

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452

Maths 2025; 10(3): 01-08

© 2025 Stats & Maths

<https://www.mathsjournal.com>

Received: 01-12-2024

Accepted: 03-01-2025

Shubhangi ChaurasiaSMS Govt. Model Science
College, Gwalior, Madhya
Pradesh, India**RK Shrivastava**Principal, Dr. Bhagwat Sahay
Govt. College, Gwalior, Madhya
Pradesh, India**Narendra Singh Thakur**Govt. Adarsh Girls College,
Sheopur, Madhya Pradesh, India

Estimation of mean with imputation of missing data using factor-type estimator under adaptive cluster sampling in sample surveys

Shubhangi Chaurasia, RK Shrivastava and Narendra Singh ThakurDOI: <https://www.doi.org/10.22271/math.2025.v10.i3a.1996>

Abstract

Sample surveys commonly have missing data, which calls for the use of statistical techniques to solve the issue. Evaluation of the missing data can be done using the data collected from the population and sample. This process is commonly referred to as the imputation technique. This publication introduces Adaptive Cluster Sampling (ACS) estimation method in the presence of missing observations of study variable in the sample. The Bias, M.S.E. and optimal M.S.E. of the proposed estimator are determined up to first order using the concept of large sample approximations.

Keywords: Imputation, adaptive cluster sampling, factor-type (F-T) estimator, bias (B.), mean squared error (M.S.E.)

1. Introduction

Adaptive cluster sampling is a sampling approach where all the units that satisfy a given criterion are included in the sample coupled with a selected sample unit. Yadav et al. (2016) [20] provides an explanation of adaptive cluster sampling (ACS). For example in epidemic time, all those units were included in the sample, which are in relation with the Corona patient for last 14 days. In the event that a Corona patient is discovered during testing, everyone who has had contact with the patient during the previous 14 days is also examined.

Let $\Omega = \{1, 2, 3, \dots, N\}$ be a population and the ordered pair $(Y_i, X_i); i = 1, 2, 3, \dots, N$ be the observations of i^{th} individual. Assume the population mean $\bar{Y} = \frac{1}{N} \sum_{i=1}^N Y_i$ is the mean of the variable under study or estimation and $\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i$ is the population mean of auxiliary variable and known in advance by past data or previous experience. Let a preliminary sample of size n is drawn by SRSWOR from Ω . Also, let $A_i; i = 1, 2, 3, \dots, n$ be the network of size m_i of unit i , that is selected in the preliminary sample. Let the means of Y and X in the i^{th} network are $\omega_{yi} = \frac{1}{m_i} \sum_{j \in A_i} y_j$ and $\omega_{xi} = \frac{1}{m_i} \sum_{j \in A_i} x_j$ respectively.

Adaptive cluster sampling may be considered as SRSWOR, if the average of A_i is considered as sample unit. Then the unbiased estimate of population mean of study and auxiliary variable is defined as $\bar{\omega}_y = \frac{1}{n} \sum_{i=1}^n \omega_{yi}$ and $\bar{\omega}_x = \frac{1}{n} \sum_{i=1}^n \omega_{xi}$ respectively. Also, $s_{wy}^2 = \frac{1}{n-1} \sum_{i=1}^n (\omega_{yi} - \bar{\omega}_y)^2$, $s_{wx}^2 = \frac{1}{n-1} \sum_{i=1}^n (\omega_{xi} - \bar{\omega}_x)^2$ are the unbiased estimator of $S_{wy}^2 = \frac{1}{N-1} \sum_{i=1}^N (\omega_{yi} - \bar{Y})^2$, $S_{wx}^2 = \frac{1}{N-1} \sum_{i=1}^N (\omega_{xi} - \bar{X})^2$ respectively.

