 1, 2, \ldots, N_d$, $j = 1, 2, \ldots, M_d$, $d = 1, 2, \ldots, D$. Our objective here is to estimate the domain mean which is given as

$$\bar{Y}_d = \frac{1}{N_d}\sum_{k=1}^{N_d}\frac{1}{M_d}\sum_{j=1}^{M_d} y_{dkj} \text{ or total } Y_d = \sum_{k=1}^{N_d}\sum_{j=1}^{M_d} y_{dkj} .$$

Case 1. Let n psus be selected from N psus by srswor design where n_d out of n psus fall in the d-th domain randomly and within each selected psu, m_d ssus are also selected from M_d ssus by srswor. Let M_d be divided into two groups M_{dk_1} and M_{dk_2} such that $M_{dk_1} + M_{dk_2} = M_d$, it is being assumed that M_{dk_1} comprises of the responding group while M_{dk_2} comprises the nonresponding groups. Let from a sample m_d ssus selected from M_d ssus, m_{dk_1} ssus respond while m_{dk_2} ssus do not respond, $m_{dk_1} + m_{dk_2} = m_d$. From the m_{dk_2} non-responding ssus a sub-sample of size h_{dk_2} ssus is selected by srswor, $m_{dk_2} = h_{dk_2} f_{dk_2}$, $k = 1, 2, \ldots, n_d$. Let $\bar{y}_{m_{dk_1}}$ denote the mean of the sample from the response class in the d-th domain while $\bar{y}_{h_{dk_2}}$ denote the mean of the sample for the nonresponse class, where

$$\bar{y}_{m_{dk_1}} = \frac{1}{m_{dk_1}}\sum_{j=1}^{m_{dk_1}} y_{dkj} \text{ and } \bar{y}_{h_{dk_2}} = \frac{1}{h_{dk_2}}\sum_{k=1}^{h_{dk_2}} y_{dkj}.$$

Theorem 2.1 An unbiased estimator of $\bar{Y}_d$ is given by,

$$\bar{y}_{1d} = \frac{N}{nN_d}\sum_{k=1}^{n_d}\frac{1}{m_d}(m_{dk_1}\bar{y}_{m_{dk_1}} + m_{dk_2}\bar{y}_{h_{dk_2}}) \quad (2.1)$$

with variance

$$V(\bar{y}_{1d}) = \frac{N(N-n)(N_d-1)}{nN_d^2(N-1)}S_{bd}^2 + \frac{N(N-n)}{nN_d^2(N-1)}N_d Q_d \bar{Y}_d^2$$

$$+\frac{N}{nN_d^2}\sum_{k=1}^{N_d}\left(\frac{1}{m_d}-\frac{1}{M_d}\right)S_{dk}^2 + \frac{N}{nN_d^2}\sum_{k=1}^{N_d}\frac{M_{dk_2}}{M_d m_d}(f_{dk_2}-1)S_{M_{dk_2}}^2 , \quad (2.2)$$

where,

$$S_{bd}^2 = \frac{1}{(N_d-1)}\sum_{k=1}^{N_d}(\bar{Y}_{dk}-\bar{Y}_d)^2,\ \bar{Y}_{dk} = \frac{1}{M_d}\sum_{j=1}^{M_d} Y_{dk_j} \text{ and}$$

$$\bar{Y}_d = \frac{1}{N_d}\sum_{k=1}^{N_d}\bar{Y}_{dk}.$$

$$S_{dk}^2 = \frac{1}{(M_d-1)}\sum_{j=1}^{M_d}(Y_{dk_j}-\bar{Y}_{dk})^2, P_d = \frac{N_d}{N}, Q_d = 1-P_d .$$

$$S_{M_{dk_2}}^2 = \frac{1}{(M_{dk_2}-1)}\sum_{j=1}^{M_{dk_2}}(Y_{dk_j}-\bar{Y}_{M_{dk_2}})^2, \bar{Y}_{M_{dk_2}} = \frac{1}{M_{dk_2}}\sum_{j=1}^{M_{dk_2}} Y_{dk_j}.$$

If the objective is to estimate the domain total, then an unbiased estimator of Y_d is given as,

$$\hat{Y}_{1d} = \frac{NM_d}{n}\sum_{k=1}^{n_d}\frac{1}{m_d}(m_{dk_1}\bar{y}_{m_{dk_1}} + m_{dk_2}\bar{y}_{h_{dk_2}}) \quad (2.3)$$

with variance

$$V(\hat{Y}_{1d}) = \frac{N(N-n)(N_d-1)}{n(N-1)}M_d^2 S_{bd}^2 + \frac{N(N-n)}{n(N-1)}M_d^2 N_d Q_d \bar{Y}_d^2$$

$$+\frac{N}{n}\sum_{k=1}^{N_d}\left(\frac{1}{m_d}-\frac{1}{M_d}\right)M_d^2 S_{dk}^2 + \frac{N}{n}\sum_{k=1}^{N_d}\frac{M_{dk_2}M_d}{m_d}(f_{dk_2}-1)S_{M_{dk_2}}^2 , \quad (2.4)$$

where $S_{bd}^2, S_{dk}^2, S_{M_{dk_2}}^2$ etc. are defined earlier.

Proof: By definition,

$$E(\bar{y}_{1d})$$
$$= E_1E_2E_3E_4E_5\left[E_6\left\{\frac{N}{nN_d}\sum_{k=1}^{n_d}\frac{1}{m_d}(m_{dk_1}\bar{y}_{m_{dk_1}} + m_{dk_2}\bar{y}_{h_{dk_2}})\right\}\right]$$
$$= E_1E_2E_3E_4\left[E_5\left\{\frac{N}{nN_d}\sum_{k=1}^{n_d}\frac{1}{m_d}(m_{dk_1}\bar{y}_{m_{dk_1}} + m_{dk_2}\bar{y}_{m_{dk_2}})\right\}\right]$$

$$= E_1E_2E_3E_4\left[E_5\left\{\frac{N}{nN_d}\sum_{k=1}^{n_d}\bar{y}_{dk}\right\}\right]$$

$$= E_1E_2E_3\left[E_4\left\{\frac{N}{nN_d}\sum_{k=1}^{n_d}\bar{y}_{dk}\right\}\right]$$

$$= E_1E_2\left[E_3\left\{\frac{N}{nN_d}\sum_{k=1}^{n_d}\bar{y}_{dk}\right\}\right]$$

$$= E_1\left[E_2\left\{\frac{N}{nN_d}\sum_{k=1}^{n_d}\bar{Y}_{dk}\right\}\right]$$

$$= E_1\left(\frac{N}{nN_d}\frac{n_d}{N_d}\sum_{k=1}^{N_d}\bar{Y}_{dk}\right)$$

$$= \bar{Y}_d.$$

Hence, $\bar{y}_{1d}$ is an unbiased estimator of domain mean $\bar{Y}_d$. Here E_6 represents conditional expectation of all possible samples of size h_{dk_2} drawn from m_{dk_2}, E_5 is the conditional expectation of all possible samples of size m_{dk_1} and m_{dk_2} respectively drawn from M_{dk_1} and M_{dk_2} by keeping m_{dk_1} and m_{dk_2} fixed, E_4 refers to the conditional expectation arising out of randomness of m_{dk_1} and m_{dk_2}, E_3 is the conditional expectation of all possible samples of size m_d drawn from M_d while E_2 refers to the conditional expectation of all possible samples of size n_d drawn from a population of size N_d keeping n_d fixed and E_1 represents the expectation arising out of randomness of n_d.

The variance of the proposed estimator can be obtained as,

$$V(\bar{y}_{1d}) = V_1E_2E_3E_4E_5E_6(\bar{y}_{1d}) + E_1V_2E_3E_4E_5E_6(\bar{y}_{1d})$$
$$+E_1E_2V_3E_4E_5E_6(\bar{y}_{1d}) + E_1E_2E_3V_4E_5E_6(\bar{y}_{1d})$$
$$+E_1E_2E_3E_4V_5E_6(\bar{y}_{1d}) + E_1E_2E_3E_4E_5V_6(\bar{y}_{1d}).$$

Here V_1, V_2, V_3, V_4, V_5, V_6 are defined similarly as E_1, E_2, E_3, E_4, E_5, E_6.

$$V_1E_2E_3E_4E_5E_6(\bar{y}_{1d})$$

$$= V_1E_2E_3E_4E_5\left[E_6\left\{\frac{N}{nN_d}\sum_{k=1}^{n_d}\frac{1}{m_d}(m_{dk_1}\bar{y}_{m_{dk_1}} + m_{dk_2}\bar{y}_{h_{dk_2}})\right\}\right]$$

$$= V_1E_2E_3E_4\left[E_5\left\{\frac{N}{nN_d}\sum_{k=1}^{n_d}\frac{1}{m_d}(m_{dk_1}\bar{y}_{m_{dk_1}} + m_{dk_2}\bar{y}_{m_{dk_2}})\right\}\right]$$

$$= V_1E_2E_3\left[E_4\left\{\frac{N}{nN_d}\sum_{k=1}^{n_d}(\bar{y}_{dk})\right\}\right]$$

$$= V_1E_2\left[E_3\left\{\frac{N}{nN_d}\sum_{k=1}^{n_d}(\bar{y}_{dk})\right\}\right]$$

$$= V_1\left[E_2\left\{\frac{N}{nN_d}\sum_{k=1}^{n_d}(\bar{y}_{dk})\right\}\right]$$

$$= V_1\left(\frac{N}{nN_d}\frac{n_d}{N_d}\sum_{k=1}^{N_d}\bar{Y}_{dk}\right)$$

$$= \frac{N(N-n)}{nN_d^2}\left\{\frac{N_dQ_d\bar{Y}_d^2}{(N-1)}\right\},$$

$$E_1V_2E_3E_4E_5E_6(\bar{y}_{1d})$$

$$= E_1V_2E_3E_4E_5\left[E_6\left\{\frac{N}{nN_d}\sum_{k=1}^{n_d}\frac{1}{m_d}(m_{dk_1}\bar{y}_{m_{dk_1}} + m_{dk_2}\bar{y}_{h_{dk_2}})\right\}\right]$$

