Communications for Statistical Applications and Methods
2017, Vol. 24, No. 3, 255–269

https://doi.org/10.5351/CSAM.2017.24.3.255
Print ISSN 2287-7843 / Online ISSN 2383-4757

Multivariate analysis of longitudinal surveys for population median

Kumari Priyanka[1,a], Richa Mittal[a]

[a]Department of Mathematics, Shivaji College, University of Delhi, India

Abstract

This article explores the analysis of longitudinal surveys in which same units are investigated on several occasions. Multivariate exponential ratio type estimator has been proposed for the estimation of the finite population median at the current occasion in two occasion longitudinal surveys. Information on several additional auxiliary variables, which are stable over time and readily available on both the occasions, has been utilized. Properties of the proposed multivariate estimator, including the optimum replacement strategy, are presented. The proposed multivariate estimator is compared with the sample median estimator when there is no matching from a previous occasion and with the exponential ratio type estimator in successive sampling when information is available on only one additional auxiliary variable. The merits of the proposed estimator are justified by empirical interpretations and validated by a simulation study with the help of some natural populations.

Keywords: longitudinal surveys, exponential ratio type estimators, population median, skewed distribution, successive sampling, bias, mean squared error, optimum replacement strategy, multi-auxiliary information

1. Introduction

Longitudinal surveys over an extensive time are increasingly important because a single time survey and analysis is inadequate to understand changes in the dynamics of economic and social process. Longitudinal surveys in which the sampling is done on successive occasions (over years or seasons or months) according to a specific rule, with partial replacement of units, are called successive (rotation) sampling. Some of examples in this area of study are in many countries, the monthly labour-force surveys conducted to estimate the number of employed and the rate of unemployment, monthly surveys in which the data on price of goods are collected to determine a consumer price index, and political opinion surveys conducted at regular intervals to measure voter preferences.

The problem of sampling on two successive occasions was first considered by Jessen (1942) that has been subsequently extended by Arnab and Okafor (1992), Eckler (1955), Gordon (1983), Narain (1953), Patterson (1950), Priyanka *et al.* (2015), Singh and Priyanka (2008a), Singh *et al.* (2013), and others. All the above efforts were devoted to the estimation of population mean or variance on two or more occasion successive sampling. The median can be used as a measure of central location when a distribution concerned with a longitudinal survey is skewed, when end-values are unknown, or when one requires reduced importance to be attached to outliers because they may be measurement errors.

A few researchers such as Martínez-Miranda *et al.* (2005), Rueda *et al.* (2008), Singh *et al.* (2007) have proposed estimators for the population median in successive sampling. Singh and Priyanka

[1] Corresponding author: Department of Mathematics, Shivaji College, University of Delhi, New Delhi 110027, India. E-mail: priyanka.ism@gmail.com

Published 31 May 2017 / journal homepage: http://csam.or.kr

(2008b) proposed an estimator to estimate population median in two-occasion successive sampling assuming that a guess value of the population median is known. In the above papers, related to the study of median, have assumed that the density functions appearing in the results are known. However, a population parameter is not generally known. Priyanka and Mittal (2014, 2016) subsequently proposed estimators for population median in successive sampling using information on the additional stable auxiliary variable available on both the occasions. In addition, they also estimated unknown density functions using the method of the generalized nearest neighbor density estimator related to the kernel estimator.

It is theoretically established that the linear regression estimator is more efficient than the ratio estimator except when the regression line of the study variable on the auxiliary variable passes through the neighborhood of the origin; in this case the efficiencies of these estimators are almost equal. Also, the ratio estimator does not perform as well as the linear regression estimator in many practical situations where the regression line does not pass through the neighborhood of the origin. The linear regression estimator always performs adequately; however, it includes a population parameter in its structure (which in case of its practical implementation) that has to be replaced by its sample estimate as the population regression coefficient is not always known. One may also consider using a difference type estimator that requires an extra constant obtained by minimization with extra cost and expertise. It has also been observed that an exponential ratio type estimator works efficiently for a low or moderate correlation of the study and auxiliary variable (Priyanka *et al.*, 2015).

Sometimes, information on several auxiliary variables may be readily available or may be made easily available by diverting a small amount of available funding for the survey. For example, to study the number (or rate) of abortions, many factors like availability of medical facilities, household income, and education level can be taken as additional auxiliary information. Likewise, one may be interested in estimating military expenditures for Asian countries, then the gross national product, average exports, and average imports may be considered as additional auxiliary information for each country.

The exponential ratio type estimator can now be considered a better alternative. Therefore inspired with advantageous property of exponential ratio type estimators and following Olkin (1958) technique of weighted ratio-type estimator, the objective of the present study is to develop a more effective and relevant estimator using exponential ratio type estimators for the population median on the current occasion in two occasion successive sampling embedding information on p-additional auxiliary variates ($p \geq 1$), which are stable over time. The properties of the proposed estimator are discussed. Optimum replacement strategies are elaborated. The proposed estimator is compared with the estimator when information on single auxiliary variable ($p = 1$) is available on both occasions and with the sample median estimator when there is no matching from the previous occasion. The dominance of the proposed estimator is justified by empirical interpretations. The results are also validated by the means of simulation studies.

2. Sample structure and notations

Let $U = (U_1, U_2, \ldots, U_N)$ be the finite population of N units, which has been sampled over two occasions. It is assumed that size of the population remains unchanged but values of units change over two occasions. Let the character under study be denoted by $x(y)$ on the first (second) occasions respectively. It is assumed that information on p-additional auxiliary variables $z_1, z_2, \ldots, z_p$, whose population median are known and stable over occasions, are readily available on both occasions and positively correlated to x and y respectively. A simple random sample (without replacement) of n

units is taken on the first occasion. A random subsample of $m = n\lambda$ units is retained for use on the second occasion. Now at the current occasion a simple random sample (without replacement) of $u = (n - m) = n\mu$ units is drawn fresh from the remaining $(N - n)$ units of the population so that the sample size on the second occasion is also n. Let the fractions of fresh and matched samples at the second (current) occasion be μ and λ ($\mu + \lambda = 1$), respectively, where $0 \leq \mu, \lambda \leq 1$. The following notations are considered for the further use:

- M_i: Population median of the variable i; $i \in \{x, y, z_1, z_2, \ldots, z_p\}$.
- $\hat{M}_i(u)$: Sample median of variable i; $i \in \{y, z_1, z_2, \ldots, z_p\}$ based on the sample size u.
- $\hat{M}_i(m)$: Sample median of variable i; $i \in \{x, y, z_1, z_2, \ldots, z_p\}$ based on the sample size m.
- $\hat{M}_i(n)$: Sample median of variable i; $i \in \{x, y, z_1, z_2, \ldots, z_p\}$ based on the sample size n.
- $f_i(M_i)$: The marginal densities of variable i; $i \in \{x, y, z_1, z_2, \ldots, z_p\}$.

