Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF
AGRICULTURAL STATISTICS 73(3) 2019 197–206

Bootstrap Variance Estimation Technique under Dual Frame Ranked Set Sampling

Pratyush Dasgupta, Tauqueer Ahmad, Anil Rai and Ankur Biswas
ICAR-Indian Agricultural Statistics Research Institute, New Delhi

Received 06 May 2019; Revised 20 July 2019; Accepted 29 July 2019

SUMMARY

Multiple frames are preferably used when a satisfactory sampling frame, covering the whole population in question, is unavailable or even if such a frame is available it may not be economically advantageous to use that frame for survey because of high cost of sampling per unit. In this paper, we dealt with the problem of variance estimation of the dual frame ranked set sample (DFRSS) estimator. We propose two rescaling Bootstrap variance estimation methods viz. strata based and cluster based, to obtain an unbiased estimator of the sampling variance of the proposed estimator. The comparison of performance of the proposed rescaled bootstrap methods with standard bootstrap methods were investigated through a simulation study. The simulation results show that the proposed methods are more stable and have lesser relative bias than the standard approaches. Among the two rescaling Bootstrap variance estimation methods, the strata based rescaling Bootstrap variance estimation approach is more powerful than its counterpart.

Keywords: Multiple frames, Ranked set sampling, Variance estimation, Bootstrap, Resampling technique.

1. INTRODUCTION

The basic theory underlying the classical approach of survey sampling assumes availability of a complete list of elements or units so that the sampler has individual access to them to have a sample representative of the predefined population. Such a list of population units is known as sampling frame for the population. It is possible that such a frame is available for all the units in the population. If not, we can redefine the population units to include them in some available frame. In many practical situations, it may be difficult to obtain a sampling frame which covers the whole population or there may not be any other satisfactory compromise for the frame. In such situations when it is not possible to designate a unique reference frame, it becomes necessary to supplement the original frame with additional frames to cover the whole population. Even if a unique frame is available, it may be very costly to select a sample from it and thus use of additional frames may be economically advantageous. Both of these situations constitute the general problem of multiple frame sampling. Hartley (1962) suggested the use of two (dual frame survey) or more than two (multiple frame survey) frames at a time in such a manner that the resulting sampling procedure is cost effective and the frames together cover the population completely. As for example, in agricultural surveys, an area frame consists of segments of land and completely covers the population. On the contrary, a list frame consists of the names and addresses of agricultural operators and, thus, the frame may be incomplete. Even though the area frame is complete but is very expensive to sample and on the other hand, list frames are usually less costly to sample. Hartley (1974) used simple random sampling to draw independent random samples from two overlapping frames and provide an estimator of the population total of characteristics under study. The problem of dual frame survey was also considered by Saxena *et al.* (1984), Skinner (1991), Lohr and Rao (2000, 2006) under different contexts. All the methods of estimation with overlapping frames proposed earlier assumed that domain membership

Corresponding author: Ankur Biswas
E-mail address: ankur.bckv@gmail.com

can be determined for every sampled unit, i.e, they are sensitive to misclassification of observations into domains. Mecatti (2007) proposed a multiple frame estimator using multiplicity approach. He showed that the single frame multiplicity estimators are insensitive to misclassification and require less information about domain membership. Singh and Mecatti (2011) proposed a generalized modified Horvitz-Thompson (GMHT) estimator for multiple frame estimation based on multiplicity approach. In 2013, they further generalized GMHT class of estimators via regression.

Dasgupta *et al.* (2018) discussed a new sampling theory under dual frame survey using ranked set sampling as a sample selection procedure in each frame. The reason for considering ranked set sampling, in each frame, was its capability to use additional information regarding the ordering of units during sample selection and its growing application as a cost-efficient alternative to simple random sampling. The basic idea behind selecting a rank set sample involves random partition of a collection of sampled units into small groups; a group size (or set-size) of two, three or four units is usually recommended. The units in each group are ranked relative to each other by visual inspections and exactly one unit of each group is chosen for quantification. Though McIntyre (1952) introduced RSS, Takahasi and Wakimoto (1968) provided the mathematical theory for McIntyre's claim. Stokes (1977), Patil *et al.* (1995) discussed the methods of RSS in the context of finite population framework. The major difficulty in using RSS is that, the expression of variance of the RSS estimator is not quite simple under finite population framework and thus its variance estimation procedure is not straight forward. Biswas *et al.* (2013) and Biswas *et al.* (2018) proposed variance estimation procedures using Jackknife and Rescaling Bootstrap techniques respectively in RSS under finite population framework. Fuller and Burmeister (1972), Saxena *et al.* (1986) considered the problem of using complex survey designs in multiple frame surveys. They further suggested that variance estimation in case of multiple frame surveys using complex designs will be very cumbersome. The estimation of variance of a RSS estimator in the context of finite population for multiple frames is much more difficult. Dasgupta *et al.* (2018) proposed Jackknife resampling techniques for that very purpose.

In this paper, we developed two rescaled variance estimation procedures for variance estimation of dual frame RSS estimator using Bootstrap resampling technique. In Section 2, we discuss dual frame ranked set sample estimator (Dasgupta *et al.*, 2018). In Section 3, we discuss proposed variance estimation of DFRSS estimator using Bootstrap resampling technique. Section 4 and 5 provide the simulation study as well as the simulation results and discussion. Concluding remarks are given in Section 6.

2. DUAL FRAME RANKED SET SAMPLE ESTIMATOR

Dasgupta *et al.* (2018) proposed DFRSS estimator under the case when neither of the frames completely covers the population of interest while their union does cover the population. This 100% coverage is sometimes achieved because the population under investigation has been redefined to match the coverage given by the two frames. So, there will be 3 non-overlapping domains, as given in Fig. 1. Therefore, $N_A = N_a + N_{ab}$ and $N_B = N_b + N_{ab}$. For simplicity, the study is restricted to the situation where the frame sizes and the domain sizes are assumed to be known and it is possible to identify the domain to which each item belongs before sampling. Therefore, selecting ranked set samples independently from the two frames will be similar to the situation in which independent sampling is restricted to the domains and then combining the information from the three domains. The choice of frame for sampling in the overlapping domain will depend upon the cost of sampling from each frame. Without loss of generalization, let us assume that C_A and C_B are the cost of selecting a unit from the frames A and B respectively and $C_A > C_B$. We propose to use frame B to select a sample from domain (ab).

