Two Stage Sampling with Two-phases at the Second Stage of Sampling for Estimation of Finite Population Mean under Random Response Mechanism

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

Received 25 January 2012; Revised 29 April 2013; Accepted 30 April 2013

SUMMARY

The problem of estimation of finite population mean in the presence of the random response has been considered when the sampling design is two-stage with two phases at the second stage. Three different types of estimators, based on subsampling of nonrespondents, collecting data on the subsample through specialized efforts, are developed. Expressions for the variances of the estimators along with unbiased variance estimators are developed. Optimum values of sample sizes are obtained by considering a suitable cost function. The percentage reduction in the expected cost of the proposed estimators is studied empirically.

Keywords: Cost function, Nonresponse, Random response, Population mean, Subsampling, 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 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 obtained from all the units in surveys. The problem of nonresponse persist even after call backs. The estimates obtained from incomplete data may be biased particularly when the respondents differ from the nonrespondents. Hansen and Hurwitz (1946) proposed a technique for adjusting for nonresponse to address the problem of bias. The technique consists of selecting a subsample of nonrespondents. Through specialized efforts data are collected from the nonrespondents so as to obtain an estimate of nonresponding units in the population. Foradori (1961) studied the subsampling of the nonrespondents technique to estimate the population total in two stages using unequal probability sampling. Srinath (1971) used a different procedure for selecting the subsample of respondents where the subsampling procedure varied according to the nonresponse rates.


Corresponding author: Kaustav Aditya
E-mail address: kaustav@iasri.res.in
context of element sampling and two-stage sampling respectively on two successive occasions. Chhikara and Sud (2009) used the approach for estimation of population and domain totals in the context of item nonresponse. Again, Sud et al. (2012) considered the problem of estimation of finite population mean in the presence of nonresponse under two stage sampling design when the response mechanism was assumed to be deterministic. It may be mentioned that the weighting and imputation procedures aim at eliminating the bias caused by nonresponse. However, these procedures are based on 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 subsampling 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 very expensive.

In what follows, different estimators of population mean using two-stage sampling designs are developed in Section 2 based on the technique of subsampling the nonrespondents, where nonresponse mechanism is assumed to follow Bernoulli distribution within each selected primary stage units (psus). However, it is assumed that the response mechanism is deterministic at the primary stage unit (psu) level i.e. the entire population of psus can be assumed to be divided into responding and nonresponding groups. Also given are expressions for variance of the estimators and unbiased variance estimators. Optimum values of sample sizes are obtained by minimizing the expected cost for a fixed variance. The results are empirically illustrated in Section 5.3.

2. THEORETICAL DEVELOPMENTS

Let the finite population \( U \) under consideration consists of \( N \) known psus labelled 1 through \( N \). Let the \( i \)-th psu comprise \( M \) second stage units (ssus). Let \( y_{ij} \) be the value of study character pertaining to \( j \)-th ssu in the \( i \)-th psu, \( i=1, 2, \ldots, N, j=1, 2, \ldots, M \). The objective is to estimate the population mean which is defined as,

\[
\bar{Y} = \frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} y_{ij}.
\]

**Case 1.** Let \( n \) psus be selected by simple random sampling without replacement (srs) from \( N \) and within each selected psu, \( m \) ssus are also selected by srs. Further, out of \( m \) ssus, \( m_1 \) ssus respond while \( m_2 \) ssus do not respond, \( m_1 + m_2 = m \). A subsample of size \( h \) is selected from \( m_2 \) by srs and data are collected on the subsampled units through specialized efforts, \( m_2 = h \frac{m_2}{m} \), \( i = 1, 2, \ldots, n \). Here at the second stage \( m_1 \) responding and \( m_2 \) nonresponding units are being generated as a result of \( m \) independent Bernoulli trials, one for each element \( i \) in \( m \) with constant probability \( \theta_i \) of “success”, i.e. the response. So, we have \( \Pr(i \in m_1 \mid m) = \theta_i\) and \( \Pr(i \in m_1 \mid m) = \theta_i^2 \).

**Theorem 2.1** An unbiased estimator of \( \bar{Y} \) is given by

\[
\bar{\bar{Y}} = \frac{1}{n} \sum_{i=1}^{n} \left( m_1 \bar{y}_{m_1} + m_2 \bar{y}_{m_2} \right),
\]  

with variance

\[
V(\bar{\bar{Y}}) = \left( \frac{1}{n} - \frac{1}{N} \right) S_b^2 + \frac{N}{Nn} \sum_{i=1}^{N} \left( m_i - 1 \right) S_{im}^2 + \frac{1}{Nnm} \sum_{i=1}^{N} \left( 1 - \theta_i \right) (f_{i2} - 1) S_{im}^2.
\]

An unbiased variance estimator of (2.2) is given as

\[
\hat{V}(\bar{\bar{Y}}) = \left( \frac{1}{n} - \frac{1}{N} \right) S_b^2 + \frac{(N-1)}{N(n-1)} \sum_{i=1}^{N} \left( m_i - 1 \right) S_{im}^2 + \frac{1}{Nmn(n-1)} \sum_{i=1}^{N} \left( D - \frac{M}{m} \right) \frac{S_{im}^2}{\alpha}.
\]

where

\[
S_b^2 = \frac{1}{(N-1)} \sum_{i=1}^{N} \left( \bar{y}_i - \bar{Y}_M \right)^2, \quad \bar{Y}_M = \frac{1}{M} \sum_{j=1}^{M} Y_{ij},
\]

\[
S_{im}^2 = \frac{1}{(M-1)} \sum_{j=1}^{M} (Y_{ij} - \bar{Y}_M)^2, \quad D = \frac{m_1^2}{m^2} + \frac{m_2^2}{m^2}
\] and
\[ \alpha = \frac{(m+1)M}{m} - \frac{m_2}{m^2(m-1)}(f_{i2} - 1) + \frac{[D-m]}{M(m-1)}. \]

Define,
\[ s_{h_2}^2 = \frac{1}{(n-1)} \left( \sum_{i=1}^{n} \bar{y}_{im}^2 - n\bar{y}_{i_2}^2 \right), \bar{y}_{im} = \frac{1}{m} \left( m_1 \bar{y}_{m_1} + m_2 \bar{y}_{m_2} \right), \]
\[ s_{im}^2 = \frac{1}{(m-1)} \left( \sum_{j=1}^{m} y_{ij}^2 + \frac{m_2}{h_2} \sum_{j=1}^{h_2} y_{ij}^2 - m\bar{y}_{im}^2 \right), \bar{y}_{im} = \frac{1}{m} \sum_{j=1}^{m} y_{ij}. \]

**Proof:** By definition, we have
\[ E(\bar{y}_r) = E_1 E_2 E_3 E_4 \left[ E_5 \left( \frac{1}{n} \sum_{i=1}^{n} \left( m_1 \bar{y}_{m_1} + m_2 \bar{y}_{m_2} \right) \right) \right] \]
\[ = E_1 E_2 E_3 \left[ E_4 \left( \frac{1}{n} \sum_{i=1}^{n} \left( m_1 \bar{y}_{m_1} + m_2 \bar{y}_{m_2} \right) \right) \right] \]
\[ = E_1 \left[ E_2 \left( \frac{1}{n} \sum_{i=1}^{n} \left( \bar{y}_{im} + (1-\theta) \bar{y}_{im} \right) \right) \right] \]
\[ = E_1 \left[ \frac{1}{n} \sum_{i=1}^{n} \left( \bar{y}_{im} + (1-\theta) \bar{y}_{im} \right) \right] \]
\[ = E_1 \left[ \frac{1}{n} \sum_{i=1}^{n} \bar{y}_{im} \right] = \frac{1}{N} \sum_{i=1}^{N} \bar{y}_{im} = \bar{Y}. \]

