Regression Estimators in Ranked Set, Median Ranked Set and Neoteric Ranked Set Sampling

Nursel Koyuncu
Department of Statistics
Hacettepe University, Turkey
nkoyuncu@hacettepe.edu.tr

Abstract

This paper proposes new regression estimators in median ranked set and neoteric ranked set sampling using one and two auxiliary variables. The proposed estimators have large gain in presicion compared to the classical ranked set sampling (RSS) design. A simulation study is designed to see the performance of suggested estimators. A real data set example is also used. In this data set, we have examined a rare endemic annual plant species which is grown in Turkey. We have found that suggested estimators are highly efficient that existing estimators.

Keywords: Median Ranked Set Sampling, Neoteric Ranked Set Sampling, Regression Estimator, Efficiency.

Introduction

Ranked set sampling (RSS) is first proposed by McIntyre (1952). He discussed this sampling design is often more efficient than simple random sampling of the same size. Many authors in sampling literature have modified this samplign design to get more efficiency such as Al-Saleh and Al-Omari (2002), Jemain and Al-Omari (2006), Jemain et al. (2007), Zamanzade and Al-Omari (2016). Moreover Al-Omari (2012), Koyuncu (2015, 2016), Zamanzade and Vock (2015) have utilized information of auxiliary variable in various median ranked set sampling. Recently, Zamanzade and Al-Omari (2016) have introduced neoteric ranked set sampling (NRSS). In this study we have proposed regression estimators in median ranked set sampling (MRSS) and neoteric ranked set sampling. To compare the efficiency we have conducted simulation study with simulated and real data sets.

Ranked Set Sampling

Ranked Set Sampling (RSS) design can be described as follows:

1. Select a simple random sample of size n^2 units from the target finite population and divide them into n samples each of size n.
2. Rank the units within each sample in increasing magnitude by using personal judgment, eye inspection or based on a concomitant variable.
3. Select the *i*th ranked unit from the *i*th $(i = 1, 2, \ldots, n)$ sample
4. Repeat steps 1 through 3 *m* times if needed to obtain a RSS of size N=*nm*.

Let $\left(X_1,Z_1,Y_1\right),\left(X_2,Z_2,Y_2\right),\ldots,\left(X_n,Z_n,Y_n\right)$ be a simple random sample of size n, then the measured ranked set sampling units are denoted by $\left\{\left(X_{(i)j},Z_{(i)j},Y_{[i]j}\right),i=1,\ldots,n,\ j=1,\ldots,m\right\}$ where $\left(X_{(i)j},Z_{(i)j},Y_{[i]j}\right)$ is the ith ranked unit from the jth cycle of two auxiliary variables and study variable respectively, $(\ )$ and $[\ ]$ indicate that the ranking of X and Z is perfect and ranking of Y has errors. In this paper we prefer to rank the units one of the most correlated auxiliary variables with study variable. The sample means of variables can be defined as in RSS:

$$\overline{y}_{RSS}=\frac{1}{nm}\sum_{j=1}^{m}\sum_{i=1}^{n}Y_{[i]j}\ ,\quad \overline{x}_{RSS}=\frac{1}{nm}\sum_{j=1}^{m}\sum_{i=1}^{n}X_{(i)j}\ ,\quad \overline{z}_{RSS}=\frac{1}{nm}\sum_{j=1}^{m}\sum_{i=1}^{n}Z_{(i)j} \qquad (1)$$

We introduce regression estimators using one and two auxiliary information, respectively, as:

$$\overline{y}_{\mathrm{Re}g1\_RSS}=\overline{y}_{RSS}+b_{RSS1}\left(\mu_x-\overline{x}_{RSS}\right) \qquad (2)$$

$$\overline{y}_{\mathrm{Re}g2\_RSS}=\overline{y}_{RSS}+b_{RSS1}\left(\mu_x-\overline{x}_{RSS}\right)+b_{RSS2}\left(\mu_z-\overline{z}_{RSS}\right) \qquad (3)$$

