Pak. J. Statist.
2018 Vol. 34(2), 163-176

RANKED SET SAMPLING APPROACH FOR ESTIMATING RESPONSE OF DEVELOPMENTAL PROGRAMS WITH LINEAR IMPACTS UNDER SUCCESSIVE PHASES

Girish Chandra[1§], Rajiv Pandey[2], D.S. Bhoj[3], Raman Nautiyal[1]
Jawaid Ashraf[2] and **Med Ram Verma[4]**

[1] Division of Forest Statistics, Indian Council of Forestry Research and Education, Dehradun-248006, India.
[2] Indian Council of Forestry Research and Education, Dehradun-248006, India.
[3] Department of Mathematical Sciences, Rutgers University, Camden, USA.
[4] ICAR-Indian Veterinary Research Institute, Bareilly, India.
[§] Corresponding author Email: gchandra23@yahoo.com

SUMMARY

The problem of estimation is motivated by some studies from the area of developmental programs management to be implemented by the government and non-government organizations under different phases. The examples of such programs are educational programs, women empowerment programs for enhancing gross enrollment ratio in a region. There are two important variables in such programs. One of the variables is called the survey variable. The survey variable changes under the impact of the program over different phases and is termed as the response variable. In our study, we consider linear impacts of the programs under successive phases and multiplicative model which relates the response and survey variables. The management of the programs is interested in estimation of the mean of the response variable. Therefore, in this paper, we propose an unbiased estimator of the population mean of the response variable by using ranked set sampling procedure on the survey variable. The estimator is then compared with the estimator based on simple random sample (SRS). The relative precisions of the proposed estimator with the estimator based on SRS are computed for some symmetric and skew distributions. It is shown that the gains in relative precisions of the proposed estimator are substantial for all sample sizes and the chosen distributions. We applied the formulas derived in this paper to one real life example from India's report, Education for all towards quality with equity, of National University of Education Planning and Administration, New Delhi, 2014.

KEYWORDS

Impact evaluation; order statistics; relative precision; skewed distribution; symmetric distribution.

© 2018 Pakistan Journal of Statistics

1. INTRODUCTION

Development programs are implemented by the government and nongovernment organisations for upgrading or improving the desired characteristic over time on the desired unit in a region or community. Main objective of such programs are to achieve the desired impact of the program's related characteristic. For example, programs, such as, mid-day meal for increasing Gross Enrollment Ratio (GER) in the school; eradication of poverty of community; eradication of polio of children up to age five, AIDS awareness program are part of developmental programs. These programs may be implemented in different phases, such as years, depending upon the volume, scope as well as the geographical spread of the units, on which the program has to be implemented.

The phased implementation of the program leads to different impacts on the units of the different phases. These different phases may be treated as different stratum. All units within a stratum may have similar impacts on the desired characteristic, and they may differ with units in another stratum. Therefore, estimating desired population parameters without consideration of lagged impacts of the units may be rather erroneous.

The problem of incorrect estimation may be addressed by giving weights to the impacts under different phases, and accordingly modifying the allocation of units for each phase. The weight may be based on the pattern of the impact under different phases. Two types of variables of the developmental program are being considered in this situation based upon the implementation with lagged impact and non-implementation without lagged impact.

The desired characteristic of the unit without implementation of the program is termed as the survey variable, S. The implementation of the program has impacted on S, and therefore the actual value of S gets influenced and attained different values depending upon the nature of the impact. More precisely, the change in values of the characteristic under the impact of program is different than the original value of S and then it is defined as the response variable, R. As an illustration, we take a Government of India's developmental program launched in 1988 named 'The National Literacy Mission' (NLM) with aims to educate 80 million adults in the age group of 15 - 35 over an eighty-year period. The voluminous program is implemented in different years across the scale of the program. One might be interested to know the response of this program on the illiterate people over a particular period of time, say 10 years. In this case, S may be taken as the amount of gain attained by the person belonging to a particular category where category may be defined with age, geographical region etc., based upon their personal experience over time but not due to the effect of implemented program. The gain may be defined by an appropriate numerical characteristic, such as score obtained in an examination conducted for these category persons. As the gain depends upon the time, the values of the S changes over the phases. R is the variable with same numerical characteristics as taken for the S in addition to considering the effects of the program's impact. Obviously, R is an increasing function with respect to the number of phases as the impact of the program increases in every successive phase.