This shows that \( \bar{y}_r \) is an unbiased estimator of the population mean \( \bar{Y} \). Here, \( E_2 \) represents the conditional expectations of all possible samples of size \( h_2 \) drawn from \( m_2 \), \( E_4 \) is the conditional expectation of all possible samples of size \( m_1 \), \( m_2 \) respectively drawn from \( m \) by keeping \( m_1 \), \( m_2 \) fixed, \( E_3 \) is the conditional expectation arising out of \( m \) independent Bernoulli trials leading to \( m_1 \) success and \( m_2 \) failures, \( m_1 + m_2 = m \), \( E_3 \) is the conditional expectation of all possible samples of size \( m \) drawn from \( M \) and \( E_1 \) refers to expectation arising out of all possible samples of size \( n \) drawn from a population of size \( N \).

Similarly, we can write
\[ V(\bar{y}_r) = V_1 \{ E_2 E_3 E_4 E_5 (\bar{y}_r) \} + E_1 V_2 \{ E_3 E_4 E_5 (\bar{y}_r) \} \]
\[ + E_1 E_2 V_3 \{ E_4 E_5 (\bar{y}_r) \} + E_1 E_2 E_3 V_4 \{ E_5 (\bar{y}_r) \} \]
\[ + E_1 E_2 E_3 E_4 \{ V_5 (\bar{y}_r) \}. \]

Various terms are expressed as below.

\[ V_1 \{ E_2 E_3 E_4 E_5 (\bar{y}_r) \} = \left( \frac{1}{n} - \frac{1}{N} \right) S_2^2, \]
\[ E_1 V_2 \{ E_3 E_4 E_5 (\bar{y}_r) \} = \frac{1}{Nn} \sum_{i=1}^{N} \left( \frac{1}{m} - \frac{1}{M} \right) S_{im}^2, \]
\[ E_1 E_2 V_3 \{ E_4 E_5 (\bar{y}_r) \} = 0, \]
\[ E_1 E_2 E_3 V_4 \{ E_5 (\bar{y}_r) \} = 0, \]
\[ E_1 E_2 E_3 E_4 \{ V_5 (\bar{y}_r) \} = \frac{1}{Nn} \sum_{i=1}^{N} (1-\theta) \left( f_{i2} - 1 \right) S_{im}^2. \]

Here, \( V_1, V_2, V_3, V_4 \) and \( V_5 \) are defined similarly as \( E_1, E_2, E_3, E_4 \) and \( E_5 \). Hence, by adding all the terms we get,
\[ V(\bar{y}_r) = \left( \frac{1}{n} - \frac{1}{N} \right) S_2^2 + \frac{1}{Nn} \sum_{i=1}^{N} \left( \frac{1}{m} - \frac{1}{M} \right) S_{im}^2 \]
\[ + \frac{1}{Nnm} \sum_{i=1}^{N} (1-\theta) \left( f_{i2} - 1 \right) S_{im}^2. \]

To obtain an unbiased variance estimator, we proceed as follows,
\[ \text{Consider, } s_{h_2}^2 = \frac{1}{(n-1)} \left( \sum_{i=1}^{n} \bar{y}_{im}^2 - n\bar{y}_{i_2}^2 \right) \text{ where, } \]
\[ \bar{y}_{im} = \frac{1}{m} \left( m_1 \bar{y}_{m_1} + m_2 \bar{y}_{m_2} \right). \]

It can be shown that,
\[ E_1 E_2 E_3 E_4 E_5 (s_{h_2}^2) = S_2^2 - \frac{1}{N(n-1)} \sum_{i=1}^{N} \left( \frac{1}{m} - \frac{1}{M} \right) S_{im}^2 \]
\[ + \frac{1}{N} \sum_{i=1}^{N} \left( \frac{1-\theta}{m} \right) (f_{i2} - 1) S_{im}^2 - \frac{n}{N(n-1)M} \]
\[ \times \sum_{i=1}^{N} \left( \frac{2\theta(1-\theta)}{m} \right) + (1-\theta)^2 + \theta^2 \frac{M}{m} \right) S_{im}^2 + \text{ and } \]
\[ E_2 E_3 E_4 E_5 (s_{im}^2) = \frac{(m+1)M}{m} S_{im}^2 - \frac{(1-\theta)}{m(m-1)} (f_{i2} - 1) S_{im}^2 \]
\[ + \frac{1}{(m-1)} \left( \frac{2\theta(1-\theta)}{m} + (1+\theta)^2 + \theta^2 - m \right) \frac{S_{im}^2}{M}. \]
Therefore,
\[
\hat{S}_b^2 = S_b^2 + \frac{1}{n(n-1)} \sum_{i=1}^{n} \left( \frac{1}{m} - 1 \right) \frac{S_{im}^2}{\alpha} \\
- \frac{1}{n} \sum_{i=1}^{n} m_i f_{i2} (f_{i2} - 1) \frac{S_{im}^2}{\alpha} + \frac{1}{M} \sum_{i=1}^{n} (D - M) \frac{S_{im}^2}{\alpha}
\]
and
\[
\hat{S}_{IM}^2 = \frac{S_{im}^2}{\alpha}
\]

Substituting the estimated values of \( S_b^2 \) and \( S_{IM}^2 \) in Eq. (2.2) we get the required result.

As the total cost of the survey is proportional to the optimum size of the sample, we determine the optimum values of \( n, m \) and \( f_{i2} \) by minimizing the expected cost for a fixed variance. To achieve this, consider the following cost function
\[
C = C_1 n + C_2 \sum_{i=1}^{n} m_i + C_3 \sum_{i=1}^{n} f_{i2}
\]
where,
\( C \) : Total cost,
\( C_1 \) : Per unit travel and miscellaneous cost between the psus,
\( C_2 \) : Cost per unit of collecting the information on the study character in the first attempt,
\( C_3 \) : Cost per unit of collecting the information by expensive method after the first attempt to obtain information failed.

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 mail. Many of the respondents may not send the required information through mails. To collect information, a subsample of nonrespondents may be obtained for data collection by specialized effort, say, personal interview.