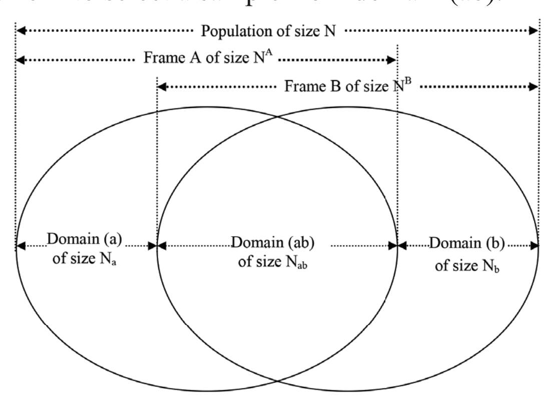

Fig. 1. Population structure resulting from application of two overlapping frames

Under the above considerations, let $\left\{y_{(j:m_a)k}\right\};\left(j=1,2,\ldots,m_a;k=1,2,\ldots,r_a\right)$ denotes the selected ranked set sample of size $n_a = m_a r_a$ belonging to domain (a) using frame A. The integers m_a and r_a are design parameters known as the set-size and replication factor (or number of cycles), respectively. The procedure of selecting a RSS from domain (a) involves the following steps:

i) Select m_a^2 units by simple random sampling without replacement (SRSWOR) from the domain and randomly partition them into m_a subsets, each containing m_a sampling units.

ii) Rank member of each subset according to the character of interest. Quantify the j[th] ranked unit from the j[th] set and denote it as $y_{(j:m_a)};\left(j=1,2,\ldots,m_a\right)$. Thus, we will have m_a quantifications out of m_a^2 selected units.

iii) Repeat the steps i) and ii) r_a times to have $m_a r_a$ quantifications out of $m_a^2 r_a$ selected units.

The RSS estimator of the domain mean $\bar{Y}_a$ is given by

$$\bar{y}_{a,RSS} = \frac{1}{m_a r_a}\sum_{k=1}^{r_a}\sum_{j=1}^{m_a} y_{(j:m_a)k} \quad . \qquad (2.1)$$

They showed that the estimator is unbiased for $\bar{Y}_a$ and its variance is given by

$$V\left[\bar{y}_{a,RSS}\right] = \frac{1}{n_a}\left[\frac{N_a - n_a - 1}{N_a - 1}\sigma_a^2 - \tilde{\gamma}_a\right], \qquad (2.2)$$

where

$$\tilde{\gamma}_a = \frac{\lfloor m_a \lfloor m_a - 1}{N_a\left(N_a-1\right)\ldots\left(N_a-2m_a+1\right)}\gamma ,$$

$$\gamma = \sum_{i=1}^{m_a}\left(\mathbf{Y}-\boldsymbol{\mu}\right)'\boldsymbol{\Gamma}\left(\mathbf{Y}-\boldsymbol{\mu}\right), \ \sum_{i=1}^{m_a}\mathbf{B}_{ii} = \frac{\boldsymbol{\Gamma}}{\begin{pmatrix} N_a \\ m_a, m_a \end{pmatrix}},$$

$$\begin{pmatrix} N_a \\ m_a, m_a \end{pmatrix} = \frac{\lfloor N_a}{\lfloor m_a \lfloor m_a \lfloor\left(N_a - 2m_a\right)},$$

and $\boldsymbol{\Gamma}$ is a $N_a \times N_a$ symmetric matrix with zeros on the diagonal as $B_{jj'}^{ii} = 0, \forall i = 1,2,\ldots,N_a$.

Let $\bar{y}_{b,RSS}$ and $\bar{y}_{ab,RSS}$ are the ranked set sample means of the domains (b) and (ab) respectively. After combining the information from three domains, the dual frame ranked set estimator of the population total Y can be given by the following expression

$$\hat{Y}_{DFRSS} = N_a\bar{y}_{a,RSS} + N_{ab}\bar{y}_{ab,RSS} + N_b\bar{y}_{b,RSS} \ . \qquad (2.3)$$

The variance of dual frame estimator $\hat{Y}_{DFRSS}$ is given by

$$V\left[\hat{Y}_{DF,RSS}\right] = \frac{N_a^2}{n_a}\left[\frac{N_a - n_a - 1}{N_a - 1}\sigma_a^2 - \tilde{\gamma}_a\right] + \frac{N_{ab}^2}{n_{ab}}\left[\frac{N_{ab} - n_{ab} - 1}{N_{ab} - 1}\sigma_{ab}^2 - \tilde{\gamma}_{ab}\right] + \frac{N_b^2}{n_b}\left[\frac{N_b - n_b - 1}{N_b - 1}\sigma_b^2 - \tilde{\gamma}_b\right]$$

$$= \left\{\frac{N_a^2}{n_a}\left(\sigma_a^2 - \tilde{\gamma}_a\right) + \frac{N_{ab}^2}{n_{ab}}\left(\sigma_{ab}^2 - \tilde{\gamma}_{ab}\right) + \frac{N_b^2}{n_b}\left(\sigma_b^2 - \tilde{\gamma}_b\right)\right\} - \upsilon_0 \qquad (2.4)$$

where σ_b^2, σ_{ab}^2, $\tilde{\gamma}_b$ and $\tilde{\gamma}_{ab}$ can be defined similarly as we defined under domain (a) and $\upsilon_0 = \frac{N_a^2}{N_a - 1}\sigma_a^2 + \frac{N_{ab}^2}{N_{ab} - 1}\sigma_{ab}^2 + \frac{N_b^2}{N_b - 1}\sigma_b^2$ is independent of sample sizes from each frame.

