Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF
AGRICULTURAL STATISTICS 64(3) 2010 417-432

Use of Imputation Methods in Two-Occasion Successive Sampling

G.N. Singh[1*] and Priyanka[2]
[1]*Indian School of Mines, Dhanbad*
[2]*Shivaji College (University of Delhi), New Delhi*

Received 18 September 2008; Revised 16 June 2010; Accepted 13 July 2010

SUMMARY

The present work intends to emphasize the role of imputation methods to deal with non-response when it occurs on both the occasions in two-occasion successive (rotation) sampling. Under imputation two difference-type estimators have been proposed for estimating the population mean on current occasion. Estimators for the current occasion are derived as a particular case when non-response occurs either on the first occasion or on the second occasion. Behaviors of the proposed estimators are examined and the optimum replacement policies are discussed. Effectiveness of the suggested imputation methods are discussed in two different situations: with and without non-response. Results are analyzed with the help of empirical studies.

Keywords: Non-response, Imputation, Successive Sampling, Chain-type, Optimum replacement policy.

1. INTRODUCTION

In many social surveys, the same population is sampled repeatedly and the same study variable is measured on each occasion, so that development over time can be observed. For example, labor force surveys are conducted monthly to estimate the numbers of people in employment, data on prices of goods are collected monthly or weekly to determine the consumer price index, political opinion surveys are conducted at regular intervals to know the voter preferences, etc. In such cases, successive (rotation) sampling may be a wise choice to generate reliable estimates on different occasions. Theory of successive (rotation) sampling appears to have started with the work of Jessen (1942). He pioneered in using the information collected in previous investigations (surveys) to produce the reliable estimates on the current occasion. This theory was further extended by Patterson (1950), Eckler (1955), Rao and Graham (1964), Gupta (1979), Das (1982), among others. Sen (1971) developed estimators for the population mean on the current occasion using information on two auxiliary variables available on previous occasion. Sen (1972, 1973) extended his work for more than two auxiliary variables. In addition to the information from previous occasion, Singh *et al.* (1991) and Singh and Singh (2001) used information on an additional auxiliary variable available only on current occasion for estimating the current population mean in two-occasion successive sampling. Singh (2003a) extended his work for h-occasions successive sampling. In many practical situations, information on an auxiliary variable may be readily available on the first as well as on the second occasion, for example numbers of academic institutions with their intake capacities are well known in an educational survey, numbers of polluting industries are known in sample surveys of environment. Utilizing the auxiliary information available for both the occasions, Feng and Zou (1997), Biradar and Singh (2001), Singh (2005), Singh and Priyanka (2006, 2007a and 2008a) have proposed several chain-type ratio, difference and regression estimators for estimating the population mean on current (second) occasion in two-occasion successive (rotation) sampling.

*Corresponding author : G.N. Singh
E-mail address : gnsingh_ism@yahoo.com

In practice non-response is one of the major problems encountered by survey statisticians. Longitudinal surveys are more prone to this problem than single-occasion surveys. For example, for an agricultural production survey it might be possible that crop on certain plots is destroyed due to some natural calamities or disease so that yield on these plots is impossible to be measured. Such non-response (incompleteness) can have different patterns and causes. Statisticians have recognized for some time that if the suitable information about the nature of non-response in the population is unknown, the inference concerning population parameters can be spoiled. Many methods are used to reduce the negative impact of non-response in sample surveys. Imputation is one which deals with the filling up method of incomplete data for adapting the standard analytic model in statistics. It is typically used when needed to substitute missing item values with certain fabricated values in the sample surveys. To deal with missing values effectively, Sande (1979) and Kalton *et al.* (1981) suggested imputation methods that make incomplete data sets structurally complete and its analysis simple. Further, Kalton *et al.* (1982) and Singh and Singh (1991) suggested useful imputation methods for surveys in which one uses an estimation procedure based on complete data set and discards data for all those units for which information is not available for at least one time stage. Imputation may also be carried out with the aid of an auxiliary variable, if such is available. For example, Lee *et al.* (1994, 1995) used the information on an available auxiliary variable for imputation purpose. Later Singh and Horn (2000) suggested a compromised method of imputation. Ahmed *et al.* (2006) and Singh (2009) have suggested several new imputation based methods utilizing the available auxiliary information to deal with the problems of non-response in sample surveys. Okafor (2001), Choudhary *et al.* (2004) and Singh and Priyanka (2007b) also discussed the treatment of non-response in successive sampling.

Motivated with the above works, Singh *et al.* (2008b) have suggested impressive imputation methods to deal with the problems of non-response on the current occasion in two-occasion successive (rotation) sampling. There may be situations where the non-response may occur on both the occasions, following this argument; the objective of the present work is to study the effect of non-response on both the occasions in two-occasion successive (rotation) sampling. Imputation methods are suggested which cope with the problems of non-response on both the occasions in two-occasion successive (rotation) sampling. Estimators for the current occasion are derived as a particular case when non-response occurs either on the first occasion or on the second occasion. Under the suggested imputation methods two estimators T and Δ are proposed for estimating the population mean of study character on current (second) occasion in two-occasion successive (rotation) sampling. The estimator T does not use any additional auxiliary information whereas the estimator Δ uses information on an auxiliary variable available on both the occasions. The performances of the proposed estimators have been examined under the assumption that non-response occurs either on previous occasion or on current occasion or on both the occasions. Suitable recommendations are made.

2. NOTATIONS AND SAMPLE STRUCTURES ON TWO-OCCASION

Let $U = (U_1, U_2, ---, U_N)$ be the N-element finite population, which has been sampled over two occasions. The character under study is denoted by x (y) on the first (second) occasion, respectively. It is assumed that information on an auxiliary variable z, with the known population mean is available on both the occasions. We assume that there is non-response on both the occasions. A simple random sample (without replacement) s_n of size n is selected on the first occasion. Let the number of responding units out of sampled n units, which are drawn on the first occasion, be denoted by r_1, the set of responding units in s_n by R_1, and that of non-responding units by R_1^c. A random sub-sample s_m of size $m = n\lambda$ is retained (matched) for its use on the second occasion from the units, which responded on the first occasion and it is assumed that these matched units are completely responding on the current (second) occasion as well. A fresh simple random sample (without replacement) s_u of size $u = (n - m) = n\mu$ is drawn on the second occasion from the non-sampled units of the population so that the sample size on the second occasion remains n. Let the number of responding units out of sampled u units, which are drawn afresh on the current occasion, be denoted by r_2, the set of responding units in s_u by R_2, and that of non-responding units by R_2^c. λ and μ ($\lambda + \mu = 1$) are the fractions of the matched and fresh sample, respectively, on the current (second) occasion. For every unit $i \in R_j$

(j = 1, 2) the value x_i (y_i) is observed, but for the units $i \in R_j^c$ (j = 1, 2) the x_i (y_i) values are missing and instead imputed values are derived.

The following notations are used hereafter:

$\bar{X}, \bar{Y}, \bar{Z}$: The population means of the variables x, y and z respectively.

$\bar{x}_n, \bar{z}_n, \bar{y}_m, \bar{x}_m, \bar{z}_m, \bar{y}_u, \bar{z}_u$: The sample means of the respective variables based on the sample sizes shown in suffices.

$\bar{x}_{r_1}, \bar{y}_{r_2}, \bar{z}_{r_1}, \bar{z}_{r_2}$: The response means of x, y and z respectively.

ρ_{yx}, ρ_{xz}, ρ_{yz} : The correlation coefficients between the variables shown in suffices.

β_{yx}, β_{xz}, β_{yz} : The population regression coefficients between the variables shown in suffices.

$S_x^2 = (N-1)^{-1} \sum_{i=1}^{N} \left(x_i - \bar{X}\right)^2$: The population mean square of the variable x.

S_y^2, S_z^2 : The population mean squares of the variables y and z respectively.

$f_1\left(=\frac{r_1}{n}\right)$, $f_2\left(=\frac{r_2}{u}\right)$: The fractions of respondents in the sample of sizes n and u respectively.

t_1 (= 1 − f_1), t_2 (= 1 − f_2) : The fractions of non-respondents in the sample of sizes n and u respectively.

