Int. J. Open Problems Compt. Math., Vol. 1, No. 1, June 2008

Estimation of Simple Linear Regression Model Using L Ranked Set Sampling

Amjad D. Al-Nasser[1] and Ahmed Radaideh[2]
[1]Department of Statistics, Yarmouk University, 21163 Irbid, Jordan
amjadn@yu.edu.jo
[2]Department of Pedagogic Sciences, Psychological and Educational, Lecce, Italy
Radaidehstat@yahoo.com

Abstract

In this article, L ranked-set sampling (LRSS) is used to estimate a simple linear regression model. We show that the estimated regression model based on LRSS is highly efficient compared to the estimators based on simple random sampling, Extreme ranked set sampling and ranked set sampling. Monte Carlo experiments are performed to assess the accuracy and the robustness of the LRSS estimates. The results are illustrated by an example.

Key Words: *Simple Linear Regression, Outliers, Ranked Set Sampling*.

1 Introduction

Regression analysis is a conceptually simple method for investigating functional relationships among variables. The relationship is expressed in the form of an equation or model connecting the response variable *(Y)* and one (*X*) or more explanatory variables. The simple true relationship can be approximated by the regression model

$$Y = \alpha + \beta X + \varepsilon$$

Where ε is assumed to be random error, α, β are unknown regression parameters to be estimated from the data.

Often in practice, data are collected on each of a number of units or cases using a simple random sample (SRS) technique. As an alternative to SRS, McIntyre [6] suggest using the ranked set sampling (RSS) for data collection by drawing *m* SRS each of size *m*, then only selecting one measurement from each SRS. This can be done by selecting the smallest observation from the first SRS, the second smallest from the second SRS, the process repeated till the largest observation is selected from the last SRS. Many authors used RSS technique in regression analysis. Patil et al. [8] compared the RSS sample and SRS sample in relation to the concomitant variable and the regression estimate. Yu and Lam [12] proposed a regression-type estimator based on RSS. They demonstrated that this estimator is always more efficient than the regression estimator using SRS and is also more efficient than the estimator proposed by Patil et al. [8] unless the correlation coefficient is low ($|\rho| < 0.4$). Muttlak [7] used RSS to estimate the parameters of the simple linear regression model treating the regressor X as a constant. Chen [2] did an extensive study on the properties of regression type estimates. Chen and Wang [4] studied the optimal RSS for the regression analysis. Samawi and Ababneh [10] and earlier Samawi et al [11], showed that the extreme ranked set sampling (ERSS) performed better than RSS at estimating model parameters.

The current study uses generalized ranked data procedure (LRSS) (Al-Nasser [1]). Specifically, a modified bivariate LRSS is proposed and used to improve parameter estimation in the simple linear regression model.

The article is organized as follows: the LRSS and the modified bivariate LRSS schemes are introduced in section 2. The definition of the modified estimator and their properties are then discussed in section 3. A Monte Carol experiment to assess the performance of the proposed scheme in estimating the regression model is given in section 4. An illustration of the proposed technique based on real data is presented in section 5. We conclude this paper by stating the need for further work.

2.1 L Ranked Set Sampling (LRSS)

In order to plan a LRSS design, *m* random samples should be selected each of size *m*. The LRSS procedure as proposed by Al-Nasser [1] is:

Step1: Select *m* random samples each of size *m*.

Step2: Rank the units within each sample with respect to the variable of interest by a visual inspection or by any other cost effective method.

Step3: Select the LRSS coefficient $k = [mp]$ such that $0 \le p < 0.5$, where $[x]$ is the largest integer value less than or equal to *x*.

Step4: For each of the first *k* ranked samples; select the unit with rank *k*+1 for actual measurement.

Step5: For the last *k* ranked samples, select the unit with rank $m - k$ for actual measurement.

Step6: For $j = k+1, k+2, ..., m-k-1$, the j^{th} ranked unit in the j^{th} ranked sample is selected for actual measurements.

The cycle may be repeated r times to obtain a sample size $n = mr$.

Based on the LRSS scheme, the estimator of the population mean when *r=1* is defined as:

$$\hat{\mu}_{LRSS} = \frac{1}{m}\left(\sum_{i=1}^{k} X_{i(k+1)} + \sum_{i=k+1}^{m-k} X_{i(i)} + \sum_{i=m-k+1}^{m} X_{i(m-k)}\right)$$

and its variance is given by:

$$\operatorname{var}\left(\hat{\mu}_{LRSS}\right) = \frac{1}{m^2}\left(\sum_{i=1}^{k} \operatorname{var}\left(X_{i(k+1)}\right) + \sum_{i=k+1}^{m-k} \operatorname{var}\left(X_{i(i)}\right) + \sum_{i=m-k+1}^{m} \operatorname{var}\left(X_{i(m-k)}\right)\right)$$

Al-Nasser [1] proved that $\hat{\mu}_{LRSS}$ is unbiased estimator of the population mean μ, and has smaller variance than $\hat{\mu}_{SRS}$ if the underlying distribution is symmetric.