The unbiased estimate of population mean is $\bar{Y}_{ac} = \bar{\omega}_y = \frac{1}{n} \sum_{i=1}^n \omega_{yi}$ (1.1)

2. Review of Literature

The authors Pochai (2008) [8], Saheli (2003) [10], Turk and Borkowski (2005) [19], Singh and Yadav (2019) [14], Yadav et al. (2016) [20], Thompson (1990) [18] and others have all done

Corresponding Author:**Shubhangi Chaurasia**SMS Govt. Model Science
College, Gwalior, Madhya
Pradesh, India

work based on “Adaptive cluster sampling to estimate the population mean”. “Auxiliary variable-based estimators of population mean under adaptive cluster sampling” have been proposed by Chao (2004) ^[2], Dryver and Chao (2007) ^[6], Chutiman and Kumphon (2008) ^[5], Chutiman (2013) ^[4] and Chaudhury and Hanif (2015) ^[3]. All of the techniques described in the literature rely on data from the sample, but they are ineffective if certain observations of the primary study variable are unintentionally or for other reasons absent. In this publication, a methodology for estimating missing data and population mean estimates was established.

In this earlier discussion of the concept of missing observation, Rubin (1976) ^[9] presented three examples of missing data patterns. The first is the missing at random (MR), the second is the observed at random (OR) and the third is parametric distribution (PD). The missing completely at random (MCR) distribution pattern, first described by Heitjan and Basu (1996) ^[7] is examined in this work. Thakur et al. (2011) ^[15] used two-phase sampling to produce missing data. The theory of modern missing data analyses is proposed by Baraldi and Enders (2010) ^[1]. Further recommendations for “Missing data in stratified random sampling” were made by Thakur, Yadav, and Pathak (2014) ^[17].

3. Proposed Imputation Methods

Shukla and Thakur (2008) ^[11] used an F-T (factor-type) estimator as a tool of imputation and then derived an estimator for population mean under SRSWOR. Shukla, et al. (2009) ^[12] - discussed factor-type estimator in two-phase sampling under imputation of missing data for estimating population mean. Some other contributions of “Factor-type estimators to estimate the population mean” are due to Thakur et al. (2016) ^[16], Thakur and Shukla (2022) ^[13] etc. the following are some suggested imputation techniques for missing data under adaptive cluster sampling based on the inspiration of these papers:

$$a) \quad (w_{yFTC1})_i = \begin{cases} w_{yi} & \text{if } i \in R \\ \frac{\bar{w}_{yr}}{(n-r)} [n\phi_1(k) - r] & \text{if } i \in R^c \end{cases} \quad (3.1)$$

$$b) \quad (w_{yFTC2})_i = \begin{cases} w_{yi} & \text{if } i \in R \\ \frac{\bar{w}_{yr}}{(n-r)} [n\phi_2(k) - r] & \text{if } i \in R^c \end{cases} \quad (3.2)$$

$$c) \quad (w_{yFTC3})_i = \begin{cases} w_{yi} & \text{if } i \in R \\ \frac{\bar{w}_{yr}}{(n-r)} [n\phi_3(k) - r] & \text{if } i \in R^c \end{cases} \quad (3.3)$$

$$\text{where } \phi_1(k) = \frac{(A+c)\bar{X}+fB\bar{w}_{xn}}{(A+fB)\bar{X}+C\bar{w}_{xn}}; \phi_2(k) = \frac{(A+c)\bar{w}_{xn}+fB\bar{w}_{xr}}{(A+fB)\bar{w}_{xn}+C\bar{w}_{xr}}; \phi_3(k) = \frac{(A+c)\bar{X}+fB\bar{w}_{xr}}{(A+fB)\bar{X}+C\bar{w}_{xr}}$$

$$A = (k-1)(k-2); B = (k-1)(k-4); C = (k-2)(k-3)(k-4);$$

$$f = \frac{n}{N} \text{ and } 0 < k < \infty \text{ is a constant.}$$

According to sections 3.1, 3.2 and 3.3 the point estimators for $\bar{Y}$ are:

$$T_{FTC1} = \bar{w}_{yr}\phi_1(k), \quad T_{FTC2} = \bar{w}_{yr}\phi_2(k), \quad T_{FTC3} = \bar{w}_{yr}\phi_3(k)$$