3. FORMULATION OF THE ESTIMATOR *T*

For the estimation of population mean $\bar{Y}$ on the second occasion, two estimators T_1 and T_2 have been formulated which do not use information on an additional auxiliary variable z, however, the information from the previous occasion on study variable may be used as an auxiliary information. Estimators T_1 and T_2 are structured to cope with the problems of non-response on both the occasions. The estimator T_1 is based on the sample s_u of size u, drawn afresh on the current occasion. Following Singh (2003b), the missing values are filled by the imputed values which are derived using the mean method of imputation. The data after imputation take the following form

$$y_{\bullet i} = \begin{cases} y_i & \text{if } i \in R_2 \\ \bar{y}_{r_2} & \text{if } i \in R_2^c \end{cases} \tag{1}$$

Under this method of imputation the estimator T_1 based on sample s_u of size u is given by

$$T_1 = \frac{1}{u}\sum_{i \in s_u} y_{\bullet i} = \frac{1}{u}\left[\sum_{i \in R_2} y_{\bullet i} + \sum_{i \in R_2^c} y_{\bullet i}\right] = \bar{y}_{r_2} \tag{2}$$

where $\bar{y}_{r_2} = \frac{1}{r_2}\sum_{i \in R_2} y_i$

The second estimator T_2 based on the sample s_m of size m, which is common to both the occasions and utilizes the information from the first occasion. Since, there is non response on the first occasion; hence, the missing values are replaced by the imputed values derived using the mean method of imputation. The data after imputation take the following form

$$x_{\bullet i} = \begin{cases} x_i & \text{if } i \in R_1 \\ \bar{x}_{r_1} & \text{if } i \in R_1^c \end{cases} \tag{3}$$

Under this method of imputation the estimator based on sample of size n of the first occasion is given by

$$\bar{x}_n = \frac{1}{n}\sum_{i \in s_n} x_{\bullet i} = \frac{1}{n}\left[\sum_{i \in R_1} x_{\bullet i} + \sum_{i \in R_1^c} x_{\bullet i}\right] = \bar{x}_{r_1} \tag{4}$$

where $\bar{x}_{r_1} = \frac{1}{r_1}\sum_{i \in R_1} x_i$.

For matched (common) sample on the current occasion, the information on study variable from the previous occasion is available and it may be utilized as auxiliary information on the current occasion. The difference/regression estimator is an explicit function of known auxiliary variable and its implementation under SRSWOR is convenient from practical point of view, viz. Sarndal *et al.* (1992). Motivated with this argument and utilizing the mean method of imputation for imputing the missing values on the first occasion, the second estimator T_2 for estimating the population mean on the current occasion is considered as difference type estimator based on sample s_m of size m (= $n\lambda$), which is common to both the occasions.

$$T_2 = \bar{y}_m + \beta_{yx}\left(\bar{x}_{r_1} - \bar{x}_m\right) \tag{5}$$

where β_{yx} is the known population regression coefficient between y and x.

Considering the convex linear combination of the estimators T_1 and T_2; the estimator T is defined as:

$$T = \varphi T_1 + (1 - \varphi)T_2 \tag{6}$$

where φ is an unknown constant to be determined by the minimization of the variance of the estimator T.

Remark 3.1 : For estimating the mean on each occasion the estimator T_1 is suitable, which implies that more belief on T_1 could be shown by choosing φ as 1 (or close to 1), while for estimating the change from one occasion to the next, the estimator T_2 could be more useful so that φ may be opted as 0 (or close to 0). For asserting both the problems simultaneously, the suitable (optimum) choice of φ is required.

4. PROPERTIES OF THE ESTIMATOR *T*

Theorem 4.1 : T is an unbiased estimator of $\bar{Y}$.

Proof : Since T_1 is a sample mean of study variable y based on r randomly responding units in the fresh sample s_u of size u, it is unbiased for the population mean $\bar{Y}$. T_2 is a difference type estimator, which is also an unbiased estimator of $\bar{Y}$. Estimator T is a convex linear combination of T_1 and T_2, hence it is an unbiased estimator of $\bar{Y}$.

Theorem 4.2 : Ignoring the finite population correction, the variance of T is

$$V(T) = \varphi^2 V(T_1) + (1 - \varphi)^2\, V(T_2) \tag{7}$$

where $$V(T_1) = \frac{1}{r_2} S_y^2 \tag{8}$$

and $$V(T_2) = \left[\frac{1}{m}\left(1-\rho_{yx}^2\right) + \frac{1}{r_1}\rho_{yx}^2\right] S_y^2 \tag{9}$$

Proof : The variance of T is given by

$$V(T) = E\left(T - \bar{Y}\right)^2 = \varphi^2 V(T_1) + (1 - \varphi)^2\, V(T_2) + 2\varphi(1 - \varphi)\,\text{cov}\,(T_1, T_2) \tag{10}$$

Variances and covariance terms in the right hand side of the equation (10) can be derived as

$$V(T_1) = V\left(\bar{y}_{r_2}\right) = \left(\frac{1}{r_2} - \frac{1}{N}\right) S_y^2 \tag{11}$$

Similarly

$$V(T_2) = E\left(T_2 - \bar{Y}\right)^2 = E\left[\bar{y}_m + \beta_{yx}\left(\bar{x}_{r_1} - \bar{x}_m\right) - \bar{Y}\right]^2$$
$$= \left[\left(\frac{1}{m} - \frac{1}{N}\right) - \left(\frac{1}{m} - \frac{1}{r_1}\right)\rho_{yx}^2\right] S_y^2 \tag{12}$$

Since, T_1 and T_2 are the unbiased estimators and based on two independent samples of sizes u and m respectively, therefore,

$$\text{cov}(T_1, T_2) = 0 \tag{13}$$

For sufficiently large population, the finite population corrections (fpc) in equations (11) and (12) can be ignored [viz., Feng and Zou (1997) and Biradar and Singh (2001)], which yields

$$V(T_1) = \frac{1}{r_2} S_y^2 \tag{14}$$

and $$V(T_2) = \left[\frac{1}{m}\left(1-\rho_{yx}^2\right) + \frac{1}{r_1}\rho_{yx}^2\right] S_y^2 \tag{15}$$

Now substituting the values of $V(T_1)$, $V(T_2)$ and $\text{cov}(T_1, T_2)$ from equations (14), (15) and (13), respectively, in equation (10), we obtain the $V(T)$ as given in equation (7).

The variance of T in equation (7) is a function of the unknown constant φ, it is minimized with respect to φ and subsequently the optimum value of φ is obtained as

$$\varphi_{\text{opt.}} = \frac{V(T_2)}{V(T_1) + V(T_2)} \tag{16}$$

Substituting this optimum value $\varphi_{\text{opt.}}$ in equation (7) we obtain the optimum variance of T as

$$V(T)_{\text{opt.}} = \frac{V(T_1)V(T_2)}{V(T_1) + V(T_2)} \tag{17}$$

Further, substituting the values from equations (8) and (9) in equation (17), which yields the simplified value of $V(T)_{\text{opt.}}$ as shown below in Theorem 4.3.

Theorem 4.3: The $V(T)_{\text{opt.}}$ is derived as

$$V(T)_{\text{opt.}} = \frac{\left[\left\{\rho_{yx}^2 + f_1\left(1-\rho_{yx}^2\right)\right\} - \mu\,\rho_{yx}^2\right]}{n\left[-f_2\,\rho_{yx}^2\,\mu^2 + \left\{f_2\,\rho_{yx}^2 + f_1\,f_2\left(1-\rho_{yx}^2\right) - f_1\right\}\mu + f_1\right]} S_y^2 \tag{18}$$

where $f_1 = \dfrac{r_1}{n}$, $f_2 = \dfrac{r_2}{u}$ and $\mu\left(=\dfrac{u}{n}\right)$

The $V(T)_{opt.}$ derived in equation (18) is the function of μ. To estimate the population mean on each occasion, a good choice for μ is 1 (the case of no matching) while for estimating the change from one occasion to the other, μ should be 0 (the case of complete matching). But to design an amicable strategy that would be efficient for both the problems simultaneously, the optimum choice of μ is desired. Since $V(T)_{opt.}$ is the function of μ (fraction of fresh sample on the current occasion), which is an important factor in reducing the cost of the survey, it is necessary to minimize $V(T)_{opt.}$ in equation (18) with respect to μ. The minimum value of μ is obtained as

$$\mu_{opt.} = \frac{A_1A_3 \pm \sqrt{A_1^2A_3^2 + A_2A_3(f_1A_2 + A_1A_4)}}{A_2A_3}$$

where $A_1 = \left\{\rho_{yx}^2 + f_1\left(1-\rho_{yx}^2\right)\right\}$, $A_2 = \left\{\rho_{yx}^2\right\}$,

$A_3 = -f_2\rho_{yx}^2$

and $A_4 = \left\{f_2\,\rho_{yx}^2 + f_1f_2\left(1-\rho_{yx}^2\right) - f_1\right\}$

The real value of $\mu_{opt.}$ exists if $\left\{A_1^2A_3^2 + A_2A_3(f_1A_2 + A_1A_4)\right\} \geq 0$. For an admissible value of μ, it should be remembered that $0 \leq \mu_{opt.} \leq 1$. Substituting the admissible value of $\mu_{opt.}$ (say μ_0) in equation (18), we have

$$V(T)_{opt.^*} = \frac{[\{\rho_{yx}^2 + f_1(1-\rho_{yx}^2)\} - \mu_0\rho_{yx}^2]}{n[-f_2\rho_{yx}^2\mu_0^2 + \{f_2\rho_{yx}^2 + f_1f_2(1-\rho_{yx}^2) - f_1\}\mu_0 + f_1]}S_y^2 \quad (19)$$

where $V(T)_{opt.^*}$ is the optimum value of T with respect to both φ and μ.