Based on the assumption $C_A > C_B$, the overall cost function can be given by $C = C_A n_a + C_B\left(n_b + n_{ab}\right)$. The optimum sample sizes in each domain, after optimizing the function $\varphi = V\left[\hat{Y}_{DFRSS}\right] - \lambda\left[C - C_A n_a + C_B\left(n_b + n_{ab}\right)\right]$, are given under the following table:

Table 1. Allocation of sample sizes over different domains after optimizing the function φ

Domain	Optimum sample size	
	Fixed overhead cost	**Fixed precision**
a (Frame A)	$n_a = \frac{C_0}{\Phi}\frac{\phi_a}{\sqrt{C_A}}$	$n_a = \frac{\Phi}{\left(V_0 + \upsilon_0\right)}\frac{\phi_a}{\sqrt{C_A}}$
ab (Frame B)	$n_{ab} = \frac{C_0}{\Phi}\frac{\phi_{ab}}{\sqrt{C_B}}$	$n_{ab} = \frac{\Phi}{\left(V_0 + \upsilon_0\right)}\frac{\phi_{ab}}{\sqrt{C_B}}$
b (Frame B)	$n_b = \frac{C_0}{\Phi}\frac{\phi_b}{\sqrt{C_B}}$	$n_b = \frac{\Phi}{\left(V_0 + \upsilon_0\right)}\frac{\phi_b}{\sqrt{C_B}}$

where

$$\Phi = \sqrt{C_A}\phi_a + \sqrt{C_A}\left(\phi_b + \phi_{ab}\right),$$

$$\phi_a = N_a\left(\sigma_a^2 - \tilde{\gamma}_a\right)^{\frac{1}{2}}, \ \phi_b = N_b\left(\sigma_b^2 - \tilde{\gamma}_b\right)^{\frac{1}{2}} \text{ and}$$

$$\phi_{ab} = N_{ab}\left(\sigma_{ab}^2 - \tilde{\gamma}_{ab}\right)^{\frac{1}{2}}.$$

If the sample is to be selected depending on the population domain sizes only, then the required size of the sample varied over different domains will be as per the following table:

Table 2. Allocation of sample sizes over different domains proportional to domain sizes

Domain (Frame used)	Sample size proportional to domain size	
	Fixed overhead cost	Fixed precision
a (Frame A)	$n_a = \frac{C_0}{\eta} N_a$	$n_a = \frac{\phi_a^2}{(V_0 + \upsilon_0)} \delta_a$
ab (Frame B)	$n_{ab} = \frac{C_0}{\eta} N_{ab}$	$n_{ab} = \frac{\phi_{ab}^2}{(V_0 + \upsilon_0)} \delta_{ab}$
b (Frame B)	$n_b \quad — N_b$	$n_b = \frac{\phi_b^2}{(V_0 + \upsilon_0)} \delta_b$

where

$$\eta = C_A N_A + C_B N_B, \ \delta_a = \frac{\phi_{ab}^2}{N_{ab}} + \frac{\phi_b^2}{N_b},$$

$$\delta_{ab} = \frac{\phi_a^2}{N_a} + \frac{\phi_b^2}{N_b} \text{ and } \delta_b = \frac{\phi_a^2}{N_a} + \frac{\phi_{ab}^2}{N_{ab}}.$$

3. VARIANCE ESTIMATION OF DFRSS ESTIMATOR USING BOOTSTRAP RESAMPLING TECHNIQUE

In this study, we propose the procedures to estimate $V\left[\hat{Y}_{DFRSS}\right]$ using resampling techniques in the context of RSS estimator in finite population. From equation (2.4), the variance of $\hat{Y}_{DFRSS}$ can be rewritten as

$$V\left[\hat{Y}_{DFRSS}\right] \cong \frac{N_a^2}{n_a}\left(\sigma_a^2 - \tilde{\gamma}_a\right) + \frac{N_{ab}^2}{n_{ab}}\left(\sigma_{ab}^2 - \tilde{\gamma}_{ab}\right) + \frac{N_b^2}{n_b}\left(\sigma_b^2 - \tilde{\gamma}_b\right) \tag{3.1}$$

In the context of RSS without replacement sampling from finite population, it is a quite difficult to apply usual rescaled Bootstrap methods. Biswas *et al.* (2018) proposed Boots trap approaches to develop the variance estimation procedures in ranked set sampling under finite population framework. Following the lines of their variance estimation procedure, we develop two different approaches to estimate the variance $V\left[\hat{Y}_{DFRSS}\right]$.

3.1 Cluster based Rescaling Bootstrap approach

Under this approach, the final ranked set sample comprising of individual samples from each frame is hypothesized as cluster samples, with cycles as clusters and each rank as its elements. Thus, all the clusters in a domain have equal number of units. The proposed Cluster Based Rescaling Bootstrap (CBRB) Method for RSS under dual frame survey is as follows:

1. Draw a resample of $r_{a,1}\left(< r_a\right)$ clusters by simple random sample without replacement (SRSWOR) from the final RSS sample that belongs to domain (a) and observe all the elements in the selected clusters to obtain a bootstrap sample, $\left\{y^*_{(j:m_a)k}\right\}; \left(j = 1, 2, \ldots, m_a;\ k = 1, 2, \ldots, r_{a,1}\right)$.
2. Compute the following terms:

$$\tilde{y}_{(j:m_a)k} = \bar{y}_{a,RSS} + f_{a,1}^{1/2}\left(y^*_{(j:m_a)k} - \bar{y}_{a,RSS}\right)$$

$$\tilde{\bar{y}}_{a,k} = \frac{1}{m_a}\sum_{j=1}^{m_a} \tilde{y}_{(j:m_a)k}$$

$$\tilde{\bar{y}}_{a,RSS,cl} = \frac{1}{r_{a,1}}\sum_{k=1}^{r_{a,1}} \tilde{\bar{y}}_{a,k} \tag{3.2}$$

where

$$\bar{y}_{a,RSS} = \frac{1}{m_a r_a}\sum_{k=1}^{r_a}\sum_{j=1}^{m_a} y_{(j:m_a)k} = \frac{1}{r_a}\sum_{k=1}^{r_a}\bar{y}_k \quad \text{and}$$

$$f_{a,1} = \frac{r_{a,1}}{r_a - r_{a,1}} \frac{N_a - n_a - 1}{N_a - 1}.$$

Similarly, compute the above functions, independently for other two domains.