4. Properties of Proposed Imputation Methods in Adaptive Cluster Sampling

Let $\varepsilon = \frac{\bar{w}_{yr}}{\bar{Y}} - 1$, $\delta = \frac{\bar{w}_{xr}}{\bar{X}} - 1$ and $\eta = \frac{\bar{w}_{xn}}{\bar{X}} - 1$, for the specified $\bar{w}_{yr}$, $\bar{w}_{xr}$ and $\bar{w}_{xn}$ respectively. Then,

$$E(\varepsilon) = E(\delta) = E(\eta) = 0$$

$$E(\varepsilon^2) = M_1 C_{wY}^2, E(\delta^2) = M_1 C_{wX}^2, E(\eta^2) = M_2 C_{wX}^2;$$

$$E(\varepsilon\delta) = M_1 \rho_{wYX} C_{wY} C_{wX}, E(\varepsilon\eta) = M_2 \rho_{wYX} C_{wY} C_{wX}, E(\delta\eta) = M_2 C_{wX}^2 \text{ and}$$

$$E(\varepsilon^i \delta^j) = E(\varepsilon^i \eta^j) = E(\delta^i \eta^j) = 0, \text{ where } i+j > 2; \quad i, j = 0, 1, 2, 3, 4, \dots$$

Remark 4.

$$\text{Define, } \theta_1 = \frac{fB}{A+fB+C}; \theta_2 = \frac{C}{A+fB+C}; \theta_3 = \frac{A+C}{A+fB+C} \text{ and } \theta_4 = \frac{A+fB}{A+fB+C}$$

where,

$$\theta = (\theta_1 - \theta_2) = -(\theta_3 - \theta_4); \quad V = \rho_{wYX} \frac{C_{wY}}{C_{wX}}, M_1 = \frac{1}{r}, M_2 = \frac{1}{n} \text{ and}$$

$$M_3 = \left(\frac{1}{r} - \frac{1}{n}\right) \text{ as } \frac{1}{N} \rightarrow 0 \text{ (negligible) and } C_{wY}^2 = \frac{S_{wY}^2}{\bar{Y}^2}, C_{wX}^2 = \frac{S_{wX}^2}{\bar{X}^2} \text{ and } \rho = \frac{S_{wYX}}{S_{wX} S_{wY}}$$

Theorem 4.1

[1. 1]: The estimator T_{FTC1} with ε , δ and η up to first order of approximation, could be written as:

$$T_{FTC1} = \bar{Y}[1 + \varepsilon + \theta(\eta + \varepsilon\eta - \theta_2\eta^2)] \quad (4.1)$$

Proof: $T_{FTC1} = \bar{w}_{yr}\phi_1(k)$

$$= \bar{Y} (1 + \varepsilon) \left[\frac{(A+C)\bar{X} + fB\bar{X}(1+\eta)}{(A+fB)\bar{X} + C\bar{X}(1+\eta)} \right]$$

Use of large sample approximation with section 4.0

$$\begin{aligned} &= \bar{Y} (1 + \varepsilon) (1 + \theta_1\eta)(1 + \theta_2\eta)^{-1} \\ &= \bar{Y} (1 + \varepsilon) (1 + \theta_1\eta) (1 - \theta_2\eta + \theta_2^2\eta^2 - \dots \dots \dots) \\ &= \bar{Y}[1 + \varepsilon + \theta(\eta + \varepsilon\eta - \theta_2\eta^2)] \end{aligned}$$

[1. 2]: Bias of T_{FTC1} up to first order approximation is:

$$B(T_{FTC1}) = -\theta M_2 \bar{Y} [\theta_2 C_{wX}^2 - \rho_{wYX} C_{wY} C_{wX}] \quad (4.2)$$

Proof: $B(T_{FTC1}) = E(T_{FTC1} - \bar{Y})$

$$\begin{aligned} &= E[\bar{Y}\{1 + \varepsilon + \theta(\eta + \varepsilon\eta - \theta_2\eta^2)\} - \bar{Y}] \\ &= -\theta M_2 \bar{Y} [\theta_2 C_{wX}^2 - \rho_{wYX} C_{wY} C_{wX}] \end{aligned}$$

[1. 3]: Mean squared error of T_{FTC1} up to first order is:

$$M(T_{FTC1}) = \bar{Y}^2 [M_1 C_{wY}^2 + M_2 \theta^2 C_{wX}^2 + 2M_2 \theta \rho_{wYX} C_{wY} C_{wX}] \quad (4.3)$$