Case 1 : When non-response occurs only on the first occasion

The estimator for the population mean on the current occasion can be obtained as $\xi = \delta\, T_1^* + (1-\delta)\, T_2$, where $T_1^* = \bar{y}_u$, T_2 is defined in equation (5) and δ is an unknown constant to be determined so as to minimize the variance of the estimator ξ.

The optimum fraction of the fresh sample to be drawn in this case is obtained as

$$\mu_{opt.} = \frac{\left[-\{\rho_{yx}^2 + f_1(1-\rho_{yx}^2)\} \pm \sqrt{\dfrac{\{\rho_{yx}^2 + f_1(1-\rho_{yx}^2)\}^2 - \rho_{yx}^2}{(f_1 + \{\rho_{yx}^2 + f_1(1-\rho_{yx}^2)\}(1-f_1))}}\right]}{-\rho_{yx}^2} = \mu_0^*$$

and the minimum variance of the estimator ξ is derived as

$$V(\xi)_{opt.^*} = \frac{\left[\left\{\rho_{yx}^2 + f_1\left(1-\rho_{yx}^2\right)\right\} - \mu_0^*\rho_{yx}^2\right]}{n\left[-\rho_{yx}^2\mu_0^{*2} + \left\{\rho_{yx}^2 + f_1\left(1-\rho_{yx}^2\right) - f_1\right\}\mu_0^* + f_1\right]}S_y^2 \quad (20)$$

Case 2 : When non-response occurs only on the current occasion

The estimator for the population mean on the current occasion can be obtained as $\xi^* = \delta^* T_1 + (1-\delta^*)T_2^*$, where T_1 is defined in equation (2), $T_2^* = \bar{y}_m + \beta_{yx}(\bar{x}_n - \bar{x}_m)$ and δ^* is an unknown constant to be determined so as to minimize the variance of the estimator ξ^*, which is same as Singh *et al.* (2008 b).

The optimum fraction of the fresh sample to be drawn in this case is obtained as

$$\mu_{opt.} = \frac{1-\sqrt{\dfrac{\left(1-\rho_{yx}^2\right)}{f_2}}}{\rho_{yx}^2} = \mu_0^{**}$$

and the minimum variance of the estimator ξ^* is given as

$$V\left(\xi^*\right)_{opt.^*} = \frac{\left[1-\mu_0^{**}\rho_{yx}^2\right]}{n[1-\mu_0^{**}(1-f_2) - f_2\,\rho_{yx}^2\,\mu_0^{**2}]}S_y^2 \quad (21)$$

5. EFFICIENCY COMPARISON

The percent relative loss in efficiency of the estimator T with respect to the estimator for the similar circumstances but under the complete response case (with no missing data) has been obtained to judge the effect of non-response on the precision of estimates under two-occasion successive sampling.

Consider the following estimator of $\bar{Y}$

$$T^* = \psi T_1^* + (1-\psi)T_2^* \tag{22}$$

where $T_1^* = \bar{y}_u$, $T_2^* = \bar{y}_m + \beta_{yx}(\bar{x}_n - \bar{x}_m)$ and ψ is the unknown constant to be determined by the minimization of the variance of the estimator T^*.

Since T^* is the unbiased estimator of $\bar{Y}$ and is based on two independent samples, the covariance term between T_1^* and T_2^* vanishes. Therefore, following Sukhatme *et al.* (1984) and ignoring fpc, the optimum variance of T^* is given by

$$V(T^*)_{\text{opt.}^*} = \left[1 + \sqrt{(1-\rho_{yx}^2)}\right]\frac{S_y^2}{2n} \tag{23}$$

The percent relative loss in precision of the estimators T, ξ and ξ^* with respect to the estimator T^* under their respective optimality conditions are given by

$$L = \frac{V(T)_{\text{opt.}^*} - V(T^*)_{\text{opt.}^*}}{V(T)_{\text{opt.}^*}} \times 100,$$

$$L^* = \frac{V(\xi)_{\text{opt.}^*} - V(T^*)_{\text{opt.}^*}}{V(\xi)_{\text{opt.}^*}} \times 100 \text{ and}$$

$$L^{**} = \frac{V(\xi^*)_{\text{opt.}^*} - V(T^*)_{\text{opt.}^*}}{V(\xi^*)_{\text{opt.}^*}} \times 100 \text{ respectively.}$$

Since, the expressions of the percent relative loss in precision are functions of ρ_{yx}^2, hence for any choice of ρ_{yx} in the range (–1, 1), the values of percent relative loss is same as the ρ_{yx} is considered in the range (0, 1). Therefore, for different choices of t_1 $(= 1 - f_1)$, t_2 $(= 1 - f_2)$ and ρ_{yx}, the percent relative loss in precisions of T, ξ and ξ^* over T^* under the optimal conditions are computed and presented below in Tables 1, 2 and 3.

Table 1. Percent Relative Loss (L) in precision of T over T^* at optimum values of μ when non-response occurs on both the occasions

ρ_{yx}		± 0.3		± 0.5		± 0.7		± 0.9	
t_1	t_2	μ_0	L	μ_0	L	μ_0	L	μ_0	L
0.05	0.05	0.2746	2.1037	0.4736	2.7463	0.5682	3.2278	0.6983	3.9412
	0.10	*	-	0.3811	4.7429	0.5293	5.5807	0.6840	6.4226
	0.15	*	-	0.2805	6.2895	0.4870	7.7596	0.6683	8.8502
	0.20	*	-	0.1707	7.3457	0.4408	9.7491	0.6513	11.2191
0.10	0.05	0.3128	2.3713	0.5013	3.1999	0.5910	4.0357	0.7142	5.4072
	0.10	0.0446	3.2530	0.4137	5.3369	0.5541	6.4903	0.7006	7.9469
	0.15	*	-	0.3184	7.0476	0.5140	8.7801	0.6858	10.4357
	0.20	*	-	0.2143	8.2938	0.4702	10.8904	0.6696	12.8689
0.15	0.05	0.3510	2.6389	0.5290	3.6534	0.6137	4.8437	0.7301	6.8732
	0.10	0.0977	3.7430	0.4462	5.9309	0.5789	7.3999	0.7172	9.4713
	0.15	*	-	0.3563	7.8057	0.5410	9.8006	0.7032	12.0212
	0.20	*	-	0.2580	9.2419	0.4997	12.0317	0.6880	14.5186
0.20	0.05	0.3891	2.9065	0.5567	4.1070	0.6364	5.6516	0.7460	8.3392
	0.10	0.1508	4.2330	0.4788	6.5249	0.6036	8.3096	0.7339	10.9956
	0.15	*	-	0.3941	8.5638	0.5680	10.8211	0.7207	13.6067
	0.20	*	-	0.3016	10.1899	0.5291	13.1731	0.7063	16.1684

Note: * indicate that μ_0 do not exist.