3. Proposed estimator T

To estimate the population median $\hat{M}_y$ on the current occasion utilizing p-additional auxiliary information, which is stable over time and readily available on both successive occasions, a multivariate weighted estimator T_u based on sample of size $u = n\mu$ drawn fresh on the current occasion is proposed as

$$T_u = \boldsymbol{W}_u' \boldsymbol{T}_{\exp}(u), \tag{3.1}$$

where $\boldsymbol{W_u}$ is a column vector of p-weights given by $\boldsymbol{W}_u = [w_{u_1}\ w_{u_2}\ \cdots\ w_{u_p}]'$, and

$$\boldsymbol{T}_{\exp}(u) = \begin{bmatrix} T(1,u) \\ T(2,u) \\ \vdots \\ T(p,u) \end{bmatrix},$$

where $T(i,u) = \hat{M}_y \exp\{(M_{z_i} - \hat{M}_{z_i}(u))/(M_{z_i} + \hat{M}_{z_i}(u))\}$ for $i = 1, 2, 3, \ldots, p$ such that $\mathbf{1}'\boldsymbol{W}_u = 1$, where $\mathbf{1}$ is a column vector of order p.

The second estimator T_m is also proposed as a weighted multivariate chain type ratio to exponential ratio estimator based on sample size $m = n\lambda$ common to both occasions and given by

$$T_m = \boldsymbol{W}_m' \boldsymbol{T}_{\exp}(m,n), \tag{3.2}$$

where $\boldsymbol{W}_m$ is a column vector of p-weights as $\boldsymbol{W}_m = [w_{m_1}\ w_{m_2}\ \cdots\ w_{m_p}]'$. and

$$\boldsymbol{T}_{\exp}(m,n) = \begin{bmatrix} T(1,m,n) \\ T(2,m,n) \\ \vdots \\ T(p,m,n) \end{bmatrix},$$

where $T(i,m,n) = \{(\hat{M}_y^*(i,m))/(\hat{M}_x^*(i,m))\}\,\hat{M}_x^*(i,n)$, where $\hat{M}_y^*(i,m) = \hat{M}_y(m)\exp\{(M_{z_i} - \hat{M}_{z_i}(m))/(M_{z_i} + \hat{M}_{z_i}(m))\}$, $\hat{M}_x^*(i,m) = \hat{M}_x(m)\exp\{(M_{z_i} - \hat{M}_{z_i}(m))/(M_{z_i} + \hat{M}_{z_i}(m))\}$ and $\hat{M}_x^*(i,n) = \hat{M}_x(n)\exp\{(M_{z_i} - \hat{M}_{z_i}(n))/(M_{z_i} + \hat{M}_{z_i}(n))\}$, for $i = 1,2,3,\ldots,p$. Such that $\mathbf{1}'\boldsymbol{W}_m = 1$, where $\mathbf{1}$ is a column vector of order p.

The optimum weights $\boldsymbol{W}_u$ and $\boldsymbol{W}_m$ in T_u and T_m are selected by minimizing their mean squared errors respectively.

Considering the convex linear combination of the two estimators T_u and T_m, we have the final estimator of the population median M_y on the current occasion as

$$T = \varphi T_u + (1-\varphi)\,T_m, \tag{3.3}$$

where φ $(0 \le \varphi \le 1)$ is an unknown constant to be determined so as to minimize the mean squared error of the estimator T.

Remark 1. The estimator T_u is suitable to estimate the median on each occasion, which implies that more belief on φ could be shown by choosing φ as 1 (or close to 1); however, to estimate the change from occasion to occasion, the estimator T_m could be more useful so φ might be chosen as 0 (or close to 0). For asserting both problems simultaneously, the suitable (optimum) choice of φ is desired.

4. Properties of the proposed estimator T

4.1. Assumptions

The properties of the proposed estimators T are derived under the following assumptions:

1) Population size is sufficiently large (i.e., $N \to \infty$), therefore finite population corrections are ignored.

2) As $N \to \infty$, the distribution of the bivariate variable (a,b), where a and $b \in \{x, y, z_1, z_2, \ldots, z_p\}$ and $a \ne b$ approaches a continuous distribution with marginal densities $f_a(\cdot)$ and $f_b(\cdot)$, respectively (Kuk and Mak, 1989).

3) The marginal densities $f_x(\cdot)$, $f_y(\cdot)$, $f_{z_1}(\cdot)$, $f_{z_2}(\cdot)$, …, f_{z_p} are positive.

4) The sample medians $\hat{M}_x(n)$, $\hat{M}_{z_i}(n)$, $\hat{M}_x(m)$, $\hat{M}_y(m)$, $\hat{M}_{z_i}(m)$, $\hat{M}_y(u)$, and $\hat{M}_{z_i}(u)$ for $i = 1,2,3,\ldots,p$; are consistent and asymptotically normal (Gross, 1980).

5) Following Kuk and Mak (1989), let P_{ab} be the proportion of elements in the population such that $a \le \hat{M}_a$ and $b \le \hat{M}_b$ where a and $b \in \{x, y, z_1, z_2, \ldots, z_p\}$, and $a \ne b$.

6) The following large sample approximations are assumed:

$$\hat{M}_y(u) = M_y(1+e_0), \quad \hat{M}_y(m) = M_y(1+e_1), \quad \hat{M}_x(m) = M_x(1+e_2), \quad \hat{M}_x(n) = M_x(1+e_3),$$
$$\hat{M}_{z_i}(u) = M_{z_i}(1+e_{4i}), \quad \hat{M}_{z_i}(m) = M_{z_i}(1+e_{5i}), \quad \hat{M}_{z_i}(n) = M_{z_i}(1+e_{6i})$$

such that $|e_k| < 1$ and $|e_{ki}| < 1$ and $\forall\ k = 0,1,2,\ldots,6$ and $i = 1,2,3,\ldots,p$.

The values of various related expectations can be seen in Singh (2003).