Proof: $M(T_{FTC1}) = E(T_{FTC1} - \bar{Y})^2$

$$\begin{aligned} &= E[\bar{Y}\{1 + \varepsilon + \theta(\eta + \varepsilon\eta - \theta_2\eta^2)\} - \bar{Y}]^2 \\ &= \bar{Y}^2 [M_1 C_{wY}^2 + M_2 \theta^2 C_{wX}^2 + 2M_2 \theta \rho_{wYX} C_{wY} C_{wX}] \end{aligned}$$

[1. 4]: The minimum m.s.e. of (T_{FTC1}) when $\theta = -V$ and expression is:

$$M(T_{FTC1})_{min.} = (M_1 - M_2 \rho_{wYX}^2) S_{wY}^2 \quad (4.4)$$

Proof: Minimum m.s.e. occurs when

$$\begin{aligned} \frac{d}{d\theta} M(T_{FTC1}) &= 0 \\ \Rightarrow 2\theta C_{wX}^2 + 2\rho_{wYX} C_{wY} C_{wX} &= 0 \text{ at } \theta = -V \\ \Rightarrow M(T_{FTC1})_{min.} &= (M_1 - M_2 \rho_{wYX}^2) S_{wY}^2 \end{aligned}$$

Theorem 4.2

[1. 1]: The estimator T_{FTC2} with ε , δ and η up to first order of approximation, could be written as:

$$T_{FTC2} = \bar{Y}[(1 + \varepsilon) + \theta\{\delta - \eta + \varepsilon\delta - \varepsilon\eta + (\theta_2 - \theta_4)\delta\eta - \delta^2\theta_2 + \eta^2\theta_4\}] \quad (4.5)$$

Proof: $T_{FTC2} = \bar{w}_{yr}\phi_2(k)$

$$= \bar{Y} (1 + \varepsilon) \left[\frac{(A+C)(1+\eta)\bar{X} + fB\bar{X}(1+\delta)}{(A+fB)(1+\eta)\bar{X} + C\bar{X}(1+\delta)} \right]$$

$$\begin{aligned} &= \bar{Y}(1 + \varepsilon)(1 + \theta_1\delta + \theta_3\eta)(1 + \theta_2\delta + \theta_4\eta)^{-1} \\ &= \bar{Y}[(1 + \varepsilon) + \theta\{\delta - \eta + \varepsilon\delta - \varepsilon\eta + (\theta_2 - \theta_4)\delta\eta - \delta^2\theta_2 + \eta^2\theta_4\}] \end{aligned}$$

[1. 2]: Bias of T_{FTC2} up to first order approximation is:

$$B(T_{FTC2}) = -\theta \bar{Y} M_3 [\theta_2 C_{wX}^2 - \rho_{wYX} C_{wY} C_{wX}] \quad (4.6)$$

Proof: $B(T_{FTC2}) = E(T_{FTC2} - \bar{Y})$

$$\begin{aligned} &= E[\bar{Y}\{1 + \epsilon + \theta\{\delta - \eta + \epsilon\delta - \epsilon\eta + (\theta_2 - \theta_4)\delta\eta - \delta^2\theta_2 + \eta^2\theta_4\}\} - \bar{Y}] \\ &= -\theta M_3 \bar{Y} [\theta_2 C_{wX}^2 - \rho_{wYX} C_{wY} C_{wX}] \end{aligned}$$

[1. 3]: Mean squared error of T_{FTC2} up to first order is:

$$M(T_{FTC2}) = \bar{Y}^2 [M_1 C_{wY}^2 + \theta M_3 \{\theta C_{wX}^2 + 2 \rho_{wYX} C_{wY} C_{wX}\}] \quad (4.7)$$

Proof: $M(T_{FTC2}) = E(T_{FTC2} - \bar{Y})^2$

$$\begin{aligned} &= E[\bar{Y}\{1 + \epsilon + \theta\{\delta - \eta + \epsilon\delta - \epsilon\eta + (\theta_2 - \theta_4)\delta\eta - \delta^2\theta_2 + \eta^2\theta_4\}\} - \bar{Y}]^2 \\ &= \bar{Y}^2 [M_1 C_{wY}^2 + \theta M_3 \{\theta C_{wX}^2 + 2 \rho_{wYX} C_{wY} C_{wX}\}] \end{aligned}$$