$$= E_1V_2E_3E_4\left[E_5\left\{\frac{N}{nN_d}\sum_{k=1}^{n_d}(\bar{y}_{dk})\right\}\right]$$

$$= E_1V_2\left[E_3\left\{\frac{N}{nN_d}\sum_{k=1}^{n_d}(\bar{y}_{dk})\right\}\right]$$

$$= E_1\left[V_2\left\{\frac{N}{nN_d}\sum_{k=1}^{n_d}(\bar{y}_{dk})\right\}\right]$$

$$= E_1\left[\frac{N^2}{n^2N_d^2}n_d^2\left(\frac{1}{n_d}-\frac{1}{N_d}\right)S_{bd}^2\right]$$

$$= \frac{N(N-n)(N_d-1)}{nN_d^2(N-1)}S_{bd}^2,$$

$$E_1E_2V_3E_4E_5E_6(\bar{y}_{1d})$$

$$= E_1E_2V_3E_4\left[E_5\left\{\frac{N}{nN_d}\sum_{k=1}^{n_d}(\bar{y}_{dk})\right\}\right]$$

$$= E_1 E_2 \left[V_3 \left\{ \frac{N}{nN_d} \sum_{k=1}^{n_d} (\bar{y}_{dk}) \right\} \right]$$

$$= E_1 \left[E_2 \left\{ \frac{N^2}{n^2 N_d^2} \sum_{k=1}^{n_d} (\frac{1}{m_d} - \frac{1}{M_d}) S_{dk}^2 \right\} \right]$$

$$= E_1 \left[\frac{N^2}{n^2 N_d^2} \frac{n_d}{N_d} \sum_{k=1}^{N_d} (\frac{1}{m_d} - \frac{1}{M_d}) S_{dk}^2 \right]$$

$$= \frac{N}{n\ N_d^2} \sum_{k=1}^{N_d} (\frac{1}{m_d} - \frac{1}{M_d}) S_{dk}^2,$$

$E_1\ E_2\ E_3\ V_4\ E_5\ E_6\ (\bar{y}_{1d})\ = 0,$

$E_1\ E_2\ E_3\ E_4\ V_5\ E_6\ (\bar{y}_{1d})\ = 0$ and

$E_1 E_2 E_3 E_4 E_5 V_6 (\bar{y}_{1d})$

$$= E_1 E_2 E_3 E_4 E_5 \left[V_6 \left\{ \frac{N}{nN_d} \sum_{k=1}^{n_d} \frac{1}{m_d} (m_{dk_1} \bar{y}_{m_{dk_1}} + m_{dk_2} \bar{y}_{h_{dk_2}}) \right\} \right]$$

$$= E_1 E_2 E_3 E_4 \left[E_5 \left(\frac{N^2}{n^2 N_d^2} \sum_{k=1}^{n_d} \left(\frac{m_{dk_2}}{m_d} \right)^2 \left\{ (\frac{1}{h_{dk_2}} - \frac{1}{m_{dk_2}}) s_{m_{dk_2}}^2 \right\} \right) \right]$$

$$= E_1 E_2 \left\{ E_3 \left[\frac{N^2}{n^2 N_d^2} \sum_{k=1}^{n_d} \frac{M_{dk_2}}{M_d m_d} (f_{dk_2} - 1) S_{M_{dk_2}}^2 \right] \right\}$$

$$= E_1 \left[\frac{N^2}{n^2 N_d^2} \frac{n_d}{N_d} \sum_{k=1}^{N_d} \frac{M_{dk_2}}{M_d m_d} (f_{dk_2} - 1) S_{M_{dk_2}}^2 \right]$$

$$= \frac{N}{n\ N_d^2} \sum_{k=1}^{N_d} \frac{M_{dk_2}}{M_d m_d} (f_{dk_2} - 1) S_{M_{dk_2}}^2 .$$

Hence, we get the required expression as specified in Equation (2.2). Next, we determine the optimum values of n, m_d and f_{dk_2} by minimising the expected cost for a fixed variance. To achieve this, we consider the following cost function

$$C = C_{1d} n_d + C_{2d} \sum_{i=1}^{n_d} m_{dk_1} + C_{3d} \sum_{i=1}^{n_d} h_{dk_2} ,$$

where,

C: Total cost.

C_{1d}: Per unit travel and miscellaneous cost in the d-th domain.

C_{2d}: Cost per unit of collecting the information on the study character in the first attempt in the d-th domain.

C_{3d}: Cost per unit of collecting the information by expensive method after the first attempt failed to collect information in the d-th domain.

The cost function considered above is suitable for situations prevailing in mail surveys. In these surveys the first attempt to collect information from the respondents is made through e-mail/postal mails. Many of the respondents may not send the required information through mails. To collect the required information, a sub-sample of non-respondents may be collected for data collection by specialized effort, say, personal interview.

The expected cost in this case is,

$$C' = E(C) = \frac{n}{N} \left\{ C_{1d} N_d + C_{2d} \sum_{k=1}^{N_d} \frac{M_{dk_1} m_d}{M_d} + C_{3d} \sum_{k=1}^{N_d} \frac{M_{dk_2} m_d}{M_d f_{dk_2}} \right\},$$

and the function to be optimised is given by $\phi = C' + \lambda \{ V(\bar{y}_{1d}) - V_0 \}$, where λ is the Lagrangian multiplier. Also, V_0 can be determined by fixing the coefficient of variation, say equal to 10% or 5%. To get closed form expression of the optimum values we assume that $m_{dk_2} = h_{dk_2} f_{2d}$, $k = 1, 2, ..., N_d$ in place of $m_{dk_2} = h_{dk_2} f_{dk_2}$, $k = 1, 2, ..., N_d$, *i.e.*, all psus in domain d, we have $f_{2d} = f_{dk_2}$. Differentiation with respect to n, m_d, λ and f_{2d}, equating the resultant derivatives to '0' and simplifying gives the optimum values as

$$n_{opt} = \frac{K_{14}}{K_{13}},\ m_{dopt} = \frac{-b_1 \pm \sqrt{b_1^2 - 4a_1 e_1}}{2a_1}, \text{ and}$$

$$f_{2dopt} = \pm \sqrt{ \frac{ C_{3d} \sum_{k=1}^{N_d} M_{dk_2} \left(\sum_{k=1}^{N_d} S_{dk}^2 - \sum_{i=1}^{N_d} \frac{M_{dk_2}}{M_d} S_{M_{dk_2}}^2 \right) }{ C_{2d} \left[\sum_{k=1}^{N_d} \frac{M_{dk_1}}{M_d} \sum_{i=1}^{N_d} M_{dk_2} S_{M_{dk_2}}^2 \right] } } .$$

As sample sizes are nonnegative values, thus,

$$m_{dopt} = \frac{-b_1 + \sqrt{b_1^2 - 4a_1 e_1}}{2a_1} \text{ and}$$

$$f_{2dopt} = \sqrt{\frac{C_{3d}\sum_{k=1}^{N_d} M_{dk_2}\left(\sum_{k=1}^{N_d} S_{dk}^2 - \sum_{i=1}^{N_d}\frac{M_{dk_2}}{M_d} S_{M_{dk_2}}^2\right)}{C_{2d}\left[\sum_{k=1}^{N_d}\frac{M_{dk_1}}{M_d}\sum_{i=1}^{N_d} M_{dk_2} S_{M_{dk_2}}^2\right]}}$$

where,

$$K_{13} = V_0 + \frac{N}{N_d^2}\left\{\frac{N_d Q_d \bar{Y}_d^2}{(N-1)}\right\} + \frac{N(N_d-1)}{N_d^2(N-1)} S_{bd}^2,$$

$$K_{14} = \left[\frac{N}{N_d} k_{12} S_{bd}^2 + \frac{N^2}{N_d^2}\left\{\frac{N_d Q_d \bar{Y}_d^2}{(N-1)}\right\} + \frac{N}{N_d^2}\sum_{k=1}^{N_d}\left(\frac{1}{m_d} - \frac{1}{M_d}\right) S_{dk}^2 + \frac{N}{N_d^2}\sum_{k=1}^{N_d}\frac{M_{dk_2}}{M_d m_d}(f_{2d}-1) S_{M_{dk_2}}^2\right],$$

$$\left[k_{12} = \left[\frac{N(N_d-1)}{N_d(N-1)}\right]\right],$$

$$a_1 = \left(C_{3d}\sum_{k=1}^{N_d}\frac{M_{dk_2}}{f_{2d}^2}\right)\left[N_d k_{12} S_{bd}^2 + N\left\{\frac{N_d Q_d \bar{Y}_d^2}{(N-1)}\right\} - \sum_{k=1}^{N_d}\frac{1}{M_d} S_{dk}^2\right]$$

$$b_1 = -\left(\{C_{2d}\sum_{k=1}^{N_d}\frac{M_{dk_1}}{M_d}\}\sum_{k=1}^{N_d} M_{dk_2} S_{M_{dk_2}}^2 - C_{3d}\sum_{k=1}^{N_d}\frac{M_{dk_2}}{f_{2d}^2}\left\{\sum_{k=1}^{N_d} S_{dk}^2 - \sum_{k=1}^{N_d}\frac{M_{dk_2}}{M_d} S_{M_{dk_2}}^2\right\}\right)$$

$$e_1 = -C_{1d} N_d \sum_{k=1}^{N_d} M_{dk_2} S_{M_{dk_2}}^2 \text{ and } V_0 = 0.0025\times\bar{Y}_d^2.$$