4.2. Bias and mean squared error of the estimator T

The estimators T_u and T_m are weighted multivariate exponential ratio and chain type ratio to exponential ratio type in nature, respectively. Hence they are biased for population median M_y. Therefore, the final estimator T defined in equation (3.3) is also a biased estimator of M_y. Bias $B(\cdot)$ and mean squared error $M(\cdot)$ of the proposed estimator have been derived to the first order of approximations that provide the following theorems:

Theorem 1. *The bias of the estimator T to the first order of approximation is obtained as*

$$B(T) = \varphi B(T_u) + (1-\varphi)B(T_m), \tag{4.1}$$

$$B(T_u) = \frac{1}{u}\boldsymbol{W}_u'\boldsymbol{B}_u, \tag{4.2}$$

$$B(T_m) = \boldsymbol{W}_m' \left(\frac{1}{m}B_{m1} + \frac{1}{n}\boldsymbol{B}_{m2}\right), \tag{4.3}$$

where $\boldsymbol{B}_u = (B_1(u), B_2(u), \ldots, B_p(u))'$, $\boldsymbol{B}_{m2} = (Bm_{21}, Bm_{22}, \ldots, Bm_{2p})'$

$$B_i(u) = \left(\frac{3\left[f_{z_i}(M_{z_i})\right]^{-2} M_y}{32M_{z_i}^2} - \frac{\left(4P_{yz_i}-1\right)\left[f_y\left(M_y\right)\right]^{-1}\left[f_{z_i}(M_{z_i})\right]^{-1}}{8M_{z_i}}\right),$$

$$B_{m1} = \left(\frac{\left[f_x(M_x)\right]^{-2} M_y}{4M_x^2} - \frac{\left(4P_{xy}-1\right)\left[f_x(M_x)\right]^{-1}\left[f_y\left(M_y\right)\right]^{-1}}{4M_x}\right),$$

where

$$Bm_{2i} = \left(\frac{\left(4P_{xy}-1\right)\left[f_x(M_x)\right]^{-1}\left[f_y\left(M_y\right)\right]^{-1}}{4M_x} - \frac{\left[3f_{z_i}(M_{z_i})\right]^{-2} M_y}{32M_{z_i}^2} - \frac{\left[f_x(M_x)\right]^{-2} M_y}{4M_x^2} - \frac{\left(4P_{yz_i}-1\right)\left[f_y\left(M_y\right)\right]^{-1}\left[f_{z_i}(M_{z_i})\right]^{-1}}{8M_{z_i}}\right)$$

for $i = 1, 2, 3, \ldots, p$.

Theorem 2. *Mean squared error of the estimator T to the first order of approximations is obtained as*

$$M(T) = \varphi^2 M(T_u) + (1-\varphi)^2 M(T_m) + 2\varphi(1-\varphi)\operatorname{cov}(T_u, T_m), \tag{4.4}$$

$$M(T_u) = \boldsymbol{W}_u'\,\boldsymbol{D}_u\boldsymbol{W}_u, \tag{4.5}$$

$$M(T_m) = (B)\boldsymbol{W}_m'\,\boldsymbol{E}\boldsymbol{W}_m + \boldsymbol{W}_m'\boldsymbol{D}_m\boldsymbol{W}_m, \tag{4.6}$$

where $\boldsymbol{W}_u = [w_{u_1}\ w_{u_2}\cdots w_{u_p}]'$, $\boldsymbol{W}_m = [w_{m_1}\ w_{m_2}\cdots w_{m_p}]'$, $\boldsymbol{E}$ *is a unit matrix of order* $p \times p$, $\boldsymbol{D}_u = (1/u - 1/N)\boldsymbol{D}_{u^*}$, $\boldsymbol{D}_m = (1/n - 1/N)\boldsymbol{D}_{m^*}$, *where*

$$\boldsymbol{D}_{u^*} = \begin{bmatrix} du_{11} & du_{12} & \ldots & du_{1p} \\ du_{21} & du_{22} & \ldots & du_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ du_{p1} & du_{p2} & \ldots & du_{pp} \end{bmatrix}_{p\times p} \quad \textit{and} \quad \boldsymbol{D}_{m^*} = \begin{bmatrix} dm_{11} & dm_{12} & \ldots & dm_{1p} \\ dm_{21} & dm_{22} & \ldots & dm_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ dm_{p1} & dm_{p2} & \ldots & dm_{pp} \end{bmatrix}_{p\times p},$$

where $B = (1/m - 1/N)B_1$,

$$B_1 = \left(\frac{[f_x(M_x)]^{-2} M_y^2}{4M_x^2} - \frac{\left(4P_{xy} - 1\right)[f_x(M_x)]^{-1}\left[f_y\left(M_y\right)\right]^{-1} M_y}{2M_x} + \frac{\left[f_y\left(M_y\right)\right]^{-2}}{4} \right),$$

$$du_{ii} = \left(\frac{\left[f_y\left(M_y\right)\right]^{-2}}{4} - \frac{\left(4P_{yz_i} - 1\right)\left[f_y\left(M_y\right)\right]^{-1}[f_{z_i}(M_{z_i})]^{-1} M_y}{4M_{z_i}} + \frac{[f_{z_i}(M_{z_i})]^{-2} M_y^2}{16M_{z_i}^2} \right),$$

$$du_{ij} = \left(\frac{\left[f_y\left(M_y\right)\right]^{-2}}{4} - \frac{\left(4P_{yz_i} - 1\right)\left[f_y\left(M_y\right)\right]^{-1}[f_{z_i}(M_{z_i})]^{-1} M_y}{8M_{z_i}} - \frac{\left(4P_{yz_j} - 1\right)\left[f_y\left(M_y\right)\right]^{-1}\left[f_{z_j}\left(M_{z_j}\right)\right]^{-1} M_y}{8M_{z_j}} + \frac{\left(4P_{z_iz_j} - 1\right)[f_{z_i}(M_{z_i})]^{-1}\left[f_{z_j}\left(M_{z_j}\right)\right]^{-1} M_y^2}{16M_{z_i}M_{z_j}} \right),$$

$$dm_{ii} = \left(\frac{[f_{z_i}(M_{z_i})]^{-2} M_y^2}{16M_{z_i}^2} + \frac{\left(4P_{xy} - 1\right)[f_x(M_x)]^{-1}\left[f_y\left(M_y\right)\right]^{-1} M_y}{2M_x} - \frac{[f_x(M_x)]^{-2} M_y^2}{4M_x^2} - \frac{\left(4P_{yz_i} - 1\right)\left[f_y\left(M_y\right)\right]^{-1}[f_{z_i}(M_{z_i})]^{-1} M_y}{4M_{z_i}} \right),$$

$$dm_{ij} = \left(-\frac{[f_x(M_x)]^{-2} M_y^2}{4M_x^2} + \frac{\left(4P_{xy} - 1\right)[f_x(M_x)]^{-1}\left[f_y\left(M_y\right)\right]^{-1} M_y}{2M_x} - \frac{\left(4P_{yz_i} - 1\right)\left[f_y\left(M_y\right)\right]^{-1}[f_{z_i}(M_{z_i})]^{-1} M_y}{8M_{z_i}} - \frac{\left(4P_{yz_j} - 1\right)\left[f_y\left(M_y\right)\right]^{-1}\left[f_{z_j}\left(M_{z_j}\right)\right]^{-1} M_y}{8M_{z_j}} + \frac{\left(4P_{z_iz_j} - 1\right)[f_{z_i}(M_{z_i})]^{-1}\left[f_{z_j}\left(M_{z_j}\right)\right]^{-1} M_y^2}{16M_{z_i}M_{z_j}} \right),$$