The expected cost in this case is,
\[
C' = E(C) = n \left[ C_1 + C_2 \sum_{i=1}^{N} m_i \theta_i + C_3 \sum_{i=1}^{N} \frac{m(1-\theta_i)}{f_{i2}} \right]
\]
To minimize the expected cost consider the function \( \phi = C' + \lambda \{ V(\mathbf{y}^*) - V_0 \} \). Here, \( \lambda \) is the Lagrangian multiplier. Also, we determine \( V_0 \) by fixing the coefficient of variation, say equal to 5%. To obtain closed form expressions for the various sample sizes we have considered \( m_i = h_{l2} f_{l2} \) in place of \( m_i = h_{l2} f_{l2} \), \( i = 1, 2, \ldots, n \). Differentiating with respect to \( n, m, \lambda \) and \( f_{l2} \) equating the resultant derivatives to "0" and simplifying give the optimum values as
\[
n_{opt} = \frac{k}{V_0 + S_b^2},
\]
\[
m_{opt} = \frac{-(B - D_1) + \sqrt{(B - D_1)^2 + 4AE}}{2A}
\]
and
\[
f_{2opt} = \pm \sqrt{C \sum_{i=1}^{N} (1-\theta_i) \left[ m_s^2 - \frac{1}{N} \sum_{i=1}^{N} (1-\theta_i) S_{im}^2 + \frac{1}{N} \sum_{i=1}^{N} \left( \frac{M - m}{M} \right) S_{im}^2 \right]}
\]
where,
\[k = S_b^2 + \frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{m} - 1 \right) S_{im}^2 + \frac{1}{N} \sum_{i=1}^{N} \left( \frac{1 - \theta_i}{m} \right) (f_{i2} - 1) S_{im}^2,
\]
\[D_1 = \left[ \frac{C_2}{N} \sum_{i=1}^{N} \theta_i + \frac{C_3}{N} \sum_{i=1}^{N} \frac{(1-\theta_i)}{f_{i2}} \right] \sum_{i=1}^{N} (1-\theta_i) S_{im}^2,
\]
\[A = C_3 \sum_{i=1}^{N} \frac{(1-\theta_i)}{f_{i2}^2} \sum_{i=1}^{N} \frac{S_{im}^2}{N},
\]
\[B = C_3 \sum_{i=1}^{N} \frac{(1-\theta_i)}{f_{i2}^2} \sum_{i=1}^{N} (f_{i2} - 1) S_{im}^2 + \sum_{i=1}^{N} S_{im}^2
\]
E = \sum_{i=1}^{N} s^2_{im} \text{ and } V_0 = 0.0025 \times \bar{Y}^2.

\textbf{Case 2.} Consider the situation that a sample of } n \text{ psus is drawn from } N \text{ and within each selected psu a sample of } m \text{ ssus is drawn by srswor design. Let there be no nonresponse in } n_1 \text{ psus. In the remaining } n_2 \text{ psus, } m_1 \text{ ssus respond while } m_2 \text{ ssus do not respond. A subsample of } h_{i2} \text{ units is selected from } m_{i2} \text{ by srswor and data are collected through specialized efforts, } m_{i2} = h_{i2}.

f_{i2}, i = 1, 2, \ldots, n_2.

In this context, we state the Theorem 2.2 as below.

\textbf{Theorem 2.2.} The estimator

\[ \bar{y}_{r} = \frac{1}{n} \left\{ \sum_{i=1}^{n} \bar{y}_{im} + \sum_{i=1}^{n} \frac{(m_1 \bar{y}_{m_1} + m_2 \bar{y}_{m_2})}{m} \right\} \]  

is unbiased for } \bar{Y}, \text{ with variance

\[ V(\bar{y}_{r}) = \left( \frac{1}{n} - \frac{1}{N} \right) S_b^2 + \frac{N-1}{Nn(n-1)} \sum_{i=1}^{n} \left( \frac{1}{m} - \frac{1}{M} \right) s^2_{im} \]

\[ + \frac{1}{n} \sum_{i=1}^{n} \frac{m_2}{m} (f_{i2} - 1) S^2_{IM} \times \left\{ \sum_{i=1}^{n} \left[ D - M \left( \frac{M}{m} \right) s^2_{im} - \sum_{j=1}^{m} \left( \frac{1}{m} - \frac{1}{M} \right) s^2_{jm} \right] \right\}, \]  

(2.6)

where

\[ s^2_{b} = \frac{1}{(n-1)} \left\{ \sum_{i=1}^{n} \bar{y}_{im}^2 + \sum_{i=1}^{n} \frac{1}{m^2} (m_1 \bar{y}_{m_1} + m_2 \bar{y}_{m_2})^2 - n\bar{y}^2 \right\} \]

and \[ s^2_{im} = \frac{1}{(m-1)} (\sum_{j=1}^{m} \bar{y}_{ij}^2 - m\bar{y}^2_{im}), \bar{y}_{im} = \frac{1}{m} \sum_{j=1}^{m} y_{ij} \] .

\textbf{Proof:} By definition, we have

\[ V(\bar{y}_{r}) = \frac{1}{n} \left\{ \sum_{i=1}^{n} \bar{y}_{im} + \sum_{i=1}^{n} \frac{(m_1 \bar{y}_{m_1} + m_2 \bar{y}_{m_2})}{m} \right\} \]

= \frac{1}{n} \left\{ \sum_{i=1}^{n} \bar{y}_{im} + \sum_{i=1}^{n} \frac{(m_1 \bar{y}_{m_1} + m_2 \bar{y}_{m_2})}{m} \right\}

= \frac{1}{n} \left\{ \sum_{i=1}^{n} \bar{y}_{im} + \sum_{i=1}^{n} \frac{(m_1 \bar{y}_{m_1} + m_2 \bar{y}_{m_2})}{m} \right\}

= \frac{1}{n} \left\{ \sum_{i=1}^{n} \bar{y}_{im} + \sum_{i=1}^{n} \frac{(m_1 \bar{y}_{m_1} + m_2 \bar{y}_{m_2})}{m} \right\}

= \frac{1}{n} \left\{ \sum_{i=1}^{n} \bar{y}_{im} + \sum_{i=1}^{n} \frac{(m_1 \bar{y}_{m_1} + m_2 \bar{y}_{m_2})}{m} \right\}

= \frac{1}{n} \left\{ \sum_{i=1}^{n} \bar{y}_{im} + \sum_{i=1}^{n} \frac{(m_1 \bar{y}_{m_1} + m_2 \bar{y}_{m_2})}{m} \right\}

= \frac{1}{n} \left\{ \sum_{i=1}^{n} \bar{y}_{im} + \sum_{i=1}^{n} \frac{(m_1 \bar{y}_{m_1} + m_2 \bar{y}_{m_2})}{m} \right\}

Thus, \[ \bar{y}_{r} \] is an unbiased estimator of the population mean } \bar{Y}. Here, } E_5 \text{ represents conditional expectations of all possible samples of size } h_{i2} \text{ drawn from } m_{i2}, } E_4 \text{ is the conditional expectation of all possible samples of size } m_{i1}, \text{ } E_3 \text{ respectively drawn from } m \text{ by keeping } m_{i1}, \text{ } m_{i2} \text{ fixed, } E_2 \text{ is the conditional expectation arising out of } m \text{ independent Bernoulli trials leading to } m_{i1} \text{ success and } m_{i2} \text{ failures, } m_{i1} + m_{i2} = m, \text{ } E_1 \text{ is the conditional expectation of all possible samples of size } m \text{ drawn from } M \text{ and } E_1 \text{ arises out of selection of all possible samples of size } n \text{ from } N.