Table 2. Percent Relative Loss (L^*) in precision of ξ over T^* at optimum values of μ when non-response occurs only on the first occasion

$t_1 \rightarrow$	0.05		0.10		0.15		0.20	
$\rho_{yx} \downarrow$	μ_0^*	L^*	μ_0^*	L^*	μ_0^*	L^*	μ_0^*	L^*
± 0.3	0.5362	0.1152	0.5606	0.2303	0.5850	0.3455	0.6094	0.4606
± 0.5	0.5591	0.3349	0.5823	0.6699	0.6055	1.0048	0.6287	1.3397
± 0.7	0.6042	0.7146	0.6250	1.4293	0.6459	2.1439	0.6667	2.8586
± 0.9	0.7116	1.4103	0.7268	2.8206	0.7420	4.2308	0.7571	5.6411

Table 3. Percent Relative Loss (L^{**}) in precision of ξ^* over T^* at optimum values of μ when non-response occurs only on the current occasion

$t_2 \rightarrow$	0.05		0.10		0.15		0.20	
$\rho_{yx} \downarrow$	μ_0^{**}	L^{**}	μ_0^{**}	L^{**}	μ_0^{**}	L^{**}	μ_0^{**}	L^{**}
± 0.3	0.2364	1.8361	*	-	*	-	*	-
± 0.5	0.4459	2.2928	0.3400	4.1489	0.2427	5.5313	0.1270	6.3977
± 0.7	0.5455	2.4199	0.5045	4.6711	0.4600	6.7391	0.4114	8.6078
± 0.9	0.6825	2.4752	0.6673	4.8983	0.6509	7.2647	0.6329	9.5694

Note : *indicate that μ_0^{**} do not exist.

For a fixed value of t_1 and ρ_{yx}, the values of μ_0 decrease while the values of L increase with the increase in the values of t_2 (Table 1). Thus, the higher the non-response rate, the smaller the fresh sample is desirable on the current occasion. However, for a fixed value of t_2 and ρ_{yx}, the values of μ_0 and L increase with the increase in the values of t_1 (Table 1). This phenomenon is obvious since the higher the non-response rate, the higher the loss in precision occurs. Similar patterns are also visible in Tables 2 and 3.

For all studied cases the loss in precision was observed, but it was not too high. This indicates that the mean method of imputation may be a nice approach to cope with the problems of non-response on the current occasion in two-occasion successive sampling.

6. FORMULATION OF ESTIMATOR Δ

The estimator T proposed in equation (6) uses information only on the character under study through the sample values over two occasions. It is well known fact that auxiliary variable assists in estimating the parameters concerned to the study variable. Sometimes, information on an auxiliary variable may be readily available on both the occasions. For example, in an environmental survey the number of polluting industries in a locality may be known, tonnage (or seat capacity) of each vehicle or ship is known in survey sampling of transportation, etc. In such cases, imputation can be carried out with the aid of the auxiliary variable. Hence onward now we assume that information on an almost stable auxiliary variable z is available on both the occasions. Under this assumption we consider two estimators of the population mean, one is the point estimator corresponding to the proposed imputation method that is based on sample s_u of size u that are drawn afresh on the current occasion such that out of these u units, r_2 units respond and remaining $(u - r_2)$ units do not respond. The study variable under proposed imputation takes the following form

$$y_{\bullet i} = \begin{cases} y_i & \text{if } i \in R_2 \\ \bar{y}_{r_2} + \dfrac{u}{(u-r_2)}\beta_{yz}\left(\bar{Z} - \bar{z}_{r_2}\right) & \text{if } i \in R_2^c \end{cases}$$

Under this method of imputation, the point estimator Δ_1 based on sample s_u of size u is defined as

$$\Delta_1 = \frac{1}{u}\sum_{i \in s_u} y_{\bullet i} = \frac{1}{u}\left[\sum_{i \in R_2} y_{\bullet i} + \sum_{i \in R_2^c} y_{\bullet i}\right]$$
$$= \bar{y}_{r_2} + \beta_{yz}\left(\bar{Z} - \bar{z}_{r_2}\right) \quad (24)$$

The second estimator based on the sample s_m of size m, which is common to both the occasions and utilizes the information from the first occasion as well. Since, there is non response on the first occasion; hence, the missing values are replaced by the imputed values, which are derived using the following method of imputation. The data after imputation take the following form

$$x_{\bullet i} = \begin{cases} x_i & \text{if } i \in R_1 \\ \bar{x}_{r_1} + \dfrac{n\,\beta_{xz}}{(n-r_1)}\left(\bar{Z} - \bar{z}_{r_1}\right) & \text{if } i \in R_1^c \end{cases}$$

Under this method of imputation, the estimator based on sample s_n of size n on the first occasion is given by

$$\bar{x}_n^* = \frac{1}{n}\sum_{i \in s_n} x_{\bullet i} = \frac{1}{n}\left[\sum_{i \in R_1} x_{\bullet i} + \sum_{i \in R_1^c} x_{\bullet i}\right]$$
$$= \bar{x}_{r_1} + \beta_{xz}\left(\bar{Z} - \bar{z}_{r_1}\right)$$

where $\bar{x}_{r_1} = \dfrac{1}{r_1}\sum_{i \in R_1} x_i$

Therefore, the second estimator for estimating the population mean on the current occasion is a chain type difference estimator based on a sample s_m of size $m\ (= n\lambda)$, which is common to both the occasions and utilizes the proposed imputation method for imputing the missing values on the first occasion. It is defined as

$$\Delta_2 = \bar{y}_m^* + \beta_{yx}\left(\bar{x}_n^* - \bar{x}_m^*\right) \tag{25}$$

where
$$\bar{y}_m^* = \bar{y}_m + \beta_{yz}\left(\bar{Z} - \bar{z}_m\right)$$
$$\bar{x}_n^* = \bar{x}_{r_1} + \beta_{xz}\left(\bar{Z} - \bar{z}_{r_1}\right)$$
$$\bar{x}_m^* = \bar{x}_m + \beta_{xz}(\bar{Z} - \bar{z}_m)$$

The final estimator of the $\bar{Y}$ is the convex linear combination of the estimators Δ_1 and Δ_2 and is defined as

$$\Delta = \varphi^*\Delta_1 + \left(1 - \varphi^*\right)\Delta_2 \tag{26}$$

where φ^* is the unknown constant to be determined so that it will minimize the variance of the estimator Δ.

6.1 Estimator Δ in Practice

The main difficulty in using the estimator Δ in practice is the knowledge of the population regression coefficients β_{yz}, β_{yx} and β_{xz}, which are more often not known. In such situations, following imputation methods may be considered :

(a) Imputation methods for imputing the missing values on the current occasion

(i)
$$y_{\bullet i} = \begin{cases} y_i & \text{if } i \in R_2 \\ \bar{y}_{r_2} + \dfrac{u}{(u-r_2)}\hat{b}_1\left(\bar{Z} - \bar{z}_{r_2}\right) & \text{if } i \in R_2^c \end{cases} \tag{27}$$

where $\hat{b}_1 = \dfrac{s_{yz}^*(r_2)}{s_z^{*2}(r_2)}$,

$$s_{yz}^*(r_2) = \frac{1}{(r_2-1)}\left[\sum_{i=1}^{r_2}\left(y_i - \bar{y}_{r_2}\right)\left(z_i - \bar{z}_{r_2}\right)\right]$$

$$s_z^{*2}(r_2) = \frac{1}{(r_2-1)}\left[\sum_{i=1}^{r_2}\left(z_i - \bar{z}_{r_2}\right)^2\right]$$

The point estimator of $\bar{Y}$ based on the imputation method given in equation (27) is

$$\Delta_1^* = \bar{y}_{r_2} + \hat{b}_1\left(\bar{Z} - \bar{z}_{r_2}\right) \tag{28}$$

(ii)
$$y_{\bullet i} = \begin{cases} y_i & \text{if } i \in R_2 \\ \bar{y}_{r_2} + \dfrac{u}{(u-r_2)}\hat{b}_2\left(\bar{Z} - \bar{z}_{r_2}\right) & \text{if } i \in R_2^c \end{cases} \tag{29}$$

where $\hat{b}_2 = \dfrac{s_{yz}^*(r_2)}{s_z^2(u)}$ and

$$s_z^2(u) = \frac{1}{(u-1)}\left[\sum_{i=1}^{u}\left(z_i - \bar{z}_u\right)^2\right]$$

The point estimator of $\bar{Y}$ based on the imputation method given in equation (29) is

$$\Delta_1^{**} = \bar{y}_{r_2} + \hat{b}_2\left(\bar{Z} - \bar{z}_{r_2}\right) \tag{30}$$