[1. 4]: The minimum m.s.e. of (T_{FTC2}) at $\theta = -V$ and expression is:

$$M(T_{FTC2})_{min.} = (M_1 - M_3 \rho_{wYX}^2) S_{wY}^2 \quad (4.8)$$

Proof: Minimum m.s.e., when

$$\begin{aligned} \frac{d}{d\theta} M(T_{FTC2}) &= 0 \\ \Rightarrow 2\theta C_{wX}^2 + 2 \rho_{wYX} C_{wY} C_{wX} &= 0 \text{ at } \theta = -V \\ \Rightarrow M(T_{FTC2})_{min.} &= (M_1 - M_3 \rho_{wYX}^2) S_{wY}^2 \end{aligned}$$

Theorem 4.3

[1. 1]: The estimator T_{FTC3} with ϵ , δ and η up to first order of approximation, could be written as:

$$T_{FTC3} = \bar{Y} [1 + \epsilon + \theta(\delta + \epsilon\delta - \delta^2\theta_2)] \quad (4.9)$$

Proof: $T_{FTC3} = \bar{w}_{yr} \phi_3(k)$

$$\begin{aligned} &= \bar{Y} (1 + \epsilon) \left[\frac{(A+C)\bar{X} + fB\bar{X}(1+\delta)}{(A+fB)\bar{X} + C\bar{X}(1+\delta)} \right] \\ &= \bar{Y} (1 + \epsilon) (1 + \theta_1\eta)(1 + \theta_2\eta)^{-1} \\ &= \bar{Y} [1 + \epsilon + \theta(\delta + \epsilon\delta - \delta^2\theta_2)] \end{aligned}$$

[1. 2]: Bias of T_{FTC3} up to first order approximation is:

$$B(T_{FTC3}) = -\theta M_1 \bar{Y} [\theta_2 C_{wX}^2 - \rho_{wYX} C_{wY} C_{wX}] \quad (4.10)$$

Proof: $B(T_{FTC3}) = E(T_{FTC3} - \bar{Y})$

$$\begin{aligned} &= E[\bar{Y}\{1 + \epsilon + \theta(\eta + \epsilon\eta - \theta_2\eta^2)\} - \bar{Y}] \\ &= -\theta M_1 \bar{Y} [\theta_2 C_{wX}^2 - \rho_{wYX} C_{wY} C_{wX}] \end{aligned}$$

[1. 3]: Mean squared error of T_{FTC3} up to first order is:

$$M(T_{FTC3}) = M_1 \bar{Y}^2 [C_{wY}^2 + \theta^2 C_{wX}^2 + 2 \theta \rho_{wYX} C_{wY} C_{wX}] \quad (4.11)$$

Proof: $M(T_{FTC3}) = E(T_{FTC3} - \bar{Y})^2$

$$= E[\bar{Y}\{1 + \epsilon + \theta(\eta + \epsilon\eta - \theta_2\eta^2)\} - \bar{Y}]^2$$

$$= M_1 \bar{Y}^2 [C_{wY}^2 + \theta^2 C_{wX}^2 + 2\theta \rho_{wYX} C_{wY} C_{wX}]$$

[1. 4]: The minimum m.s.e. of (T_{FTC3}) when $\theta = -V$ and expression is:

$$M(T_{FTC3})_{min.} = (1 - \rho_{wYX}^2) M_1 S_{wY}^2 \text{ at } \theta = -V \quad (4.12)$$

Proof: Minimum m.s.e. occurs when $\frac{d}{d\theta} M(T_{FTC3}) = 0$

$$M(T_{FTC3})_{min.} = (1 - \rho_{wYX}^2) M_1 S_{wY}^2 \text{ at } \theta = -V$$

5. Values of k with optimality condition

The optimality condition $\theta = -V$ provides the equation

$$AV + (V + 1)fB + (V - 1)C = 0 \quad (5.1)$$

This is a cubic equation in terms of k , and for which M.S.E. is optimal, we obtain a maximum of three k values, such as k_1 , k_2 and k_3 . The best criterion for determining k 's value is:

Step 1: For $i, j = 1, 2, 3$, compute $|B(T_{FTC})_{k_j}|$

Step 2: For given values of i , choose k_j as

$$|B(T_{FTC})_{k_j}| = \min_{j=1,2,3} [|B(T_{FTC})_{k_j}|]$$

In the end, this provides bias control at the ideal m.s.e.