$\forall\ i \neq j = 1, 2, 3, \ldots, p$ *and* $\operatorname{cov}(T_u, T_m) = 0$ *as they are based on two independent samples.*

Remark 2. The mean squared errors of the estimators T in equation (4.4) depend on the population parameters P_{xy}, P_{yz_i}, P_{xz_i}, $P_{z_iz_j}$, $f_x(M_x)$, $f_y(M_y)$, and $f_{z_i}(M_{z_i})$; $(i \neq j = 1, 2, 3, \ldots, p)$. The properties of the proposed estimators can be easily studied if the parameters are known. Otherwise (which is the most common in practice) the unknown population parameters are replaced by the sample estimates. The population proportions P_{xy}, P_{yz_i}, P_{xz_i}, and $P_{z_iz_j}$ can be replaced by the sample estimate $\hat{P}_{xy}$, $\hat{P}_{yz_i}$, $\hat{P}_{xz_i}$, and $\hat{P}_{z_iz_j}$ and the marginal densities $f_y(M_y)$, $f_x(M_x)$, and $f_{z_i}(M_{z_i})$; $(i \neq j = 1, 2, 3, \ldots, p)$ can be substituted by their kernel estimator or nearest neighbor density estimator or generalized nearest neighbor density estimator related to the kernel estimator (Silverman, 1986). Here, the marginal densities $f_y(M_y)$, $f_x(M_x)$ and $f_{z_i}(M_{z_i})$ are replaced by $\hat{f}_y(\hat{M}_y(m))$, $\hat{f}_x(\hat{M}_x(n))$, and $\hat{f}_{z_i}(\hat{M}_{z_i}(n))$; $i = 1, 2, \ldots, p$, respectively, which are obtained by the method of generalized nearest neighbor density estimator related to the kernel estimator.

To estimate $f_y(M_y)$, $f_x(M_x)$, and $f_{z_i}(M_{z_i})$; $i = 1, 2, \ldots, p$ by generalized nearest neighbor density estimator related to the kernel estimator, the following procedure has been adopted:

Choose an integer $h \approx n^{1/2}$ and define the distance $\delta(x_1, x_2)$ between two points on the line to be $|x_1 - x_2|$. For $\hat{M}_x(n)$, define $\delta_1(\hat{M}_x(n)) \leq \delta_2(\hat{M}_x(n)) \leq \cdots \leq \delta_n(\hat{M}_x(n))$ to be the distances, arranged in ascending order, from $\hat{M}_x(n)$ to the points of the sample. The generalized nearest neighbor density

estimate is defined by

$$\hat{f}\left(\hat{M}_x(n)\right) = \frac{1}{n\delta_h\left(\hat{M}_x(n)\right)} \sum_{i=1}^{n} K\left[\frac{\hat{M}_x(n) - x_i}{\delta_h\left(\hat{M}_x(n)\right)}\right],$$

where the kernel function K, satisfies the condition $\int_{-\infty}^{\infty} K(x)dx = 1$. Here, the kernel function is chosen as Gaussian Kernel given by $K(x) = (1/2\pi)e^{-[(1/2)x^2]}$. The estimate of $f_y(M_y)$ and $f_{z_i}(M_{z_i})$; $i = 1, 2, \ldots, p$ can be obtained by the above explained procedure in similar manner.

5. Choice of optimal weights

To find the optimum weight vector $\boldsymbol{W}_u = [w_{u_1}\ w_{u_2}\ \cdots\ w_{u_p}]'$, the mean squared error $M(T_u)$ given in equation (4.5) is the minimized subject to the condition $\mathbf{1}'\boldsymbol{W}_u = 1$ using the method of Lagrange's multiplier explained as:

To find the extrema using Lagrange's multiplier technique, we define Q_1 as

$$Q_1 = \boldsymbol{W}_u'\boldsymbol{D}_u\boldsymbol{W}_u - \lambda_u\left(\mathbf{1}'\boldsymbol{W}_u - 1\right), \tag{5.1}$$

where $\mathbf{1}$ is a unit column vector of order p and $\lambda_u$is the Lagrangian multiplier.

Now, by differentiating equation (5.1) partially with respect to $\boldsymbol{W}_u$ and equating it to zero we have

$$\frac{\partial Q_1}{\partial \boldsymbol{W}_u} = \frac{\partial}{\partial \boldsymbol{W}_u}\left[\boldsymbol{W}_u'\boldsymbol{D}_u\boldsymbol{W}_u - \lambda_u(\mathbf{1}'\boldsymbol{W}_u - 1)\right] = 0.$$

This implies that, $2\boldsymbol{D}_u\boldsymbol{W}_u - \lambda_u\mathbf{1} = 0$, which yields

$$\boldsymbol{W}_u = \frac{\lambda_u}{2}\boldsymbol{D}_u^{-1}\mathbf{1}. \tag{5.2}$$

Now pre-multiplying equation (5.2) by $\mathbf{1}'$, we get

$$\frac{\lambda_u}{2} = \frac{1}{\mathbf{1}'\boldsymbol{K}_u^{-1}\mathbf{1}}. \tag{5.3}$$

Thus, using equation (5.3) in equation (5.2), we obtain the optimal weight vector as

$$\boldsymbol{W}_{u_{\text{opt.}}} = \frac{\boldsymbol{D}_u^{-1}}{\mathbf{1}'\boldsymbol{D}_u^{-1}\mathbf{1}}. \tag{5.4}$$

In similar manners, the optimal of the weight $\boldsymbol{W}_m = [w_{m_1}\ w_{m_2}\ \cdots\ w_{m_p}]'$ is obtained by minimizing $M(T_m)$ subject to the constraint $\mathbf{1}'\boldsymbol{W}_m = 1$ using the method of Lagrange's multiplier, for this we define

$$Q_2 = (B)\boldsymbol{W}_m'\boldsymbol{E}\ \boldsymbol{W}_m + \boldsymbol{W}_m'\boldsymbol{D}_m\boldsymbol{W}_m - \lambda_m(\mathbf{1}'\boldsymbol{W}_m - 1),$$

where λ_m is the Lagrangian multiplier.