(b) Imputation methods for imputing the missing values on the first occasion

(i)
$$x_{\bullet i} = \begin{cases} x_i & \text{if } i \in R_1 \\ \bar{x}_{r_1} + \dfrac{n}{(n-r_1)}\hat{b}_3\left(\bar{Z} - \bar{z}_{r_1}\right) & \text{if } i \in R_1^c \end{cases} \tag{31}$$

where $\hat{b}_3 = \dfrac{s_{xz}^*(r_1)}{s_z^{*2}(r_1)}$,

$$s_{xz}^{*}(r_1) = \frac{1}{(r_1-1)}\left[\sum_{i=1}^{r_1}(x_i-\bar{x}_{r_1})(z_i-\bar{z}_{r_1})\right]$$

$$s_{z}^{*2}(r_1) = \frac{1}{(r_1-1)}\left[\sum_{i=1}^{r_1}(z_i-\bar{z}_{r_1})^2\right]$$

Under this method of imputation the estimator based on sample s_n of size n on the first occasion is given by

$$\bar{x}_n^{**} = \bar{x}_{r_1} + \hat{b}_3(\bar{Z}-\bar{z}_{r_1}) \quad (32)$$

$$\text{(ii)}\quad x_{\bullet i} = \begin{cases} x_i & \text{if } i \in R_1 \\ \bar{x}_{r_1} + \dfrac{n}{(n-r_1)}\hat{b}_4(\bar{Z}-\bar{z}_{r_1}) & \text{if } i \in R_1^c \end{cases} \quad (33)$$

where $\hat{b}_4 = \dfrac{s_{xz}^{*}(r_1)}{s_z^2(n)}$

$$s_z^2(n) = \frac{1}{(n-1)}\left[\sum_{i=1}^{n}(z_i-\bar{z}_n)^2\right]$$

The point estimator based on the method of imputation given in equation (33) is

$$\bar{x}_n^{***} = \bar{x}_{r_1} + \hat{b}_4(\bar{Z}-\bar{z}_{r_1}) \quad (34)$$

Now, replacing the unknown regression parameters by their estimators in equation (25), we obtain the following two estimators of $\bar{Y}$ based on matched sample s_m of size m corresponding to the two imputation methods proposed in equations (31) and (33).

$$\text{(i)}\quad \Delta_2^{*} = \bar{y}_m^{**} + b_{yx}(m)\left(\bar{x}_n^{**}-\bar{x}_m^{**}\right) \quad (35)$$

where $\bar{y}_m^{**} = \bar{y}_m + b_{yz}(m)(\bar{Z}-\bar{z}_m)$ and

$$\bar{x}_m^{**} = \bar{x}_m + b_{xz}(m)(\bar{Z}-\bar{z}_m)$$

$$\text{(ii)}\quad \Delta_2^{**} = \bar{y}_m^{**} + b_{yx}(m)\left(\bar{x}_n^{***}-\bar{x}_m^{**}\right) \quad (36)$$

where $b_{yx}(m)$, $b_{yz}(m)$ and $b_{xz}(m)$ are the sample regression coefficients between the variables shown in suffices and based on the sample sizes indicated in the braces.

Thus, we may suggest the following two different structures of estimator Δ for estimating population mean $\bar{Y}$ on the current occasion :

$$\Delta^{*} = \varphi^{**}\Delta_1^{*} + \left(1-\varphi^{**}\right)\Delta_2^{*} \quad (37)$$

and $$\Delta^{**} = \varphi^{***}\Delta_1^{**} + \left(1-\varphi^{***}\right)\Delta_2^{**} \quad (38)$$

where φ^{**} and φ^{***} are the unknown constants to be determined in such a way so that the mean square errors of Δ^{*} and Δ^{**} are minimum.

7. PROPERTIES OF THE ESTIMATORS Δ, Δ^{*} AND Δ^{**}

Theorem 7.1 : Δ is an unbiased estimator of $\bar{Y}$.

Proof : Δ_1 and Δ_2 both are difference type estimators and they are unbiased for $\bar{Y}$. Since, the final estimator Δ is a convex linear combination of Δ_1 and Δ_2, therefore, Δ is also an unbiased estimator of $\bar{Y}$.

Theorem 7.2 : Variance of Δ, ignoring the finite population correction is given by

$$V(\Delta) = \varphi^{*2}V(\Delta_1) + \left(1-\varphi^{*}\right)^2 V(\Delta_2) \quad (39)$$

where $$V(\Delta_1) = \frac{1}{r_2}A\,S_y^2 \quad (40)$$

and $$V(\Delta_2) = \left[\left(\frac{1}{m}\right)A + \left(\frac{1}{m}-\frac{1}{r_1}\right)B\right]S_y^2 \quad (41)$$

where $A = 1-\rho_{yz}^2$ and

$$B = 2\,\rho_{yz}^2\,\rho_{yx} - \rho_{yx}^2\left(1+\rho_{yz}^2\right)$$

Remark 7.1 : The results of Theorem 7.2 are derived under the following assumptions:

(i) A population size N is sufficiently large so that the finite population corrections (fpc) may be ignored.

(ii) $\rho_{xz} = \rho_{yz}$, which is an intuitive assumption, considered by Cochran (1977) and Feng and Zou (1997).

(iii) Δ_1 and Δ_2 are difference type estimators, they are unbiased. These two estimators are based on two independent samples s_u and s_m of sizes u and m respectively, hence, the covariance term between them is vanished.

Since, Δ_1^{*} and Δ_1^{**} are simple linear regression type estimators and Δ_2^{*} and Δ_2^{**} are chain type regression estimators, hence, all they are biased estimators of $\bar{Y}$. Therefore, the resulting estimators Δ^* and Δ^{**}, defined in equations (37) and (38) respectively, are also biased. The bias B(.) and mean square errors MSE(.), up to the first order of approximations and under the conditions that the fpcs are ignored, are derived under the large sample approximations. Thus, we have the following theorems:

Theorem 7.3 : $B(\Delta^*) = \varphi^{**} B\left(\Delta_1^*\right) + \left(1-\varphi^{**}\right) B\left(\Delta_2^*\right)$ (42)

where $$B\left(\Delta_1^*\right) = -\frac{\beta_{yz}}{r_2}\left[\frac{C_{012}}{S_{yz}} - \frac{C_{003}}{S_z^2}\right]$$

and

$$B\left(\Delta_2^*\right) = -\frac{\beta_{yz}}{m}\left[\frac{C_{012}}{S_{yz}} - \frac{C_{003}}{S_z^2}\right] + \left(\frac{1}{m}-\frac{1}{r_1}\right)\left[\beta_{yx}\left\{\frac{C_{300}}{S_x^2} - \frac{C_{210}}{S_{yx}}\right\} + \beta_{yx}\beta_{xz}\left\{\frac{C_{111}}{S_{yx}} - \frac{C_{201}}{S_x^2} - \frac{C_{003}}{S_z^2} + \frac{C_{102}}{S_{xz}}\right\}\right]$$

where $C_{abc} = E\left[\left(x_i - \bar{X}\right)^a \left(y_i - \bar{Y}\right)^b \left(z_i - \bar{Z}\right)^c\right]$; $a \geq 0$, $b \geq 0$, $c \geq 0$

Theorem 7.4 : $B(\Delta^{**}) = \varphi^{***} B\left(\Delta_1^{**}\right) + \left(1-\varphi^{***}\right) B\left(\Delta_2^{**}\right)$ (43)

where $$B\left(\Delta_1^{**}\right) = -\beta_{yz}\left[\frac{1}{r_2}\left(\frac{C_{012}}{S_{yz}}\right) - \frac{1}{u}\left(\frac{C_{003}}{S_z^2}\right)\right]$$

and $$B\left(\Delta_2^{**}\right) = -\frac{\beta_{yz}}{m}\left[\frac{C_{012}}{S_{yz}} - \frac{C_{003}}{S_z^2}\right] + \left(\frac{1}{m}-\frac{1}{r_1}\right)\left[\beta_{yx}\left\{\frac{C_{300}}{S_x^2} - \frac{C_{210}}{S_{yx}}\right\} + \beta_{yz}\beta_{xz}\left\{\frac{C_{111}}{S_{yx}} - \frac{C_{201}}{S_x^2} + \frac{C_{102}}{S_{xz}}\right\}\right] + \left(\frac{1}{n}-\frac{1}{m}\right)\beta_{yz}\beta_{xz}\left\{\frac{C_{003}}{S_z^2}\right\}$$

Theorem 7.5 : $MSE\left(\Delta^*\right) = \phi^{**2} MSE\left(\Delta_1^*\right) + \left(1-\varphi^{**}\right)^2 MSE\left(\Delta_2^*\right)$ (44)

where $$MSE\left(\Delta_1^*\right) = \frac{1}{r_2} A\, S_y^2 \quad (45)$$

and $$MSE\left(\Delta_2^*\right) = \left[\left(\frac{1}{m}\right)A + \left(\frac{1}{m}-\frac{1}{r_1}\right)B\right]S_y^2 \quad (46)$$

Theorem 7.6 : $MSE\left(\Delta^{**}\right) = \varphi^{***2} MSE\left(\Delta_1^{**}\right) + \left(1-\varphi^{***}\right)^2 MSE\left(\Delta_2^{**}\right)$ (47)

where $$MSE\left(\Delta_1^{**}\right) = \frac{1}{r_2} A\, S_y^2 \quad (48)$$

and $$MSE\left(\Delta_2^{**}\right) = \left[\left(\frac{1}{m}\right)A + \left(\frac{1}{m}-\frac{1}{r_1}\right)B\right]S_y^2 \quad (49)$$

Remark 7.2 : Theorems are derived under the assumptions given in Remark 7.1.