Median Ranked Set Sampling

Al-Omari (2012) has introduced (MRSS) as follows:

1. Select n random samples each of size n bivariate units from the finite population of interest.
2. The units within each sample are ranked by visual inspection or any other cost free method with respect to a variable of interest.
3. If n is odd, select the $\left(\frac{n+1}{2}\right)$th-smallest ranked unit X together with the associated Y from each set, i.e., the median of each set. If n is even, from the first $\frac{n}{2}$ sets select the $\left(\frac{n}{2}\right)$ th ranked unit X together with the associated Y and from the other $\frac{n}{2}$ sets select the $\left(\frac{n+2}{2}\right)$ th ranked unit X together with the associated Y.
4. The whole process can be repeated m times if needed to obtain a sample of size nm units.

Let $\left(X_{i(1)},Z_{i(1)},Y_{i[1]}\right),\left(X_{i(2)},Z_{i(2)},Y_{i[2]}\right),\ldots,\left(X_{i(n)},Z_{i(n)},Y_{i[n]}\right)$ be the order statistics of $X_{i1},X_{i2},\ldots,X_{in}$, $Z_{i1},Z_{i2},\ldots,Z_{in}$ and the judgment order of $Y_{i1},Y_{i2},\ldots,Y_{in}$ $\left(i=1,2,\ldots,n\right)$, For odd and even sample sizes the units measured using MRSS are denoted by MRSSO and MRSSE, respectively. For odd sample size let $\left(X_{1\left(\frac{n+1}{2}\right)}, Z_{1\left(\frac{n+1}{2}\right)}, Y_{1\left[\frac{n+1}{2}\right]}\right)$, $\left(X_{2\left(\frac{n+1}{2}\right)}, Z_{2\left(\frac{n+1}{2}\right)}, Y_{2\left[\frac{n+1}{2}\right]}\right), \ldots, \left(X_{n\left(\frac{n+1}{2}\right)}, Z_{n\left(\frac{n+1}{2}\right)}, Y_{n\left[\frac{n+1}{2}\right]}\right)$ denote the observed units by MRSSO.

$$\bar{x}_{MRSSO} = \frac{1}{n}\sum_{i=1}^{n} X_{i\left(\frac{n+1}{2}\right)}, \quad \bar{z}_{MRSSO} = \frac{1}{n}\sum_{i=1}^{n} Z_{i\left(\frac{n+1}{2}\right)} \text{ and } \bar{y}_{MRSSO} = \frac{1}{n}\sum_{i=1}^{n} Y_{i\left[\frac{n+1}{2}\right]} \tag{4}$$

be the sample mean of X, Z and Y respectively in (MRSS).

For even sample size let $\left(X_{1\left(\frac{n}{2}\right)}, Z_{1\left(\frac{n}{2}\right)}, Y_{1\left[\frac{n}{2}\right]}\right)$, $\left(X_{2\left(\frac{n}{2}\right)}, Z_{2\left(\frac{n}{2}\right)}, Y_{2\left[\frac{n}{2}\right]}\right), \ldots, \left(X_{\frac{n}{2}\left(\frac{n}{2}\right)}, Z_{\frac{n}{2}\left(\frac{n}{2}\right)}, Y_{\frac{n}{2}\left[\frac{n}{2}\right]}\right)$, $\left(X_{\frac{n+2}{2}\left(\frac{n+2}{2}\right)}, Z_{\frac{n+2}{2}\left(\frac{n+2}{2}\right)}, Y_{\frac{n+2}{2}\left[\frac{n+2}{2}\right]}\right)\left(X_{\frac{n+4}{2}\left(\frac{n+2}{2}\right)}, Z_{\frac{n+4}{2}\left(\frac{n+2}{2}\right)}, Y_{\frac{n+4}{2}\left[\frac{n+2}{2}\right]}\right), \left(X_{n\left(\frac{n+2}{2}\right)}, Z_{n\left(\frac{n+2}{2}\right)}, Y_{n\left[\frac{n+2}{2}\right]}\right)$ denote the observed units by MRSSE.