The proportion of the gain attained with and without the program is considered as the 'impact' of the program (Pandey, 2010) for any particular phase. For the different purposes, various cases of pattern of the impacts had been reported such as additive impact (Pandey and Verma, 2008), multiplicative impact (Pandey, 2010), Binomial impact (Pandey *et al*., 2012), exponential impact (Verma *et al*., 2012).

The present case considers that the program has been implemented over the various phases and the impacts are linearly proportional to the phases. More precisely, the linear impact is being considered as following the arithmetic progression with respect to successive phases for the sake of easy calculations and generating further advanced ideas. However, other advanced linear or non-linear model can also be used for the improvement of response estimates. Therefore, the estimation of characteristics of R will be based on the realised impacts under successive phases. Our aim is to estimate R over k phases of the program with the known nature of the S, such as, lognormal distribution with known parameters.

For the estimation purposes various sampling techniques depending upon the nature of the population and objectives had been reported in literature (Cochran, 1977). Ranked set sampling (RSS), introduced by McIntyre (1952), is one of the techniques to estimate the population parameters when measurements of all sampling units is costly, but small sets of units can be ranked according to the value of the characteristic under study by means of visual inspection or other methods not requiring actual measurements. The present case is a suitable candidate for RSS, as the rank may be easily assigned to the units based upon some concomitant variable like impact.

RSS has been satisfactorily used in estimating pasture and forage yield (McIntyre, 1952, 1978), bone mineral density in a human population (Nahhas *et al*., 2002), health and nutritional examination (Chen *et al*., 2003; 2005); single family homes sales prices (Chen *et al*., 2004), forests, grassland, and other vegetation resources (Johnson *et al*., 1993); crop production (Husby *et al*., 2005); market and consumer surveys (Kowalczyk, 2005). A method of RSS in the first phase of adaptive cluster sampling is introduced by Chandra *et al*. (2011) for obtaining precise estimates pertaining to rare and endangered species. For some recent developments in RSS, one may refer to Khan *et al*. (2016), Zamanzade and Mahdizadeh (2016) and Khan and Shabbir (2016).

In the following Sections, the assumptions about the program having the successive impacts over different phases and estimation methods of R using S are discussed. In section 2, we derived the relations between the means and variances of S and R under SRS procedure. The RSS procedure is also reviewed in the context of the present problem. In Section 3, we derived an unbiased estimator of the population mean of R based on RSS procedure on S. We also derived the formula of the relative precision of our estimator with estimator based on RSS. Section 4 gives the numerical values of relative precisions for symmetric and skew distributions. In Section 4 we also give the applications of our formulas to the real world development program. Section 5, summarizes the results of the paper with discussion.

2. ESTIMATION OF THE MEAN OF R AND S

Suppose, a particular development program is implemented over k phases and the impacts of the program are assumed to follow linear model over the phases. More specifically we consider that the successive impacts of the program follows arithmetic progression, that is, the impact of the i^{th} phase out of k phases is defined by

$$I_{[i:k]} = a + (i\text{-}1)d \text{ , for any real positive numbers } a \text{ and } d \tag{1}$$

Impacts have cumulative effect with respect to the successive phases and therefore will be in ascending order with lowest impact at first phase which is implementation year of the program and the highest impact at the k^{th} phase which is concluding year of the program. Here we deal with the main variable, R and concomitant variable, S and impact variable, I . Then the multiplicative model with concomitant variable is chosen by following Chen *et al*. (2004) as:

$$R = SI + \varepsilon \tag{2}$$

where, S , R and I , represents the vector of S , R and I , respectively for all successive phases of the program and ε is a vector of random error with mean 0 and unknown variance σ_e^2 and is independent of S .

With the model (2), the NLM example discussed in Section 1 can be understood with the following hypothetical values for 5 phased program and error term as zero:

Table 2.1

An Empirical Example showing the Values of S , R and I with $d = 0.50$.

Phase No. i	R	S	I
1	25	25	1
2	39	26	1.50
3	52	26	2.00
4	70	28	2.50
5	90	30	3.00

With this example the impact value for the first phase of R and S are the same and i equal to 1 implies that $a = 1$. The values given in the tables show the hypothetical scores obtained in the presence and absence of the program for the well-defined category of persons.

Let $\left(S_1, S_2, ...S_k\right)$ be a simple random sample of size k on S with population mean μ_S and a finite population variance σ_S^2 irrespective of the phases.