2.2 Bivariate L Ranked Set Sampling (LRSS)

In order to have a bivariate L ranked set sample, the following steps should be performed:

Step1 Randomly draw *m* independent sets each containing *m* bivariate sample units.

Step2 Rank the units within each sample with respect to the $X's$ by visual inspection or any other cost effective method.

Step3 Select LRSS coefficient, $K = [mp]$ such that $0 \le p < 0.5$, and $[X]$ the largest integer value less than or equal to X.

Step4 For each of the first $(k+1)$ ranked samples; select the unit with rank $k+1$ and measure the Y value that corresponding to $x_{(k+1)i}$ and denote it by $y_{[k+1]i}$.

Step5 For $j = k+2, ..., m-k-1$, the unit with rank j in the j^{th} ranked sample is selected and measures the y value that corresponds.

Step6 The procedure continued until $(m-k)^{th}$ unit selected from the each of the last $(m-k)^{th}$ ranked samples, with respect to the first characteristic and measure the correspond y value.

For example, if $k=1$ and $m=5$ then the selected ranked sample will be as given in Table.1.

Table 1: Selected Bivariate LRSS when m = 5 and k = 1.

$x_{(1)1}, y_{[1]1}$	$x_{(1)2}, y_{[1]2}$	$x_{(1)3}, y_{[1]3}$	$x_{(1)4}, y_{[1]4}$	$x_{(1)5}, y_{[1]5}$
$x_{(2)1}, y_{[2]1}$	$x_{(2)2}, y_{[2]2}$	$x_{(2)3}, y_{[2]3}$	$x_{(2)4}, y_{[2]4}$	$x_{(2)5}, y_{[2]5}$
$x_{(3)1}, y_{[3]1}$	$x_{(3)2}, y_{[3]2}$	$x_{(3)3}, y_{[3]3}$	$x_{(3)4}, y_{[3]4}$	$x_{(3)5}, y_{[3]5}$
$x_{(4)1}, y_{[4]1}$	$x_{(4)2}, y_{[4]2}$	$x_{(4)3}, y_{[4]3}$	$x_{(4)4}, y_{[4]4}$	$x_{(4)5}, y_{[4]5}$
$x_{(5)1}, y_{[5]1}$	$x_{(5)2}, y_{[5]2}$	$x_{(5)3}, y_{[5]3}$	$x_{(5)4}, y_{[5]4}$	$x_{(5)5}, y_{[5]5}$

3 Estimating Simple Linear Regression Parameters

In completion of the sampling, let $d^{x}_{(i)j}$ and $d^{y}_{[i]j}$ be, respectively, X with rank k and the corresponding value of Y obtained from the i^{th} set in the j^{th} cycle. Then, the regression equation based on bivariate LRSS can be modeled as:

$$\left.\begin{aligned}
&d^{y}_{[i]j} = \alpha + \beta d^{x}_{(i)j} + d^{\varepsilon}_{[i]j} \qquad i=1,\ldots,m \quad j=1,\ldots,r \\
&d^{y}_{[i]j} = \begin{cases} Y_{[k+1]j} & i \le k \\ Y_{[i]j} & k+1 \le i \le m-k \quad ;\ j=1,2,\ldots,r \\ Y_{[m-k]j} & m-k+1 \le i \le m \end{cases} \\
&d^{x}_{(i)j} = \begin{cases} X_{(k+1)j} & i \le k \\ X_{(i)j} & k+1 \le i \le m-k \quad ;\ j=1,2,\ldots,r \\ X_{(m-k)j} & m-k+1 \le i \le m \end{cases}
\end{aligned}\right\} \tag{1}$$

where $d_{[i]j}^{\varepsilon}$ is the random error.