To obtain the variance we proceed as follows:

\[ V(\bar{y}_{r}) = V_1 \{E_2E_3E_4(\bar{y}_{r})\} + E_1V_2 \{E_3E_4E_5(\bar{y}_{r})\} + E_1E_2V_3 \{E_4E_5(\bar{y}_{r})\} + E_1E_2E_3V_4 \{E_5(\bar{y}_{r})\} \]

where,

\[ V_1 \{E_2E_3E_4E_5(\bar{y}_{r})\} = \frac{1}{n} \left\{ \sum_{i=1}^{n} \bar{y}_{im} + \sum_{i=1}^{n} \frac{(m_1 \bar{y}_{m_1} + m_2 \bar{y}_{m_2})}{m} \right\} \]

\[ E_1V_2 \{E_3E_4E_5(\bar{y}_{r})\} = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{1}{m} - \frac{1}{M} \right) s^2_{IM} \]

\[ E_1E_2V_3 \{E_4E_5(\bar{y}_{r})\} = 0, \]

\[ E_1E_2E_3V_4 \{E_5(\bar{y}_{r})\} = 0 \]

and \[ E_1E_2E_3E_4 \{E_5(\bar{y}_{r})\} = \frac{1}{n} \sum_{i=1}^{n} \frac{(1-\theta)(f_{i2} - 1)}{m} S^2_{IM}. \]
Here, $V_1, V_2, V_3, V_4$ and $V_5$ are defined similarly as $E_1, E_2, E_3, E_4$ and $E_5$. By adding the above three terms we get the required result. To obtain an unbiased variance estimator, consider,

$$s_b^2 = \frac{1}{(n-1)} \left[ \sum_{i=1}^{n} \bar{y}_{im}^2 + \sum_{i=1}^{n} \frac{1}{m} (m_1 \bar{y}_{m_1} + m_2 \bar{y}_{m_2})^2 - n \bar{y}_r^2 \right]$$

Taking the expectations and simplifying we get,

$$E_1 E_2 E_3 E_4 E_5 (s_b^2) = S_b^2 - \frac{1}{N(n-1)} \sum \left( \frac{1}{m} - \frac{1}{M} \right) S_M^2$$

$$+ \frac{n}{N(n-1)} \left[ \sum_{i=1}^{N} \left( \frac{1}{m} - \frac{1}{M} \right) S_M^2 \right]$$

and also we have

$$E_3 E_4 E_5 (s_{im}^2) = S_{im}^2,$$

$$E_2 E_4 E_5 (s_{im}^2) = \frac{(m+1)M}{m} S_M^2 - \frac{(1-\theta)}{m(m-1)} (f_{12} - 1) S_M^2$$

Here $E_6$ represents conditional expectations of all possible samples of size $h_{i2}$ drawn from $m_{i2}, E_5$ is the conditional expectation of all possible samples of size $m_{i1}, m_{i2}$ respectively drawn from $m$ by keeping $m_{i1}, m_{i2}$ fixed, $E_4$ is the conditional expectation arising out of $m$ independent Bernoulli trials leading to $m_{i1}$ success and $m_{i2}$ failures, $m_{i1} + m_{i2} = m$, $E_3$ is the conditional expectation of all possible samples of size $m$ drawn from $M, E_2$ arising out of selection of all possible samples of size $n_1, n_2$ drawn from $N_1, N_2$ keeping $n_1, n_2$ fixed and $E_1$ is the expectation arises out of randomness of $n_1$ and $n_2, n_1 + n_2 = n, N_1 + N_2 = N$.

Thus,

$$\hat{S}^2_b = s_b^2 + \frac{1}{n(n-1)} \sum \left( \frac{1}{m} - \frac{1}{M} \right) S_M^2$$

$$+ \frac{1}{(n-1)} \left[ \sum_{i=1}^{n} \left( \frac{1}{m} - \frac{1}{M} \right) S_{im}^2 \right].$$

For the psus with nonresponse $\hat{S}_{im}^2 = s_{im}^2$ while in psus with nonresponse problem $\hat{S}_{im}^2 = s_{im}^2/\alpha$. Substituting the estimated values in the variance expression in Eq. (2.5) we obtain the required result.

To determine the optimum values of $n, m$ and $f_{i2}$ we proceed as earlier i.e. minimization of expected cost subject to fixed variance. The relevant cost function in this case is,

$$C = C_1 n_2 + C_2 n_1 m + C_2 \sum_{i=1}^{n_1} m_1 + C_3 \sum_{i=1}^{n_2} h_{i2},$$

where, $C_1, C_2$ and $C_3$ are same as defined earlier. The expected cost is,

$$C'' = E(C)$$

$$= \frac{n}{N} \left[ C_1 N_2 + C_2 N_1 m + C_2 \sum_{i=1}^{n_1} m_1 + C_3 \sum_{i=1}^{n_2} (1-\theta)m \right]$$

To minimize the expected cost consider the function $\phi = E(C) + \lambda \{ V(\bar{y}_r^2) V_0 \}$, where $\lambda$ is the Lagrangian multiplier. To obtain closed form expressions for the various sample sizes we have considered $m_{i2} = h_{i2}f_{i2}$ in place of $m_{i2} = h_{i2}f_{i2}$, $i = 1, 2, ..., n_2$. The optimum values obtained through minimization are as follows:

$$n_{opt} = \frac{k_1}{(V_0 + S_b^2/N)}, m_{opt} = \sqrt{B_2/A_2} \quad \text{and} \quad f_{2opt} = \frac{B_1}{\sqrt{A_1}}.$$
\[ B_1 = C_3 \sum_{i=1}^{N_3} (1-\theta_i) \left[ \sum_{i=1}^{N_3} S_{iM}^2 - \sum_{i=1}^{N_3} (1-\theta_i) S_{iM}^2 \right], \]
\[ A_2 = C_2 N_1 + C_2 \sum_{i=1}^{N_3} \theta_i + C_3 \sum_{i=1}^{N_3} (1-\theta_i) f_2 \left[ N S_{iM}^2 - \frac{1}{M} \sum_{i=1}^{N} S_{iM}^2 \right], \]
\[ B_2 = C_1 N_2 \left[ \sum_{i=1}^{N} S_{iM}^2 + \sum_{i=1}^{N_2} (1-\theta_i) (f_2 - 1) S_{iM}^2 \right], \]
and \[ V_0 = 0.0025 \times \varphi^2. \]

**Case 3.** Let a sample of \( n \) psus is drawn from \( N \), within each selected psu a sample of \( m \) ssus is drawn by srswr. Let there be no nonresponse in \( n_1 \) psus. In the \( n_2 \) psus \( m_1 \) ssus respond while \( m_2 \) ssus do not respond. A subsample of \( h_2 \) units is selected by srswr from \( m_2 \) and data are collected through specialized efforts, let there be complete nonresponse in the \( n_3 \) psus. Further a subsample of \( h_3 \) psus is drawn out of \( n_3 \) psus and data are collected through specialized efforts on each of \( m \) ssus in the selected \( h_3 \) psus. Here \( n_3 = f_2 h_3 \) and \( m_2 = h_2 f_{22} \), \( i = 1, 2, \ldots, n_2 \). In this context we state the following theorem.