The standard unbiased estimator of μ_S is

$$\hat{\mu}_S = \frac{1}{k}\sum_{i=1}^{k} S_i \text{ with } Var\left(\hat{\mu}_S\right) = \sigma_S^2 / k .$$

Let μ_R denote the population mean of R. Then μ_R can be written using (2) as

$$\mu_R = C\mu_S , \tag{3}$$

where, $C = \frac{\left(2a+(k-1)d\right)}{2}$ denotes the average impact for all k phases.

Then, the unbiased estimator of μ_R in terms of μ_S can be written as:

$$\hat{\mu}_R = \frac{1}{k}\sum_{i=1}^{k}\left(a+(i-1)d\right)S_i ,$$

with

$$Var\left(\hat{\mu}_R\right) = \frac{\sigma_R^2}{k} = \frac{\sum_{i=1}^{k}\left(a+(i-1)d\right)^2}{k}\sigma_S^2$$

$$= \frac{\sigma_S^2}{k}D , \tag{4}$$

where, $D = a^2 + (k-1)ad + \frac{(k-1)(2k-1)}{6}d^2$.

In practice, it is difficult to measure variable S as measuring the units are costly and time consuming. Therefore, we use the RSS procedure on the survey variable, S, and then estimate the mean of the response variable, R.

2.1 Ranked Set Sampling (RSS) Procedure

RSS is a method for improving precision in estimation of population mean using ranking of the units based on some concomitant variable. Ranking of the units are rather easy and inexpensive in comparison to the actual measurement of the units (McIntyre, 1952). The RSS approach facilitates for the impact evaluation of R by considering ranking of observations based on the realized impacts of the phases. A brief procedure of RSS on the variable S is explained in the next paragraph.

A simple random sample of k units from the population of the survey variable is selected and the units are ranked on the basis of some concomitant variable including the impact variable I, without actual measurements. Then the unit with the lowest rank is quantified. Then the second simple random sample of size k is selected and the units are ranked as before. Then the unit with the second lowest rank is quantified. This procedure is continued until we quantify the observation with the highest rank from the k^{th} simple random sample. This completes one cycle of RSS procedure. The cycle then may be repeated m times to get a balanced ranked set sample of size $n = km$. Without loss of generality, it is assumed that there is perfect ranking i.e. within each cycle one unit belongs from each rank order statistics (phase) and hence m ranked units from each phase are included in the measurement of S. For the sake of easy calculations we have chosen $m = 1$.

The k^2 ordered observations in the k samples can be displayed as

$$\begin{array}{l} S_{(11)}, S_{(12)}, \ldots S_{(1k)} \\ S_{(21)}, S_{(22)}, \ldots S_{(2k)} \\ \qquad \vdots \\ S_{(k1)}, S_{(k2)}, \ldots S_{(kk)} \end{array}$$

The observations $S_{(11)}, S_{(22)}, \ldots S_{(kk)}$ are actually measured accurately on S. Takahasi and Wakimoto (1968) proved that the relative precision (RP) of balanced RSS with respect to SRS lies between 1 and $(k+1)/2$. Dell and Clutter (1972) also showed that the RSS estimator is more precise than SRS estimator even in the presence of ranking errors. The issues on unbalanced RSS are also discussed by various authors in various aspects such as Kaur *et al*. (1997), Bhoj (2001), Chen *et al*. (2004) and Tiwari and Chandra (2011).

Let $S_{(i:k)}\left(\equiv S_{(ii)}\right)$ and $R_{(i:k)}$, $i=1(1)k$, denote the values of S and R, respectively for the unit taken for measurement belonging to i^{th} rank order. Under the multiplicative model (2), we have for the i^{th} phase:

$$R_{(i:k)} = S_{(i:k)} \times I_{[i:k]} + \varepsilon_{(i:k)} \tag{5}$$

where, $\varepsilon_{(i:k)}$ is the random error term with mean 0 and unknown variance σ_e^2 and independent of $S_{(i:k)}$.

Suppose, $\mu_{R(i:k)}$ and $\mu_{S(i:k)}$ denote the population mean and $\sigma^2_{R(i:k)}$ and $\sigma^2_{S(i:k)}$ denote the population variances of R and S respectively for the i^{th} phase. For fixed i, realizations corresponding to the i^{th} phase of both the variables R and S are independently and identically distributed with respective means $\mu_{R(i:k)}$, $\mu_{S(i:k)}$ and variances $\sigma^2_{R(i:k)}$, $\sigma^2_{S(i:k)}$ respectively.