Case 2. Let n psus be selected from N psus by simple random sampling without replacement (srswor) design where n_d out of n psus fall in the d-th domain and within each selected psu, m_d ssus are also selected from M_d ssus by srswor. Let there be no nonresponse in n_{1d} psus and partial nonresponse in n_{2d} psus. In the n_{2d} psus m_{dk_1} ssus responds and m_{dk_2} ssus do not respond, $m_{dk_1} + m_{dk_2} = m_d$. From the m_{dk_2} nonresponding ssus a subsample of h_{dk_2} ssus is selected by srswor and data is collected through specialized effort, $m_{dk_2} = h_{dk_2} f_{dk_2}$, $k = 1, 2, ..., n_{2d}$. In this context we state the Theorem 2.2 as below:

Theorem 2.2 An unbiased estimator of $\bar{Y}_d$ is given by,

$$\bar{y}_{2d} = \frac{N}{nN_d}\left\{\sum_{k=1}^{n_{1d}}\bar{y}_{dk} + \sum_{k=1}^{n_{2d}}\frac{1}{m_d}(m_{dk_1}\bar{y}_{m_{dk_1}} + m_{dk_2}\bar{y}_{h_{dk_2}})\right\}, \tag{2.5}$$

and the variance of the estimator is

$$V(\bar{y}_{2d}) = \frac{N(N-n)(N_d-1)}{nN_d^2(N-1)} S_{bd}^2 + \frac{N(N-n)}{nN_d^2(N-1)} N_d Q_d \bar{Y}_d^2 + \frac{N}{nN_d^2}\sum_{k=1}^{N_d}\left(\frac{1}{m_d} - \frac{1}{M_d}\right) S_{dk}^2 + \frac{N}{nN_d^2}\sum_{k=1}^{N_{2d}}\frac{M_{dk_2}}{M_d m_d}(f_{dk_2}-1) S_{M_{dk_2}}^2. \tag{2.6}$$

Here, S_{dk}^2, S_{bd}^2, $S_{M_{dk_2}}^2$ etc. have been defined earlier. Note that the domain sample size n_d is a random variable. If the objective is to estimate the domain total, then an unbiased estimator of Y_d is given as,

$$\hat{Y}_{2d} = \frac{NM_d}{n}\left\{\sum_{k=1}^{n_{1d}}\bar{y}_{dk} + \sum_{k=1}^{n_{2d}}\frac{1}{m_d}(m_{dk_1}\bar{y}_{m_{dk_1}} + m_{dk_2}\bar{y}_{h_{dk_2}})\right\} \tag{2.7}$$

with variance

$$V(\hat{Y}_{2d}) = \frac{N(N-n)(N_d-1)}{n(N-1)} M_d^2 S_{bd}^2 + \frac{N(N-n)}{n(N-1)} M_d^2 N_d Q_d \bar{Y}_d^2 + \frac{N}{n}\sum_{k=1}^{N_d}\left(\frac{1}{m_d} - \frac{1}{M_d}\right) M_d^2 S_{dk}^2 + \frac{N}{n}\sum_{k=1}^{N_{2d}}\frac{M_{dk_2} M_d}{m_d}(f_{dk_2}-1) S_{M_{dk_2}}^2. \tag{2.8}$$

The various terms defined in the above expressions are defined earlier.

Proof: By definition,

$$E(\bar{y}_{2d}) = E_1E_2E_3E_4E_5\left[E_6\frac{N}{nN_d}\left\{\sum_{k=1}^{n_{1d}}\bar{y}_{dk} + \sum_{k=1}^{n_{2d}}\frac{1}{m_d}(m_{dk_1}\bar{y}_{m_{dk_1}} + m_{dk_2}\bar{y}_{h_{dk_2}})\right\}\right]$$

$$= E_1E_2E_3E_4\left[E_5\frac{N}{nN_d}\left\{\sum_{k=1}^{n_{1d}}\bar{y}_{dk} + \sum_{k=1}^{n_{2d}}\bar{y}_{dk}\right\}\right]$$

$$= E_1E_2E_3E_4\left[E_5\frac{N}{nN_d}\{\sum_{k=1}^{n_d}\bar{y}_{dk}\}\right]$$

$$= E_1E_2E_3\left[E_4\frac{N}{nN_d}\{\sum_{k=1}^{n_d}\bar{Y}_{dk}\}\right]$$

$$= E_1E_2\left[E_3\frac{N}{nN_d}\{\sum_{k=1}^{n_d}\bar{Y}_{dk}\}\right]$$

$$= E_1\left[E_2\frac{N}{nN_d}\{\sum_{k=1}^{n_d}\bar{Y}_{dk}\}\right]$$

$$= E_1\left[\frac{N}{nN_d}\frac{n_d}{N_d}\sum_{k=1}^{N_d}\bar{Y}_{dk}\right]$$

$$= \frac{1}{N_d}\sum_{k=1}^{N_d}\bar{Y}_{dk} = \bar{Y}_d.$$

Thus, $\bar{y}_{2d}$ is an unbiased estimator of domain mean $\bar{Y}_d$. Here, E_6 represents conditional expectations of all possible samples of size h_{dk_2} drawn from m_{dk_2}, E_5 is the conditional expectation of all possible samples of size m_d drawn from M_d, E_4 arising out of selection of all possible samples of size n_{1d} and n_{2d} from N_{1d} and N_{2d} keeping n_{1d} and n_{2d} fixed, E_3 refers to conditional expectation arising out of randomness of n_{1d} and n_{2d}, E_2 arising out of selection of all possible samples of size n_d drawn from N_d keeping n_d fixed whereas E_1 is the expectation arising out of randomness of n_d.

To obtain the variance we proceed as follows:

By definition,

$$V(\bar{y}_{2d}) = V_1E_2E_3E_4E_5E_6(\bar{y}_{2d}) + E_1V_2E_3E_4E_5E_6(\bar{y}_{2d}) + E_1E_2V_3E_4E_5E_6(\bar{y}_{2d}) + E_1E_2E_3V_4E_5E_6(\bar{y}_{2d}) + E_1E_2E_3E_4V_5E_6(\bar{y}_{2d}) + E_1E_2E_3E_4E_5V_6(\bar{y}_{2d}).$$

Here, V_1, V_2, V_3, V_4, V_5, V_6 are defined in the same way as E_1, E_2, E_3, E_4, E_5, E_6. The expressions for the variance have been derived in the same way as in Case 1.

$$V_1E_2E_3E_4E_5E_6(\bar{y}_{2d}) = \frac{N(N-n)}{nN^2{}_d}\frac{N_d}{(N-1)}Q_d\bar{Y}_d^2,$$

$$E_1V_2E_3E_4E_5E_6(\bar{y}_{2d}) = \frac{N(N-n)(N_d-1)}{nN_d^2(N-1)}S_{bd}^2,$$

$$E_1E_2V_3E_4E_5E_6(\bar{y}_{2d}) = 0,$$

$$E_1E_2E_3V_4E_5E_6(\bar{y}_{2d}) = 0,$$

$$E_1E_2E_3E_4V_5E_6(\bar{y}_{2d}) = \frac{N}{nN_d^2}\sum_{k=1}^{N_d}(\frac{1}{m_d}-\frac{1}{M_d})S_{dk}^2, \text{ and}$$

$$E_1E_2E_3E_4E_5V_6(\bar{y}_{2d}) = \frac{N}{nN_d^2}\sum_{k=1}^{N_{2d}}\frac{M_{dk_2}}{M_dm_d}\left(f_{dk_2}-1\right)S_{M_{dk_2}}^2.$$

By adding the above three terms we get the required variance of the estimator. We determine the optimum values of n, m_d and f_{dk_2} by minimizing the expected cost for a fixed variance. The relevant cost function in this case is,

$$C = C_{1d}n_{2d} + C_{2d}n_{1d}m_{dk} + C_{2d}\sum_{k=1}^{n_{2d}}m_{dk_1} + C_{3d}\sum_{k=1}^{n_{2d}}h_{dk_2},$$

where C, C_{1d}, C_{2d} and C_{3d} are same as defined earlier. The expected cost is,

$$E(C) = \frac{n}{N}\left\{C_{1d}N_{2d} + C_{2d}N_{1d}m_d + C_{2d}\sum_{k=1}^{N_{2d}}\frac{M_{dk_1}m_d}{M_d} + C_{3d}\sum_{k=1}^{N_{2d}}\frac{M_{dk_2}m_d}{M_d f_{dk_2}}\right\}.$$

Consider the following function

$\phi = E(C) + \lambda\{V(\bar{y}_{2d})-V_0\}$, where λ is the Lagrangian multiplier. To get closed form expression of the optimum values we assume that $m_{dk_2} = h_{dk_2}f_{2d}$, $k = 1, 2, ..., n_{2d}$ in place of $m_{dk_2} = h_{dk_2}f_{dk_2}, k = 1, 2, ..., n_{2d}$. The optimum values are as follows:

$$n_{dopt} = \frac{K_{15}}{K_{16}},$$

$$m_{dopt} = \frac{-b_2 \pm \sqrt{b_2{}^2-4a_2e_2}}{2a_2} \text{ and } f_{2dopt} = \pm\sqrt{\frac{B_2}{B_1}}.$$