$$\bar{x}_{MRSSE} = \frac{1}{n}\left(\sum_{i=1}^{\frac{n}{2}} X_{i\left(\frac{n}{2}\right)} + \sum_{i=\frac{n+2}{2}}^{n} X_{i\left(\frac{n+2}{2}\right)}\right), \quad \bar{z}_{MRSSE} = \frac{1}{n}\left(\sum_{i=1}^{\frac{n}{2}} Z_{i\left(\frac{n}{2}\right)} + \sum_{i=\frac{n+2}{2}}^{n} Z_{i\left(\frac{n+2}{2}\right)}\right) \text{ and}$$

$$\bar{y}_{MRSSE} = \frac{1}{n}\left(\sum_{i=1}^{\frac{n}{2}} Y_{i\left[\frac{n}{2}\right]} + \sum_{i=\frac{n+2}{2}}^{n} Y_{i\left[\frac{n+2}{2}\right]}\right) \tag{5}$$

be the sample mean of X, Z and Y respectively.

In MRSS we can re-write estimators in defined in section two as given by, respectively

$$\bar{y}_{\mathrm{Re}g1\_MRSS} = \bar{y}_{MRSS} + b_{MRSS1}\left(\mu_x - \bar{x}_{MRSS}\right) \tag{6}$$

$$\bar{y}_{\mathrm{Re}g2\_MRSS} = \bar{y}_{MRSS} + b_{MRSS1}\left(\mu_x - \bar{x}_{MRSS}\right) + b_{MRSS2}\left(\mu_z - \bar{z}_{MRSS}\right) \tag{7}$$

Neoteric Ranked Set Sampling

Zamanzade and Al-Omari (2016) have defined a new neoteric ranked set sampling (NRSS). The NRSS scheme can be described as follows:

1. Select a simple random sample of size n^2 units from the target finite population.
2. Ranked the n^2 selected units in an increasing magnitude based on a concomitant variable, personal judgment or any inexpensive method.

3. If n is an odd, then select the $\left[\frac{n+1}{2}+(i-1)n\right]$th ranked unit for $i=1,\ldots,n$. If n is an even, then select the $\left[\ell+(i-1)n\right]$th ranked unit, where $\ell=\frac{n}{2}$ if i is an even and $\ell=\frac{n+2}{2}$ if i is an odd for $i=1,\ldots,n$.
4. Repeat steps 1 through 3 *m* times if needed to obtain a NRSS of size N=*nm*.

Let $\left(X_{1j},Z_{1j},Y_{1j}\right),\left(X_{2j},Z_{2j},Y_{2j}\right),\ldots,\left(X_{n^2j},Z_{n^2j},Y_{n^2j}\right)$ be n^2 simple random units selected from the population of interest and let $\left(X_{(1)j},Z_{(1)j},Y_{[1]j}\right)$, $\left(X_{(2)j},Z_{(2)j},Y_{[2]j}\right)$, …, $\left(X_{(n^2)j},Z_{(n^2)j},Y_{[n^2]j}\right)$ be order statistics of $\left(X_{1j},Z_{1j},Y_{1j}\right),\left(X_{2j},Z_{2j},Y_{2j}\right),\ldots,\left(X_{n^2j},Z_{n^2j},Y_{n^2j}\right)$ for $j=1,\ldots,m$.

$$\bar{y}_{NRSS}=\frac{1}{nm}\sum_{j=1}^{m}\sum_{i=1}^{n}Y_{[(i-1)n+\ell]j}\,,\quad \bar{x}_{NRSS}=\frac{1}{nm}\sum_{j=1}^{m}\sum_{i=1}^{n}X_{((i-1)n+\ell)j}\,,$$

$$\bar{z}_{NRSS}=\frac{1}{nm}\sum_{j=1}^{m}\sum_{i=1}^{n}Z_{((i-1)n+\ell)j} \qquad (8)$$

be the sample means of study and auxiliary variables in (NRSS). And now we can define regression estimators in (NRSS) as follows:

$$\bar{y}_{\mathrm{Re}g1\_\mathrm{N}RSS}=\bar{y}_{NRSS}+b_{NRSS1}\left(\mu_x-\bar{x}_{NRSS}\right) \qquad (9)$$

$$\bar{y}_{\mathrm{Re}g2\_\mathrm{N}RSS}=\bar{y}_{NRSS}+b_{NRSS1}\left(\mu_x-\bar{x}_{NRSS}\right)+b_{NRSS2}\left(\mu_z-\bar{z}_{NRSS}\right) \qquad (10)$$