**Theorem 2.2.** The unbiased estimator of \( \bar{y} \) is
\[ \bar{y}_r = \frac{1}{n} \sum_{i=1}^{n} \bar{y}_{im} + \frac{1}{n} \left( m_1 \bar{y}_{m1} + m_2 \bar{y}_{h2} \right) + \frac{n_3}{h_3} \sum_{i=1}^{n_3} \bar{y}_{im} \]
with variance
\[ V(\bar{y}_r) = \left( \frac{1}{n} - \frac{1}{N} \right) S_b^2 + \left( \frac{1}{n} \right) \left( \frac{1}{m} - \frac{1}{M} \right) \left[ \sum_{i=1}^{N} S_{iM}^2 + \sum_{i=1}^{N_2} (1-\theta_i) S_{iM}^2 \right] + \frac{f_3}{n} \sum_{i=1}^{N_3} S_{iM}^2 \]
\[ + \frac{1}{N_3} \sum_{i=1}^{N_3} (1-\theta_i) (f_2 - 1) S_{iM}^2 + \frac{N_3}{N} (f_3 - 1) S_{N3}^2, \]
where \( S_{N3}^2 = \frac{1}{N_3-1} \sum_{i=1}^{N_3} (\bar{y}_{im} - \bar{y}_{N3})^2 \), where,
\[ \bar{y}_{N3} = \frac{1}{N_3} \sum_{i=1}^{N_3} \bar{y}_{im}, \]
An unbiased estimator of variance is,
\[ \hat{V}(\bar{y}_r) = \left( \frac{1}{n} - \frac{1}{N} \right) S_b^2 + \left( \frac{1}{n} \right) \left( \frac{1}{m} - \frac{1}{M} \right) \sum_{i=1}^{n} S_{im}^2 + \frac{1}{N} \sum_{i=1}^{N} m_2 (f_2 - 1) \frac{s_{im}^2}{\alpha} + \frac{(N-1)}{n} \left( \frac{1}{m} - \frac{1}{M} \right) \sum_{i=1}^{n} S_{im}^2 \]
\[ + \frac{(N-n)}{N(n-1)} \sum_{i=1}^{n} D - \frac{M}{m} \sum_{i=1}^{n} S_{im}^2 + n_3 (f_3 - 1) \frac{(N-1)}{n(n-1)} \times \left[ S_{bh0} - \left( \frac{1}{m} - \frac{1}{M} \right) \sum_{i=1}^{n} S_{im}^2 \right] + \frac{1}{n} \left( f_3 - (N-n)(n-f_3) \right) \]
\[ \times \sum_{i=1}^{n_3} \left( \frac{1}{m} - \frac{1}{M} \right) S_{im}^2, \]
Hence, \( \bar{Y}_r \) is an unbiased estimator of the population mean \( \bar{Y} \). Where, \( E_6 \) represents conditional expectations of all possible samples of size \( h_2 \) drawn from, \( E_2 \) is the conditional expectation of all possible samples of size \( m_1, m_2 \) respectively drawn from \( m \) by keeping \( m_1, m_2 \) fixed, \( E_4 \) is the conditional expectation arising out of \( m \) independent Bernoulli trials leading to \( m_1 \) success and \( m_2 \) failures, \( m_1 + m_2 = m \), \( E_4 \) is the conditional expectation of all possible samples of size \( m \) drawn from \( M \), \( E_6 \) arises out of selection of all possible samples of size \( n \) from \( N \).

To obtain the variance we proceed as follows:

\[
V(\bar{Y}_r) = V\{E_2\bar{E}_2E_4\bar{E}_4E_6(\bar{Y}_r)\} + E_1V_1(\bar{E}_2\bar{E}_4\bar{E}_5E_6(\bar{Y}_r)) + E_1E_2V_2(\bar{E}_2\bar{E}_4E_5E_6(\bar{Y}_r)) + E_1E_2E_3V_3(\bar{E}_3E_4\bar{E}_5E_6(\bar{Y}_r)) + E_1E_2E_3E_4V_4(\bar{E}_3E_4\bar{E}_5(\bar{Y}_r)) + E_1E_2E_3E_4E_5V_5(\bar{E}_6(\bar{Y}_r)).
\]

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

Hence, we have,

\[
V_1(\bar{E}_2\bar{E}_4\bar{E}_5E_6(\bar{Y}_r)) = \frac{1}{n} - \frac{1}{N}S^2_b,
\]

\[
E_1V_1(\bar{E}_2\bar{E}_4\bar{E}_5E_6(\bar{Y}_r)) = \frac{N_3}{Nn}(f_3 - 1)S^2_{bn},
\]

\[
E_1E_2V_2(\bar{E}_2\bar{E}_4E_5E_6(\bar{Y}_r)) = \frac{1}{Nn}\left(1 - \frac{1}{m}\right)\left(\frac{1}{N}S^2_{bM} + \frac{N_2}{N}S^2_{bM} + \frac{N_3}{N}S^2_{bN}\right),
\]

\[
E_1E_2E_3V_3(\bar{E}_3E_4\bar{E}_5E_6(\bar{Y}_r)) = 0,
\]

\[
E_1E_2E_3E_4V_4(\bar{E}_3E_4\bar{E}_5(\bar{Y}_r)) = 0,
\]

\[
E_1E_2E_3E_4E_5V_5(\bar{E}_6(\bar{Y}_r)) = \frac{1}{Nn}\sum_{i=1}^{N_3}(1 - \theta_i)m(f_1 - 1)S^2_{bM}.
\]

Thus, by adding all the terms we obtain the required variance of the estimator i.e.

\[
V(\bar{Y}_r) = \left(\frac{1}{n} - \frac{1}{N}\right)S^2_b + \frac{1}{Nn}\left(1 - \frac{1}{m}\right)\left(\frac{1}{N}S^2_{bM} + \frac{N_2}{N}S^2_{bM} + \frac{N_3}{N}S^2_{bN}\right) + \frac{N_3}{Nn}\sum_{i=1}^{N_3}(1 - \theta_i)m(f_1 - 1)S^2_{bM}.
\]

To obtain the unbiased estimator of variance, consider,

\[
s^2_b = \frac{1}{(n - 1)}\left[\sum_{i=1}^{n} \bar{y}_{im}^2 + \sum_{i=1}^{n} \frac{1}{m}(m_1\bar{y}_{m_1} + m_2\bar{y}_{m_2})^2 + \frac{n_3}{h_3}\sum_{i=1}^{n_3} \bar{y}_{m_3}^2 - n\bar{y}_r^2\right],
\]

where,

\[
E_1E_2E_3E_4E_5E_7(s^2_b) = \frac{S^2_b + \frac{1}{Nn}\sum_{i=1}^{N_3}(1 - \theta_i)m(f_1 - 1)S^2_{bM} + \frac{n - f_3}{N(n - 1)}\frac{1}{Nn}\sum_{i=1}^{N_3}(1 - \theta_i)(f_1 - 1)S^2_{bM}}{N(n - 1)}.
\]

Here, \( E_7 \) represents conditional expectations of all possible samples of size \( h_2 \) drawn from \( m_2 \), \( E_6 \) is the conditional expectation of all possible samples of size \( m_1, m_2 \) respectively drawn from \( m \) by keeping \( m_1, m_2 \) fixed, \( E_5 \) is the conditional expectation arising out of \( m \) independent Bernoulli trials leading to \( m_1 \) success and \( m_2 \) failures, \( m_1 + m_2 = m \), \( E_4 \) is the conditional expectation of all possible samples of size \( m \) drawn from \( M \), \( E_6 \) arises out of selection of all possible samples of size \( n \) from \( N \) and \( E_7 \) refers to the expectation arises out of selection of all possible samples of size \( n_1, n_2 \) and \( n_3 \) drawn from \( N_1, N_2 \) and \( N_3 \) keeping \( n_1, n_2 \) and \( n_3 \) fixed.
randomness of \( n_1, n_2 \) and \( n_3 \), where \( n_1 + n_2 + n_3 = n \) and \( N_1 + N_2 + N_3 = N \).