Now, differentiating Q_2 with respect to $\boldsymbol{W}_m$ and equating to 0, we get

$$\boldsymbol{W}_{m_{\text{opt.}}} = \frac{\boldsymbol{D}_m^{-1}}{\mathbf{1}'\boldsymbol{D}_m^{-1}\mathbf{1}}. \tag{5.5}$$

Then substituting the optimum values of $\boldsymbol{W}_u$ and $\boldsymbol{W}_m$ in equations (4.5) and (4.6), respectively, the optimum mean squared errors of the estimators are obtained as:

$$M(T_u)_{\text{opt.}} = \left(\frac{1}{u} - \frac{1}{N}\right) \frac{1}{\mathbf{1}' \boldsymbol{D}_{u^*}^{-1} \mathbf{1}}, \tag{5.6}$$

$$M(T_m)_{\text{opt.}} = \left(\frac{1}{m} - \frac{1}{N}\right) B_1 + \left(\frac{1}{n} - \frac{1}{N}\right) \frac{1}{\mathbf{1}' \boldsymbol{D}_{m^*}^{-1} \mathbf{1}}. \tag{5.7}$$

□

6. Minimum mean squared errors of the proposed estimator T

The mean squared error of the estimator given in equation (4.4) is a function of unknown constants φ, therefore, it is minimized with respect to φ so that the optimum values of φ is obtained as

$$\varphi_{\text{opt.}} = \frac{M(T_m)_{\text{opt.}}}{M(T_u)_{\text{opt.}} + M(T_m)_{\text{opt.}}}. \tag{6.1}$$

Now substituting the value of $\varphi_{opt.}$ in equation (4.4), we obtain the optimum mean squared error of the estimator T as

$$M(T)^*_{\text{opt.}} = \frac{M(T_u)_{\text{opt.}}.M(T_m)_{\text{opt.}}}{M(T_u)_{\text{opt.}} + M(T_m)_{\text{opt.}}}. \tag{6.2}$$

Further, substituting the optimum values of the mean squared error of the estimators T_u and T_m obtained in equations (5.6) and (5.7) in equations (6.1) and (6.2), respectively, the simplified values $\varphi_{\text{opt.}}$ and $M(T)^*_{\text{opt.}}$ is obtained as

$$\varphi_{\text{opt.}} = \frac{\mu\left[\mu C - (B_1 + C)\right]}{\left[\mu^2 C - \mu(B_1 + C - A) - A\right]}, \tag{6.3}$$

$$M(T)^*_{\text{opt.}} = \frac{1}{n} \frac{\left[\mu D_1 - D_2\right]}{\left[\mu^2 C - \mu D_3 - A\right]}, \tag{6.4}$$

where $A = 1/\mathbf{1}' \boldsymbol{K}_{u^*}^{-1} \mathbf{1}$, $C = 1/\mathbf{1}' \boldsymbol{K}_{m^*}^{-1} \mathbf{1}$, $D_1 = AC$, $D_2 = AB_1 + AC$, $D_3 = B_1 + C - A$,

$$B_1 = \left(\frac{\left[f_y\left(M_y\right)\right]^{-2}}{4} + \frac{\left[f_x\left(M_x\right)\right]^{-2} M_y^2}{4M_x^2} - \frac{\left(4P_{xy} - 1\right)\left[f_x\left(M_x\right)\right]^{-1}\left[f_y\left(M_y\right)\right]^{-1} M_y}{2M_x} \right)$$

and μ is the fraction of the sample drawn fresh at the current occasion.

7. Optimum replacement strategy for the estimator T

The key design parameter affecting the estimates of change is the overlap between successive samples. Maintaining high overlap between the repeats of a survey is operationally convenient since many sampled units have been located and have some experience in the survey. Hence to decide about the optimum value of μ (fractions of samples to be drawn fresh on current occasion) so that M_y may be estimated with maximum precision and minimum cost, we minimize the mean squared error $M(T)^*_{\text{opt.}}$

in equation (6.4) with respect to μ. The optimum value of μ so obtained is one of the two roots given by

$$\hat{\mu} = \frac{G_2 \pm \sqrt{G_2^2 - G_1 G_3}}{G_1}, \tag{7.1}$$

where $G_1 = CD_1$, $G_2 = CD_2$, and $G_3 = AD_1 + D_2D_3$.

The real value of $\hat{\mu}$ exist, iff $G_2^2 - G_1G_3 \geq 0$. For any situation, which satisfies this condition, two real values of $\hat{\mu}$ may be possible; hence, we choose a value of $\hat{\mu}$ such that $0 \leq \hat{\mu} \leq 1$. All other values of $\hat{\mu}$ are inadmissible. If both the real values of $\hat{\mu}$ are admissible, the lowest one will be the best choice because it reduces the total cost of the survey. Substituting the admissible value of $\hat{\mu}$ say μ_0 from equation (7.1) in to the equation (6.4), we get the optimum value of the mean squared error of the estimator T with respect to φ as well as μ which, is given as

$$M(T)^{**}_{\text{opt.}} = \frac{1}{n}\frac{[\mu_0 D_1 - D_2]}{\left[\mu_0^2 C - \mu_0 D_3 - A\right]}. \tag{7.2}$$

8. Efficiency with increased number of auxiliary variables

Increasing the number of auxiliary variables typically increases the precision of the estimates. This section verifies this property for the proposed estimator as under: Let $T_{|p}$ and $T_{|q}$ be two proposed estimators based on p and q auxiliary variables, respectively such that $p < q$, then $M(T_p) \geq M(T_q)$, i.e.,

$$M(T_p) - M(T_q) \geq 0, \tag{8.1}$$

$$\frac{1}{n}\frac{\left[\mu A_p C_p - A_p(B + C_p)\right]}{\left[\mu^2 C_p - \mu(B + C_p + A_p) - A_p\right]} - \frac{1}{n}\frac{\left[\mu A_q C_q - A_q(B + C_q)\right]}{\left[\mu^2 C_q - \mu(B + C_q + A_q) - A_q\right]} \geq 0.$$

On simplification, we get

$$\left(A_p - A_q\right)\left[(\mu - 1)^2\left(\mu C_p C_q + \frac{A_p A_q\left(C_p - C_q\right)}{\left(A_p - A_q\right)}\right) - \mu B\left(\left(C_p - C_q\right)(\mu - 1) - B\right)\right] \geq 0.$$

This reduces to the condition

$$(A_p - A_q) \geq 0. \tag{8.2}$$

From Section 6, we get

$$\frac{1}{\mathbf{1}'\boldsymbol{D}_p^{-1}\mathbf{1}} - \frac{1}{\mathbf{1}'\boldsymbol{D}_q^{-1}\mathbf{1}} \geq 0, \qquad \mathbf{1}'\boldsymbol{D}_q^{-1}\mathbf{1} \geq \mathbf{1}'\boldsymbol{D}_p^{-1}\mathbf{1}.$$

Following Rao (2002), the matrix $\boldsymbol{D}_q$ can be partitioned and can be written as

$$\boldsymbol{D}_q = \begin{pmatrix} \boldsymbol{D}_p & \boldsymbol{F} \\ \boldsymbol{F}' & \boldsymbol{G} \end{pmatrix},$$

where $\boldsymbol{F}$, $\boldsymbol{F}'$, and $\boldsymbol{G}$ are matrices deduced from $\boldsymbol{D}_q$ such that their order never exceeds $q-p$ and always greater than or equal to 1. Then,

$$\boldsymbol{D}_q^{-1} = \begin{bmatrix} \boldsymbol{D}_p^{-1} + \boldsymbol{HJH}' & -\boldsymbol{HJ} \\ -\boldsymbol{JH}' & \boldsymbol{J} \end{bmatrix}, \tag{8.3}$$

where $\boldsymbol{J} = (\boldsymbol{G} - \boldsymbol{F}'\boldsymbol{D}_p^{-1}\boldsymbol{F})^{-1}$ and $\boldsymbol{H} = \boldsymbol{D}_p^{-1}\boldsymbol{F}$ (Olkin, 1958; Rao, 2002)