Under the regular assumptions of simple linear regression model Draper and Smith [5], the least square estimates of the regression parameters in (1) are given by:

$$\hat{\beta}_{LRSS} = \frac{\sum_{j=1}^{r}\left[\sum_{i=1}^{m}(d_{(i)j}^{x} - d^{\bar{x}})(d_{[i]j}^{y} - d^{\bar{y}})\right]}{\sum_{j=1}^{r}\left[\sum_{i=1}^{m}(d_{(i)j}^{x} - d^{\bar{x}})^2\right]} \tag{2}$$

and

$$\hat{\alpha}_{LRSS} = d^{\bar{y}} - \hat{\beta}_{LRSS} d^{\bar{x}} \tag{3}$$

where

$$d^{\bar{y}} = \frac{1}{rm}\sum_{j=1}^{r}\sum_{i=1}^{m} d_{[i]j}^{y} \quad \text{and} \quad d^{\bar{x}} = \frac{1}{rm}\sum_{j=1}^{r}\sum_{i=1}^{m} d_{(i)j}^{x}$$

Hereafter, the fitted model will be:

$$d_{[i]j}^{\hat{y}} = \hat{\alpha}_{LRSS} + \hat{\beta}_{LRSS} d_{(i)j}^{x} \tag{4}$$

note that, the estimated residuals are given by

$$d_{[i]j}^{\varepsilon} = d_{[i]j}^{y} - d_{[i]j}^{\hat{y}} \tag{5}$$

Theorem 1 Assume that (1) is satisfied then:

1- $E(\hat{\beta}_{LRSS}) = \beta$

2- $E(\hat{\alpha}_{LRSS}) = \alpha$

3- $\operatorname{var}\left(\hat{\beta}_{LRSS}\right)=\sum_{j=1}^{r}\left[\sum_{i}^{m}\left(\sigma_{[i]j}^{2}\times\frac{\left(d_{(i)j}^{x}-d^{\bar{x}}\right)^{2}}{Sxx^{2}}\right)\right]$

4- $\operatorname{var}\left(\hat{\alpha}_{LRSS}\right)=\frac{1}{m^{2}}\sum_{j=1}^{r}\sum_{i=1}^{m}\sigma_{d_{[i]j}^{e}}^{2}+\sum_{j=1}^{r}\left[\sum_{i}^{m}\left(\sigma_{[i]j}^{2}\times\frac{\left(d_{(i)j}^{x}-d^{\bar{x}}\right)^{2}}{Sxx^{2}}\right)\right]$

Proof: Without loss of generality suppose that *r=1* then $\hat{\beta}$ given (2) can be rewritten as

$$\hat{\beta}_{LRSS}=\frac{S_{d^{x}d^{y}}}{S_{d^{x}d^{x}}}=\sum_{i=1}^{m}c_{i}\times d_{[i]}^{y}$$

Where: $c_{i}=\frac{\left(d_{(i)}^{x}-d^{\bar{x}}\right)}{S_{d^{x}d^{x}}}$, $S_{d^{x}d^{x}}=\sum_{i=1}^{m}\left(d_{(i)}^{x}-d^{\bar{x}}\right)^{2}$ and

$$S_{d^{x}d^{y}}=\sum_{i=1}^{m}\left(d_{(i)}^{x}-d^{\bar{x}}\right)\left(d_{(i)}^{y}-d^{\bar{y}}\right)$$

Now $E\left(\hat{\beta}_{LRSS}\right)=E\left(\sum_{i=1}^{m}c_{i}\times d_{[i]}^{y}\right)=\sum_{i=1}^{m}\left(c_{i}\times E\left(d_{[i]}^{y}\right)\right)$

$$=\sum_{i=1}^{m}\left(c_{i}\,E\left(\alpha+\beta d_{(i)}^{x}+d_{[i]}^{\varepsilon}\right)\right)=\alpha\sum_{i=1}^{m}\left(c_{i}\right)+\beta\sum_{i=1}^{m}c_{i}\,d_{(i)}^{x}$$

Note that; $\sum_{i=1}^{m}c_{i}=\sum_{i=1}^{m}\left(\frac{\left(d_{(i)}^{x}-d^{\bar{x}}\right)}{S_{d^{x}d^{x}}}\right)=\frac{1}{S_{d^{x}d^{x}}}\sum_{i=1}^{m}\left(d_{(i)}^{x}-d^{\bar{x}}\right)=0$

and $\sum_{i=1}^{m}\left(c_i \times d_{(i)}^{x}\right)=\frac{1}{S_{d^x d^x}}\left(\sum_{i=1}^{m}\left(d_{(i)}^{x}-d^{\bar{x}}\right)d_{(i)}^{x}\right)$

$$=\frac{1}{S_{d^x d^x}}\left(\sum_{i=1}^{m}\left(\left(d_{(i)}^{x}\right)^2-d_{(i)}^{x}d^{\bar{x}}\right)\right)$$

$$=\frac{1}{S_{d^x d^x}}\left(\sum_{i=1}^{m}\left(d_{(i)}^{x}\right)^2\right)-m\left(d^{\bar{x}}\right)^2=\frac{S_{d^x d^x}}{S_{d^x d^x}}=1.$$