Remark 7.3 : From equations (44) and (47) it follows that up to the first order of approximation, the mean square errors of the estimators Δ^* and Δ^{**} are equal, which are also same as the variance of the estimator Δ obtained in equation (39). Estimators Δ^* and Δ^{**} are based on two different sample estimates and up to the first order of approximations they are equally precise as Δ.

8. MINIMUM VARIANCE OF Δ

Since the variance of the estimator Δ in equation (39) is a function of the unknown constant φ^*, therefore, it is minimized with respect to φ^*; subsequently, the optimum value of φ^* is obtained as

$$\varphi^*_{\text{opt.}} = \frac{V(\Delta_2)}{V(\Delta_1) + V(\Delta_2)} \quad (50)$$

Substituting the value of $\varphi^*_{\text{opt.}}$ in (39) we obtain the optimum variance of Δ as

$$V(\Delta)_{\text{opt.}} = \frac{V(\Delta_1)V(\Delta_2)}{V(\Delta_1) + V(\Delta_2)} \quad (51)$$

Further, substituting the values from equations (40) and (41) in equation (51), which yields the simplified value of $V(\Delta)_{\text{opt.}}$ as

$$V(\Delta)_{opt.} = \frac{A(C+\mu B)}{n\left[f_2 B\mu^2 + (f_2C - f_1A)\mu + f_1A\right]} S_y^2 \quad (52)$$

where $C = f_1 (A + B) - B$

9. OPTIMUM REPLACEMENT POLICY

To determine the optimum value of μ (fraction of sample to be taken afresh on the current (second) occasion) so that $\bar{Y}$ may be estimated with the maximum precision, we minimize $V(\Delta)_{opt.}$ in equation (52) with respect to μ. It yields the optimum value of μ as

$$\mu_{opt.} = \frac{-f_2C \pm \sqrt{f_1 f_2 A(B+C)}}{f_2 B} \quad (53)$$

The real values of $\mu_{opt.}$ exist, if $\{f_1 f_2 A (B + C\} \geq 0$. Since A, B and C are the functions of the correlation coefficients ρ_{yz} and ρ_{yx}, for any choices of ρ_{yz} and ρ_{yx}, if $\{f_1 f_2 A (B + C)\} \geq 0$, two real values of $\mu_{opt.}$ from equation (53) may be obtained. To choose the admissible value of $\mu_{opt.}$ (say μ_1), it should be remembered that $0 \leq \mu_1 \leq 1$. Substituting the admissible value μ_1 from equation (53) in equation (52), we have

$$V(\Delta)_{opt.^*} = \frac{A(C+\mu_1 B)}{n\left[f_2 B\mu_1^2 + (f_2C - f_1A)\mu_1 + f_1A\right]} S_y^2 \quad (54)$$

where $V(\Delta)_{opt.^*}$ is the optimum value of Δ with respect to both φ^* and μ.

10. SPECIAL CASES

Case 10.1 : When non-response occurs only on the first occasion

The estimator for the population mean on the current occasion can be obtained as $\tau = \theta T_{1u} + (1 - \theta)\Delta_2$, where $T_{1u} = \bar{y}_u + \beta_{yz}\left(\bar{Z} - \bar{z}_u\right)$, Δ_2 is defined in equations (25) and θ is an unknown constant to be determined so as to minimize the variance of the estimator τ.

The optimum fraction of the fresh sample to be drawn in this case is obtained as

$$\mu_{opt.} = \frac{-C \pm \sqrt{f_1 A(B+C)}}{B} = \mu_1^*$$

Hence, the minimum variance of the estimator τ is calculated as

$$V(\tau)_{opt.^*} = \frac{A\left(C+\mu_1^* B\right)}{n\left[B\mu_1^{*2} + (C - f_1A)\mu_1^* + f_1A\right]} S_y^2 \quad (55)$$

Case 10.2 : When non-response occurs only on the current occasion

The estimator for the population mean on the current occasion can be obtained as

$$\tau^* = \theta^*\Delta_1 + (1 - \theta^*)T_{2m}$$

where Δ_1 is defined in equation (24),

$$T_{2m} = \bar{y}_m^* + \beta_{yx}\left(\hat{\bar{x}}_n - \bar{x}_m^*\right)$$

where $\bar{y}_m^* = \bar{y}_m + \beta_{yz}\left(\bar{Z} - \bar{z}_m\right)$

$\hat{\bar{x}}_n = \bar{x}_n + \beta_{xz}\left(\bar{Z} - \bar{z}_n\right)$

and $\bar{x}_m^* = \bar{x}_m + \beta_{xz}(\bar{Z} - \bar{z}_m)$

θ^* is an unknown constant to be determined so as to minimize the variance of the estimator τ^*, which is same as Singh *et al.* (2008 b) work.

The optimum fraction of the fresh sample to be drawn in this case is obtained as

$$\mu_{opt.} = \frac{-f_2 A \pm \sqrt{f_2 A(A+B)}}{f_2 B} = \mu_1^{**}$$

Hence, the minimum variance of the estimator τ^* is given as

$$V(\Delta)_{opt.^*} = \frac{A\left(A+\mu_1^{**} B\right)}{n\left[f_2 B\mu_1^{**2} + A(f_2 - 1)\mu_1^{**} + A\right]} S_y^2 \quad (56)$$

11. EFFICIENCY COMPARISON

To examine the loss in efficiency of the estimator Δ owing to non-response, the percent relative loss in efficiency of estimator Δ with respect to Singh and Priyanka (2008 a) estimator T_{sp} has been computed for different choices of ρ_{yz} and ρ_{yx}. The estimator T_{sp} is defined under the similar circumstances as the estimator Δ, but in the absence of non-response. The estimator T_{sp} is reproduced as

$$T_{sp} = \psi^* T_{1u} + \left(1 - \psi^*\right)T_{2m} \quad (57)$$

where $T_{1u} = \bar{y}_u + \beta_{yz}\left(\bar{Z} - \bar{z}_u\right)$ and

$$T_{2m} = \bar{y}_m^* + \beta_{yx}\left(\bar{x}_n^* - \bar{x}_m^*\right)$$

where
$$\bar{y}_m^* = \bar{y}_m + \beta_{yz}\left(\bar{Z} - \bar{z}_m\right)$$
$$\bar{x}_n^* = \bar{x}_n + \beta_{xz}\left(\bar{Z} - \bar{z}_n\right)$$
$$\bar{x}_m^* = \bar{x}_m + \beta_{xz}(\bar{Z} - \bar{z}_m)$$

and ψ^* is the unknown constant to be determined by the minimization of the variance of T_{sp}. Since T_{1u} and T_{2m} are unbiased estimators of $\bar{Y}$, the minimum (optimum) variance of T_{sp} is given by

$$V\left(T_{sp}\right)_{opt.^*} = \frac{A[A + \mu_2 B]}{n[A + \mu_2^2 B]} S_y^2 \quad (58)$$

μ_2 is the admissible optimum value of μ_{opt}.

where
$$\mu_{\text{opt.}} = \frac{-A \pm \sqrt{A(A+B)}}{B} = \mu_2 \text{ (say)} \quad (59)$$

Remark 11.1 : The admissible value of μ_{opt}. (say μ_2) in equation (59) is obtained in the similar manner as that of μ_1.

Remark 11.2 : The estimator T_{sp} is based on population regression coefficients, which may not be always known. In such situations the unknown population regression coefficients may be estimated by their corresponding sample estimates. It is interesting to see that up to the first order of approximations the mean square error of the estimator based on sample estimates of population regression coefficients is similar to the variance of T_{sp} viz. Singh and Priyanka (2008 a). Therefore, for the comparison purpose simply the estimator T_{sp} has been considered.