Thus,

\[
\hat{S}_{ib}^2 = s_{ib}^2 - \frac{1}{n}
\left( \frac{1}{m} - \frac{1}{M} \right)
\sum_{i=1}^{n_1} \sum_{j=1}^{m_1} \frac{m_{ij}^2}{m^2} (f_{ij} - 1) \frac{s_{im}^2}{\alpha}
\]

\[
+ \frac{1}{n(n-1)} \sum_{i=1}^{n_1} \left( D - \frac{M}{m} \right) s_{im}^2
\frac{1}{M \alpha} + \frac{1}{n(n-1)} \sum_{i=1}^{n_2} \left( \frac{1}{m} - \frac{1}{M} \right) s_{im}^2
\]

\[
+ \frac{n_3}{n(n-1)} (f_3 - 1) \left( s_{ih3}^2 - \frac{1}{h_3} \frac{1}{(1 - \frac{1}{M})} \sum_{i=1}^{h_3} s_{im}^2 \right)
\]

\[
- \frac{n - f_3}{n(n-1)} \sum_{i=1}^{n_3} \left( \frac{1}{m} - \frac{1}{M} \right) s_{im}^2.
\]

Here again, for the psus with no nonresponse \( \hat{S}_{ib}^2 = s_{im}^2 \) while in psus when there is nonresponse problem \( \hat{S}_{ib}^2 = s_{im}^2 \) and \( \hat{S}_{ibh_3}^2 = s_{ih3}^2 - \frac{1}{h_3} \frac{1}{(1 - \frac{1}{M})} \sum_{i=1}^{h_3} s_{im}^2 \).

Substituting the estimated values in the Eq. (2.8) we get the required expression in Eq. (2.9).

To determine the optimum values of \( n, m, f_1, \) and \( f_3 \) we proceed as follows.

The cost function in this case is given as,

\[
C = C_1 nm + C_2 n_1 m + C_2 \sum_{i=1}^{n_1} m_1 + C_3 \left( \sum_{i=1}^{h_2} n_2 + n_3 h_3 \right)
\]

where, \( C, C_1, C_2 \) and \( C_3 \) are same as defined earlier.

The expected cost is,

\[
C'' = E(C)
= n \left[ C_1 m + C_2 \frac{N_1}{N} m + C_2 \sum_{i=1}^{n_1} m_1 + C_3 \frac{N_1}{N} \frac{(1 - \theta)m}{f_2} + C_3 \frac{N_3}{N} f_3 \right].
\]

To minimize the expected cost consider the function, \( \phi = E(C) + \lambda \{ V(Y^2) - V_0 \} \). To obtain closed form expressions for the various sample sizes we have considered \( m_1 = h_2 f_2 \) in place of \( m_1 = h_2 f_2, i = 1, 2, \ldots, n \), and also to overcome the problem arising due to simultaneous minimisation of \( n, m, f_2, f_3 \) we assume that \( n_3 = f_2 h_3 \). Thus, minimization gives the optimum values as

\[
n_{opt} = \frac{k_2}{(V_0 + S_{ib}^2)}, \quad m_{opt} = \pm \frac{G_2}{\sqrt{D_2}} \text{ and } f_{2opt} = -B_2 \pm \sqrt{B_2 - 4A_2G_3}
\]

where,

\[
k_2 = S_{ib}^2 + \frac{N_1}{N} (f_2 - 1) S_{ih3}^2 + \frac{1}{N} \left( \frac{1}{m} - \frac{1}{M} \right) \times \left[ \sum_{i=1}^{n_1} S_{im}^2 + \sum_{i=1}^{n_2} S_{im}^2 + f_3 \sum_{i=1}^{n_3} S_{im}^2 \right] + \frac{1}{N} \sum_{i=1}^{n_3} \frac{(1 - \theta_i)}{m} (f_2 - 1) S_{im}^2,
\]

\[
A_2 = N(C_1 + C_2 \frac{N_1}{N} + C_2 \sum_{i=1}^{n_1} \theta_i) \sum_{i=1}^{n_3} (1 - \theta_i) S_{im}^2,
\]

\[
B_2 = -C_3 \sum_{i=1}^{n_3} (1 - \theta_i) S_{im}^2,
\]

\[
G_3 = -C_3 \sum_{i=1}^{n_3} (1 - \theta_i) \left( \sum_{i=1}^{n_1} S_{im}^2 + \sum_{i=1}^{n_2} S_{im}^2 - \sum_{i=1}^{n_3} (1 - \theta_i) S_{im}^2 \right),
\]

\[
G_2 = \sum_{i=1}^{n_3} (1 - \theta_i) \left( N_3 S_{ih3}^2 + \frac{1 - \theta_i}{m} S_{im}^2 \right),
\]

\[
D_2 = N_3 n_3 \sum_{i=1}^{n_3} (1 - \theta_i) S_{im}^2 \text{ and } V_0 = 0.0025 \times \bar{Y}^2.
\]

**Control Case.** The following estimator was also considered for efficiency comparison purpose. Here we assume that a srswor sample of \( n \) psus is selected from \( N \) and within each selected psu a srswor sample of \( m \) ssus are selected. Data are collected through specialised efforts i.e. there is no nonresponse. Then we give the following Theorem.
Theorem 2.4. The estimator
\[ \bar{Y} = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} y_{ij} = \frac{1}{n} \sum_{i=1}^{n} \bar{Y}_{im} \]  
(2.10)
is unbiased for \( \bar{Y} \), with variance
\[ V(\bar{Y}) = \left( \frac{1}{n} - \frac{1}{N} \right) S_b^2 + \frac{1}{Nn} \sum_{i=1}^{N} \left( \frac{1}{m} - \frac{1}{M} \right) S_{im}^2. \]  
(2.11)
An unbiased estimator of variance is,
\[ \hat{V}(\bar{Y}) = \left( \frac{1}{n} - \frac{1}{N} \right) S_b^2 + \frac{1}{Nn} \sum_{i=1}^{N} \left( \frac{1}{m} - \frac{1}{M} \right) S_{im}^2, \]  
(2.12)
where,
\[ s_b^2 = \frac{1}{(n-1)} \sum_{i=1}^{n} (\bar{Y}_{im} - \bar{Y})^2, \bar{Y}_{im} = \frac{1}{m} \sum_{j=1}^{m} y_{ij}, \text{ and} \]
\[ s_{im}^2 = \frac{1}{(m-1)} \sum_{j=1}^{m} (y_{ij} - m\bar{Y}_{im})^2. \]

Proof: By definition,
\[ E(\bar{Y}) = \frac{1}{n} E_1 \left[ E_2 \left( \frac{1}{n} \sum_{i=1}^{n} \bar{Y}_{im} \right) \right] = E_1 \left( \frac{1}{n} \sum_{i=1}^{n} \bar{Y}_{im} \right), \]
\[ = \frac{1}{N} \sum_{i=1}^{N} \bar{Y}_{im} = \bar{Y}. \]

Thus, \( \bar{Y} \) is an unbiased estimator of the population mean \( \bar{Y} \). Here, \( E_2 \) denotes the conditional expectation pertaining to all possible samples of size \( m \) drawn from \( M \) and \( E_1 \) is the conditional expectation pertaining to all possible samples of size \( n \) drawn from \( N \).