Therefore; $E\left(\hat{\beta}_{LRSS}\right)=\beta$

2- Now for the intercept estimator we have

$$E\left(\hat{\alpha}_{LRSS}\right)=E\left(d^{\bar{y}}\right)-d^{\bar{x}}E\left(\hat{\beta}_{LRSS}\right)=\left(\alpha+\beta d^{\bar{x}}\right)-\beta d^{\bar{x}}=\alpha$$

3- $\operatorname{var}\left(\hat{\beta}_{LRSS}\right)=\operatorname{var}\left(\sum_{i=1}^{m}c_i \times d_{[i]}^{y}\right)$

$$=\sum_{i=1}^{m}c_i^2 \operatorname{var}\left(d_{[i]}^{y}\right)=\sum_{i=1}^{m}\sigma_{[i]}^2 c_i^2=\sum_{i=1}^{m}\left(\sigma_{[i]}^2\frac{\left(d_{(i)}^{x}-d^{\bar{x}}\right)^2}{S_{d^x d^x}^2}\right)$$

4- $\operatorname{var}\left(\hat{\alpha}_{LRSS}\right)=\operatorname{var}\left(d^{\bar{y}}\right)+\left(d^{\bar{x}}\right)^2\operatorname{var}\left(\hat{\beta}_{LRSS}\right)$

$$=\frac{\sigma_{d^{\varepsilon}_{[i]}}^2}{m}+\left(d^{\bar{x}}\right)^2\left(\sum_{i=1}^{m}\left(\sigma_{[i]}^2\frac{\left(d_{(i)}^{x}-d^{\bar{x}}\right)^2}{S_{d^x d^x}^2}\right)\right)$$

$$= \frac{1}{m^2} \sum_{i=1}^{m} \sigma^2_{d^{\varepsilon}_{[i]}} + \left(d^{\bar{x}}\right)^2 \left(\sum_{i=1}^{m} \left(\sigma^2_{[i]} \frac{\left(d^{x}_{(i)} - d^{\bar{x}}\right)^2}{S^2_{d^x d^x}} \right) \right)$$

Following Yu and Lam [12] the LRSS regression estimator is given by

$$\hat{\mu}_{LRSS.\mathrm{Re}g} = d^{\bar{y}} + \hat{\beta}(\overline{X} - d^{\bar{x}})$$

Moreover, under model (1) and the above assumptions, then for fixed value of r we have

$$\frac{\left(\hat{\beta} - \beta\right)}{\sqrt{Var(\hat{\beta})}} \xrightarrow{L} N(0,1),\ as\, m \to \infty$$

and

$$\frac{\left(\hat{\alpha} - \alpha\right)}{\sqrt{Var(\hat{\alpha})}} \xrightarrow{L} N(0,1),\ as\, m \to \infty$$

The proof of these results are concluded directly using the ideas of RSS (Chen et al [3]).

4 Simulation Study

To illustrate the performance of the LRSS estimator's Monte Carlo simulation studies were conducted considering two cases inliers and outlier cases. The simulation plan has the following assumptions:

- Generate 10000 random samples using SRS, RSS, ERSS and LRSS (with k=1, 2).
- Set the number of cycles $r = 5, 10, 20$, and set size $m = 5, 6, 7, 8$.
- Initiate the strength of the association between the two variables by $\rho = 0.1, 0.5$ and 0.9.
- The intercept and the slope are initialled as $\alpha = 0$ and $\beta = \rho$.

- The error term is generated from $N(0,1-\rho^2)$ and the regressor from N (0, 1).
- Also, we consider an outlier case, by generating an outlier (one observation). For this observation we generate the error term from $N(0,5^2)$.
-The relative efficiency (RE) for the estimated model based on LRSS is computed according to the following expression:

$$RE = \frac{MSE(\hat{\mu}_{SRS.\mathrm{Re}\,g})}{MSE(\hat{\mu}_{LRSS.\mathrm{Re}\,g})} \tag{6}$$

The results of the MSE for the SLR model for inliers case is given in Table.2 – Table.4; and the results for outlier cases are given in Table 5 – Table 7.