3. Independently replicate step 1-2 a large number, say B, times and calculate the corresponding $\tilde{\bar{y}}^1_{a,RSS,cl}, \tilde{\bar{y}}^2_{a,RSS,cl}, \ldots, \tilde{\bar{y}}^B_{a,RSS,cl}$.
4. Bootstrap variance estimator of $\tilde{\bar{y}}_{a,RSS,cl}$ is given by

$$\hat{\hat{V}}_{a,cl(b)} = V_*\left[\tilde{\bar{y}}_{a,RSS,cl}\right] = E_*\left[\tilde{\bar{y}}_{a,RSS,cl} - E_*\left(\tilde{\bar{y}}_{a,RSS,cl}\right)\right]^2 \tag{3.3}$$

with its Monte-Carlo approximation, $\hat{\hat{V}}'_{a,cl(b)}$ as

$$\hat{\hat{V}}'_{a,cl(b)} = \frac{1}{B-1}\sum_{b=1}^{B}\left(\tilde{\bar{y}}^b_{a,RSS,cl} - \bar{\tilde{\bar{y}}}_{a,RSS,cl}\right)^2 \tag{3.4}$$

where

$$\bar{\tilde{\bar{y}}}_{a,RSS,cl} = \frac{1}{B}\sum_{b=1}^{B}\tilde{\bar{y}}^b_{a,RSS,cl}$$

5. Similarly, repeat the steps 1-4 and obtain the above functions for domains (ab) and (b) respectively.

Using the equations (3.2) and (3.3), the variance of $\tilde{\tilde{y}}_{a,RSS,cl}$ for domain (a) can be simplified as

$$\hat{V}_{a,cl(b)} = V_*\left[\tilde{\tilde{y}}_{a,RSS,cl}\right] = V_*\left[\frac{f_{a,1}^{1/2}}{m_a r_{a,1}}\sum_{k=1}^{r_{a,1}}\sum_{j=1}^{m_a} y^*_{(j:m_a)k} + \left(1-f_{a,1}^{1/2}\right)\bar{y}_{a,RSS}\right]$$

$$= \frac{f_{a,1}}{r_a-1}\left(\frac{1}{r_{a,1}}-\frac{1}{r_a}\right)\sum_{k=1}^{r_a}\left(\tilde{\tilde{y}}_{a,k}-\bar{y}_{a,RSS}\right)^2$$

$$= \frac{N_a-n_a-1}{r_a\left(r_a-1\right)\left(N_a-1\right)}\sum_{k=1}^{r_a}\left(\tilde{\tilde{y}}_{a,k}-\bar{y}_{a,RSS}\right)^2 \qquad (3.5)$$

Taking expectation on equation (3.5), we get

$$E\left[\hat{V}_{a,cl(b)}\right] = \frac{N_a-n_a-1}{r_a\left(r_a-1\right)\left(N_a-1\right)}E\left[\sum_{k=1}^{r_{a,1}}\left(\tilde{\tilde{y}}_{a,k}-\bar{y}_{a,RSS}\right)^2\right]$$

$$= \frac{N_a-n_a-1}{r_a\left(r_a-1\right)\left(N_a-1\right)}\left[\frac{\left(r_a-1\right)\left(N_a-1\right)}{m_a\left(N_a-1\right)}\sigma_a^2 - \frac{\left(r_a-1\right)}{m_a}\tilde{\gamma}_a\right]$$

$$= \frac{N_a-n_a-1}{n_a\left(N_a-1\right)}\left(\sigma_a^2-\tilde{\gamma}_a\right) \qquad (3.6)$$

Combing the bootstrap estimates from individual domains using equation (2.3), we finally obtain the bootstrap estimate of population total Y as,

$$\hat{\tilde{Y}}_{DFRSS,cl} = N_a\tilde{\tilde{\bar{y}}}_{a,RSS,cl} + N_{ab}\tilde{\tilde{\bar{y}}}_{ab,RSS,cl} + N_b\tilde{\tilde{\bar{y}}}_{b,RSS,cl} \qquad (3.7)$$

and expected value of its variance estimate is given by

$$E\left[V\left(\hat{\tilde{Y}}_{DFRSS,cl}\right)\right] = \frac{N_a^2\left(N_a-n_a-1\right)}{n_a\left(N_a-1\right)}\left(\sigma_a^2-\tilde{\gamma}_a\right) + \frac{N_{ab}^2\left(N_{ab}-n_{ab}-1\right)}{n_{ab}\left(N_{ab}-1\right)}\left(\sigma_{ab}^2-\tilde{\gamma}_{ab}\right) + \frac{N_b^2\left(N_b-n_b-1\right)}{n_b\left(N_b-1\right)}\left(\sigma_b^2-\tilde{\gamma}_b\right)$$

$$\cong \frac{N_a^2}{n_a}\left(\sigma_a^2-\tilde{\gamma}_a\right) + \frac{N_{ab}^2}{n_{ab}}\left(\sigma_{ab}^2-\tilde{\gamma}_{ab}\right) + \frac{N_b^2}{n_b}\left(\sigma_b^2-\tilde{\gamma}_b\right)$$

$$\cong V\left[\hat{Y}_{DFRSS}\right]$$

Hence, the proposed estimator of variance following Cluster Based Rescaling Bootstrap (CBRB) method is approximately unbiased for the variance of proposed RSS estimator $\hat{Y}_{DF,RSS}$ under dual frame survey.

3.2 Strata based Rescaling Bootstrap approach

In this approach, the ranks in final ranked set sample are considered to form artificial strata, with cycle observations as its elements. Thus, all the strata in a particular domain have equal number of units. For example, the m_a ranks in the final sample belonging to domain (a) forms m_a artificial strata, and in each stratum there will be r_a observations. The proposed Strata Based Rescaling Bootstrap (SBRB) Method for RSS under dual frame survey is as follows:

1. Draw $r_{a,2}\left(<r_a\right)$ units from each stratum (independently) by SRSWOR from the final RSS sample that belongs to domain (a) to obtain a bootstrap sample $\left\{y^*_{(j:m_a)k}\right\};\left(j=1,2,\ldots,m_a;\ k=1,2,\ldots,r_{a,2}\right)$.
2. Compute the following functions for j^{th} stratum

$$\tilde{y}_{(j:m_a)k} = \bar{y}_{a,(j)} + f_{a,2}^{1/2}\left(y^*_{(j:m_a)k} - \bar{y}_{a,(j)}\right)$$

$$\tilde{\tilde{y}}_{a,(j)} = \frac{1}{r_{a,2}}\sum_{k=1}^{r_{a,2}}\tilde{y}_{(j:m_a)k} \qquad (3.8)$$