Since positive values are relevant here so,

$$m_{dopt} = \frac{-b_2 + \sqrt{b_2{}^2-4a_2e_2}}{2a_2} \text{ and } f_{2opt} = \sqrt{\frac{B_2}{B_1}},$$

where,

$$K_{15} = \left[\frac{N}{N_d}k_{12}S_{bd}^2 + \frac{N^2}{N_d^2}\left\{\frac{N_dQ_d\bar{Y}_d^2}{(N-1)}\right\}\right.$$

$$+\frac{N}{N_d^2}\sum_{k=1}^{N_d}(\frac{1}{m_d}-\frac{1}{M_d})S_{dk}^2+\frac{N}{N_d^2}\sum_{k=1}^{N_{2d}}\frac{M_{dk_2}}{M_d m_d}(f_{2d}-1)S_{M_{dk_2}}^2\Bigg]$$

$$K_{16}=V_0+\frac{N}{N_d^2}\left\{\frac{N_d Q_d \bar{Y}_d^2}{(N-1)}\right\}+\frac{N(N_d-1)}{N_d^2(N-1)}S_{bd}^2,$$

$$\left[k_{12}=\left[\frac{N(N_d-1)}{N_d(N-1)}\right]\right],$$

$$a_2=\left(C_{3d}\sum_{k=1}^{N_{2d}}\frac{M_{dk_2}}{f_{2d}^2}\right)\left[N_d k_{12}S_{bd}^2+N\left\{\frac{N_d Q_d \bar{Y}_d^2}{(N-1)}\right\}\right.$$

$$\left.-\sum_{k=1}^{N_d}\frac{1}{M_d}S_{dk}^2\right],$$

$$b_2=-\left[\left(\{C_{2d}N_{1d}+C_{2d}\sum_{k=1}^{N_{2d}}\frac{M_{dk_1}}{M_d}\}\sum_{k=1}^{N_{2d}}M_{dk_2}S_{M_{dk_2}}^2\right)\right.$$

$$\left.-C_{3d}\sum_{k=1}^{N_{2d}}\frac{M_{dk_2}}{f_{2d}^2}\left\{\sum_{k=1}^{N_d}S_{dk}^2-\sum_{k=1}^{N_{2d}}\frac{M_{dk_2}}{M_d}S_{M_{dk_2}}^2\right\}\right],$$

$$c_2=-C_{1d}N_{2d}\sum_{k=1}^{N_{2d}}M_{dk_2}S_{M_{dk_2}}^2,$$

$$B_1=\left(C_{2d}N_{1d}+C_{2d}\sum_{k=1}^{N_{2d}}\frac{M_{dk_1}}{M_d}\right)\sum_{i=1}^{N_{2d}}M_{dk_2}S_{M_{dk_2}}^2,$$

$$B_2=C_{3d}\sum_{k=1}^{N_{2d}}M_{dk_2}\left[\sum_{k=1}^{N_d}S_{dk}^2-\sum_{k=1}^{N_{2d}}\frac{M_{dk_2}}{M_d}S_{M_{dk_2}}^2\right]$$ and

$V_0 = 0.0025\times\bar{Y}_d^2$.

Case 3. Let n psus be selected from N psus by srswor design where n_d out of n psus fall in the d-th domain randomly and within each selected psu, m_d ssus are also selected from M_d ssus by srswor. Let there be no nonresponse in n_{1d} psus and partial nonresponse in n_{2d} psus and complete nonresponse in the n_{3d} psus, $n_{1d}+n_{2d}+n_{3d}=n_d$. In n_{2d} psus m_{dk_1} units responds and m_{dk_2} units do not respond, $m_{dk_1}+m_{dk_2}=m_d$. From the m_{dk_2} nonresponding units a subsample of h_{dk_2} units is selected by srswor, $m_{dk_2}=h_{dk_2}f_{dk_2}$, $k=1, 2, ..., n_{2d}$. Further a sub-sample of h_{3d} psus is drawn out of n_{3d} psus and data are collected through specialized efforts on each of m_d ssus in the selected h_{3d} psus, $n_{3d}=f_{3d}h_{3d}$. In this context we state the Theorem 2.3 as below,

Theorem 2.3 An unbiased estimator of $\bar{Y}_d$ is given by,

$$\bar{y}_{3d}=\frac{N}{nN_d}\left\{\sum_{k=1}^{n_{1d}}\bar{y}_{dk}+\sum_{k=1}^{n_{2d}}\frac{1}{m_d}(m_{dk_1}\bar{y}_{m_{dk_1}}+m_{dk_2}\bar{y}_{h_{dk_2}})\right.$$

$$\left.+\frac{n_{3d}}{h_{3d}}\sum_{k=1}^{h_{3d}}\bar{y}_{dk}\right\} \tag{2.9}$$

The variance of the estimator (2.9) is given as,

$$V(\bar{y}_{3d})=\frac{N(N-n)(N_d-1)}{nN_d^2(N-1)}S_{bd}^2+\frac{N(N-n)}{nN_d^2(N-1)}N_d Q_d \bar{Y}_d^2$$

$$+\frac{N}{nN_d^2}\left[\sum_{k=1}^{N_{1d}+N_{2d}}(\frac{1}{m_d}-\frac{1}{M_d})S_{dk}^2+f_{3d}\sum_{k=1}^{N_{3d}}(\frac{1}{m_d}-\frac{1}{M_d})S_{dk}^2\right]$$

$$+\frac{NN_{3d}}{nN_d^2}(f_{3d}-1)S_{b_{N_{3d}}}^2+\frac{N}{nN_d^2}\sum_{k=1}^{N_{2d}}\frac{M_{dk_2}}{M_d m_d}(f_{dk_2}-1)S_{M_{dk_2}}^2. \tag{2.10}$$

Here $S_{b_{N_{3d}}}^2=\frac{1}{N_{3d}-1}\sum_{k=1}^{N_{3d}}(\bar{Y}_{dk}-\bar{Y}_{N_{3d}})^2$ and $\bar{Y}_{N_{3d}}=\frac{1}{N_{3d}}\sum_{k=1}^{N_{3d}}\bar{Y}_{dk}$

and rest of the terms are defined earlier. If the objective is to estimate the domain total, then an unbiased estimator of Y_d is given as,

$$\hat{Y}_{3d}=\frac{NM_d}{n}\left\{\sum_{k=1}^{n_{1d}}\bar{y}_{dk}+\sum_{k=1}^{n_{2d}}\frac{1}{m_d}(m_{dk_1}\bar{y}_{m_{dk_1}}+m_{dk_2}\bar{y}_{h_{dk_2}})\right.$$

$$\left.+\frac{n_{3d}}{h_{3d}}\sum_{k=1}^{h_{3d}}\bar{y}_{dk}\right\} \tag{2.11}$$

with variance

$$V(\hat{Y}_{3d})=\frac{N(N-n)(N_d-1)}{n(N-1)}M_d^2S_{bd}^2$$

$$+\frac{N(N-n)}{n(N-1)}M_d^2N_dQ_d\bar{Y}_d^2+\frac{N}{n}\left[\sum_{k=1}^{N_{1d}+N_{2d}}(\frac{1}{m_d}-\frac{1}{M_d})M_d^2S_{dk}^2\right.$$

$$\left.+f_{3d}\sum_{k=1}^{N_{3d}}(\frac{1}{m_d}-\frac{1}{M_d})M_d^2S_{dk}^2\right]+\frac{NN_{3d}M_d^2}{n}(f_{3d}-1)S_{b_{N_{3d}}}^2$$

$$+\frac{N}{n}\sum_{k=1}^{N_{2d}}\frac{M_{dk_2}M_d}{m_d}(f_{dk_2}-1)S_{M_{dk_2}}^2\ . \tag{2.12}$$

The terms defined in the above expressions are defined earlier.

Proof: By definition,

$$E(\bar{y}_{3d}) = E_1E_2E_3E_4E_5E_6E_7\frac{N}{nN_d}\times$$

$$\left\{\sum_{k=1}^{n_{1d}}\bar{y}_{dk} + \sum_{k=1}^{n_{2d}}\frac{1}{m_d}(m_{dk_1}\bar{y}_{m_{dk_1}} + m_{dk_2}\bar{y}_{h_{dk_2}}) + \frac{n_{3d}}{h_{3d}}\sum_{k=1}^{h_{3d}}\bar{y}_{dk}\right\}$$

$$= E_1E_2E_3E_4E_5\left[E_6\frac{N}{nN_d}\left\{\sum_{k=1}^{n_{1d}}\bar{y}_{dk} + \sum_{k=1}^{n_{2d}}\bar{y}_{dk} + \frac{n_{3d}}{h_{3d}}\sum_{k=1}^{h_{3d}}\bar{y}_{dk}\right\}\right]$$

$$= E_1E_2E_3E_4\left[E_5\left\{\frac{N}{nN_d}\left(\sum_{k=1}^{n_{1d}}\bar{Y}_{dk} + \sum_{k=1}^{n_{2d}}\bar{Y}_{dk} + \frac{n_{3d}}{h_{3d}}\sum_{k=1}^{h_{3d}}\bar{Y}_{dk}\right)\right\}\right]$$

$$= E_1E_2E_3\left[E_4\left\{\frac{N}{nN_d}\left(\sum_{k=1}^{n_{1d}}\bar{Y}_{dk} + \sum_{k=1}^{n_{2d}}\bar{Y}_{dk} + \sum_{k=1}^{n_{3d}}\bar{Y}_{dk}\right)\right\}\right]$$

$$= E_1E_2\left[E_3\left\{\frac{N}{nN_d}\sum_{k=1}^{n_d}\bar{Y}_{dk}\right\}\right]$$

$$= E_1\left[E_2\left\{\frac{N}{nN_d}\sum_{k=1}^{n_d}\bar{Y}_{dk}\right\}\right]$$

$$= E_1\left[\frac{N}{nN_d}\{\frac{n_d}{N_d}\sum_{k=1}^{N_d}\bar{Y}_{dk}\}\right]$$

$$= \frac{1}{N_d}\sum_{k=1}^{N_d}\bar{Y}_{dk} = \bar{Y}_d.$$

Hence, it can be seen that $\bar{y}_{3d}$ is an unbiased estimator of domain mean $\bar{Y}_d$. Here, E_7 represents conditional expectations of all possible samples of size h_{dk_2} drawn from m_{dk_2}, E_6 is the conditional expectation of all possible samples of size m_d drawn from M_d, E_5 refers to conditional expectation arising out of selection of all possible samples of size h_{3d} drawn from n_{3d}, E_4 arising out of selection of all possible samples of size n_{1d}, n_{2d} and n_{3d} drawn from N_{1d}, N_{2d} and N_{3d} keeping n_{1d}, n_{2d} and n_{3d} fixed, E_3 is the conditional expectation arising out of randomness of n_{1d}, n_{2d} and n_{3d}, E_2 arising out of selection of all possible samples of size n_d drawn from N_d keeping n_d fixed whereas E_1 is the expectation arising out of randomness of n_d.