Table 2: RE for Regression model with ρ = 0.1

r	m	ERSS	RSS	$LRSS_1$	$LRSS_2$
5	5	.562	.984	1.964	3.467
	6	.486	.978	1.759	3.446
	7	.452	.974	1.612	3.001
	8	.413	.982	1.526	2.651
10	5	.567	.981	1.959	3.438
	6	.493	.988	1.772	3.441
	7	.458	.991	1.638	3.042
	8	.418	.990	1.538	2.659
20	5	.573	.997	1.972	3.462
	6	.497	.995	1.783	3.463
	7	.460	.994	1.642	3.042
	8	.421	.998	1.550	2.680

Table 3: RE for Regression model with ρ = 0.5

r	m	ERSS	RSS	$LRSS_1$	$LRSS_2$
5	5	.659	.978	1.723	2.869
	6	.601	.973	1.560	2.831
	7	.580	.983	1.470	2.536
	8	.550	.984	1.397	2.246
10	5	.672	.989	1.732	2.874
	6	.612	.988	1.581	2.854
	7	.588	.993	1.483	2.549
	8	.557	.992	1.409	2.266
20	5	.675	.996	1.738	2.861
	6	.618	.994	1.591	2.860
	7	.591	.995	1.487	2.547
	8	.559	.993	1.409	2.262

Table 4: RE for Regression model with ρ = 0.9

r	m	ERSS	RSS	$LRSS_1$	$LRSS_2$
5	5	.907	.990	1.179	1.465
	6	.898	.989	1.137	1.464
	7	.893	.996	1.116	1.385
	8	.887	.995	1.097	1.317
10	5	.913	.991	1.182	1.476
	6	.903	.996	1.148	1.471
	7	.890	.995	1.117	1.389
	8	.885	.996	1.101	1.320
20	5	.918	.998	1.188	1.471
	6	.902	.999	1.149	1.470
	7	.895	.996	1.120	1.390
	8	.886	.997	1.102	1.319

Table 5: RE for Regression model with ρ = 0.1: outlier case

r	m	ERSS	RSS	$LRSS_1$	$LRSS_2$
5	5	0.474	0.980	3.537	6.668
	6	0.404	0.987	2.992	6.205
	7	0.368	0.988	2.635	5.184
	8	0.330	0.979	2.361	4.286
10	5	0.477	0.986	3.576	6.681
	6	0.406	0.989	3.019	6.238
	7	0.372	0.996	2.663	5.210
	8	0.335	0.990	2.404	4.361
20	5	0.483	0.996	3.594	6.709
	6	0.411	0.996	3.056	6.297
	7	0.372	0.998	2.669	5.219
	8	0.335	0.996	2.404	4.359

Table 6: RE for Regression model with ρ = 0.5: outlier case

r	m	ERSS	RSS	LRSS1	LRSS2
5	5	0.491	0.976	3.240	5.689
	6	0.423	0.975	2.758	5.276
	7	0.389	0.977	2.431	4.396
	8	0.363	0.989	2.226	3.753
10	5	0.498	0.989	3.263	5.694
	6	0.434	0.995	2.794	5.300
	7	0.397	0.993	2.484	4.468
	8	0.365	0.996	2.250	3.760
20	5	0.503	1.000	3.286	5.723
	6	0.435	0.995	2.803	5.297
	7	0.398	0.992	2.475	4.440
	8	0.366	0.998	2.259	3.784

Table 7: RE for Regression model with ρ = 0.9: outlier case

r	m	ERSS	RSS	LRSS1	LRSS2
5	5	0.612	0.984	1.922	2.727
	6	0.554	0.975	1.718	2.544
	7	0.542	0.993	1.616	2.270
	8	0.506	0.983	1.508	2.014
10	5	0.618	0.998	1.928	2.724
	6	0.565	0.993	1.748	2.567
	7	0.536	0.991	1.620	2.268
	8	0.513	0.997	1.537	2.050
20	5	0.621	1.000	1.941	2.724
	6	0.571	0.998	1.745	2.554
	7	0.538	0.995	1.614	2.267
	8	0.513	1.002	1.538	2.039

The simulation results indicate that estimation of the simple linear regression model using LRSS is more efficient than using the traditional sampling techniques; SRS, ERSS or RSS. Moreover, when the data contains outliers the LRSS is shown to be a robust technique, and as the value of K increases the RE increases. Moreover, the RE of regression estimators decreases as the set size or the cycle size increases. Also, for fixed r and m, the RE decreases whenever ρ increases. It seems that, for a moderate or large sample size, the RE is slightly different when using either RSS or ERSS. However, using LRSS is generally more efficient than using SRS, ERSS or RSS for regression analysis.

5 Illustration Using Real Data

In this section, an illustration of the LRSS procedure in estimation using simple linear regression is discussed based on a real data set from Platt et al [9].

5.1 Real Data Set

The original data were collected on seven variables about tree characteristics of which we have used only two: X, the diameter in centimeters at breast height and Y, the entire height in feet. The regression model is analyzed assuming that the population is consists of *375* trees. The summary statistics of the data are reported in Table.8.