The percent relative loss in precision of the estimators Δ, τ and τ^* with respect to T_{sp} under their respective optimality conditions are given by

$$L_1 = \frac{V(\Delta)_{\text{opt.}^*} - V\left(T_{sp}\right)_{\text{opt.}^*}}{V(\Delta)_{\text{opt.}^*}} \times 100$$

$$L_1^* = \frac{V(\tau)_{\text{opt.}^*} - V\left(T_{sp}\right)_{\text{opt.}^*}}{V(\tau)_{\text{opt.}^*}} \times 100 \text{ and}$$

$$L_1^{**} = \frac{V\left(\tau^*\right)_{\text{opt.}^*} - V\left(T_{sp}\right)_{\text{opt.}^*}}{V\left(\tau^*\right)_{\text{opt.}^*}} \times 100$$

The expressions of the optimum μ (μ_1, μ_1^*, μ_1^{**}, and μ_2) and the percent relative losses are given in terms of the population correlation coefficients ρ_{yz} and ρ_{yx}. Therefore, the percent relative losses have been computed for different choices of correlations ρ_{yz} and ρ_{yx}. Percent relative losses in precision of the estimators Δ, τ and τ^* over T_{sp} have been calculated for different choices of t_1, t_2, ρ_{yz} and ρ_{yx}, and compiled in Tables 4 – 6.

Table 4. Percent Relative Loss (L_1) in precision of Δ over T_{sp} at optimum values of μ when non response occurs on both the occasions and for positive values of ρ_{yx} and ρ_{yz}

ρ_{yz} →			0.5			0.7			0.9		
t_1	t_2↓	ρ_{yx}↓	μ_1	μ_2	L_1	μ_1	μ_2	L_1	μ_1	μ_2	L_1
0.05	0.05	0.3	*	-	-	0.58	0.47	2.52	0.43	0.38	1.17
		0.5	0.25	0.51	2.04	0.62	0.47	2.67	0.42	0.37	1.04
		0.7	0.52	0.55	2.96	0.26	0.51	2.07	0.46	0.40	1.49
		0.9	0.67	0.67	3.78	0.62	0.62	3.52	*	-	-
	0.1	0.3	*	-	-	0.68	0.47	5.90	0.46	0.38	4.12
		0.5	*	-	-	0.75	0.47	6.27	0.45	0.37	3.99
		0.7	0.46	0.55	5.17	*	-	-	0.49	0.40	4.45
		0.9	0.65	0.67	6.25	0.59	0.62	5.95	*	-	-
	0.15	0.3	*	-	-	0.79	0.47	9.83	0.49	0.38	7.25
		0.5	*	-	-	0.89	0.47	10.58	0.47	0.37	7.10
		0.7	0.39	0.55	7.08	*	-	-	0.52	0.40	7.63
		0.9	0.63	0.67	8.65	0.57	0.62	8.28	*	-	-

ρ_{yz} →			0.5			0.7			0.9		
t_1	t_2↓	ρ_{yx}↓	μ_1	μ_2	L_1	μ_1	μ_2	L_1	μ_1	μ_2	L_1
0.10	0.05	0.3	*	-	-	0.60	0.47	2.28	0.46	0.38	-0.23
		0.5	0.29	0.51	2.30	0.64	0.47	2.52	0.45	0.37	-0.49
		0.7	0.54	0.55	3.55	0.30	0.51	2.34	0.48	0.40	0.38
		0.9	0.68	0.67	5.10	0.64	0.62	4.59	0.02	0.51	1.54
	0.10	0.3	*	-	-	0.70	0.47	5.77	0.49	0.38	2.91
		0.5	0.005	0.51	3.03	0.76	0.47	6.29	0.48	0.37	2.65
		0.7	0.49	0.55	5.88	0.02	0.51	3.14	0.52	0.40	3.52
		0.9	0.67	0.67	7.63	0.62	0.62	7.11	*	-	-
	0.15	0.3	*	-	-	0.80	0.47	9.78	0.51	0.38	6.22
		0.5	*	-	-	0.89	0.47	10.57	0.50	0.37	5.95
		0.7	0.42	0.55	7.93	*	-	-	0.55	0.40	6.86
		0.9	0.65	0.67	10.10	0.59	0.62	9.52	*	-	-
0.15	0.05	0.3	*	-	-	0.63	0.47	2.05	0.49	0.38	-1.64
		0.5	0.33	0.51	2.57	0.66	0.47	2.37	0.48	0.37	-2.03
		0.7	0.57	0.55	4.14	0.34	0.51	2.60	0.51	0.40	-0.73
		0.9	0.70	0.67	6.41	0.66	0.62	5.67	0.07	0.51	1.79
	0.10	0.3	*	-	-	0.71	0.47	5.64	0.52	0.38	1.69
		0.5	0.05	0.51	3.52	0.77	0.47	6.15	0.51	0.37	1.31
		0.7	0.51	0.55	6.59	0.08	0.51	3.63	0.54	0.40	2.60
		0.9	0.69	0.67	9.01	0.64	0.62	8.26	*	-	-
	0.15	0.3	*	-	-	0.81	0.47	9.72	0.54	0.38	5.19
		0.5	*	-	-	0.90	0.47	10.56	0.53	0.37	4.80
		0.7	0.45	0.55	8.78	*	-	-	0.57	0.40	6.10
		0.9	0.67	0.67	11.55	0.61	0.62	10.77	*	-	-

Note : "*" indicate that μ_1 do not exist.

From Table 4 it is visible that for fixed values of t_2, ρ_{yz} and ρ_{yx}, μ_1 and L_1 increase with the increase in the values of t_1. For fixed values of t_1, ρ_{yz} and ρ_{yx}, μ_1 decreases while L_1 increases with the increase in the values of t_2. This indicates that the higher the non-response rate, the smaller the fresh sample on the current occasion is desired. For fixed values of t_2 , t_1 and ρ_{yx}, the values of μ_1 and L_1 decrease with the increase in the values of ρ_{yz}. This indicates that higher the correlation between the study and auxiliary variable, lower the amount of fresh sample required on the current occasion and the amount of loss in precision also decreases. When the non-response occurs either on the first occasion or on the current occasion similar type of patterns may be seen from the Tables 5-6.

12. CONCLUSION

The losses in precision of the proposed estimators were observed owing to the presence of non-response on both the occasions in two-occasion successive sampling. The imputation methods considered in the present work may be quite acceptable as the losses in precision are not much appreciable due to non-response for positive correlations (i) between the study variables over two occasions (i.e. ρ_{yx}) and (ii) between the study and auxiliary variables (i.e. ρ_{yz}). Since the non-response is a common phenomenon in sample surveys, the application of the proposed imputation methods may be justifiable for controlling the negative effects of non-response on the precision of estimates especially in the positively correlated situations.

Table 5. Percent Relative Loss (L_1^*) in precision of τ over T_{sp} at optimum values of μ when non-response occurs on the first occasion and for positive values of ρ_{yx} and ρ_{yz}

ρ_{yz} →		0.5			0.7			0.9		
t_1	ρ_{yx}↓	μ_1^*	μ_2	L_1^*	μ_1^*	μ_2	L_1^*	μ_1^*	μ_2	L_1^*
0.05	0.3	0.5192	0.4939	-0.0620	0.4927	0.4660	-0.3650	0.4094	0.3783	-0.378
	0.5	0.5353	0.5109	0.1064	0.5004	0.4741	-0.2730	0.4022	0.3707	-1.744
	0.7	0.5760	0.5536	0.4844	0.5357	0.5113	0.1105	0.4275	0.3973	-1.292
	0.9	0.6833	0.6667	1.2500	0.6428	0.6240	0.9939	0.5302	0.5054	0.054
0.10	0.3	0.5445	0.4939	-0.1230	0.5194	0.4660	-0.7300	0.4405	0.3783	-3.217
	0.5	0.5598	0.5109	0.2129	0.5267	0.4741	-0.5460	0.4336	0.3707	-3.487
	0.7	0.5983	0.5536	0.9689	0.5602	0.5113	0.2211	0.4576	0.3973	-2.583
	0.9	0.7000	0.6667	2.5000	0.6616	0.6240	1.9878	0.5549	0.5054	0.107
0.15	0.3	0.5698	0.4939	-0.1850	0.5461	0.4660	-1.0960	0.4716	0.3783	-4.825
	0.5	0.5842	0.5109	0.3193	0.5530	0.4741	-0.8190	0.4651	0.3707	-5.231
	0.7	0.6206	0.5536	1.4533	0.5846	0.5113	0.3316	0.4877	0.3973	-3.876
	0.9	0.7167	0.6667	3.7500	0.6804	0.6240	2.9817	0.5796	0.5054	0.161
0.20	0.3	0.5951	0.4939	-0.2470	0.5728	0.4660	-1.4610	0.5026	0.3783	-6.433
	0.5	0.6087	0.5109	0.4257	0.5793	0.4741	-1.0920	0.4966	0.3707	-12.142
	0.7	0.6429	0.5536	1.9377	0.6090	0.5113	0.4421	0.5179	0.3973	-5.167
	0.9	0.7333	0.6667	5.0000	0.6992	0.6240	3.9755	0.6044	0.5054	0.215