To obtain the variance we proceed as follows,

$$V(\bar{y}_{3d}) = V_1E_2E_3E_4E_5E_6E_7(\bar{y}_{3d}) + E_1V_2E_3E_4E_5E_6E_7(\bar{y}_{3d})$$
$$+E_1E_2V_3E_4E_5E_6E_7(\bar{y}_{3d}) + E_1E_2E_3V_4E_5E_6E_7(\bar{y}_{3d})$$
$$+E_1E_2E_3E_4V_5E_6E_7(\bar{y}_{3d}) + E_1E_2E_3E_4E_5V_6E_7(\bar{y}_{3d})$$
$$+E_1E_2E_3E_4E_5E_6V_7(\bar{y}_{3d}).$$

Here, V_1, V_2, V_3, V_4, V_5, V_6, V_7 are defined similarly as E_1, E_2, E_3, E_4, E_5, E_6, E_7. The terms in the variance are developed in the same way as in Case 1.

$$V_1E_2E_3E_4E_5E_6E_7(\bar{y}_{3d}) = \frac{N(N-n)}{nN^2_d}\frac{N_d}{(N-1)}Q_d\bar{Y}_d^2,$$

$$E_1V_2E_3E_4E_5E_6E_7(\bar{y}_{3d}) = \frac{N(N-n)(N_d-1)}{nN_d^2(N-1)}S_{bd}^2,$$

$$E_1E_2V_3E_4E_5E_6E_7(\bar{y}_{3d}) = 0,$$

$$E_1E_2E_3V_4E_5E_6E_7(\bar{y}_{3d}) = 0 \text{ and}$$

$$E_1E_2E_3E_4V_5E_6E_7(\bar{y}_{3d})$$

$$= E_1E_2E_3E_4V_5\left[E_6\frac{N}{nN_d}\left\{\sum_{k=1}^{n_{1d}}\bar{y}_{dk} + \sum_{k=1}^{n_{2d}}\bar{y}_{dk} + \frac{n_{3d}}{h_{3d}}\sum_{k=1}^{h_{3d}}\bar{y}_{dk}\right\}\right]$$

$$= E_1E_2E_3E_4\left(V_5\left[\frac{N}{nN_d}\{\sum_{k=1}^{n_{1d}}\bar{Y}_{dk} + \sum_{k=1}^{n_{2d}}\bar{Y}_{dk} + \frac{n_{3d}}{h_{3d}}\sum_{k=1}^{h_{3d}}\bar{Y}_{dk}\}\right]\right)$$

$$= E_1E_2E_3\left\{E_4\left[\frac{N^2}{n^2N^2_d}\frac{n_{3d}^2}{h_{3d}^2}h_{3d}^2\left(\frac{1}{h_{3d}} - \frac{1}{n_{3d}}\right)s^2_{b_{N_{3d}}}\right]\right\}$$

$$= E_1E_2\left[E_3\frac{N^2}{n^2N^2_d}n_{3d}(f_{3d}-1)S^2_{b_{N_{3d}}}\right]$$

$$= E_1\left\{E_2\left(\frac{N^2}{n\,N^2_d}\frac{N_{3d}}{N_d}\frac{n_d}{n}(f_{3d}-1)S^2_{b_{N_{3d}}}\right)\right\}$$

$$= E_1\left\{\frac{N^2}{n\,N^2_d}\frac{N_{3d}}{N_d}\frac{n_d}{n}(f_{3d}-1)S^2_{b_{N_{3d}}}\right\}$$

$$= \frac{NN_{3d}}{nN_d^2}(f_{3d}-1)S^2_{b_{N_{3d}}},$$

$$E_1E_2E_3E_4E_5V_6E_7(\bar{y}_{3d})$$

$$=\frac{N}{nN_d^2}\left[\sum_{k=1}^{N_{1d}+N_{2d}}\left(\frac{1}{m_d}-\frac{1}{M_d}\right)S_{dk}^2+f_{3d}\sum_{k=1}^{N_{3d}}\left(\frac{1}{m_d}-\frac{1}{M_d}\right)S_{dk}^2\right],$$

$$E_1E_2E_3E_4E_5E_6V_7(\bar{y}_{3d})=\frac{N}{nN_d^2}\sum_{k=1}^{N_{2d}}\frac{M_{dk_2}}{M_dm_d}\left(f_{dk_2}-1\right)S_{M_{dk_2}}^2.$$

Thus, by adding all the terms we obtain the required result.

To determine the optimum values of n, m_d, f_{dk_2} and f_{3d} we proceed as follows,

The cost function in this case is,

$$C=C_{1d}(n_{2d}+h_{3d})+C_{2d}n_{1d}m_d+C_{2d}\sum_{k=1}^{n_{2d}}m_{dk_1}$$

$$+C_{3d}\left(\sum_{k=1}^{n_{2d}}h_{dk_2}+h_{3d}m_d\right),$$

where, C, C_{1d}, C_{2d} and C_{3d} are same as defined earlier. The expected cost is,

$$E(C)=\frac{n}{N}\left\{C_{1d}N_{2d}+C_{1d}\frac{N_{3d}}{f_{3d}}+C_{2d}N_{1d}m_d\right.$$

$$\left.+C_{2d}\sum_{k=1}^{N_{2d}}\frac{M_{dk_1}m_d}{M_d}+C_{3d}\sum_{k=1}^{N_{2d}}\frac{M_{dk_2}m_d}{M_df_{dk_2}}+C_{3d}\frac{N_{3d}m_d}{f_{3d}}\right\}.$$

To minimize the expected cost subject to fixed variance consider the function $\phi = E(C) + \lambda\{V(\bar{y}_{3d})-V_0\}$. To get closed form expression of the optimum value we assume that $m_{dk_2}=h_{dk_2}f_{2d}$, k = 1, 2, ..., n_{2d} in place of $m_{dk_2}=h_{dk_2}f_{dk_2}$, k = 1, 2, ..., n_{2d}. To overcome the problem arising due to simultaneous minimization of n, m_d, f_{2d} and f_{3d} we assume that $n_{3d}=f_{2d}h_{3d}$. Thus solving the above equations gives the optimum values as,

$$n_{opt}=\frac{K_{17}}{K_{18}},\ m_{dopt}=\frac{-b_2\pm\sqrt{b_2^2-4a_2e_2}}{2a_2}\quad\text{and}$$

$$f_{2dopt}=\frac{-b_3\pm\sqrt{b_3^2-4a_3e_3}}{2a_3}.$$

Since positive values are relevant here, so,

$$m_{dopt}=\frac{-b_2+\sqrt{b_2^2-4a_2e_2}}{2a_2}\quad\text{and}$$

$$f_{2opt}=\frac{-b_3+\sqrt{b_3^2-4a_3e_3}}{2a_3},$$

where,

$$K_{17}=\frac{N}{N_d}k_{12}S_{bd}^2+\frac{N^2}{N_d^2}\left\{\frac{N_dQ_d\bar{Y}_d^2}{(N-1)}\right\}$$

$$+\frac{N}{N_d^2}\left(\sum_{k=1}^{N_{1d}+N_{2d}}\left(\frac{1}{m_d}-\frac{1}{M_d}\right)S_{dk}^2+f_{3d}\sum_{k=1}^{N_{3d}}\left(\frac{1}{m_d}-\frac{1}{M_d}\right)S_{dk}^2\right)$$

$$+\frac{N}{N_d^2}\sum_{k=1}^{N_{2d}}\frac{M_{dk_2}}{M_dm_d}(f_{2d}-1)S_{M_{dk_2}}^2+\frac{NN_{3d}}{N_d^2}(f_{3d}-1)S_{b_{N3d}}^2,$$

$$K_{18}=V_0+\frac{N}{N_d^2}\left\{\frac{N_dQ_d\bar{Y}_d^2}{(N-1)}\right\}+\frac{N(N_d-1)}{N_d^2(N-1)}S_{bd}^2,$$

$$\left[k_{12}=\left[\frac{N(N_d-1)}{N_d(N-1)}\right]\right],$$

$$a_2=C_{3d}\sum_{k=1}^{N_{2d}}\frac{M_{dk_2}}{M_d}\left(N_{3d}S_{b_{N3d}}^2-\sum_{k=1}^{N_{3d}}\frac{1}{M_d}S_{dk}^2\right),$$

$$b_2=C_{3d}\sum_{k=1}^{N_{2d}}\frac{M_{dk_2}}{M_d}\sum_{k=1}^{N_{3d}}S_{dk}^2-C_{3d}N_{3d}\sum_{k=1}^{N_{2d}}\frac{M_{dk_2}}{M_d}S_{M_{dk_2}}^2$$

$$e_2=-C_{1d}N_{3d}\sum_{k=1}^{N_{2d}}\frac{M_{dk_2}}{M_d}S_{M_{dk_2}}^2,$$

$$e_3=-C_{3d}\sum_{k=1}^{N_{2d}}\frac{M_{dk_2}}{M_d}\left(\sum_{k=1}^{N_{1d}+N_{2d}}S_{dk}^2-\sum_{k=1}^{N_{2d}}\frac{M_{dk_2}}{M_d}S_{M_{dk_2}}^2\right)$$

$$a_3=\left(C_{2d}N_{1d}+C_{2d}\sum_{k=1}^{N_{2d}}\frac{M_{dk_1}}{M_d}\right)\sum_{k=1}^{N_{2d}}\frac{M_{dk_2}}{M_d}S_{M_{dk_2}}^2,$$

$$b_3=C_{3d}\left\{N_{3d}\sum_{k=1}^{N_{2d}}\frac{M_{dk_2}}{M_d}S_{M_{dk_2}}^2-\sum_{k=1}^{N_{2d}}\frac{M_{dk_2}}{M_d}\sum_{k=1}^{N_{3d}}S_{dk}^2\right\},$$

and $V_0 = 0.0025 \times \bar{Y}_d^2$.