Simulation Study

In this section, we conducted a simulation study to investigate the properties of proposed estimators. In the simulation study, we consider finite populations of size $N=10000$ generated from a bivariate normal distribution. The samples were generated from a bivariate normal distribution using mvrnorm function in R programme. In the simulation, we considered $\mu_x=2$, $\mu_z=3$, $\mu_y=4$, $\rho_{xy}=0.80$, $\rho_{zy}=0.75$. We have computed mean square errors (MSEs) and percent relative efficiencies (PREs) of estimators with respect to $\bar{y}_{RSS}$ for $n=3,4,5,6,7$ on the basis of 60.000 replications and displayed in Table1. From the Table1 we can say that regression estimators are more efficient in (NRSS) and (MRSS) compared to (RSS). We can also see that when the sample size is odd, the regression estimators in (NRSS) are more efficient than regression estimators in (MRSS).

Real Data Set

In this data set, we have examined a rare endemic annual plant species which is grown in Ankara-Turkey. 896 seeds of endemic plant's weight, height and papus are measured. Using this data set as a population by setting height as study variable (Y) and weight (X) and papus (Z) as auxiliary variables. We have selected 10000 sample under median ranked set and neoteric ranked set sampling design. From each sample we have estimated the height of plant with suggested estimators and calculated mean square error. According to mean square error value we have decided that which sampling plan is suitable for this data set. Descriptive statistics of data set: $\mu_x = 1.29$, $\mu_z = 1.97$, $\mu_y = 2.99$, $\rho_{xy} = 0.834$, $\rho_{zy} = 0.703$. The result of real data set is summarized in Table2.

Conclusion

In this study, we have introduced regression estimators in median and neoteric ranked set sampling designs which are recently introduced by Al-Omari (2012) and Zamanzade and Al-Omari (2016). We have compared the efficiency of our suggested estimators with Al-Omari (2012) and Zamanzade and Al-Omari (2016) estimators. We have found that our suggested regression estimators are highly efficient than existing estimators.

References

1. Al-Omari, A.I. (2012). Ratio estimation of the population mean using auxiliary information in simple random sampling and median ranked set sampling. Statistics and Probability Letters, 82, 1883-1890.
2. Al-Saleh, M.F. and Al-Omari, A.I. (2002). Multistage ranked set sampling. Journal of Statistical Planning and Inference, 102, 273-286.
3. Jemain, A.A. and Al-Omari, A.I. (2006). Multistage median ranked set samples for estimating the population mean. Pakistan Journal of Statistics, 22, 195-207
4. Jemain, A.A., Al-Omari, A.I. and Ibrahim, K. (2007). Multistage extreme ranked set sampling for estimating the population mean. Journal of Statistical Theory and Application, 6(4), 456-471.
5. Koyuncu N. (2016). New difference-cum-ratio and exponential type estimators in median ranked set sampling. Hacettepe Journal of Mathematics and Statistics, 45, 1, 207-225.
6. Koyuncu N. (2015). Ratio estimation of the population mean in extreme ranked set and double robust extreme ranked set sampling. International Journal of Agricultural and Statistical Sciences, 11(1), 21-28.
7. McIntyre, G.A. (1952). A method for unbiased selective sampling using ranked sets. Australian Journal of Agricultural Research, 3, 385–390.
8. Zamanzade, E. and Al-Omari, A.I. (2016). New ranked set sampling for estimating the population mean and variance. Hacettepe Journal of Mathematics and Statistics, 45(6), 1891-1905.
9. Zamanzade, E. and Vock, M. (2015). Variance estimation in ranked set sampling using concomitant variable. Statistics & Probability Letters, 105, 1-5.