Now rewriting $\mathbf{1}'\boldsymbol{D}_q^{-1}\mathbf{1}$ by putting the value of $\boldsymbol{D}_q^{-1}$ from equation (8.3), we get

$$\begin{aligned}
\mathbf{1}'\boldsymbol{D}_q^{-1}\mathbf{1} &= \begin{bmatrix} \mathbf{1}_p & \mathbf{1}_{q-p} \end{bmatrix}' \begin{bmatrix} \boldsymbol{D}_p^{-1} + \boldsymbol{HJH}' & -\boldsymbol{HJ} \\ -\boldsymbol{JH}' & \boldsymbol{J} \end{bmatrix} \begin{bmatrix} \mathbf{1}_p \\ \mathbf{1}_{q-p} \end{bmatrix} \\
&= \begin{pmatrix} \mathbf{1}_p' \left(\boldsymbol{D}_p^{-1} + \boldsymbol{HJH}'\right) - \mathbf{1}_{q-p}'\boldsymbol{JH}' & -\mathbf{1}_p'\boldsymbol{HJ} + \mathbf{1}_{q-p}'\boldsymbol{J} \end{pmatrix} \begin{bmatrix} \mathbf{1}_p \\ \mathbf{1}_{q-p} \end{bmatrix} \\
&= \mathbf{1}_p' \left(\boldsymbol{D}_p^{-1} + \boldsymbol{HJH}'\right)\mathbf{1}_p - \mathbf{1}_{q-p}'\boldsymbol{JH}'\mathbf{1}_p - \mathbf{1}_p'\boldsymbol{HJ}\mathbf{1}_{q-p} + \mathbf{1}_{q-p}'\boldsymbol{J}\mathbf{1}_{q-p}
\end{aligned}$$

implies

$$\begin{aligned}
\mathbf{1}'\boldsymbol{D}_q^{-1}\mathbf{1} - \mathbf{1}_p'\left(\boldsymbol{D}_p^{-1}\right)\mathbf{1}_p &= \mathbf{1}_p'\left(\boldsymbol{HJH}'\right)\mathbf{1}_p - \mathbf{1}_{q-p}'\boldsymbol{JH}'\mathbf{1}_p - \mathbf{1}_p'\boldsymbol{HJ}\mathbf{1}_{q-p} + \mathbf{1}_{q-p}'\boldsymbol{J}\mathbf{1}_{q-p}, \\
\mathbf{1}'\boldsymbol{D}_q^{-1}\mathbf{1} - \mathbf{1}_p'\left(\boldsymbol{D}_p^{-1}\right)\mathbf{1}_p &= \begin{bmatrix} \mathbf{1}_p & \mathbf{1}_{q-p} \end{bmatrix}' \begin{bmatrix} \boldsymbol{HJH}' & -\boldsymbol{HJ} \\ -\boldsymbol{JH}' & \boldsymbol{J} \end{bmatrix} \begin{bmatrix} \mathbf{1}_p \\ \mathbf{1}_{q-p} \end{bmatrix}, \\
\mathbf{1}'\boldsymbol{D}_q^{-1}\mathbf{1} - \mathbf{1}_p'\left(\boldsymbol{D}_p^{-1}\right)\mathbf{1}_p &= \mathbf{1}' \begin{bmatrix} \boldsymbol{H} \\ -\boldsymbol{I} \end{bmatrix} \boldsymbol{J} \begin{bmatrix} \boldsymbol{H} & -\boldsymbol{I} \end{bmatrix} \mathbf{1} \geq 0.
\end{aligned}$$

The latter follows since $\boldsymbol{J}$ is positive definite so that $\boldsymbol{R}'\boldsymbol{JR} \geq 0$ for all $\boldsymbol{R}$, where $\boldsymbol{R} = \begin{pmatrix} \boldsymbol{H} & -\boldsymbol{I} \end{pmatrix}\mathbf{1}$.

Hence, this leads to the result that utilizing more auxiliary variables provides more efficient estimates in terms of mean squared error for the proposed estimator.

9. Efficiency comparison

To evaluate the performance of the proposed estimator, the estimator T at optimum conditions is compared to sample median estimator $\hat{M}_y(n)$, when there is no matching from the previous occasion. For empirical investigations the proposed estimator has been considered for the case $p = 1$ and $p = 2$. Since, $\hat{M}_y(n)$ is unbiased for population median, so the variance of $\hat{M}_y(n)$ is given as

$$V\left(\hat{M}_y(n)\right) = \frac{1}{n}\frac{[f_y(M_y)]^{-2}}{4}. \tag{9.1}$$

The percent relative efficiencies $E_{T_{|p=1}}$ and $E_{T_{|p=2}}$ of the estimator T (under their respective optimum conditions) with respect to $\hat{M}_y(n)$ is given by

$$E_{T_{|p=1}} = \frac{V\left(\hat{M}_y(n)\right)}{M\left(T_{|p=1}\right)^{**}_{\text{opt.}}} \times 100 \qquad \text{and} \qquad E_{T_{|p=2}} = \frac{V\left(\hat{M}_y(n)\right)}{M\left(T_{|p=2}\right)^{**}_{\text{opt.}}} \times 100 \tag{9.2}$$

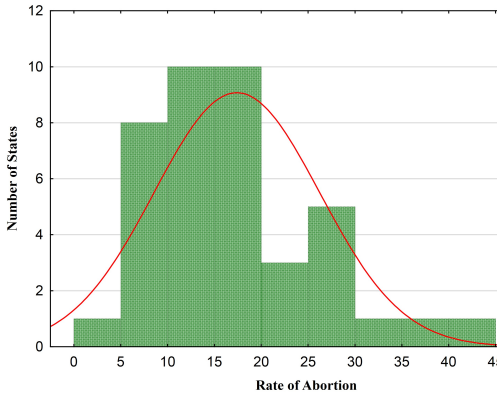

Figure 1: *Rate of abortion versus different U.S. states in 2007.*

10. Numerical illustrations

To justify the practical use of the proposed multivariate estimator T, a completely known population has been considered with illustrations that suppose two auxiliary variables (i.e., $p = 2$) are available. The real population has been taken from the Statistical Abstracts of the United States.