Table 8: Summary Statistics for the Tree Data

	Diameter(x) in cm	Entire Height (y) in feet
N	375	375
Mean	21.8971	54.83
Std. Deviation	17.63671	57.656
Range	73.2	242

Table 9: Summary Statistics for the selected samples of size 75

		Range	Minimum	Maximum	Mean	Std. Deviation
SRS	x	66.90	2.30	69.20	20.1227	17.79634
	y	219.00	4.00	223.00	48.8933	58.30896
RSS	x	66.90	2.30	69.20	21.4427	18.96384
	y	219.00	4.00	223.00	55.8400	64.10023
$LRSS_1$	x	48.70	4.20	52.90	18.5333	13.57199
	y	205.00	6.00	211.00	42.7600	44.77718
$LRSS_2$	x	41.40	5.10	46.50	17.3213	11.50245
	y	203.00	8.00	211.00	43.3467	48.53516
ERSS	x	66.90	2.30	69.20	28.4773	23.89840
	y	219.00	4.00	223.00	79.3200	79.59614

Based on the entire measurements a random sample of size 75 is drawn by using different sampling schemes, SRS, RSS, ERSS, and LRSS ($k=1, 2$). In RSS, ERSS and LRSS procedure we use m sets each of size m, where $m=5$, and repeat this cycle fifteen times "i.e., r = 15" to achieve a sample of size *75*. The summary statistics of the selected random samples is presented in Table.9.

It can be noted that, the average of *regressor* varied from *17.3-28.5* and response from *42.7-79.3* depends on which sampling scheme is used.

5.2 Data Analysis

In order to form the regression model based on different sampling scheme, a visual inspection using scatter plot is used (Figure.1).

The scatter plots in Figure.1 suggested that the relationship between both variables is not linear. Therefore, both variables are re-expressed by a natural logarithmic transformation. After here, the least square method is used for model fitting; the estimates of the regression parameters are given in Table. 10.