Control Case. In this case it is assumed that there is no nonresponse and complete response is obtained through incurring extra cost. This case was considered for efficiency comparison purpose. Let n

psus be selected from N psus by srswor design where n_d out of n psus fall in the d-th domain and within each selected psu, m_d ssus are also selected from M_d ssus by srswor. Here, the domain sample size n_d is random variable. Data are collected through specialized efforts *i.e.* there is no nonresponse. Then we give the following Theorem,

Theorem 2.4. The unbiased estimator for $\bar{Y}_d$ is given as,

$$\bar{y}_{4d} = \frac{N}{nN_d}\sum_{k=1}^{N_d}\bar{y}_{dk}, \tag{2.13}$$

and the variance of the estimator is,

$$V(\bar{y}_{4d}) = \frac{N(N-n)(N_d-1)}{nN_d^2(N-1)}S_{bd}^2 + \frac{N(N-n)}{nN_d^2(N-1)}N_dQ_d\bar{Y}_d^2$$
$$+\frac{N}{nN_d^2}\sum_{k=1}^{N_d}\left(\frac{1}{m_d}-\frac{1}{M_d}\right)S_{dk}^2, \tag{2.14}$$

where S_{bd}^2, S_{dk}^2, Q_d etc. are defined earlier. If the objective is to estimate the domain total, then an unbiased estimator of Y_d is given as,

$$\hat{Y}_{4d} = \frac{NM_d}{n}\sum_{k=1}^{N_d}\bar{y}_{dk} \tag{2.15}$$

with variance,

$$V(\hat{Y}_{4d}) = \frac{N(N-n)(N_d-1)}{n(N-1)}M_d^2S_{bd}^2$$
$$+\frac{N(N-n)}{n(N-1)}M_d^2N_dQ_d\bar{Y}_d^2$$
$$+\frac{N}{n}\sum_{k=1}^{N_d}\left(\frac{1}{m_d}-\frac{1}{M_d}\right)M_d^2S_{dk}^2. \tag{2.16}$$

The terms defined in the above expression are defined earlier.

Proof:

$$E(\bar{y}_{4d}) = E_1E_2\left[E_3\left(\frac{N}{nN_d}\sum_{k=1}^{n_d}\bar{y}_{dk}\right)\right]$$
$$= E_1\left[E_2\left(\frac{N}{nN_d}\sum_{k=1}^{n_d}\bar{Y}_{dk}\right)\right]$$
$$= E_1\left[\frac{N}{nN_d}\frac{n_d}{N_d}\sum_{k=1}^{N_d}\bar{Y}_{dk}\right]$$
$$= \frac{1}{N_d}\sum_{k=1}^{N_d}\bar{Y}_{dk} = \bar{Y}_d.$$

Hence $\bar{y}_{4d}$ is an unbiased estimator of $\bar{Y}_d$. Here, E_3 is conditional expectation pertaining to all possible samples of size m_d drawn from M_d, E_2 is expectation pertaining to all possible samples of size n_d drawn from N_d keeping n_d fixed and E_1 is the expectations arising out of randomness of n_d.

The variance is given as,

$$V(\bar{y}_{4d}) = V_1E_2E_3(\bar{y}_{4d}) + E_1V_2E_3(\bar{y}_{4d}) + E_1E_2V_3(\bar{y}_{4d})$$

$$V_1E_2E_3(\bar{y}_{4d}) = \frac{N(N-n)}{nN_d^2(N-1)}N_dQ_d\bar{Y}^2,$$

$$E_1V_2E_3(\bar{y}_{4d}) = \frac{N(N-n)(N_d-1)}{nN_d^2(N-1)}S_{bd}^2,$$

$$E_1E_2V_3(\bar{y}_{4d}) = \frac{N}{nN_d^2}\sum_{k=1}^{N_d}\left(\frac{1}{m_d}-\frac{1}{M_d}\right)S_{dk}^2.$$

For optimization the relevant cost function in this case is,

$$C = C_{1d}n_d + C_{3d}m_d.$$

Here expected cost is given as,

$$E(C) = \frac{n}{N}\left(C_{1d}N_d + C_{3d}N_dm_d\right),$$

where, C, C_{1d} and C_{3d} have been defined earlier. To obtain optimum values of n and m_d we minimize the expected cost by fixing the variance. The optimum values are obtained in the same way as earlier.

$$n_{opt} = \frac{\left[\frac{N^2}{N_d^2}\frac{(N_d-1)}{(N-1)}S_{bd}^2 + \frac{N^2}{N_d^2}\left\{\frac{N_dQ_d\bar{Y}_d^2}{(N-1)}\right\} + \frac{N}{N_d^2}\sum_{k=1}^{N_d}\left(\frac{1}{m_d}-\frac{1}{M_d}\right)S_{dk}^2\right]}{\left(V_0 + \frac{N}{N_d^2}\left\{\frac{N_dQ_d\bar{Y}_d^2}{(N-1)}\right\} + \frac{N(N_d-1)}{N_d^2(N-1)}S_{bd}^2\right)} \text{ and}$$

m_{dopt}

$$= \pm \sqrt{\frac{C_{1d} \sum_{k=1}^{N_d} S_{dk}^2}{C_{3d} \left(\frac{N(N_d - 1)}{(N-1)} S_{bd}^2 + \left\{ \frac{N N_d Q_d \bar{Y}_d^2}{(N-1)} \right\} - \frac{1}{M_d} \sum_{k=1}^{N_d} S_{dk}^2 \right)}}.$$

We avoid negative values, therefore,

m_{dopt}

$$= \sqrt{\frac{C_{1d} \sum_{k=1}^{N_d} S_{dk}^2}{C_{3d} \left(\frac{N(N_d - 1)}{(N-1)} S_{bd}^2 + \left\{ \frac{N N_d Q_d \bar{Y}_d^2}{(N-1)} \right\} - \frac{1}{M_d} \sum_{k=1}^{N_d} S_{dk}^2 \right)}}.$$

3. EMPIRICAL ILLUSTRATION

For empirical illustration, first a population of size 1000 was generated from normal distribution with mean 22 and variance 2.5 and then from this population N = 100 psus were formed by combining the adjacent 10 units and allocating them to the respective psus. The population was divided into three domains. The first domain is of size N_1 = 25 psus each of size M_1 = 10 ssus, the second domain is of size N_2 = 35 psus each of size M_2 = 10 ssus and the third domain is of size N_3 = 40 psus each of size M_3 = 10 ssus. The mean and variance of the character under study for the first, second and third domains are 21.90 and 4.36, 22.05 and 4.17, and 21.98 and 3.79 respectively.

From N psus a sample of n = 50 psus each of size 10 ssus are drawn using srswor. As all the domains are of similar nature, hence, we considered different response rate in different domains *i.e.* $M_{dk_1} = M_{dk_2} = 5$ *i.e.* nonresponse rate as 50% for the first domain, $M_{dk_1} = 6,\ M_{dk_2} = 4$ in the second domain *i.e.* nonresponse rate as 40% and $M_{dk_1} = 7,\ M_{dk_2} = 3,$ *i.e.* nonresponse rate as 30% in the third domain. Further, we used various combinations of C_{1d}, C_{2d} and C_{3d}. For computing the optimum values of sample sizes, we used CV of 5 and 10%. The percentage reduction in expected cost (%RIEC) of $\bar{y}_{1d}, \bar{y}_{2d}, \bar{y}_{3d}$ over $\bar{y}_{4d}$ along with optimum values of sample sizes for different combinations of C_{1d}, C_{2d} and C_{3d} are given in Tables 3.1 and 3.4 for domain 1, in Tables 3.2 and 3.5 for domain 2 and in Tables 3.3 and 3.6 for domain 3. Here,

Table 3.1. The optimum values of sample sizes (for CV of 5%) along with percentage reduction in expected cost of $\bar{y}_{11}, \bar{y}_{21}, \bar{y}_{31}$ over $\bar{y}_{41}$ in domain 1.