Table 6. Percent Relative Loss (L_1^{**}) in precision of τ^* over T_{sp} at optimum values of μ when non-response occurs only on current occasion and for positive values of ρ_{yx} and ρ_{yz}

ρ_{yz} →		0.5			0.7			0.9		
t_2↓	ρ_{yx}↓	μ_1^{**}	μ_2	L_1^{**}	μ_1^{**}	μ_2	L_1^{**}	μ_1^{**}	μ_2	L_1^{**}
0.05	0.3	*	-	-	0.5609	0.4660	2.7515	0.4034	0.3783	2.582
	0.5	0.2124	0.5109	1.7791	0.5992	0.4741	2.8255	0.3941	0.3707	2.578
	0.7	0.4938	0.5536	2.3666	0.2242	0.5113	1.8070	0.4276	0.3973	2.594
	0.9	0.6493	0.6667	2.4679	0.5995	0.6240	2.4514	*	-	-
0.10	0.3	*	-	-	0.6637	0.4660	6.0329	0.4306	0.3783	5.336
	0.5	*	-	-	0.7345	0.4741	6.3371	0.4195	0.3707	5.320
	0.7	0.4290	0.5536	4.4522	*	-	-	0.4604	0.3973	5.386
	0.9	0.6306	0.6667	4.8683	0.5729	0.6240	4.8005	*	-	-
0.15	0.3	*	-	-	0.7754	0.4660	9.8892	0.4601	0.3783	8.278
	0.5	*	-	-	0.8817	0.4741	10.5930	0.4471	0.3707	8.241
	0.7	0.3586	0.5536	6.2329	*	-	-	0.4961	0.3973	8.394
	0.9	0.6102	0.6667	7.1954	0.5440	0.6240	7.0385	*	-	-
0.20	0.3	*	-	-	0.8974	0.4660	14.372	0.4924	0.3783	11.420
	0.5	*	-	-	*	-	-	0.4772	0.3707	11.350
	0.7	0.2817	0.5536	7.6814	*	-	-	0.5350	0.3973	11.640
	0.9	0.5880	0.6667	9.4427	0.5124	0.6240	9.1555	*	-	-

Note : * indicate that μ_1^{**} do not exist.

ACKNOWLEDGEMENT

Authors are thankful to the referee for his valuable and inspiring suggestions. Authors are also obliged to the Indian School of Mines, Dhanbad for providing the financial assistance to carry out the present work.

REFERENCES

Ahmed, M.S., Al-Titi, O., Al-Rawi, Z. and Abu-Dayyeh, W. (2006). Estimation of a population mean using different imputation methods. *Statistics in Transition,* **7(6)**, 1247-1264.

Biradar, R.S. and Singh, H.P. (2001). Successive sampling using auxiliary information on both occasions. *Cal. Stat. Assoc. Bull*., **51,** 243-251.

Choudhary, R.K., Bathla, H.V.L. and Sud, U.C. (2004). On non-response in sampling on two occasions. *J. Ind. Soc. Agril. Statist.*, **58(3)**, 331-343.

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

Das, A.K. (1982). Estimation of population ratio on two occasions. *J. Ind. Soc. Agril. Statist.,* **34,** 1-9.

Eckler, A.R. (1955). Rotation sampling. *Ann. Math. Stat.*, **26,** 664-685.

Feng, S. and Zou, G. (1997). Sample rotation method with auxiliary variable. *Commun. Stat. Theo-Meth.,* **26(6)**, 1497-1509.

Gupta, P.C. (1979). Sampling on two successive occasions. *Jour. Stat.. Res.,* **13**, 7-16.

Jessen, R.J. (1942). Statistical investigation of a sample survey for obtaining farm facts. *Ames, USA. in Iowa Agricultural Experiment Station Road Bulletin No.* **30**, 1-104.

Kalton, G., Kasprzyk, D. and Santos, R. (1981). Issues of non-response and imputation of income and program participation. Current Topics in Survey Sampling (D. Krevoski, R. Platek and J.N.K. Rao eds.), *Acad. Press, New York,* 455-480.

Kalton, G. and Kasprzyk, D. (1982). Imputing for missing survey responses. *Proceedings of section Survey Research Method, American Stat. Assoc.,* 22-31.

Lee, H., Rancourt, E. and Särndal, C. E. (1994). Experiments with variance estimation from survey data with imputed values. *Jour. Official Stat.,* **10(3)**, 231-243.

Lee, H., Rancourt, E. and Särndal, C.E. (1995). Variance estimation in the presence of imputed data for the generalized estimation system. *Proc. of the American Stat. Assoc. (Social Survey Research Methods Section),* 384-389.

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

Patterson, H.D. (1950). Sampling on successive occasions with partial replacement of units. *Jour. Royal Stat. Assoc., Ser. B*, **12,** 241-255.

Rao, J.N.K. and Graham, J.E. (1964). Rotation design for sampling on repeated occasions. *Jour. American Stat. Assoc.,* **59,** 492-509.

Sande, I.G. (1979). A personal view of hot-deck imputation procedures. *Survey Methodology, Statistics Canada*, 238-247.

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

Sen, A.R. (1971). Successive sampling with two auxiliary variables. *Sankhya,* **B33,** 371-378.

Sen, A.R. (1972). Successive sampling with p ($p \geq 1$) auxiliary variables. *Ann. Math. Stat.,* **43,** 2031-2034.

Sen, A.R. (1973). Theory and application of sampling on repeated occasions with several auxiliary variables. *Biometrics,* **29,** 381-385.

Singh, G.N. and Singh, V.K. (2001). On the use of auxiliary information in successive sampling. *Jour. Ind. Soc. Agri. Stat.*, **54(1)**, 1-12.

Singh, G.N. (2003a). Estimation of population mean using auxiliary information on recent occasion in h-occasion successive sampling. *Statistics in Transition,* **6**, 523-532.

Singh, G.N. (2005). On the use of chain-type ratio estimator in successive sampling. *Statistics in Transition,* **7**, 21-26.

Singh, G.N. and Priyanka, K. (2006). On the use of chain-type ratio to difference estimator in successive sampling. *Int. Jour. App. Math. Stat.,* **5**, S06, 41-49.

Singh, G.N. and Priyanka, K. (2007a). On the use of auxiliary information in search of good rotation patterns on successive occasions. *Bulletin of Statistics and Economics,* **1**, A07, 42-60.

Singh, G.N. and Priyanka, K. (2007 b). Effect of non-response on current occasion in search of good rotation patterns on successive occasions. *Statistics in Transition- new series,* **8(2)**, 273-292.

Singh, G.N. and Priyanka, K. (2008 a). Search of good rotation patterns to improve the precision of estimates at current occasion. *Commun. Stat. Theo-Meth.,* **37(3)**, 337-348.

Singh, G.N., Priyanka, K. and Kozak, M. (2008 b). Use of imputation methods at current occasion in two-occasion rotation patterns. *Model Assisted Statistics and Applications*, **3(2)**, 99-112.

Singh, R. and Singh, N. (1991). Imputation methods in two-dimensional surveys. In: *Recent Advances in Agricultural Statistics Research*, Wiley Eastern Ltd., New Delhi.

Singh, S. and Horn, S. (2000). Compromised imputation in survey sampling. *Metrika,* **51,** 267-276.

Singh, S. (2003b). *Advanced Sampling Theory with applications*. Kluwer Academic Publishers, London.

Singh, S. (2009). A new method of imputation in survey sampling, *Statistics*, **43(5)**, 499-511.

Singh, V.K., Singh, G.N. and Shukla, D. (1991). An efficient family of ratio-cum-difference type estimators in successive sampling over two occasions. *Jour. Sci. Res.,* **41C**, 149-159.

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