3. PROPOSED ESTIMATOR FOR THE MEAN OF RESPONSE VARIABLE

In analogous to McIntyre (1952), we propose the following estimator of μ_R :

$$\bar{R}_{(k)RSS} = \frac{1}{k}\sum_{i=1}^{k} C S_{(i:k)} , \tag{6}$$

where C is given by (3).

Using the property that $\sum_{i=1}^{k} E\left(S_{(i:k)}\right) = k\mu_S$ of RSS, it is easy to verify that

$$E\left(\bar{R}_{(k)RSS}\right)=\mu_R .$$

The variance of the estimator is

$$Var\left(\bar{R}_{(k)RSS}\right)=\frac{C^2}{k}\frac{\sum_{i=1}^{k}\sigma^2_{S(i:k)}}{k} .$$

The RP of the proposed estimator $\bar{R}_{(k)RSS}$ with SRS estimator $\hat{\mu}_R$ is

$$RP=\frac{Var\left(\hat{\mu}_R\right)}{Var\left(\bar{R}_{(k)RSS}\right)}=\frac{D}{C^2}\left(\frac{\sigma^2_S}{\frac{\sum_{i=1}^{k}\sigma^2_{S(i:k)}}{k}}\right).$$

Substituting the values of D and C from (3) and (4), we get

$$RP=\frac{a^2+(k-1)ad+\frac{(k-1)(2k-1)}{6}d^2}{a^2+(k-1)ad+\frac{(k-1)^2}{4}d^2}\left(\frac{\sigma^2_S}{\frac{\sum_{i=1}^{k}\sigma^2_{S(i:k)}}{k}}\right). \tag{7}$$

It is clear that $\frac{a^2+(k-1)ad+\frac{(k-1)(2k-1)}{6}d^2}{a^2+(k-1)ad+\frac{(k-1)^2}{4}d^2}>1$, since the first two terms in the numerator and denominator are same and the third term of numerator is greater than the third term of denominator with difference $\frac{\left(k^2-1\right)d^2}{12}$.

Further, under RSS we have the result that $\sigma^2_S>\frac{\sum_{i=1}^{k}\sigma^2_{S(i:k)}}{k}$.

This implies that the RP is the product of two terms, where each term is greater than one. RP increases as k increases for fixed d, and it also increases as d increases with fixed k. In our study we already assume that the value of d is fixed. To see the relation between d and RP for fixed values of a and k, we have plotted the graph for standard normal distribution $N\left(0,1\right)$ as given in Fig. 3.1. From this figure it is seen that as the value of d increases the corresponding RP also increases rapidly. However, increment in RP is minimal when d reaches certain value. Similar graphs have been observed for the other distributions of S.