Table 1: MSE and Efficiency of Estimators

	$n=3$		$n=4$		$n=5$		$n=6$		$n=7$	
Estimator	**MSE**	**Efficiency**	**MSE**	**Efficiency**	**MSE**	**Efficiency**	**MSE**	**Efficiency**	**MSE**	**Efficiency**
$\bar{y}_{RSS}$	0,2299	100,00	0,1567	100,00	0,1188	100,00	0,0941	100,00	0,0771	100,00
$\bar{y}_{\text{Re}g1\_RSS}$	0,1210	190,08	0,0907	172,79	0,0732	162,29	0,0608	154,81	0,0518	148,69
$\bar{y}_{\text{Re}g2\_RSS}$	0,0440	522,78	0,0317	493,83	0,0248	478,37	0,0202	464,56	0,0172	449,33
$\bar{y}_{NRSS}$	0,1979	116,15	0,1419	110,47	0,0989	120,17	0,0811	115,97	0,0654	117,86
$\bar{y}_{\text{Re}g1\_\text{N}RSS}$	0,1206	190,55	0,0907	172,78	0,0716	166,05	0,0603	155,95	0,0522	147,80
$\bar{y}_{\text{Re}g2\_\text{N}RSS}$	0,0415	**553,75***	0,0302	519,61	0,0235	**506,54***	0,0196	479,05	0,0163	**472,10***
$\bar{y}_{MRSS}$	0,2121	108,39	0,0910	172,25	0,1079	110,15	0,0526	178,95	0,0693	111,31
$\bar{y}_{\text{Re}g1\_\text{M}RSS}$	0,1207	190,42	0,0450	348,48	0,0720	164,96	0,0302	311,41	0,0506	152,30
$\bar{y}_{\text{Re}g2\_\text{M}RSS}$	0,0427	538,06	0,0168	**934,85***	0,0238	498,88	0,0106	**883,92***	0,0166	464,62

***Represent most efficient estimator**

Table 2: MSE and Efficiency of Estimators

	$n=3$		$n=4$		$n=5$		$n=6$		$n=7$	
Estimator	**MSE**	**Efficiency**	**MSE**	**Efficiency**	**MSE**	**Efficiency**	**MSE**	**Efficiency**	**MSE**	**Efficiency**
$\bar{y}_{RSS}$	0,1068	100,00	0,0711	100,00	0,0517	100,00	0,0404	100,00	0,0315	100,00
$\bar{y}_{\text{Re}g1\_RSS}$	0,0485	220,17	0,0370	192,28	0,0293	176,41	0,0249	161,95	0,0200	157,44
$\bar{y}_{\text{Re}g2\_RSS}$	0,0656	162,69	0,0477	148,90	0,0376	137,46	0,0316	127,87	0,0261	120,31
$\bar{y}_{NRSS}$	0,0768	138,98	0,0479	148,53	0,0318	162,86	0,0235	171,92	0,0182	172,45
$\bar{y}_{\text{Re}g1\_\text{N}RSS}$	0,0279	**382,42**	0,0194	365,58	0,0148	**348,72**	0,0118	342,49	0,0095	**331,30**
$\bar{y}_{\text{Re}g2\_\text{N}RSS}$	0,0479	222,91	0,0353	201,23	0,0264	195,80	0,0216	186,62	0,0183	171,74
$\bar{y}_{MRSS}$	0,1416	75,41	0,0696	102,11	0,0893	57,88	0,0615	65,69	0,0822	38,27
$\bar{y}_{\text{Re}g1\_\text{M}RSS}$	0,0396	269,70	0,0164	**434,29**	0,0278	186,29	0,0159	254,63	0,0256	122,94
$\bar{y}_{\text{Re}g2\_\text{M}RSS}$	0,0427	249,96	0,0171	414,66	0,0220	234,76	0,0098	**414,22**	0,0145	217,02

***Represent most efficient estimator**