cost			Control ($\bar{y}_{41}$)		First estimator ($\bar{y}_{11}$)				Second estimator ($\bar{y}_{21}$)				Third estimator ($\bar{y}_{31}$)			
C_{11}	C_{21}	C_{31}	n	m_1	n	m_1	f_{21}	% RIEC	n	m_1	f_{21}	% RIEC	n	m_1	f_{21}	% RIEC
25	1	45	24	6	6	9	1.37	65.33	6	8	5.21	98.71	6	7	1.92	97.25
25	1	50	24	6	6	9	1.45	65.44	6	7	5.48	98.77	6	6	2.02	97.49
25	1	55	24	6	6	9	1.52	65.54	6	7	5.75	98.82	6	6	2.12	97.68
25	2	45	24	6	6	8	0.97	55.47	6	6	3.68	98.27	6	7	1.38	96.34
25	2	50	24	6	6	8	1.02	55.55	6	6	3.87	98.36	6	6	1.45	96.67
25	2	55	24	6	6	8	1.07	55.61	6	6	4.06	98.44	6	6	1.52	96.94
30	1	45	24	7	6	9	1.37	67.10	6	8	5.21	98.69	6	7	1.92	97.38
30	1	50	24	7	6	9	1.45	67.23	6	8	5.48	98.75	6	6	2.02	97.60
30	1	55	24	6	6	9	1.52	67.35	6	8	5.75	98.80	6	6	2.12	97.78
30	2	45	24	7	6	9	0.97	58.29	6	7	3.68	98.25	6	7	1.38	96.55
30	2	50	24	7	6	9	1.02	58.39	6	7	3.87	98.34	6	6	1.45	96.85
30	2	55	24	6	6	9	1.07	58.47	6	7	4.06	98.42	6	6	1.52	97.10
35	1	45	24	7	6	9	1.37	68.40	6	9	5.21	98.67	6	7	1.92	97.47
35	1	50	24	7	6	9	1.45	68.55	6	9	5.48	98.73	6	6	2.02	97.68
35	1	55	24	7	6	9	1.52	68.68	6	9	5.75	98.78	6	6	2.12	97.85
35	2	45	24	7	6	9	0.97	60.41	6	8	3.68	98.24	6	7	1.38	96.70
35	2	50	24	7	6	9	1.02	60.52	6	7	3.87	98.33	6	6	1.45	96.99
35	2	55	24	7	6	9	1.07	60.61	6	7	4.06	98.40	6	6	1.52	97.20

Table 3.2. The optimum values of sample sizes (for CV of 5%) along with percentage reduction in expected cost of $\bar{y}_{12}, \bar{y}_{22}, \bar{y}_{32}$ over $\bar{y}_{42}$ in domain 2.

cost			Control ($\bar{y}_{42}$)		First estimator ($\bar{y}_{12}$)				Second estimator ($\bar{y}_{22}$)				Third estimator ($\bar{y}_{32}$)			
C_{12}	C_{22}	C_{32}	n	m_2	n	m_2	f_{22}	% RIEC	n	m_2	f_{22}	% RIEC	n	m_2	f_{22}	% RIEC
25	1	45	30	7	4	7	1.05	81.89	4	6	1.00	98.25	4	3	3.34	97.49
25	1	50	30	7	4	7	1.11	81.96	4	6	1.06	98.33	4	3	3.53	97.64
25	1	55	30	6	4	7	1.16	82.03	4	6	1.11	98.40	4	3	3.70	97.76
25	2	45	30	7	4	6	0.74	76.02	4	5	0.71	97.58	4	3	2.36	97.25
25	2	50	30	7	4	6	0.78	76.07	4	5	0.75	97.70	4	3	2.49	97.42
25	2	55	30	6	4	6	0.82	76.11	4	5	0.78	97.80	4	3	2.61	97.57
30	1	45	30	8	4	7	1.05	82.97	4	7	1.00	98.25	4	3	3.34	97.35
3 0	1	50	30	7	4	7	1.11	83.05	4	7	1.06	98.33	4	3	3.53	97.51
3 0	1	55	30	7	4	7	1.16	83.12	4	7	1.11	98.40	4	3	3.70	97.63
30	2	45	30	8	4	7	0.74	77.67	4	6	0.71	97.58	4	3	2.36	97.13
3 0	2	50	30	7	4	7	0.78	77.73	4	6	0.75	97.70	4	3	2.49	97.30
30	2	55	30	7	4	7	0.82	77.79	4	6	0.78	97.80	4	3	2.61	97.45
35	1	45	30	8	4	7	1.05	83.77	4	8	1.00	98.24	4	3	3.34	97.23
35	1	50	30	8	4	7	1.11	83.86	4	7	1.06	98.32	4	3	3.53	97.38
35	1	55	30	7	4	7	1.16	83.94	4	7	1.11	98.39	4	3	3.70	97.51
35	2	45	30	8	4	7	0.74	78.92	4	6	0.71	97.58	4	3	2.36	97.01
35	2	50	30	8	4	7	0.78	79.00	4	6	0.75	97.70	4	3	2.49	97.19
35	2	55	30	7	4	7	0.82	79.06	4	6	0.78	97.80	4	3	2.61	97.34

Table 3.3. The optimum values of sample sizes (for CV of 5%) along with percentage reduction in expected cost of $\bar{y}_{13}, \bar{y}_{23}, \bar{y}_{33}$ over $\bar{y}_{43}$ in domain 3.

cost			Control ($\bar{y}_{43}$)		First estimator ($\bar{y}_{13}$)				Second estimator ($\bar{y}_{23}$)				Third estimator ($\bar{y}_{33}$)			
C_{13}	C_{23}	C_{33}	n	m_3	n	m_3	f_{23}	% RIEC	n	m_3	f_{23}	% RIEC	n	m_3	f_{23}	% RIEC
25	1	45	20	9	4	13	0.96	62.40	4	6	1.55	97.41	4	1	3.00	96.42
25	1	50	20	8	4	13	1.02	62.52	4	5	1.63	97.57	4	1	3.15	96.60
25	1	55	20	8	4	13	1.07	62.62	4	5	1.71	97.70	4	1	3.30	96.77
25	2	45	20	9	3	13	0.68	48.65	4	5	1.09	96.13	4	1	2.15	96.28
25	2	50	20	8	3	13	0.72	48.73	4	5	1.15	96.37	4	1	2.26	96.49
25	2	55	20	8	3	13	0.75	48.80	4	4	1.21	96.58	4	1	2.37	96.66
30	1	45	20	9	4	13	0.96	65.05	4	6	1.55	97.40	4	1	3.00	96.13
30	1	50	20	9	4	13	1.02	65.19	4	6	1.63	97.55	4	1	3.15	96.33
30	1	55	20	9	4	13	1.07	65.31	4	6	1.71	97.68	4	1	3.30	96.51
30	2	45	20	9	3	13	0.68	52.57	4	5	1.09	96.12	4	1	2.15	96.00
30	2	50	20	9	3	13	0.72	52.67	4	5	1.15	96.36	4	1	2.26	96.22
30	2	55	20	9	3	13	0.75	52.76	4	5	1.21	96.57	4	1	2.37	96.40
35	1	45	20	9	4	13	0.96	67.06	4	7	1.55	97.39	4	1	3.00	95.87
35	1	50	20	10	4	13	1.02	67.22	4	6	1.63	97.54	4	1	3.15	96.08
35	1	55	20	10	4	13	1.07	67.35	4	6	1.71	97.67	4	1	3.30	96.27
35	2	45	20	9	3	13	0.68	55.57	4	6	1.09	96.12	4	1	2.15	95.75
35	2	50	20	10	3	13	0.72	55.68	4	5	1.15	96.36	4	1	2.26	95.97
35	2	55	20	10	3	13	0.75	55.78	4	5	1.21	96.56	4	1	2.37	96.17

Table 3.4. The optimum values of sample sizes (for CV of 10%) along with percentage reduction in expected cost of $\bar{y}_{11}, \bar{y}_{21}, \bar{y}_{31}$ over $\bar{y}_{41}$ in domain 1.

cost			Control ($\bar{y}_{41}$)		First estimator ($\bar{y}_{11}$)				Second estimator ($\bar{y}_{21}$)				Third estimator ($\bar{y}_{31}$)			
C_{11}	C_{21}	C_{31}	n	m_1	n	m_1	f_{21}	% RIEC	n	m_1	f_{21}	% RIEC	n	m_1	f_{21}	% RIEC
25	1	45	19	6	6	9	1.37	56.01	6	8	5.20	98.36	6	7	1.92	96.51
25	1	50	19	6	6	9	1.45	56.15	6	7	5.48	98.44	6	6	2.02	96.81
25	1	55	19	6	6	9	1.52	56.28	6	7	5.75	98.51	6	6	2.12	97.06
25	2	45	19	6	6	8	0.97	43.51	6	6	3.68	97.81	6	7	1.38	95.36
25	2	50	19	6	6	8	1.02	43.60	6	6	3.87	97.92	6	6	1.45	95.77
25	2	55	19	6	6	8	1.07	43.68	6	6	4.06	98.02	6	6	1.52	96.12
30	1	45	19	7	6	9	1.37	58.26	6	8	5.20	98.33	6	7	1.92	96.68
30	1	50	19	7	6	9	1.45	58.42	6	8	5.48	98.41	6	6	2.02	96.96
30	1	55	19	6	6	9	1.52	58.57	6	8	5.75	98.48	6	6	2.12	97.19
30	2	45	19	7	6	9	0.97	47.08	6	7	3.68	97.78	6	7	1.38	95.62
30	2	50	19	7	6	9	1.02	47.20	6	7	3.87	97.90	6	6	1.45	96.01
30	2	55	19	6	6	9	1.07	47.31	6	7	4.06	98.00	6	6	1.52	96.30
35	1	45	19	7	6	9	1.37	59.91	6	9	5.20	98.31	6	7	1.92	96.80
35	1	50	19	7	6	9	1.45	60.10	6	9	5.48	98.39	6	6	2.02	97.06
35	1	55	19	7	6	9	1.52	60.26	6	9	5.75	98.46	6	6	2.12	97.28
35	2	45	19	7	6	9	0.97	49.76	6	8	3.68	97.76	6	7	1.38	95.81
35	2	50	19	7	6	9	1.02	49.90	6	7	3.87	97.88	6	6	1.45	96.18
35	2	55	19	7	6	9	1.07	50.03	6	7	4.06	97.98	6	6	1.52	96.48

Table 3.5. The optimum values of sample sizes (for CV of 10%) along with percentage reduction in expected cost of $\bar{y}_{12}, \bar{y}_{22}, \bar{y}_{32}$ over $\bar{y}_{42}$ in domain 2.