Figure 1: Scatter Plot of Tree Data by Using Different Sampling Schemes

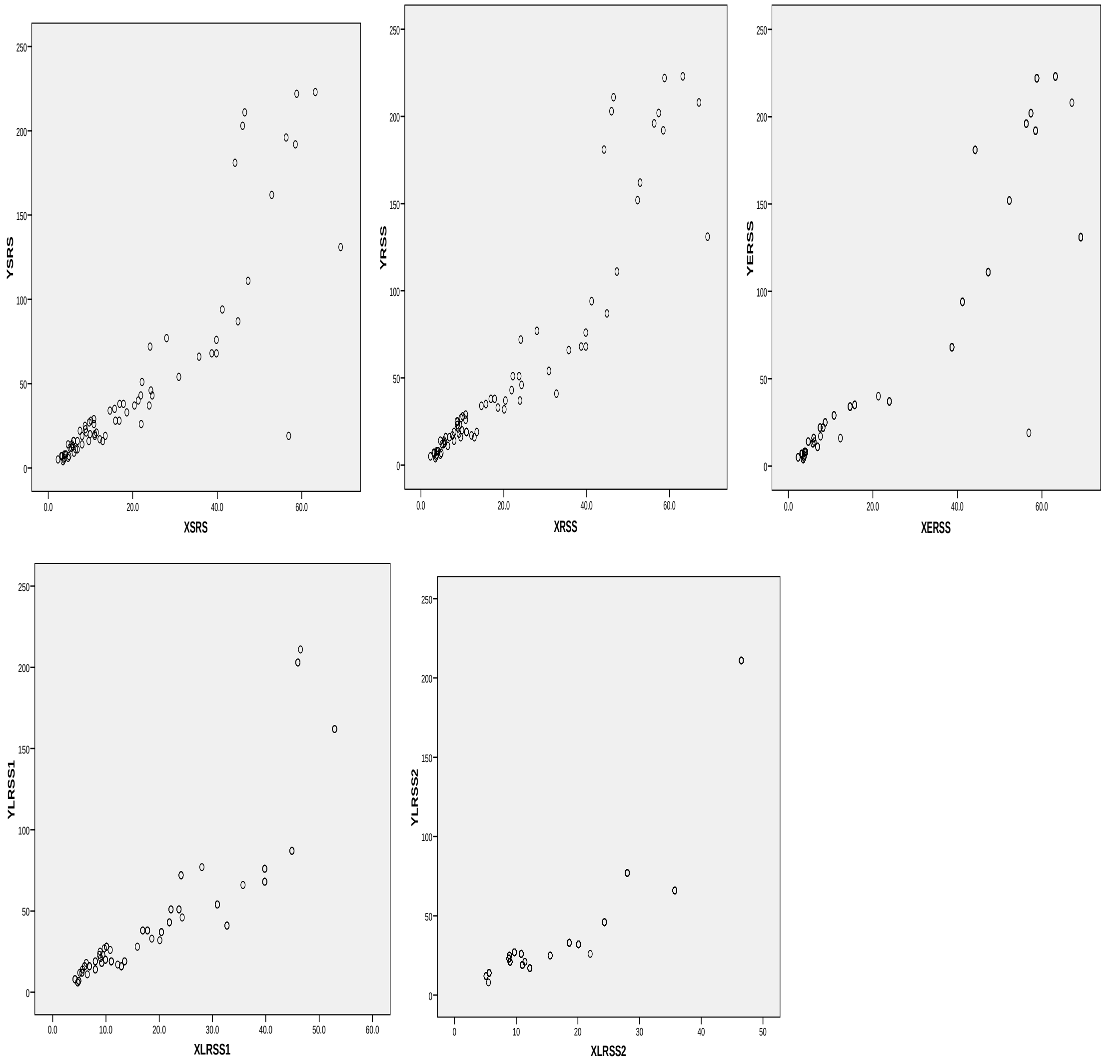


Table 10: Regression Analysis of Tree data

Method	Constant	Log(Diameter)	Adj(R^2)	MSE
SRS	0.556* (0.130)	1.066* (0.047)	0.875	0.134
RSS	0.468* (0.099)	1.120* (0.035)	0.932	0.080
ERSS	0.525* (0.116)	1.112* (0.38)	0.920	0.138
LRSS1	0.531* (0.124)	1.073* (0.045)	0.885	0.080
LRSS2	0.614* (0.154)	1.063* (0.057)	0.826	0.090

Note: Standard Errors in parentheses; * Statistically Significant at 1%.

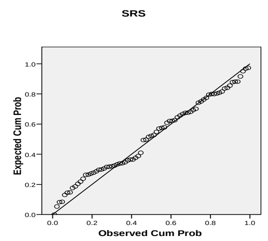
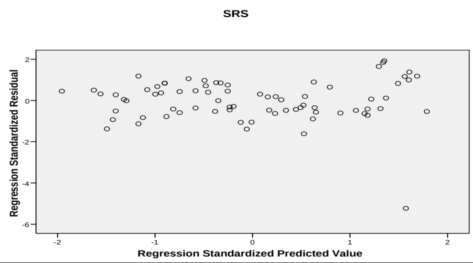
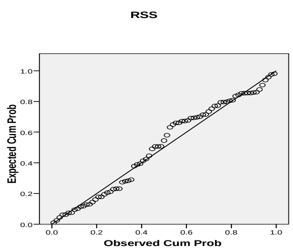
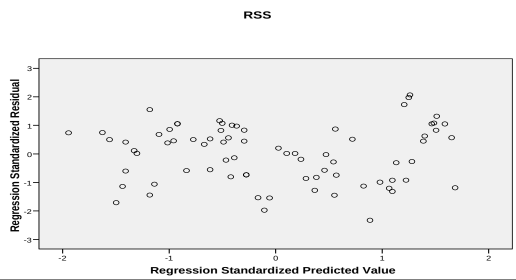
The results suggest that the RSS, LRSS1 and LRSS2 perform well compared to the SRS and ERSS in regards MSE point of view. Also, it can be noted that using RSS the intercept and slope have the minimum standard error and the highest fitting measure (i.e., 93.2%). Moreover, the residual plot and the normality p-p plot Figure.2 suggest that the model reasonably fits the data using these methods. In conclusion, from the data analysis and simulation results; the LRSS produced a satisfactory estimation for simple linear regression compared to the SRS and the other ranked data sampling schemes.