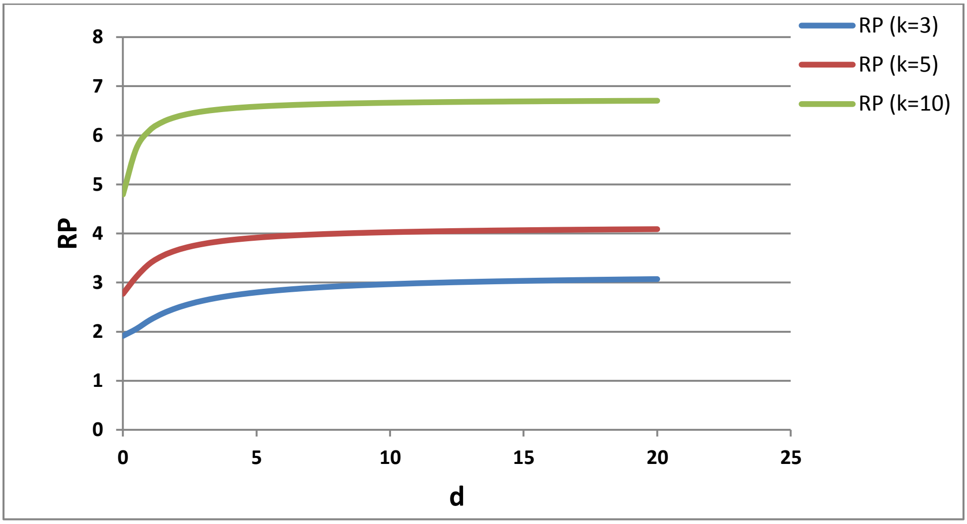

Fig 3.1: Relation between RP and d for N (0, 1) for $k=3, 5, 10$ when $a=1$

4. NUMERICAL EXAMPLES

4.1 Application to Real Developmental Program

We have considered one real life example from India's report, "Education for all towards quality with equity", of National University of Educational Planning and Administration, New Delhi, 2014. This report was based on the developmental program launched with the aim of providing free and compulsory education to all children of India. Under this program the GER of Scheduled Tribes (ST) students in secondary education during the period 2004 and 2008 was recorded by the government based upon the estimates given by the schools of the country. However, in actual practice, it is difficult to get the information from all the schools of India in time, and it is time consuming also. The GER performance for boys and girls of Class IX between the ages 14-15 years are shown in the Table 4.1.

Table 4.1
GER in Secondary Education (2004 to 2008)

Phase no.	GER of ST students	
	Boys	Girls
2004 (i=1)	43.3	30.5
2005 (i =2)	44.7	33.0
2006 (i =3)	47.5	35.6
2007 (i = 4)	48.8	37.2
2008 (i = 5)	51.8	40.9

The values of impact of the program in each successive phases can be found using equation (1). From Table 4.1, we transformed all GER values in the form of impacts by dividing all values of GER by its first phase impact value. Then the average of transformed values may be equated to average impact value, $C = \frac{\left(2a + (k-1)d\right)}{2}$. This gives the value of d at $a = 1$ which are shown in Table 4.2.

Table 4.2
GER Transformation into Impact

Phase No. (*i*)	**Transformed GER for Calculating Impacts**	
	Boys	**Girls**
1	1	1
2	1.032	1.082
3	1.097	1.167
4	1.127	1.220
5	1.196	1.341
Average	1.091	1.162
Value of *d*	0.045	0.080

Now, using the above impact of the program we require survey the schools using RSS procedure. We use the simulated samples for the values of S which have been generated for five phases from the range 43.300 to 43.898 for boys and 30.500 to 30.891 for girls. In actual practice, the number of schools can be randomly selected on yearly basis for measuring the GER for this particular category. For this example, we have taken $m = 3$ with sample size $mk = 15$. Units taken for measurements are shown by the **bold** figures in Table 4.3.

Table 4.3
Ranked Set Sample of Size 15 with $k = 5$, $m = 3$ for GER (Boy)

Cycle	**Set I (for $S_{(11)}$)**	**Set II (for $S_{(22)}$)**	**Set III (for $S_{(33)}$)**	**Set IV (for $S_{(44)}$)**	**Set V (for $S_{(55)}$)**
	43.71	**43.48**	43.49	**43.60**	43.71
	43.84	43.87	43.86	43.40	43.49
I	43.83	43.67	43.51	43.52	**43.81**
	43.45	43.64	43.86	43.63	43.37
	43.84	43.42	**43.68**	43.49	43.62
	43.47	43.82	**43.64**	43.42	43.57
	43.46	**43.35**	43.86	43.81	43.57
II	43.63	43.64	43.50	43.44	43.69
	43.42	43.83	43.80	43.37	**43.84**
	43.48	43.35	43.44	**43.57**	43.48
	43.58	43.68	43.66	43.88	43.78
	43.61	43.86	43.65	**43.87**	43.85
III	43.60	43.37	43.76	43.33	43.40
	43.50	43.69	43.34	43.82	43.57
	43.62	**43.56**	**43.65**	43.42	**43.89**

Table 4.4
Ranked Set Sample of Size 15 with $k=5$, $m=3$ for GER (Girl)

Cycle	Set I (for $S_{(11)}$)	Set II (for $S_{(22)}$)	Set III (for $S_{(33)}$)	Set IV (for $S_{(44)}$)	Set V (for $S_{(55)}$)
I	**30.50**	30.61	30.84	30.60	**30.84**
	30.86	**30.58**	30.52	30.72	30.74
	30.79	30.58	**30.78**	30.80	30.68
	30.55	30.66	30.68	30.52	30.76
	30.81	30.57	30.87	**30.74**	30.74
II	30.82	30.79	30.71	30.62	**30.88**
	30.74	30.82	30.60	30.81	30.69
	30.76	30.55	**30.71**	30.68	30.56
	30.78	30.51	30.52	30.72	30.53
	30.89	**30.53**	30.80	**30.79**	30.79
III	30.73	30.60	**30.77**	30.61	30.53
	30.51	**30.59**	30.77	30.87	30.53
	30.77	30.80	30.84	**30.64**	30.64
	30.80	30.54	30.86	30.60	**30.79**
	30.62	30.78	30.61	30.60	30.66