The population comprise of $N = 40$ states of the United States. Let y_i represent the rate of abortions during 2008 in the i^{th} state of U.S., x_i be the rate of abortions during 2007 in the i^{th} state of U. S., z_{1_i} denote the rate of abortions in 2005 in the i^{th} state of U.S. and z_{2_i} denote the rate of abortions during 2004 in the i^{th} state of U.S. Data are presented in Figure 1.

The graph in Figure 1 shows that the rate of abortions are almost skewed towards the right. One reason for the skewness may be the distribution of population in different states. It is believed that states having a larger population are expected to have larger rate of abortions. Thus skewness of data indicates that the use of median may be a good measure for central location than the mean in this situation.

For the considered population, the optimum value of μ defined in equation (7.1) and percent relative efficiencies $E_{T_{|p=1}}$ and $E_{T_{|p=1}}$ defined in equation (9.2) of T (for $p = 1$ and $p = 2$ under their optimal conditions) with respect to $\hat{M}_y(n)$ have been computed (Table 1). To validate the above empirical results, Monte Carlo simulation was performed for the considered population.

10.1. Simulation algorithm

1) Choose 5,000 samples of size $n = 15$ using simple random sampling without replacement on first occasion for both the study and auxiliary variables.

2) Calculate sample median $\hat{M}_{x|k}(n)$, $\hat{M}_{z_1|k}(n)$, and $\hat{M}_{z_2|k}(n)$ for $k = 1, 2, \ldots, 5000$.

3) Retain $m = 13$ units out of each $n = 15$ sample units of the study and auxiliary variables at the first occasion.

4) Calculate sample median $\hat{M}_{x|k}(m)$, $\hat{M}_{y|k}(m)$, $\hat{M}_{z_1|k}(m)$, and $\hat{M}_{z_2|k}(m)$ for $k = 1, 2, \ldots, 5000$.

5) Select $u = 2$ units using simple random sampling without replacement from $N - n = 25$ units of the population for study and auxiliary variables at the second occasion.

Table 1: Comparison of the proposed estimators $T_{|p=1}$ and $T_{|p=2}$ (at respective optimum conditions) with respect to the estimator $\hat{M}_y(n)$

	Optimum value of μ_0	Percent relative efficiency $E_{T_{\|p=i}}$
$p = 1$	0.5478	171.16
$p = 2$	0.5229	199.54

Table 2: Estimated values of population median using the proposed estimators $T_{|p=1}$ and $T_{|p=2}$ at optimum conditions

Actual value	$n = 10$		$n = 15$		$n = 20$	
$M_y = 15.50$	Estimated value	RSE	Estimated value	RSE	Estimated value	RSE
$T_{\|p=1}$	14.83	8.57%	15.10	7.85%	16.16	6.95%
$T_{\|p=2}$	15.01	8.47%	15.47	7.49%	15.98	6.83%

RSE = relative standard error of the estimator.

6) Calculate sample median $\hat{M}_{y|k}(u)$, $\hat{M}_{z_1|k}(u)$, and $\hat{M}_{z_2|k}(u)$ for $k = 1, 2, \ldots, 5000$.

7) Iterate the parameter φ from 0.1 to 0.9 with a step size of 0.1.

8) Calculate the percent relative efficiencies of the proposed estimator T with the case $p = 1$ and $p = 2$ (i.e., $T_{|p=1}$ and $T_{|p=2}$) with respect to the sample median estimator $\hat{M}_y(n)$ as:

$$E_1(\text{sim}) = \frac{\sum_{k=1}^{5000}\left[\hat{M}_{y|k}(n) - M_y\right]^2}{\sum_{k=1}^{5000}\left[T_{p=1|k} - M_y\right]^2} \times 100 \qquad \text{and} \qquad E_2(\text{sim}) = \frac{\sum_{k=1}^{5000}\left[\hat{M}_{y|k}(n) - M_y\right]^2}{\sum_{k=1}^{5000}\left[T_{p=2|k} - M_y\right]^2} \times 100.$$

For better analysis, the above simulation experiments were repeated for different choices of μ. For convenience the different choices of μ are considered as different sets for the considered population which is:

$$\text{Set I} : n = 15, \mu = 0.10,\ (m = 13, u = 2), \qquad \text{Set II} : n = 15, \mu = 0.20,\ (m = 12, u = 3),$$
$$\text{Set III} : n = 15, \mu = 0.30,\ (m = 10, u = 5), \qquad \text{Set IV} : n = 15, \mu = 0.40,\ (m = 9, u = 6).$$

Table 2 presents the simulation results.

11. Mutual comparison of the estimators $T_{|p=1}$ and $T_{|p=2}$

The performances of the estimator $T_{|p=1}$ and $T_{|p=2}$ have been elaborated empirically as well as through simulation studies in Section 10, with the results presented in Tables 1–3. The mutual comparison of the estimators for the cases when $p = 1$ and $p = 2$ has been elaborated graphically and presented in Figure 2.

12. Interpretation of results

1) It is clear from Table 1 that optimum values of μ_0 (for $p = 1$ and $p = 2$) exist for the considered population and μ_0 $(p = 2)$ $< \mu_0$ $(p = 1)$. This indicates that less fraction of a fresh sample is required when more numbers of auxiliary variables are used. Hence, the total cost of the survey will also be reduced when more numbers of additional auxiliary variables are considered.

Table 3: Monte Carlo simulation results when the proposed estimators $T_{|p=1}$ and $T_{|p=2}$ are compared to $\hat{M}_y(n)$

φ		Set			
		I	II	III	IV
0.1	E_1(sim)	307.79	536.79	316.69	422.16
	E_2(sim)	528.09	750.31	503.03	572.87
0.2	E_1(sim)	304.69	523.28	352.51	452.88
	E_2(sim)	538.70	742.33	545.61	624.11
0.3	E_1(sim)	294.64	505.08	370.03	460.89
	E_2(sim)	521.40	727.42	556.61	621.96
0.4	E_1(sim)	277.73	470.01	366.51	471.37
	E_2(sim)	480.55	669.88	521.47	618.86
0.5	E_1(sim)	260.27	426.75	355.05	459.98
	E_2(sim)	431.18	588.43	479.63	582.28
0.6	E_1(sim)	241.34	381.47	328.24	443.04
	E_2(sim)	379.41	506.80	418.77	542.37
0.7	E_1(sim)	222.24	339.33	301.81	413.75
	E_2(sim)	329.89	433.81	366.84	488.27
0.8	E_1(sim)	204.09	298.86	272.14	378.01
	E_2(sim)	285.40	366.39	316.65	431.25
0.9	E_1(sim)	184.42	263.30	239.41	337.58
	E_2(sim)	243.33	312.30	268.82	373.59