6. Comparison

[1]: The minimum M.S.E. of $\bar{y}_{FTC1}$ is

$$M(T_{FTC1})_{min.} = (M_1 - M_2 \rho_{wYX}^2) S_{wY}^2$$

and the minimum M.S.E. of $\bar{y}_{FTC2}$ is

$$M(T_{FTC2})_{min.} = (M_1 - M_3 \rho_{wYX}^2) S_{wY}^2$$

Now, let $D_1 = [M(T_{FTC1})_{min.} - M(T_{FTC2})_{min.}]$

$$= \rho^2 S_{wY}^2 (M_3 - M_2) \quad (6.1)$$

So, $\bar{y}_{FTC2}$ is better than $\bar{y}_{FTC1}$ if $D_1 > 0$

$$\text{i.e. } \rho^2 S_{wY}^2 (M_3 - M_2) > 0$$

$$\Rightarrow \rho^2 S_{wY}^2 \left(\frac{1}{r} - \frac{2}{n} \right) > 0$$

$$\Rightarrow r < \frac{n}{2}$$

Hence $D_1 > 0$, therefore $\bar{y}_{FTC2}$ is better than $\bar{y}_{FTC1}$.

[2]: Let, $D_2 = [M(T_{FTC1})_{min.} - M(T_{FTC3})_{min.}] > 0$

$$= (M_1 - M_2 \rho_{wYX}^2) S_{wY}^2 - (1 - \rho_{wYX}^2) M_1 S_{wY}^2 > 0 \quad (6.2)$$

$$= M_3 \rho^2 S_{wY}^2 > 0$$

Since $D_2 > 0$, therefore $\bar{y}_{FTC3}$ is always better than $\bar{y}_{FTC1}$.

[3]: $D_3 = [M(T_{FTC2})_{min.} - M(T_{FTC3})_{min.}] > 0$

$$= (M_1 - M_3 \rho_{wYX}^2) S_{wY}^2 - (1 - \rho_{wYX}^2) M_1 S_{wY}^2 \quad (6.3)$$

$$= M_2 \rho^2 S_{wY}^2$$

Since $D_3 > 0$, therefore $\bar{y}_{FTc3}$ is always better than $\bar{y}_{FTc2}$.

7. Empirical Study

An artificial population in Table 7.1(a) and Table 7.1(b) generated the values of main variable and auxiliary variable with size $N = 400$. The summary regarding population is given below:

$$\mu_Y = 1.2275; \quad \mu_X = .56500; \quad S_{wY}^2 = 12.6791; \quad S_{wX}^2 = 3.79790$$

$$\rho_{wYX} = 0.80710; \quad C_{wX} = 3.44920; \quad C_{wY} = 2.9008; \quad K_w = \rho_{wYX} \frac{C_{wY}}{C_{wX}} = 0.67877$$

Given a random sample of size $n = 20$; $f = 0.05$ by SRSWOR calculated values of the constant k , the three degree equation in terms of k yields three k values by solving the optimum condition $\theta = -V$

$$k_1 = 7.3283849; \quad k_2 = 1.7091097; \quad k_3 = 2.3370733$$

Using 30,000 repeated samples selected by SRSWOR from population $N = 400$, Bias (B.) and M.S.E. (M.) of the proposed estimators are calculated. Tables 7.2, 7.3 and 7.4 provide the results of the calculations. The following steps are part of the simulation process:

1. By using SRSWOR, select a random sample of size $n = 20$ from the population of $N = 400$.
2. Dropout 2,4,6,8 units randomly from each sample corresponding to Y .
3. Using the suggested and existing methods, compute and impute the dropped units of y .
4. To obtain numerous sample-based estimations, repeat the previous steps 30,000 times.
5. The Bias of $\hat{y}_1$ is obtained by

$$B(\hat{y}) = \frac{1}{30,000} \sum_{i=1}^{30,000} [(\hat{y}_i) - \bar{Y}]$$

6. The M.S.E. of $\hat{y}_1$ is obtained by

$$M(\hat{y}) = \frac{1}{30,000} \sum_{i=1}^{30,000} [(\hat{y}_i) - \bar{Y}]^2$$

0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	2	24	5	0	0	0	0	0	0	0	0	0	0	0
0	0	1	22	5	4	5	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	2	0	4	8	0	0	0	0	0	33	0	0	0	27	0	0
0	0	0	0	0	0	0	0	0	0	0	0	7	6	7	1	0	5	0	0
0	0	0	0	0	0	0	21	0	0	1	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	4	0	5	7	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	5	7	0	7	7	6	3	0	0	0	0	0	0
0	0	0	5	4	3	0	5	8	4	5	1	0	5	0	0	0	0	0	0
0	7	65	0	4	5	0	9	0	0	0	0	0	3	1	0	0	0	0	0
0	1	4	5	0	7	3	3	0	0	0	0	0	0	0	0	0	0	0	0
0	1	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	27	0	0	21
0	0	0	0	0	0	0	0	0	0	0	0	0	29	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 7.1(a): Population of study variable (Y)

0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	1	11	3	0	0	0	0	0	0	0	0	0	0	0
0	0	0	11	2	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	2	4	0	0	0	0	0	12	0	0	0	15	0	0
0	0	0	0	0	0	0	0	0	0	0	0	3	2	3	0	0	2	0	0
0	0	0	0	0	0	0	16	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	2	0	2	3	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	2	2	0	3	3	2	1	0	0	0	0	0	0
0	0	0	2	2	1	0	2	3	2	2	0	0	2	0	0	0	0	0	0
0	3	18	0	2	2	0	4	0	0	0	0	0	1	0	0	0	0	0	0
0	0	2	2	0	3	1	1	0	0	0	0	0	0	0	0	0	0	0	0
0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	0	0	12
0	0	0	0	0	0	0	0	0	0	0	0	0	27	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 7.1(b): Population of auxiliary variable (X)

Estimator	Bias				Min M (.)			
	r=2	r=4	r=6	r=8	r=2	r=4	r=6	r=8
$(T_{FTC1})_{k_1}$	0.142428	-0.03776	-0.21925	-0.40034	0.032949	0.021061	0.07296	0.187201154
$(T_{FTC1})_{k_2}$	0.610169	0.367497	0.125334	-0.11653	0.405108	0.174662	0.064641	0.068279
$(T_{FTC1})_{k_3}$	-1.56612	-1.51794	-1.47796	-1.43713	2.89209	2.636008	2.418857	2.2209

Table 7.2: (Bias and M.S.E. of Proposed Imputed Estimator T_{FTC1})

Estimator	Bias				Min M (.)			
	r=2	r=4	r=6	r=8	r=2	r=4	r=6	r=8
$(T_{FTC2})_{k_1}$	7.389107	6.95252	6.430321	5.773983	58.7837	52.03236	44.9283	36.80497
$(T_{FTC2})_{k_2}$	7.394771	6.978003	6.486873	5.878859	58.86653	52.3883	45.6707	38.07254
$(T_{FTC2})_{k_3}$	7.374499	6.889691	6.296189	5.540064	58.57071	51.16066	43.19261	34.06973

Table 7.3: (Bias and M.S.E. of Proposed Imputed Estimator T_{FTC2})

Estimator	Bias				Min M (.)			
	r=2	r=4	r=6	r=8	r=2	r=4	r=6	r=8
$(T_{FTC3})_{k_1}$	0.28103	0.258219	0.225305	0.189001	0.097953	0.085576	0.069752	0.064273
$(T_{FTC3})_{k_2}$	0.731155	0.62623	0.515273	0.403244	0.557686	0.409248	0.285152	0.191899
$(T_{FTC3})_{k_3}$	-1.33298	-1.06788	-0.80854	-0.56601	2.3185	1.584344	0.987845	0.589158

Table 7.4: (Bias and M.S.E. of Proposed Imputed Estimator T_{FTC3})