In this example, the values of $\sigma^2_{S(i:k)}$ corresponding to the five rank order statistics are computed as 0.0016, 0.0112, 0.0004, 0.0273 and 0.0016 for boys and 0.0184, 0.0010, 0.0017, 0.0058 and 0.0020 for girls, respectively. σ^2_S based upon the 225 unordered observations were estimated at 0.0283 (boys) and 0.0132 (girls). The proposed estimator for the population mean was computed with $C=1.09$ (Boys) and $C=1.16$ (Girls) and are given as $\bar{R}_{(k)RSS}=47.54$ for boys and $\bar{R}_{(k)RSS}=35.62$ for girls. The computed values of RP using equation (7) are given for boys and girls as 3.361 and 2.311, respectively.

With this example, we also deal with the problem of imperfect ranking i.e. the judgment-ranked order of the values of S which do not match with their true numerical orders. Different imperfect models have been discussed by various researchers. Some of these include: non parametric Mann-Whitney-Wilcoxon statistics and Sign test (Bohn and Wolfe, 1994; Hettmansperger, 1995); Ranking error probability matrix method for some class of models (Aragon *et al*., 1999); Monotone likelihood ratio principle method (Fligner and MacEachern, 2006); Empirical assessment of the ranking accuracy in RSS for data from the Third National Health and Nutrition Examination Study (Chen *et al*., 2006). In the present example, we take the simple imperfect ranking model in which the middle observation of S i.e. third observation is in each cycle is taken for the measurement instead of perfect ranking of each order statistics. These observations are shown by underlined figures in Table 4.3 and Table 4.4. We wish to see when the ranking is imperfect, does an RSS sample still performs better than SRS sample?

Under this imperfect ranking scenario, the values of $\sigma^2_{S(i:k)}$ corresponding to the five rank order statistics becomes 0.0156, 0.0273, 0.0217, 0.0091 and 0.0444 for boys and 0.0002, 0.0186, 0.0042, 0.0069 and 0.0039 for girls, respectively. The estimators are found to be $\bar{R}_{(k)RSS} = 47.50$ for boys and $\bar{R}_{(k)RSS} = 35.62$ for girls. More importantly, we found RP for boys is 1.201 and RP for girls is 1.969. This proves the arguments suggested by Dell and Clutter (1972) that even in the imperfect ranking, the gain of RSS over SRS may still present. However, the possibility of better performance of imperfect ranking over perfect ranking cannot be denied.

4.2 Calculating RPs based on Variances of Order Statistics of Skewed and Symmetric Distributions

In this section, the numerical values of RP have been calculated for four different distributions of S for $k = 2(1)10$ and four different values of $d = 0.25$, 0.50, 0.75 and 1.00. We have considered two skewed distributions viz. standard lognormal $LN(0,1)$ and standard Gamma $G(1)$ and two symmetric distributions viz. $U(0,1)$ and standard Normal $N(0,1)$. The values of variances of order statistics for these distributions are readily available in Harter and Balakrishnan (1996).

Table 4.5

***RPs* for Two Skewed Distributions at $k = (1)10$ and $d = 0.25(0.25)1$.**

k	LN (0,1)				G(1)			
	$d = 0.25$	$d = 0.50$	$d = 0.75$	$d = 1$	$d = 0.25$	$d = 0.50$	$d = 0.75$	$d = 1$
2	1.2018	1.2347	1.2755	1.3191	1.3498	1.3867	1.4325	1.4815
3	1.3751	1.4386	1.5033	1.5626	1.6800	1.7576	1.8367	1.9091
4	1.5319	1.6212	1.7002	1.7653	1.9993	2.1159	2.2190	2.3040
5	1.6774	1.7878	1.8752	1.9423	2.3114	2.4635	2.5839	2.6764
6	1.8143	1.9416	2.0340	2.1012	2.6180	2.8017	2.9351	3.0321
7	1.9441	2.0849	2.1802	2.2467	2.9201	3.1317	3.2748	3.3747
8	2.0679	2.2196	2.3164	2.3818	3.2182	3.4544	3.6050	3.7066
9	2.1866	2.3470	2.4444	2.5084	3.5128	3.7705	3.9270	4.0297
10	2.3006	2.4681	2.5655	2.6280	3.8040	4.0808	4.2419	4.3453