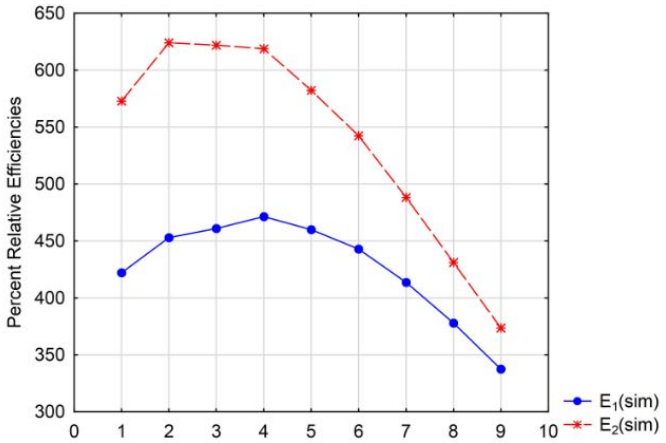

Figure 2: *Mutual comparison of the proposed estimators $T_{|p=1}$ and $T_{|p=2}$ when compared with estimator $\hat{M}_y(n)$.*

2) Table 1 also explains that the value of $E_{T|p=2} > E_{T|p=1}$, this also indicates that efficiency is significantly increased when more numbers of auxiliary variates are considered, which also resembles in accordance with the theory.

3) In Table 2, the estimates of population median have been computed using the proposed estimator T for $p = 1$ and $p = 2$ at their respective optimum conditions. We see that the estimates for the population median are quite near the original value of the population median.

4) In Table 2, the relative standard error of the estimators (RSE) has also been computed; in addition, it is observed that RSE is reduced as sample size is increased. RSE also decreases when more

numbers of auxiliary variables are used. The value of RSE is considerably low and this indicates that the estimates are quite reliable.

5) From simulation study in Table 3 and from Figure 2, we observe that the value of $E_1(\text{sim})$ and $E_2(\text{sim})$ exists for all choices of φ and for all different sets. As φ increases the value of $E_1(\text{sim})$ and $E_2(\text{sim})$ decreases for all sets, which indicates that the efficiency of the estimator T gets reduced if more weight is given to the estimator defined on current occasion and is in accordance with the results of Shukhatme *et al.* (1984). The big difference in the two lines in Figure 2 shows that the performance of estimator drastically enhances when more numbers of auxiliary variables are considered.

6) From Table 3 we also observe that for set II, the estimators $T_{|p=1}$ and $T_{|p=2}$ prove extensively better than the sample median estimator. No fixed pattern is observed in the efficiencies of the proposed estimators if the value of fraction of fresh sample to be drawn on current occasion increases.

13. Conclusion

From the preceding interpretations, it may be concluded that the use of multivariate exponential ratio type estimators for the estimation of a population median at the current occasion in two occasions; therefore, successive sampling is highly appreciable as vindicated through empirical and simulation results. The mutual comparison of the proposed estimators indicates that the estimators utilizing more auxiliary variables perform better in terms of cost and precision. Hence, the proposed multivariate estimator T may be recommended for its practical use in longitudinal surveys to estimate the population median by survey practitioners.

Acknowledgement

Authors are thankful to UGC, New Delhi, India for providing the financial assistance under the grant for MRP No. 42-42 (2013)/SR to carry out the present work. Authors also acknowledge the free access to the data from Statistical Abstracts of the United States available on the Internet.

References

Arnab R and Okafor FC (1992). A note on double sampling over two occasions, *Pakistan Journal of Statistics*, **8**, 9–18.

Eckler AR (1955). Rotation sampling, *Annals of Mathematical Statistics*, **26**, 664–685.

Gordon L (1983). Successive sampling in large finite populations, *The Annals of Statistics*, **11**, 702–706.

Gross ST (1980). Median estimation in sample surveys. In *Proceedings of the section on Survey Research Methods at the Annual Meeting of the American Statistical Association*, Houston TX, 181–184.

Jessen RJ (1942). *Statistical Investigation of a Sample Survey for Obtaining Farm Facts*, Agricultural Experiment Station of the Iowa State College of Agriculture and Mechanic Arts, Ames.

Kuk AYC and Mak TK (1989). Median estimation in the presence of auxiliary information, *Journal of the Royal Statistical Society Series B (Methodological)*, **51**, 261–269.

Martínez-Miranda MD, Rueda-Garcia M, Arcos-Cebrián A, Román-Montoya Y, and Gonzalez-Aguilera S (2005). Quintile estimation under successive sampling, *Computational Statistics*, **20**, 385–399.

Narain RD (1953). On the recurrence formula in sampling on successive occasions, *Journal of the Indian Society of Agricultural Statistics*, **5**, 96–99.

Olkin I (1958). Multivariate ratio estimation for finite populations, *Biometrika*, **45**, 154–165.

Patterson HD (1950). Sampling on successive occasions with partial replacement of units, *Journal of the Royal Statistical Society Series B (Methodological)*, **12**, 241–255.

Priyanka K and Mittal R (2014). Effective rotation patterns for median estimation in successive sampling, *Statistics in Transition New Series*, **15**, 197–220.

Priyanka K and Mittal R (2016). Searching effective rotation patterns for population median using exponential type estimators in two-occasion rotation sampling, *Communications in Statistics - Theory and Methods*, **45**, 5443–5460.

Priyanka K, Mittal R, and Kim JM (2015). Multivariate rotation design for population mean in sampling on successive occasions, *Communications for Statistical Applications and Methods*, **22**, 445–462.

Rao CR (2002). *Linear Statistical Inference and its Applications* (2nd ed), John Wiley & Sons, New York.

Rueda MDM, Muñoz JF, and Arcos A (2008). Successive sampling to estimate quantiles with P-Auxiliary Variables, *Quality and Quantity*, **42**, 427–443.

Silverman BW (1986). *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London.

Singh GN and Priyanka K (2008a). Search of good rotation patterns to improve the precision of estimates at current occasion, *Communications in Statistics - Theory and Methods*, **37**, 337–348.

Singh GN and Priyanka K (2008b). On the use of guess value for the estimation of population median on current occasion in two occasions rotation patterns, *Statistics in Transition New Series*, **9**, 215–232.

Singh GN, Priyanka K, Prasad S, Singh S, and Kim JM (2013). A class of estimators for population variance in two occasion rotation patterns, *Communications for Statistical Applications and Methods*, **20**, 247–257.

Singh HP, Tailor R, Singh S, and Kim JM (2007). Quintile estimation in successive sampling, *Journal of the Korean Statistical Society*, **36**, 543–556.

Singh S (2003). *Advanced Sampling Theory with Applications; How Michael 'Selected' Amy*, Kluwer Academic Publishers, Dordrecht.

Sukhatme PV, Sukhatme BV, Sukhatme S, and Asok C (1984). *Sampling Theory of Surveys with Applications*, Iowa State University Press and Indian Society of Agricultural Statistics, New Delhi.

Received February 17, 2017; Revised March 17, 2017; Accepted April 7, 2017