1. Independently obtain functions in equation (3.8) for all ranks and compute

$$\tilde{\tilde{y}}_{a,RSS,st} = \frac{1}{m_a}\sum_{j=1}^{m_a}\tilde{\tilde{y}}_{a,(j)}$$

where

$$\bar{y}_{a,(j)} = \frac{1}{r_{a,2}}\sum_{k=1}^{r_{a,2}}y_{(j:m_a)k} \quad \text{and}$$

$$f_{a,2} = \frac{r_{a,2}}{r_a-r_{a,2}}\frac{N_a-n_a-1}{N_a-1}.$$

2. Independently replicate step 1-2 a large number, say B, times and calculate the corresponding $\tilde{\tilde{y}}^1_{a,RSS,st},\tilde{\tilde{y}}^2_{a,RSS,st},\ldots,\tilde{\tilde{y}}^b_{a,RSS,st},\ldots,\tilde{\tilde{y}}^B_{a,RSS,st}$.
3. Bootstrap variance estimator of $\tilde{\tilde{y}}_{a,RSS,st}$ is given by

$$\hat{V}_{a,st(b)} = V_*\left[\tilde{\tilde{y}}_{a,RSS,st}\right] = E_*\left[\tilde{\tilde{y}}_{a,RSS,st} - E_*\left(\tilde{\tilde{y}}_{a,RSS,st}\right)\right]^2 \qquad (3.9)$$

with its Monte-Carlo approximation, $\hat{V}'_{a,st(b)}$ as

$$\hat{V}'_{a,st(b)} = \frac{1}{B-1}\sum_{b=1}^{B}\left(\tilde{\tilde{y}}^b_{a,RSS,st} - \tilde{\tilde{\bar{y}}}_{a,RSS,st}\right)^2 \qquad (3.10)$$

where

$$\tilde{\tilde{\bar{y}}}_{a,RSS,st} = \frac{1}{B}\sum_{b=1}^{B}\tilde{\tilde{y}}^b_{a,RSS,st}.$$

Similarly, repeat the steps 1-4 and obtain the above functions for domains (ab) and (b) respectively.

Using the equations (3.8) and (3.9), the variance of $\tilde{\tilde{\bar{y}}}_{a,RSS,\,st}$ for domain (a) can be simplified as

$$\hat{\hat{V}}_{a,st(b)} = V_*\left[\tilde{\tilde{\bar{y}}}_{a,RSS,\,st}\right] = V_*\left[\frac{f_{a,2}^{1/2}}{m_a r_{a,2}}\sum_{k=1}^{r_{a,2}}\sum_{j=1}^{m_a} y^*_{(j:m_a)k} + \left(1-f_{a,2}^{1/2}\right)\bar{y}_{a,RSS}\right]$$

$$= \frac{f_{a,2}}{m_a^2\left(r_a-1\right)}\left(\frac{1}{r_{a,2}}-\frac{1}{r_a}\right)\sum_{j=1}^{m_a}\left[\sum_{k=1}^{r_{a,2}} y^2_{(j:m_a)k} - r_{a,2}\,y^2_{a,(j)}\right]$$

$$= \frac{N_a-n_a-1}{m_a^2 r_a\left(r_a-1\right)\left(N_a-1\right)}\sum_{j=1}^{m_a}\left[\sum_{k=1}^{r_{a,2}} y^2_{(j:m_a)k} - r_{a,2}\,y^2_{a,(j)}\right] \quad (3.11)$$

Taking expectation on both sides of equation (3.11), we get

$$E\left[\hat{\hat{V}}_{a,st(b)}\right] = \frac{N_a-n_a-1}{m_a^2 r_a\left(r_a-1\right)\left(N_a-1\right)} E\left[\sum_{j=1}^{m_a}\left(\sum_{k=1}^{r_{a,2}} y^2_{(j:m_a)k} - r_{a,2}\,y^2_{a,(j)}\right)\right]$$

$$= \frac{N_a-n_a-1}{n_a\left(N_a-1\right)}\left(\sigma_a^2-\tilde{\gamma}_a\right) \quad (3.12)$$

Similar to cycle based bootstrap method, after combining the individual estimates from each domain, $\hat{\hat{Y}}_{DFRSS,st} = N_a\tilde{\tilde{\bar{\bar{y}}}}_{a,RSS,\,st} + N_{ab}\tilde{\tilde{\bar{\bar{y}}}}_{ab,RSS,\,st} + N_b\tilde{\tilde{\bar{\bar{y}}}}_{b,RSS,\,st}$, it can be easily shown that the proposed bootstrap variance estimate $V\left(\hat{\hat{Y}}_{DFRSS,st}\right)$ is unbiased for the variance of proposed DFRSS estimator $\hat{Y}_{DF,RSS}$.

4. SIMULATION STUDY

The performance of the proposed variance estimation procedures using Bootstrap method in multiple frame approach using RSS in each frame was examined by carrying out a simulation study. Under the simulation study, a univariate normal population was generated using SAS (Statistical Analysis System) software of size 4000. The mean and variance of the generated univariate normal population are 320 and 28.086 respectively. Size of the domain (a), (ab) and (b) are chosen as 1000, 2000 and 1000 respectively. Further, 500 RSS samples of different sample sizes with different combination of number of cycles and ranks were drawn from the simulated population under dual frame survey. Then, the estimates of proposed RSS estimator under dual frame survey as well as its variance, % CV, skewness and kurtosis were obtained based on estimates from these 500 samples for each sample size separately. At the same time, 500 SRSWOR samples were generated under dual frame survey to compare the RSS scheme with usual SRSWOR scheme for each RSS sample size. Further, percentage gain in efficiency of the RSS estimator dual frame survey with respect to SRSWOR estimator dual frame survey of population mean was obtained using the following expression

$$\%GE = \left[\frac{V\left(\bar{y}_{DFSRS}\right)-V\left(\bar{y}_{DFRSS}\right)}{V\left(\bar{y}_{DFRSS}\right)}\right]\times 100 \quad (4.1)$$

where, $V\left(\bar{y}_{DFRSS}\right)$ and $V\left(\bar{y}_{DFSRS}\right)$ are the variance obtained based on 500 samples for the RSS estimator dual frame survey and usual SRSWOR estimator dual frame survey, respectively. Further, to study the performance of developed variance estimation procedures using Bootstrap method, these procedures were applied on each selected RSS sample for different combination of number of cycles and number of ranks. For this, percentage Relative Bias (%RB) and Relative Stability (RS) of the estimates of variance of RSS estimator of population mean, $v_1(\hat{\bar{\mu}})$, were computed for the proposed approaches. The formula for RB and RS are given by

$$\%RB = \left[\frac{\frac{1}{s}\sum_s v_{1s}(\hat{\bar{\mu}}^*) - V(\bar{y}_{DFRSS})}{V(\bar{y}_{DFRSS})}\right]\times 100 \quad (4.2)$$

and

$$RS = \frac{\sqrt{MSE\left\{v_1(\hat{\bar{\mu}}^*)\right\}}}{MSE(\bar{y}_{DFRSS})} = \frac{\left[\frac{1}{s}\sum_s\left\{v_{1s}(\hat{\bar{\mu}}^*) - V(\bar{y}_{DFRSS})\right\}^2\right]^{1/2}}{V(\bar{y}_{DFRSS})} \quad (4.3)$$

where MSE denotes the mean square error and *s* denotes the number of samples selected for variance estimation. SAS codes were written for selection of ranked set samples and for obtaining variance of the RSS estimator of population mean, estimates of variance, RB and RS for both the approaches in the context of finite population.