Again we see that, \( V(\bar{Y}) = V_1 E_2 (\bar{Y}) + E_1 V_2 (\bar{Y}), \) where
\[ V_1 E_2 (\bar{Y}) = \left( \frac{1}{n} - \frac{1}{N} \right) S_b^2 \]  
and
\[ E_1 V_2 (\bar{Y}) = \frac{1}{Nn} \sum_{i=1}^{N} \left( \frac{1}{m} - \frac{1}{M} \right) S_{im}^2. \]
To obtain unbiased estimator of variance, we have,
\[ E(s_b^2) = E_1 E_2 (s_b^2) = S_b^2 + \frac{1}{Nn} \sum_{i=1}^{N} \left( \frac{1}{m} - \frac{1}{M} \right) S_{im}^2 \]  
and
\[ E(s_{im}^2) = S_{im}^2. \]

Now substituting \( \hat{s}_b^2 = s_b^2 - \frac{1}{n} \sum_{i=1}^{n} (\frac{1}{m} - \frac{1}{M}) s_{im}^2 \) and \( \hat{s}_{im}^2 = \bar{s}_{im}^2 \) in the variance expression we obtain the required result.

The cost function in this case is, \( C = C_1 n + C_p m \), where, \( C, C_1 \) and \( C_3 \) have been defined earlier. To obtain optimum values of \( n \) and \( m \) we minimize the cost by fixing the variance. The optimum values are as follows,
\[ n_{opt} = \frac{S_b^2 + \frac{1}{n} \sum_{i=1}^{N} \left( \frac{1}{m} - \frac{1}{M} \right) S_{im}^2}{(V_0 + \frac{S_b^2}{N})}, \]  
and
\[ m_{opt} = \sqrt{\frac{C_1 S_b^2 + \frac{1}{MN} \sum_{i=1}^{N} S_{im}^2}{C_3}}, \]

3. EMPIRICAL ILLUSTRATION

For the purpose of empirical illustration we consider the MU284 data given in Sarjal et al. (1992). Using this data a population with \( N=27 \) psus and \( M=10 \) ssus was generated by combining the adjacent 10 units and allocating them to the respective psus. In our analysis we considered two target variables from this MU284 data. These variables are denoted by P85 and P75, and described as the human population (in thousand) of 270 municipalities of Sweden in 1985 and 1975 respectively. Here, we used four different values of \( \theta \) for each psus. Further, various combinations of cost components \( C_1, C_2 \) and \( C_3 \) were considered. We computed the percentage reduction in expected cost (%RIEC) as well as optimum values of sample sizes of different estimators described in previous with respect to controlled estimator \( \bar{Y} \). The values of \%RIEC, optimum sample sizes and various combinations of cost components \( C_1, C_2 \) and \( C_3 \) are reported in Table 3.1 and Table 3.2. In particular, Table 3.1 and Table 3.2 present the values for P85 and P75 respectively. Note that the percentage reduction in expected cost for case \( i(i=1, 2, 3) \) is computed as
\[ \%RIEC = \frac{(C - C^{(i)})}{C} \times 100 \]  
where \( C \) is the total cost for Control Case and \( C^{(i)} \) \( (i = 1, 2, 3) \) is the expected cost.
for the case i. That is, $C^{(1)} = C^*$, $C^{(2)} = C^*$ and $C^{(3)} = C^*$ are the expected cost for case 1, 2 and 3 respectively. The empirical analysis reported in the paper was done using SAS 9.3 software.

The results for variable P85 reported in Table 3.1 reveal that the %RIEC is maximum for the second estimator followed by the third estimator and least in the first estimator. The %RIEC increases with increase in travel and miscellaneous cost ($C_1$) for the first and third estimator and decreases in case of the second estimator. The %RIEC decreases with increase in data collection cost at first attempt ($C_2$) for all the estimators. The %RIEC increases for all the estimators with the increase in cost per unit of collecting the information by expensive method after the first attempt to obtain information failed ($C_3$). It is also seen that, for given $C_1$ and $C_2$ as $C_3$ increases the rate of increase of %RIEC is maximum in the second estimator followed by the third estimator and least in the first estimator.

Turning now to Table 3.2 for the results of variable P75. In Table 3.2 again similar to variable P85 we observe that the %RIEC is maximum for the second estimator followed by the third estimator and least in the first estimator. However, in contrast the %RIEC

Table 3.1. The optimum values of sample sizes along with percentage reduction in expected cost (%RIEC) of $\bar{y}_r$, $\bar{y}_r$, $\bar{y}_r$ over $\bar{y}_r$ for the variable P85.

<table>
<thead>
<tr>
<th>Cost</th>
<th>Control ($\bar{y}_r$)</th>
<th>First estimator ($\bar{y}_r$)</th>
<th>Second estimator ($\bar{y}_r$)</th>
<th>Third estimator ($\bar{y}_r$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_1$</td>
<td>$C_2$</td>
<td>$C_3$</td>
<td>$n$</td>
<td>$m$</td>
</tr>
<tr>
<td>25</td>
<td>2</td>
<td>45</td>
<td>14</td>
<td>3</td>
</tr>
<tr>
<td>25</td>
<td>2</td>
<td>50</td>
<td>14</td>
<td>3</td>
</tr>
<tr>
<td>25</td>
<td>2</td>
<td>55</td>
<td>14</td>
<td>3</td>
</tr>
<tr>
<td>25</td>
<td>4</td>
<td>45</td>
<td>14</td>
<td>3</td>
</tr>
<tr>
<td>25</td>
<td>4</td>
<td>50</td>
<td>14</td>
<td>3</td>
</tr>
<tr>
<td>25</td>
<td>4</td>
<td>55</td>
<td>14</td>
<td>3</td>
</tr>
<tr>
<td>30</td>
<td>2</td>
<td>45</td>
<td>14</td>
<td>4</td>
</tr>
<tr>
<td>30</td>
<td>2</td>
<td>50</td>
<td>14</td>
<td>3</td>
</tr>
<tr>
<td>30</td>
<td>2</td>
<td>55</td>
<td>14</td>
<td>3</td>
</tr>
<tr>
<td>30</td>
<td>4</td>
<td>45</td>
<td>14</td>
<td>4</td>
</tr>
<tr>
<td>30</td>
<td>4</td>
<td>50</td>
<td>14</td>
<td>3</td>
</tr>
<tr>
<td>30</td>
<td>4</td>
<td>55</td>
<td>14</td>
<td>3</td>
</tr>
<tr>
<td>35</td>
<td>2</td>
<td>45</td>
<td>14</td>
<td>4</td>
</tr>
<tr>
<td>35</td>
<td>2</td>
<td>50</td>
<td>14</td>
<td>4</td>
</tr>
<tr>
<td>35</td>
<td>2</td>
<td>55</td>
<td>14</td>
<td>3</td>
</tr>
<tr>
<td>35</td>
<td>4</td>
<td>45</td>
<td>14</td>
<td>4</td>
</tr>
<tr>
<td>35</td>
<td>4</td>
<td>50</td>
<td>14</td>
<td>4</td>
</tr>
<tr>
<td>35</td>
<td>4</td>
<td>55</td>
<td>14</td>
<td>3</td>
</tr>
</tbody>
</table>
Table 3.2. The optimum values of sample sizes along with percentage reduction in expected cost (%RIEC) of $\bar{y}^r, \bar{y}^f, \bar{y}^r$ over $\bar{y}^r$ for the variable P75.