Table 4.6

***RPs* for Two Symmetric Distributions at $k = 2(1)10$ and $d = 0.25(0.25)1$.**

k	U(0,1)				N(0,1)			
	$d = 0.25$	$d = 0.50$	$d = 0.75$	$d = 1$	$d = 0.25$	$d = 0.50$	$d = 0.75$	$d = 1$
2	1.5184	1.5599	1.6155	1.6667	1.4850	1.5256	1.5760	1.6299
3	2.0533	2.1481	2.2449	2.3333	1.9648	2.0555	2.1481	2.2327
4	2.6033	2.7551	2.8892	3.0000	2.4439	2.5864	2.7124	2.8163
5	3.1667	3.3750	3.5400	3.6667	2.9241	3.1165	3.2688	3.3858
6	3.7416	4.0041	4.1947	4.3333	3.4056	3.6445	3.8180	3.9442
7	4.3266	4.6400	4.8521	5.0000	3.8884	4.1701	4.3607	4.4937
8	4.9199	5.2809	5.5112	5.6666	4.3722	4.6930	4.8977	5.0357
9	5.5206	5.9257	6.1717	6.3331	4.8568	5.2132	5.4295	5.5716
10	6.1280	6.5740	6.8335	7.0000	5.3419	5.7307	5.9569	6.1020

5. CONCLUSIONS AND DISCUSSION

In this paper we proposed the RSS procedure on the survey variable to get the unbiased estimator for the population mean of the response variable. The proposed estimator is compared with the estimator based on SRS procedure. It is shown that the gains in the relative precisions the proposed estimator using RSS over SRS estimator are high for chosen symmetric and skew distributions and sample sizes. The relative precisions are also shown for one real life experimental data on GER. Moreover, the relative precision increases with sample size.

Organizations are implementing developmental programs in phased manner across the geographical region or a particular region or community and are interested to know the impact of the program in the form of response variable. This study addresses cases with (i) measurements of units of survey variable of development program are very costly or tedious and (ii) impact of the program is known in general and follows a linear trend in particular. The phased implementation ruled out the use of the simple random sampling for estimation of response variable. In such situations, RSS, a cost-effective and precise method, of sample selection provides better estimate of the characteristic under study. RSS contains information across phases of the program. The proposed procedure seems to be a better alternative procedure compared to SRS procedure for developmental programs. The appropriate allocation of units from each phase of the program under RSS scheme may be taken more than one way. This article also demonstrates theoretically the patterns of arithmetic progression of impacts when impact of the first phase is known. This may facilitate to attain the optimum result of the program. The limitation of the proposed model is imperfect ranking of the units. The decrease accuracy in the ranking may be addressed by increasing the number of phases to retain the efficiency. Relative precision of RSS compared with SRS is an increasing function of the number of phases.

ACKNOWLEDGMENT

The authors thank the unknown Reviewers and the Editor for their useful comments and suggestions on the earlier version of this paper.