5. SIMULATION RESULT AND DISCUSSION

Important statistical properties such as variance, % CV, skewness, kurtosis and percentage gain in efficiency (%GE) of the proposed dual frame ranked set sampling estimator of population total with respect to Hartley's usual SRS estimator were obtained for different set size and number of cycle over two overlapping frames and are presented in Table 3. The graphical representation in terms of variance is shown in the Fig. 2.

The proposed DFRSS estimator of population total is found to be almost unbiased in all the cases considered here. It is evident from Table 3 that the variance of the proposed DFRSS estimator decreases with the increase of sample size as well as when the set size increases for a fixed sample size. In contrary, the percentage gain in efficiency of the proposed DFRSS estimator increases with the increase of sample size as well as when the set size increases for a fixed sample size. With the increase of sample size as well as with the increase of set size for a fixed sample size, the DFRSS estimator becomes more stable in terms of % CV. These results ensure the superiority of DFRSS over Hartley's usual SRS estimator in case of finite population.

The %RB and RS of the proposed Bootstrap variance estimation approaches were obtained for different sample sizes with different combination of set size and number of cycles and are presented in Table 4.

The graphs presented in Fig. 3 and 4 depict the comparison between all the variance estimators in terms of RB and RS:

Table 4 and Fig. 3 show that a considerable reduction in absolute percentage relative bias had been achieved by using the proposed rescaling factors for both the Bootstrap variance estimation procedure as compared with the standard Bootstrap methods. Both the proposed rescaling Bootstrap procedure had shown similar absolute %RB. It is evident from Table 4 that both the rescaled based methods are at par with each other as far as RS is concerned. But with the increase of set size for a fixed sample size, the estimator of the variance obtained following cluster based rescaling methods becomes less stable in most of the cases, however there was no significant fluctuation in RS under strata based rescaled Bootstrap method. It was also observed that, RS decreases as number of cycle increases for a fixed set size (Fig. 4). Therefore, it can be concluded that estimator of the variance obtained following both the proposed rescaled Bootstrap methods are almost comparable with respect to absolute %RB

Table 3. Statistical properties of the DFRSS estimator for different combinations of set size and number of cycle

Set	Cycle	Est. of Pop. Total	Variance	Skewness	Kurtosis	%CV	%GE
2	**30**	1282361.91	73595488.95	0.01	0.05	0.67	0.97
3	**20**	1281702.32	64444783.90	0.63	-0.25	0.63	23.03
4	**15**	1281845.28	47707083.26	0.01	0.32	0.54	29.67
2	**60**	1281864.97	32860853.10	0.00	0.08	0.45	13.07
3	**40**	1281114.18	23931474.13	0.00	0.11	0.38	35.91
4	**30**	1281647.26	23666660.50	0.00	-0.03	0.38	36.82
2	**90**	1281965.14	18777895.68	0.00	-0.14	0.34	27.96
3	**60**	1280859.10	15984961.42	0.00	0.00	0.31	36.20
4	**45**	1282049.99	14961113.42	0.00	0.11	0.30	46.56

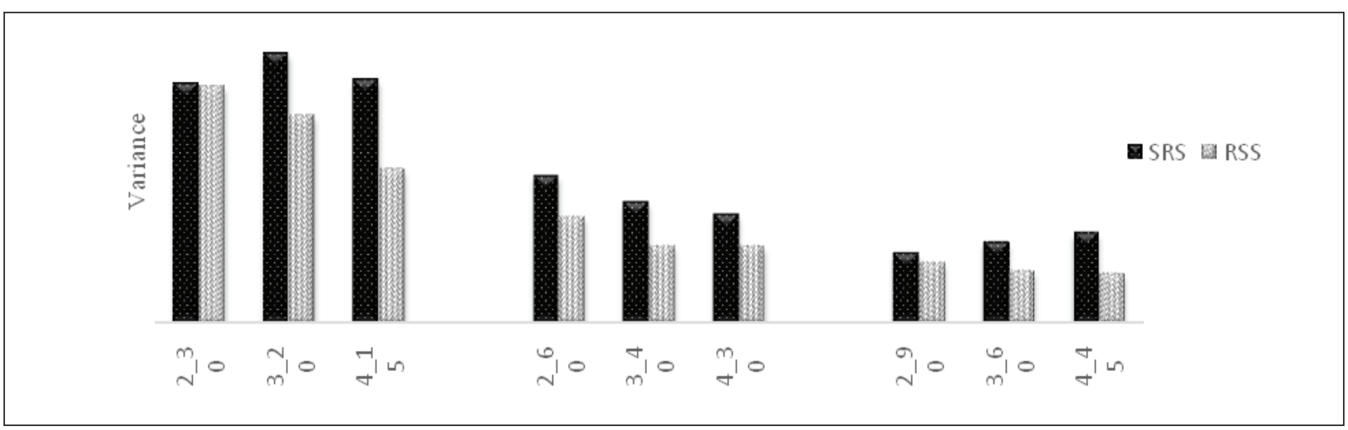

Fig. 2. Variance of the RSS estimator for different combinations of set size and different number of cycles in comparison with Hartley's estimator under dual frame RSS sampling

for different sample sizes with different combination of set size and number of cycles. Proposed strata based rescaling Bootstrap method is superior than proposed cluster based rescaling Bootstrap method as its RS doesn't fluctuate much with variation in sample sizes.