cost			Control ($\bar{y}_{42}$)		First estimator ($\bar{y}_{12}$)				Second estimator ($\bar{y}_{22}$)				Third estimator ($\bar{y}_{32}$)			
C_{12}	C_{22}	C_{32}	n	m_2	n	m_2	f_{22}	% RIEC	n	m_2	f_{22}	% RIEC	n	m_2	f_{22}	% RIEC
25	1	45	22	7	4	7	1.05	66.97	4	6	1.00	96.81	4	3	3.34	95.42
25	1	50	22	7	4	7	1.11	67.10	4	6	1.06	96.95	4	3	3.53	95.69
25	1	55	22	6	4	7	1.16	67.21	4	6	1.11	97.08	4	3	3.70	95.92
25	2	45	22	7	4	6	0.74	56.25	4	5	0.71	95.58	4	3	2.36	94.98
25	2	50	22	7	4	6	0.78	56.34	4	5	0.75	95.80	4	3	2.49	95.29
25	2	55	22	6	4	6	0.82	56.42	4	5	0.78	95.99	4	3	2.61	95.56
30	1	45	22	8	4	7	1.05	68.93	4	7	1.00	96.80	4	3	3.34	95.17
30	1	50	22	7	4	7	1.11	69.08	4	7	1.06	96.95	4	3	3.53	95.45
30	1	55	22	7	4	7	1.16	69.21	4	7	1.11	97.07	4	3	3.70	95.68
30	2	45	22	8	4	7	0.74	59.28	4	6	0.71	95.58	4	3	2.36	94.76
30	2	50	22	7	4	7	0.78	59.39	4	6	0.75	95.80	4	3	2.49	95.08
30	2	55	22	7	4	7	0.82	59.48	4	6	0.78	95.99	4	3	2.61	95.35
35	1	45	22	8	4	7	1.05	70.39	4	8	1.00	96.79	4	3	3.34	94.94
35	1	50	22	8	4	7	1.11	70.55	4	7	1.06	96.94	4	3	3.53	95.22
35	1	55	22	7	4	7	1.16	70.70	4	7	1.11	97.07	4	3	3.70	95.46
35	2	45	22	8	4	7	0.74	61.56	4	6	0.71	95.59	4	3	2.36	94.56
35	2	50	22	8	4	7	0.78	61.69	4	6	0.75	95.80	4	3	2.49	94.88
35	2	55	22	7	4	7	0.82	61.80	4	6	0.78	95.99	4	3	2.61	95.15

Table 3.6. The optimum values of sample sizes (for CV of 10%) along with percentage reduction in expected cost of $\bar{y}_{13}, \bar{y}_{23}, \bar{y}_{33}$ over $\bar{y}_{43}$ in domain 3.

cost			Control ($\bar{y}_{43}$)		First estimator ($\bar{y}_{13}$)				Second estimator ($\bar{y}_{23}$)				Third estimator ($\bar{y}_{33}$)			
C_{13}	C_{23}	C_{33}	n	m_3	n	m_3	f_{23}	% RIEC	n	m_3	f_{23}	% RIEC	n	m_3	f_{23}	% RIEC
25	1	45	12	9	3	13	0.96	40.03	3	6	1.55	95.9	4	1	3.00	94.28
25	1	50	12	8	3	13	1.02	40.22	3	5	1.63	96.1	4	1	3.15	94.59
25	1	55	12	8	3	13	1.07	40.39	3	5	1.71	96.3	4	1	3.30	94.84
25	2	45	12	9	3	13	0.68	18.11	3	5	1.09	93.8	4	1	2.15	94.07
25	2	50	12	8	3	13	0.72	18.24	3	5	1.15	94.2	4	1	2.26	94.39
25	2	55	12	8	3	13	0.75	18.35	3	4	1.21	94.5	4	1	2.37	94.67
30	1	45	12	9	3	13	0.96	44.27	3	6	1.55	95.9	4	1	3.00	93.83
30	1	50	12	9	3	13	1.02	44.49	3	6	1.63	96.1	4	1	3.15	94.15
30	1	55	12	9	3	13	1.07	44.68	3	6	1.71	96.3	4	1	3.30	94.43
30	2	45	12	10	3	13	0.68	24.37	3	5	1.09	93.8	4	1	2.15	93.63
30	2	50	12	9	3	13	0.72	24.52	3	5	1.15	94.2	4	1	2.26	93.97
30	2	55	12	9	3	13	0.75	24.66	3	5	1.21	94.5	4	1	2.37	94.26
35	1	45	12	9	3	13	0.96	47.47	3	7	1.55	95.8	4	1	3.00	93.42
35	1	50	12	10	3	13	1.02	47.72	3	6	1.63	96.1	4	1	3.15	93.76
35	1	55	12	10	3	13	1.07	47.93	3	6	1.71	96.3	4	1	3.30	94.05
35	2	45	12	10	3	13	0.68	29.14	3	6	1.09	93.8	4	1	2.15	93.22
35	2	50	12	10	3	13	0.72	29.33	3	5	1.15	94.2	4	1	2.26	93.58
35	2	55	12	10	3	13	0.75	29.49	3	5	1.21	95	4	1	2.37	93.89

Tables 3.1, 3.2 and 3.3 report the results with respect to CV of 5% and Tables 3.4, 3.5 and 3.6 for the CV of 10%. The empirical analysis is carried out using SAS 9.3 software package.

4. DISCUSSION AND CONCLUSION

A close perusal of all the results in Tables 3.1-3.6 show that in terms of the criteria of the percentage reduction in expected cost (%RIEC), the proposed estimators of domain mean based on sub-sampling of the non-respondents technique for the three different cases of occurrence of nonresponse are better than the estimator based only on interview method of data collection resulting in 100% response. Further, in Tables 3.1 and 3.4 for domain 1 where we assumed the response rate as 50%, the %RIEC increases with increase in travel and miscellaneous cost (C_{11}) for case 1 and 3 of nonresponse but it decreases with increase in the same for the case 2. Tables 3.2 and 3.5 for domain 2, where we assumed the response rate as 40% and Tables 3.3 and 3.6 for domain 3 where we assumed the response rate as 30%, it can be seen that, the %RIEC increases with increase in travel and miscellaneous cost (C_{11}) for case 1, remains almost constant for the case 2 and decreases in case of case 3. Again from all the Tables and for all the domains it can be seen that, for all the cases the %RIEC decreases with increase in data collection cost at first attempt (C_{21}). Again, for all the three cases the %RIEC increases with the increase in the cost per unit of collecting the information by expensive method after the first attempt to obtain information failed (C_{31}) for all the three domains. Further, the rate of increase of %RIEC is maximum, as cost of data collection by expensive method increases, for case 3 while in case 1 and 2, the increase in the same is almost equal. Finally, the %RIEC is maximum for case 1 in domain 1 followed by domain 3 and it is least in domain 2. The %RIEC is maximum for the case 2 in domain 2 followed by domain 3 and it is least in domain 1 while the same is maximum for the case 3 in domain 3 followed by domain 2 and it is least in domain 1.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the valuable comments and suggestions of the Associate Editor and the referee. These led to a considerable improvement in the paper.

REFERENCES

Aditya, K., Sud, U.C., and Chandra H. (2013). Estimation of domain total for unknown domain size in the presence of nonresponse. *Statist. appl.*, **10**,13-25.

Agrawal, M.C., and Mrida, C.K. (2007). Some efficient estimators of the domain parameters. *Statist. Probab. Lett.*, **77**, 704-709.

Chhikara, Raj S., and Sud, U.C. (2009). Estimation of population and domain totals under two-phase sampling in the presence of non-response. *J. Ind. Soc. Agril. Statist.,* **63(3)**, 297-304.

Cochran, W.G. (1977). *Sampling Techniques*, 3rd Edition. John Wiley & Sons, Inc., New York.

Durbin, J. (1954). Non-response and call-backs in surveys. *Bull. Intern. Statist. Inst.,* **34,** 72-86.

Foradori, G.T. (1961). Some non-response sampling theory for two stage designs. Institute of Statistics, North Carolina State College.

Hansen, M.H. and Hurwitz, W.N. (1946). The problem of non-response in sample surveys. *J. Amer. Statist. Assoc.*, **41**, 517-529.

Okafor, F.C. (2001). Treatment of non-response in successive sampling. *Statistica,* **61(2)**, 195-204.

Okafor, F.C. (2005). Sub-sampling the non-respondents in two-stage sampling over successive occasions. *J. Ind. Statist. Assoc.*, **43**, 33-49.

Okafor, F.C., and Lee, H. (2000). Double sampling for ratio and regression estimation with sub-sampling the non-respondents. *Survey Methodology*, **26(2),** 183-188.

Kalton, G. and Kasprzyk, D. (1986). The treatment of missing survey data. *Survey Methodology*, **12**, 1-16.

Oh, H.L., and Scheuren, F.J. (1983). Weighting adjustment for unit non-response. In: W.G. Madow, I. Olkin, and B. Rubin (eds.), *Incomplete Data in Sample Surveys*, Vol.2. Academic press, New York, 143-184.

Rancourt, E., Lee, H. and Särndal, C.E. (1994). Bias corrections for survey estimates from data with ratio imputed values for confounded non-response. *Survey Methodology*, **20**, 137-147.

Singh, R. and Mangat, N.P.S. (1996). *Elements of Survey Sampling*. Kluwer Academic publishers.

Sarndal, C.E., Swensson, B. and Wretman, J. (1992). *Model Assisted Survey Sampling*, Springer-Verlag, New York.

Srinath, K.P. (1971). Multiphase sampling in non-response problems. *J. Amer. Statist. Assoc.*, **66,** 583-586.

Sud, U.C., Chandra H. and Chikkara, R.S. (2010). Domain estimation in the presence of non-response. *J. Ind. Soc. Agril. Statist.*, **64**, 343-347.

Tripathi, T.P. and Khare, B.B. (1997). Estimation of mean vector in presence of non-response. *Commu. Statist.-Theory Methods*, **26(9)**, 2255-2269.