REFERENCES

1. Aragon, M.E.D., Patil, G.P. and Taillie, C. (1999). A performance indicator for ranked set sampling using ranking error probability matrix, *Environmental and Ecological Statistics*, 6(1), 75-89.
2. Bhoj, D.S. (2001). Ranked set sampling with unequal samples. *Biometrics*, 57, 957- 962.
3. Bohn, L.L. and Wolfe, D.A. (1994). The effect of imperfect judgment rankings on properties of procedures based on the ranked-set samples analog of the Mann-Whitey-Wilcoxon statistic. *Journal of the American Statistical Association*, 89, 168-176.
4. Chandra, G., Tiwari, N. and Chandra, H. (2011). Adaptive cluster sampling based on ranked sets. *Metodoloski Zvezki*, 8(1), 39-55.
5. Chen, H., Stasny, E.A. and Wolfe, D.A. (2006). An empirical assessment of ranking accuracy in ranked set sampling, *Computational Statistics and Data Analysis*, 51(2), 1411-1419.
6. Chen, Z., Bai, Z.D. and Sinha, B.K. (2004). *Ranked set sampling: theory and applications*. New York: Springer-Verlag.
7. Chen, Z., Stasny, E.A. and Wolfe, D.A. (2003). Improved estimation of disease prevalence using ranked set sampling. *Technical Report Number 718*, Department of Statistics, The Ohio State University.
8. Chen, Z., Stasny, E.A. and Wolfe, D.A. (2005). Ranked set sampling for efficient estimation of a population proportion. *Statistics in Medicine*, 24(21), 3319-29.
9. Cochran, W.G. (1977). Sampling techniques, Wiley Publication.
10. Dell, T.R. and Clutter, J.L. (1972). Ranked set sampling theory with order statistics background. *Biometrics*, 28, 545-555.
11. Fligner, M.A. and MacEachern, S.N. (2006). Nonparametric two-sample methods for ranked-set sample data. *Journal of the American Statistical Association*, 101(475), 1107-1118.
12. Harter, H.L. and Balakrishnan, N (1996). CRC handbook of tables for the use of order statistics in estimation, CRC Press, Boca Raton.
13. Hettmansperger, T.P. (1995). The ranked-set sampling sign test. *Nonparametric Statistics*, 4, 263-270.
14. Husby, C.E., Stansy, E.A. and Wolfe, D.A. (2005). An application of ranked set sampling for mean and median estimation using USDA crop production data. *Journal of Agricultural, Biological and Environmental Statistics*, 10(3), 354-373.
15. Johnson, G.D., Paul, G.P. and Sinha, A.K. (1993). Ranked set sampling for vegetation research. *Abstracta Botanica*, 17, 87-102.
16. Kaur, A., Patil, G.P. and Taillie, C. (1997). Unequal allocation models for ranked set sampling with skew distributions. *Biometrics*, 53, 123-130
17. Khan, L. and Shabbir, J. (2016). Improved ratio-type estimators of population mean in ranked set sampling using two concomitant variables. *Pakistan Journal of Statistics and Operational Research*, 12(3), 507-518.

18. Khan, L. and Shabbir, J. and Kadilar, C. (2016). Efficient classes of ratio-type estimators of population mean under stratified median ranked set sampling. *Pakistan Journal of Statistics*, 32(6), 475-496.
19. Kowalczyk, B. (2005). Alternative sampling designs some applications of qualitative data in survey sampling. *Statistics in Transition*, 7(2), 427-443.
20. McIntyre, G.A. (1952). A method for unbiased selective sampling using ranked sets. *Australian Journal of Agricultural Research,* 3, 385-390.
21. McIntyre, G.A. (1978). Statistical Aspects of Vegetation Sampling. In: Mannetje, L.T. (ed.), *Measurement of Grassland Vegetation and Animal Production Bulletin* 52, *Commonwealth Bureau of Pastures and Field Crops*. Hurley, Berkshire, U.K., 8-21.
22. Nahhas, R.W., Wolfe, D.A. and Chen, H. (2002). Ranked set sampling: cost and optimal set size. *Biometrics,* 58, 964-971.
23. Pandey, R. (2010). Samples allocation decisions for program evaluation: a case of multiplicative effect on units due to developmental initiatives in phased manner. *Folia Forestalia Polonica*, Series A, 52 (2), 83-88.
24. Pandey, R. and Verma, M.R. (2008). Samples allocation in different strata for impact evaluation of developmental programme. *Rev. Mat. Estat*., 26(4), 103-112.
25. Pandey, R., Ashraf, J. and Verma, M.R. (2012). Proportional allocation stratum weight for population units under external impact attributed to binomial functions in successive strata. *International Journal of Mathematics Sciences and Applications*, 2(3), 867-872.
26. 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, 1-31.
27. Tiwari, N. and Chandra, G. (2011). A systematic approach for unequal allocations in ranked set sampling with skew distributions. *Journal of the Indian Society of Agricultural Statistics*, 65(3), 331-338.
28. Verma, M.R., Singh, B., Pandey, R. and Joorel, J.P.S. (2014). Sample allocation procedure for evaluation of the impact of development programme under exponential phase effect. *Journal of the Indian Statistical Association*, 52(1), 151-160.
29. Zamanzade, E. and Mahdizadeh, M. (2016). Entropy estimation from judgment post stratified data. *Pakistan Journal of Statistics and Operation Research*, 4(12), 625-637.