6. CONCLUSION

In this paper, we proposed a Bootstrap variance estimation procedure for dual frame ranked set sample estimator. The standard Bootstrap technique failed to provide an unbiased estimate of sampling variance

Table 4. %RB and RS of Bootstrap estimates of variance of proposed DFRSS estimator for different combinations of set size and different number of cycles

Set	Cycle	Cluster Based				Strata Based			
		Simple		Rescaled		Simple		Rescaled	
		%RB	RS	%RB	RS	%RB	RS	%RB	RS
2	**30**	32.01	0.41	-15.53	0.22	34.38	0.39	-14.89	0.19
	60	52.76	0.69	-6.33	0.04	52.21	0.83	-6.66	0.12
	90	75.86	0.78	3.10	0.12	78.96	0.81	4.97	0.11
3	**20**	36.43	0.50	-20.00	0.29	37.49	0.43	-19.30	0.23
	40	79.84	0.84	10.28	0.20	80.84	0.83	10.86	0.17
	60	69.57	0.74	6.58	0.17	71.72	0.74	7.94	0.13
4	**15**	61.21	0.77	3.16	0.30	63.19	0.67	4.41	0.15
	30	66.66	0.76	2.23	0.23	66.89	0.70	2.32	0.13
	45	77.22	0.83	3.93	0.18	75.80	0.78	3.08	0.11

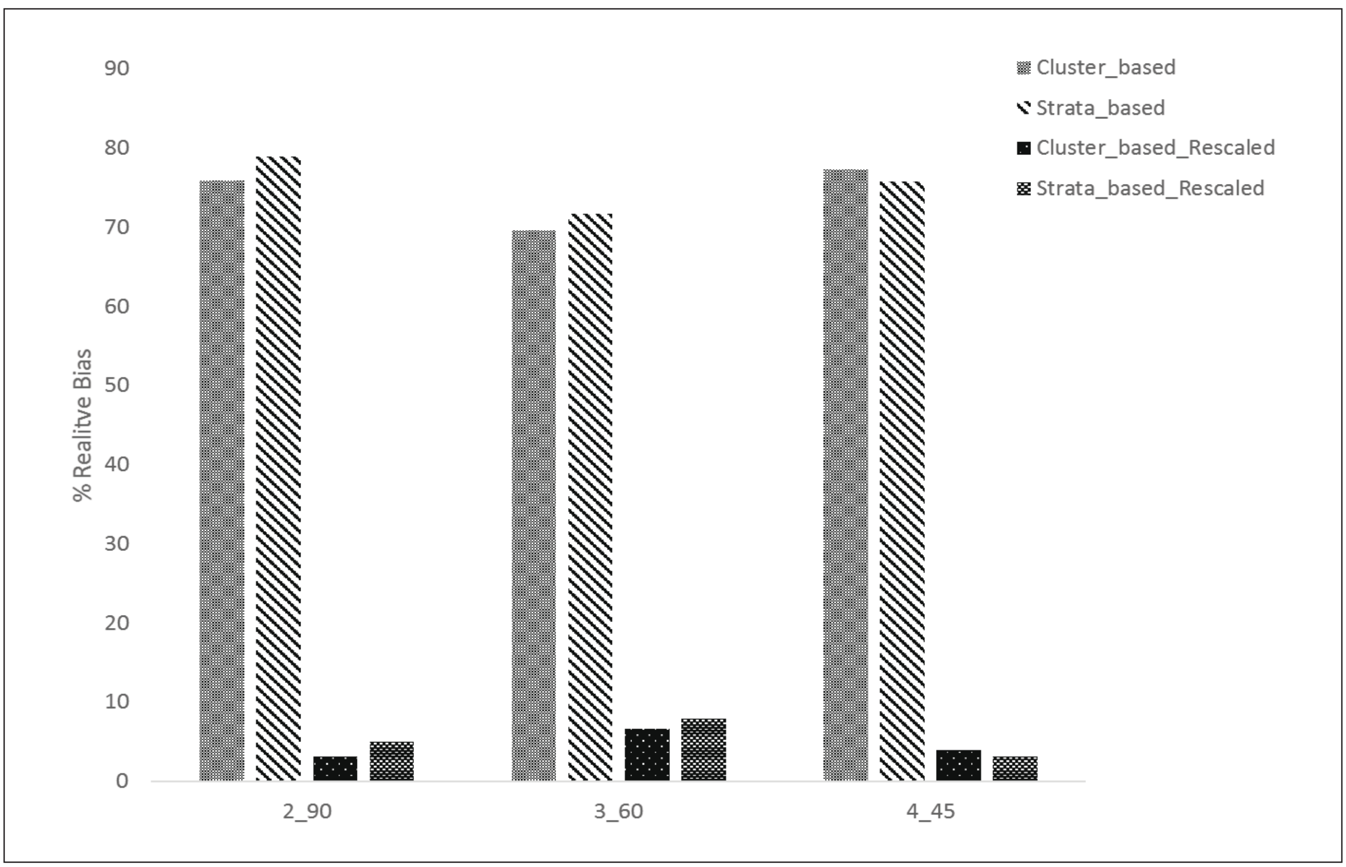

Fig. 3. %RB of Bootstrap estimates of variance of DFRSS estimator for different combinations of set size and different number of cycles when sample allocation to each domain is 180