8. Conclusion

In the present paper some new factor type imputation procedures are advocated in adaptive cluster sampling scheme and their theoretical results are obtained under large sample approximations. A simulation study perform over 30,000 samples for comparison purpose while considering missing observations as 2, 4, 6 and 8 respectively in the sample of size 20 and obtained the bias and m.s.e. of proposed estimators. From the table 7.2, 7.3 and 7.4 the estimator $(T_{FTC3})_{k_1}$ has minimum bias and mean squared error for all values of r . Hence we conclude that the estimator $(T_{FTC3})_{k_1}$ is best for imputation purpose. Again the property of this estimator in the multiple choices of k , so that we can select the optimum value of k , for which we obtain minimum m.s.e. with its bias.

References

1. Baraldi AN, Enders CK. An introduction to modern missing data analyses. Journal of School Psychology. 2010;48(1):5–37.
2. Chao CT. Ratio estimation on adaptive cluster sampling. Journal of Chinese Statistical Association. 2004; 42:307–327.
3. Chaudhary MS, Hanif M. Generalized exponential-cum-exponential estimator in adaptive cluster sampling. Pakistan Journal of Statistics and Operation Research. 2015;11(4):553–574.

4. Chutiman N. Adaptive cluster sampling using auxiliary variable. *Journal of Mathematics and Statistics*. 2013;9(3):249–255.
5. Chutiman N, Kumphon B. Ratio estimator using two auxiliary variables for adaptive cluster sampling. *Thailand Statistician*. 2008;6(2):241–256.
6. Dryver AL, Chao CT. Ratio estimators in adaptive cluster sampling. *Environmetrics*. 2007; 18:607–620.
7. Heitjan DF, Basu S. Distinguishing ‘missing at random’ and ‘missing completely at random’. *The American Statistician*. 1996; 50:207–213.
8. Pochai N. Ratio estimator using two auxiliary variables for adaptive cluster sampling. *Journal of Thailand Statistical Association*. 2008; 6:241–256.
9. Rubin DB. Inference and missing data. *Biometrika*. 1976; 63:581–593.
10. Salehi MM. Comparison between Hansen-Hurwitz and Horvitz-Thompson estimators for adaptive cluster sampling. *Environmental and Ecological Statistics*. 2003; 10:115–127.
11. Shukla D, Thakur NS. Estimation of mean with imputation of missing data using factor-type estimator. *Statistics in Transition*. 2008;9(1):33–48.
12. Shukla D, Thakur NS, Thakur DS. Utilization of non-response auxiliary population mean in imputation for missing observations. *Journal of Reliability and Statistical Studies*. 2009; 2:28–40.
13. Thakur NS, Shukla D. Missing data estimation based on the chaining technique in survey sampling. *Statistics in Transition*. 2022; 23(4):91–111. DOI:10.2478/stattrans-2022-0044.
14. Singh HP, Yadav A. A class of estimators for estimating the population mean and variance using auxiliary information under adaptive cluster sampling in sample surveys. *Bulletin of Pure and Applied Sciences-Mathematics and Statistics*. 2019; 38E (1):176–192.
15. Thakur NS, Yadav K, Pathak S. Estimation of mean in presence of missing data under two-phase sampling scheme. *Journal of Reliability and Statistical Studies*. 2011;4(2):93–104.
16. Thakur NS, Yadav K, Singh R. A family of factor-type estimators for estimation of population variance. *Journal of Reliability and Statistical Studies*. 2016;9(2):21–31.
17. Thakur NS, Yadav K, Pathak S. Estimation of mean with imputation of missing data in stratified random sampling. *Research & Reviews: A Journal of Statistics*. 2014;3(1):17–30.
18. Thompson SK. Adaptive cluster sampling. *Journal of the American Statistical Association*. 1990; 85:1050–1059. DOI:10.1080/01621459.1990.10474975.
19. Turk P, Borkowski JJ. A review of adaptive cluster sampling, 1990–2005. *Environmental and Ecological Statistics*. 2005; 12:55–94.
20. Yadav SK, Mishra S, Chutiman N. Improved ratio estimators of population mean in adaptive cluster sampling. *Journal of Statistics Applications and Probability Letters*. 2016; 3(1):1–6.