<table>
<thead>
<tr>
<th>Cost $C_1$</th>
<th>Cost $C_2$</th>
<th>Cost $C_3$</th>
<th>Control $(\bar{y}^r)$ $n, m$</th>
<th>First estimator $(\bar{y}^f)$ $n, m, f_2$</th>
<th>%RIEC</th>
<th>Second estimator $(\bar{y}^r)$ $n, m, f_2$</th>
<th>%RIEC</th>
<th>Third estimator $(\bar{y}^r)$ $n, m, f_2$</th>
<th>%RIEC</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>2</td>
<td>45</td>
<td>15 8</td>
<td>23 13</td>
<td>2.12</td>
<td>45.10</td>
<td>23 8</td>
<td>2.78</td>
<td>66.92</td>
</tr>
<tr>
<td>25</td>
<td>2</td>
<td>50</td>
<td>15 8</td>
<td>23 13</td>
<td>2.23</td>
<td>45.32</td>
<td>23 8</td>
<td>2.93</td>
<td>67.41</td>
</tr>
<tr>
<td>25</td>
<td>2</td>
<td>55</td>
<td>15 8</td>
<td>24 13</td>
<td>2.34</td>
<td>45.52</td>
<td>24 8</td>
<td>3.08</td>
<td>67.84</td>
</tr>
<tr>
<td>25</td>
<td>4</td>
<td>45</td>
<td>15 8</td>
<td>21 13</td>
<td>1.84</td>
<td>34.09</td>
<td>21 7</td>
<td>1.97</td>
<td>63.42</td>
</tr>
<tr>
<td>25</td>
<td>4</td>
<td>50</td>
<td>15 8</td>
<td>21 13</td>
<td>1.94</td>
<td>34.48</td>
<td>21 7</td>
<td>2.07</td>
<td>64.08</td>
</tr>
<tr>
<td>25</td>
<td>4</td>
<td>55</td>
<td>15 8</td>
<td>22 13</td>
<td>2.04</td>
<td>34.82</td>
<td>22 7</td>
<td>2.18</td>
<td>64.65</td>
</tr>
<tr>
<td>30</td>
<td>2</td>
<td>45</td>
<td>15 9</td>
<td>22 13</td>
<td>1.99</td>
<td>46.71</td>
<td>22 9</td>
<td>2.78</td>
<td>66.81</td>
</tr>
<tr>
<td>30</td>
<td>2</td>
<td>50</td>
<td>15 9</td>
<td>22 13</td>
<td>2.10</td>
<td>46.92</td>
<td>22 9</td>
<td>2.93</td>
<td>67.30</td>
</tr>
<tr>
<td>30</td>
<td>2</td>
<td>55</td>
<td>15 8</td>
<td>23 13</td>
<td>2.20</td>
<td>47.11</td>
<td>23 9</td>
<td>3.08</td>
<td>67.74</td>
</tr>
<tr>
<td>30</td>
<td>4</td>
<td>45</td>
<td>15 9</td>
<td>20 13</td>
<td>1.75</td>
<td>37.07</td>
<td>20 8</td>
<td>1.97</td>
<td>63.34</td>
</tr>
<tr>
<td>30</td>
<td>4</td>
<td>50</td>
<td>15 9</td>
<td>21 13</td>
<td>1.85</td>
<td>37.43</td>
<td>21 8</td>
<td>2.07</td>
<td>64.00</td>
</tr>
<tr>
<td>30</td>
<td>4</td>
<td>55</td>
<td>15 8</td>
<td>21 13</td>
<td>1.94</td>
<td>37.75</td>
<td>21 8</td>
<td>2.18</td>
<td>64.57</td>
</tr>
<tr>
<td>35</td>
<td>2</td>
<td>45</td>
<td>15 10</td>
<td>21 13</td>
<td>1.88</td>
<td>47.89</td>
<td>21 9</td>
<td>2.78</td>
<td>66.71</td>
</tr>
<tr>
<td>35</td>
<td>2</td>
<td>50</td>
<td>15 9</td>
<td>22 13</td>
<td>1.98</td>
<td>48.09</td>
<td>22 9</td>
<td>2.93</td>
<td>67.21</td>
</tr>
<tr>
<td>35</td>
<td>2</td>
<td>55</td>
<td>15 9</td>
<td>22 13</td>
<td>2.08</td>
<td>48.27</td>
<td>22 9</td>
<td>3.08</td>
<td>67.64</td>
</tr>
<tr>
<td>35</td>
<td>4</td>
<td>45</td>
<td>15 10</td>
<td>20 13</td>
<td>1.68</td>
<td>39.28</td>
<td>20 7</td>
<td>1.97</td>
<td>63.26</td>
</tr>
<tr>
<td>35</td>
<td>4</td>
<td>50</td>
<td>15 9</td>
<td>20 13</td>
<td>1.77</td>
<td>39.63</td>
<td>20 7</td>
<td>2.07</td>
<td>63.92</td>
</tr>
<tr>
<td>35</td>
<td>4</td>
<td>55</td>
<td>15 9</td>
<td>21 13</td>
<td>1.86</td>
<td>39.92</td>
<td>21 7</td>
<td>2.18</td>
<td>64.49</td>
</tr>
</tbody>
</table>

increases with increase in travel and miscellaneous cost $(C_1)$ for the first estimator and decreases in case of the second and third estimator. Moreover, like for P85, the %RIEC decreases with increase in data collection cost at first attempt $(C_2)$ for the all the estimators in case of P75 too. We also noticed a identical pattern between P85 and P75 with respect to $C_3$. Overall, results for two variable are almost identical.

4. DISCUSSION AND CONCLUSION

The %RIEC is maximum for the second estimator because there is partial nonresponse in the second stage for the first estimator for whole sample size $n$, whereas for the second estimator, there is partial nonresponse in the second stage only for a part of the sample size (i.e. $n_2$ psus) and there is complete response in the other part (i.e. $n_1$ psus) where as for the third estimator, there is full response in $n_1$ psus, partial nonresponse in the second stage for $n_2$ psus and complete nonresponse at the first stage for $n_3$ psus ($n_1 + n_2 + n_3 = n$). Thus the first estimator is more costly than other two estimators and the third estimator is more costly than the second estimator.

To summarize, all the three estimators, of population mean in the presence of nonresponse based on subsampling of the nonrespondents, have better % RIEC as compared to the estimator based only on interview method of data collection resulting in 100%
response. Among all the three estimators, the second estimator has the maximum %RIEC followed by the third estimator and %RIEC is least in the first estimator. Hence, the second estimator was found best among all the three estimators in respect of the criterion of %RIEC.

ACKNOWLEDGEMENTS
Authors are grateful to the Associate Editor and referees for their valuable comments.

REFERENCES