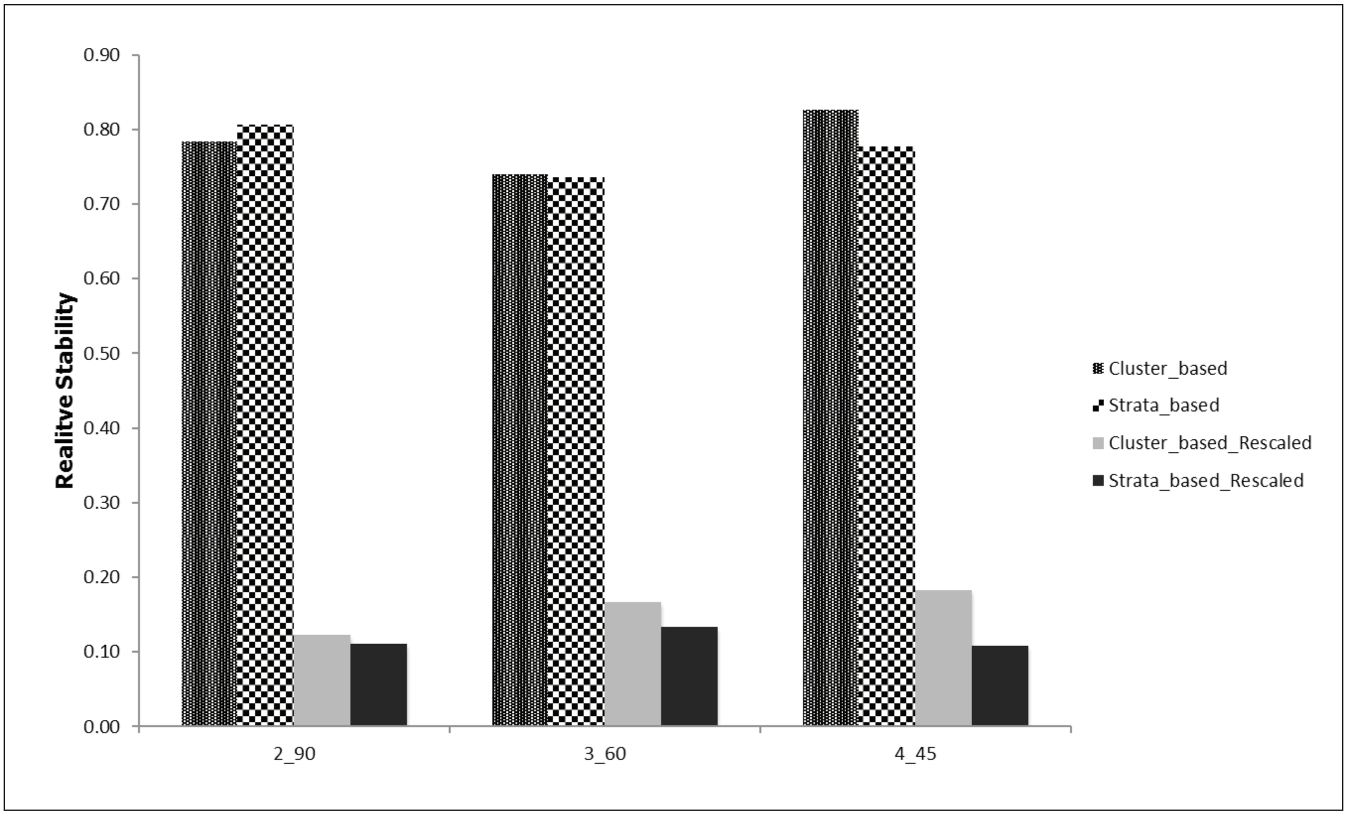

Fig. 4. RS of Bootstrap estimates of variance of DFRSS estimator for different combinations of set size and different number of cycles when sample allocation to each domain is 180

of the DFRSS estimator. We proposed two rescaled Bootstrap methods, viz. cluster based and strata based rescaled bootstrap methods, which are unbiased. The performance of the proposed rescaled Bootstrap methods was investigated through simulation. Both the rescaled Bootstrap methods are more stable than the standard Bootstrap methods in terms of %RB and RS for all tested set of set sizes and number of cycles. However, strata based rescaling Bootstrap method showed better stability than cluster based rescaling Bootstrap method.

ACKNOWLEDGEMENT

The authors want to acknowledge DST, Ministry of Science and Technology, Govt. of India for their financial support to carry out this research.

REFERENCES

Biswas, A., Ahmad, T. and Rai, A. (2013). Variance Estimation using Jackknife Method in Ranked Set Sampling under Finite Population Framework. *J. Ind. Soc. Agril. Statist.*, **67(3)**, 345-353.

Biswas, A., Rai, A. and Ahmad, T. (2018). Rescaling Bootstrap Technique for Variance Estimation for Ranked Set Samples in Finite Population. *Communications in Statistics - Simulation and Computation.* DOI: 10.1080/03610918.2018.1527349.

Dasgupta, P., Ahmad, T., Biswas, A. and Rai, A. (2018). A dual frame approach for estimating finite population total using ranked set sampling. *Int. J. Agril. Statist. Sci.*, **14(1)**, 409-418, 2018.

Fuller, W.A. and Burmeister, L.F. (1972). Estimators for samples selected from two overlapping frames. *Proceedings of the Social Statistics Section*, *Amer. Statist. Assoc.*. 245-249.

Hartley, H.O. (1962). Multiple Frame Surveys. *Proceedings of the Social Statistics Section*, *Amer. Statist. Assoc.*. 203-206.

Hartley, H.O. (1974). Multiple Frame Methodology and Selected Applications. *Sankhya*, Ser. C. **36**, 99-118.

Lohr, S.L. and Rao, J.N.K. (2000). Inference in dual frame surveys. *J. Amer. Statist. Assoc.*. **95**, 271-280.

Lohr, S.L. and Rao, J.N.K. (2006). Estimation in Multiple-Frame Surveys. *J. Amer. Statist. Assoc.*. **101**, 1019-1030.

Mecatti, F. (2007). A single frame multiplicity estimator for multiple frame surveys. *Survey Methodology*, **33**, 151-157.

McIntyre, G.A. (1952). A method for unbiased selective sampling, using ranked sets. *Aust. J. Agril. Res.*. **3**, 385-390.

Patil, G.P., Sinha, A.K. and Taillie, C. (1995). Finite population corrections for ranked set sampling. *Annals of the Institute of Statistical Mathematics*. **47**, 621-636.

Saxena, B.C., Narain, P. and Srivastava, A.K. (1984). Multiple Frame Surveys in two stage sampling. *Sankhya*, Ser. B. **46**, 75-82.

Saxena, B.C., Narain, P. and Srivastava, A.K. (1986). Robustness of Hartley's estimator for multiple frame surveys. *J. Ind. Soc. Agril. Statist.*. **38(3)**, 370-378.

Singh, A.C. and Mecatti, F. (2011). Generalized multiplicity-adjusted Horvitz-Thompson type estimation as a unified approach to multiple frame surveys. *J. Off. Statist.*, **27**, 633-650.

Singh, A.C. and Mecatti, F. (2013). Generalized multiplicity-adjusted multiple frame estimation via regression. Proceedings of Society for Technical Communication Symposium.

Skinner, C.J. (1991). On the efficiency of raking ratio estimation for multiple frame surveys. *J. Amer.* S*tatist. Assoc.*. **86**, 779-784.

Stokes, S.L. (1977). Ranked set sampling with concomitant variables. *Communication in Statistics* (*Theory and Methods*). **6**, 1207-1211.

Takahasi, K. and Wakimoto, K. (1968). On unbiased estimates of the population mean based on the sample stratified by means of ordering. *Annals of the Institute of Statistical Mathematics*. **20(1)**, 1-31.