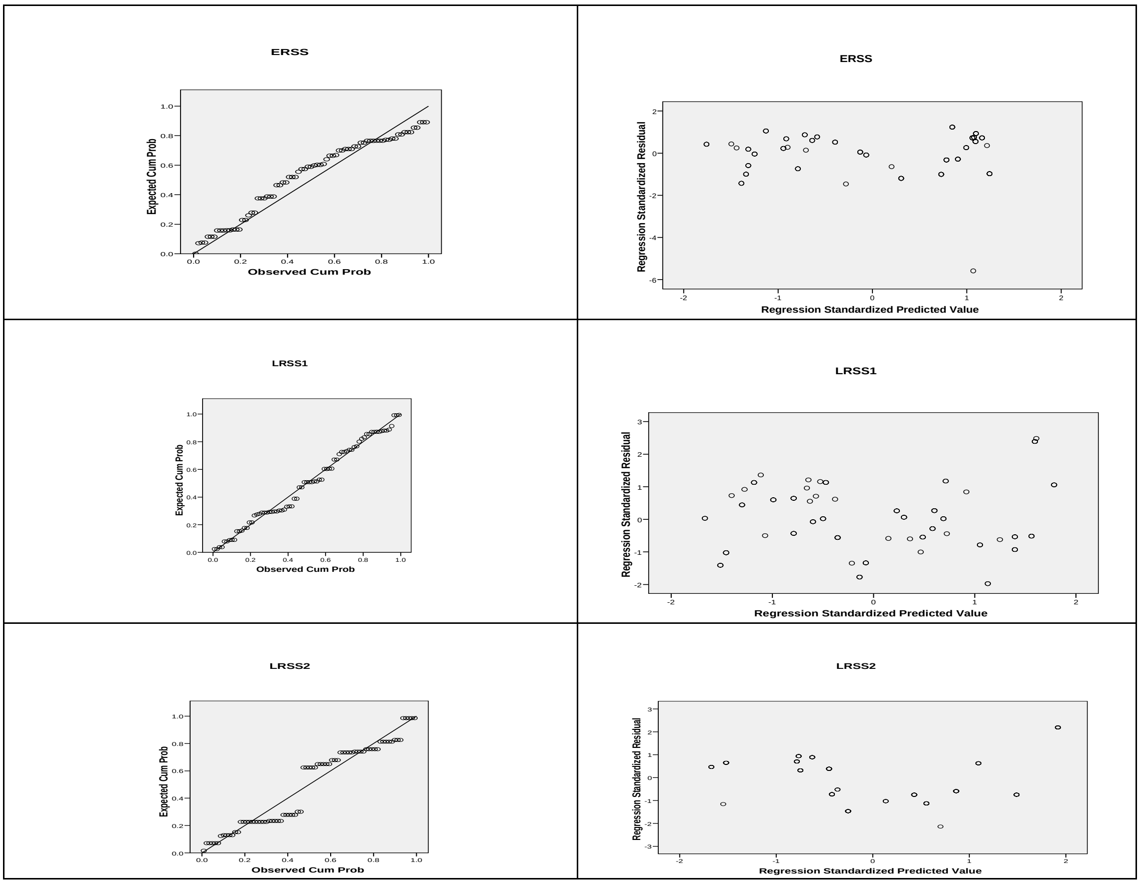
6 Future Works

In this paper, we suggest the use of LRSS to estimate the simple linear regression parameters. However, a modification of a ranked set sampling technique is still needed to improve estimation of the regression parameters. Also, the problem considered in this article can be extended to fit a multiple linear regression equation.

Moreover, there is a need to explore the usefulness of ranked sampling techniques for higher dimensional regression models such as semi parametric or non-linear models. This will be considered in future work.

Figure. 2 Residual Analysis using Different Sampling Scheme: “Response is Ln(Height)”

Normal P-P plot of the standardized residuals	Scatter plot of Residual *vs.* Predicted values
	
	



References

[1] D. A. Al-Nasser. L Ranked set sampling: A generalization procedure for robust visual sampling. Communications in Statistics: Simulation and Computation. 36(1). 2007, 33 – 43.

[2] Z. Chen. Ranked-set sampling with regression-type estimators. Journal of Statistical Planning and Inference. 92(2001), 181-192.

[3] Z. Chen. Bai, Z., and Sinha, B. K. Ranked Set Sampling: Theory and Applications. *Springer: New York.* 2004.

[4] Z. Chen and Wang. Y. Efficient Regression Analysis with Ranked-Set Sampling. Biometrics 60(2004), 997–1004.

[5] N. Draper, and Smith, H. Applied Regression Analysis. 2nd edition. USA: John Wiley & sons, Inc. 1981.

[6] G. A. McIntyre. A method of unbiased selective sampling, using ranked sets. *Australian J. Agricultural Research.* 3(1952), 385–390.

[7] H. A. Muttlak. Parameters Estimation in a simple linear regression using rank set sampling, *Biometrical. J.*, 37(7) (1995), 799–810.

[8] G. P. Patil, Sinha, A.K., and Taillie, C. Observational economy of ranked set sampling: comparison with the regression estimator. Environmetrics 4(1993), 399-412.

[9] W. J. Platt, Evans, G. W, and Rathbun, S. L. The population dynamics of a long-lived conifer, *The Amer. Naturalist.* 131(1988), 391–525.

[10] H. M. Samawi, and Ababneh, F. On regression analysis using ranked set sample, *Journal of Statistical Research (JSR)*, 35 (2) (2001), 93–105.

[11] H. M. Samawi, Ahmed, M. S., and Abu-Dayyeh, W. Estimating the population mean using extreme ranked set sampling, *Biometrical. J.*, 38 (5) (1996), 577–586.

[12] P. L. H. Yu., Lam, K., Regression estimator in ranked set sampling. Biometrics 53(1997), 1070-1080.