Decile Ranked Set Sampling

Dr .Nasr Ibrahim Rashwan

Dept. Of Statistics & Mathematics

Faculty of Comm., Tanta University

Decile Ranked Set Sampling

Dr .Nasr Ibrahim Rashwan

Dept. Of Statistics & Mathematics
Faculty of Comm., Tanta University

Abstract

The Ranked Set Sampling Method (RSS) as suggested by McIntyre (1952) may be modified to yield new sampling methods with improved. Since the efficiency of this method is susceptible to errors in ranking, Several modifications for the RSS are introduced such as extreme ranked set sampling (ERSS), median ranked set sampling (MRSS), Double ranked set sampling (DRSS), percentile ranked set sampling (PRSS) and quartile ranked set sampling (QRSS).

In this study decile ranked set sampling (CRSS) Method is suggested for estimating the population mean (μ), this method is compared with the simple random sample (SRS) and other ranked set sampling. The CRSS estimator is unbiased of the population mean for symmetric distributions about its mean in addition the CRSS method are more efficient than SRS and RSS for all symmetric and asymmetric distribution.

Key words: Ranked set sampling; simple random sample; *decile ranked set sample*; unbiased estimator; symmetric distributions; asymmetric distributions.

1- ***Introduction***

The Ranked Set Sampling (RSS) was suggested by McIntyre (1952) for estimating mean pasture yields with greater efficiency than SRS but without Mathematical theory to support his suggestion. (Patil et al., 1994)

Takahasi and Wakimoto (1968) provided the necessary mathematical theory of RSS and they proved that the sample mean of the RSS is an unbiased estimator of the population mean with smaller variance than the sample random sample with the same size of sample. Dell and clutter (1972) showed that the mean of RSS is unbiased of the population mean regardless of error in ranking. Several modifications for the RSS are introduced by Samawi et al. (1996); Muttlak (1997); Al-Saleh et al. (2000) and Muttlak (2003). They showed that these modifications (Methods) estimators are unbiased estimator of the population mean. In this paper, we introduce the considered modification of RSS method of decile Ranked Set Sampling (CRSS). The newly suggested method is compared with SRS and RSS. Also the properties of CRSS for estimating the population mean will be discussed.

2- *Sampling Methods*

2-1 Simple Random Sample:

The SRS as a method of selecting (n) observations out of the population of size (N) such that every one of the $\binom{N}{n}$ distinct samples is equally Likely to be chosen an equally to 1) $\binom{N}{n}$ /(if we draw the elements without replacement. To draw a simple random sample of size from a population of size N, the units of the entire population are listed from 1 to N. A unit of the population is selected to be included in the sample based on the outcome from the table of random numbers or a computer program produces such a table. Sampling could be with replacement or without replacement (Cochran, 1977).

Let $X_1, X_2, \ldots, X_n$ be a sample random sample of size (n). Then the unbiased estimator of the population mean is the sample mean which given by

$$\overline{X}_{SRS} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \dots (1)$$

and the variance of $\overline{X}_{SRS}$ for infinite population is defined as

$$Var(\overline{X}_{SRS}) = \frac{\sigma^2}{n} \qquad \dots (2)$$

where σ^2 is the population variance and usually estimated the sample variance

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n} (X_i - \overline{X}_{SRS})^2 \qquad \dots (3)$$

2-2 Ranked Set Sampling

The RSS may be summarized as follows:

- Randomly selected n^2 sample units from the target population.
- Randomly partition the sample into n sets, each of size n
- The units within set are then ranked with respect to a variable of interest.
- Choose a sample for actual analysis by including the smallest ranked unit in the first set, then the second smallest ranked unit in the second set, the procedure continues in this manner until the largest unit is been selected from the n^{th} set.
- We can repeat this process m times, I needed to obtain a set (sample) of size nm from initial n^2m units.

Let $X_1, X_2, \dots, X_n$ be a random sample with probability density function f(x) with mean μ and variance σ^2.

Let $X_{11}, X_{12},\ldots,X_{1n}, X_{21}, X_{22},\ldots,X_{2n};\ldots; X_{n1}, X_{n2},\ldots,X_{nn}$ be independent random variables all with the cumulative distribution function F(x).

Let $X_{(i:n)}$ dentote the i^{th} order statistic from the i^{th} sample of size n (i = 1,2,…,n). The unbiased estimator of the population mean using RSS (assume that m = 1) is defined as

$$\overline{X}_{RSS} = \frac{1}{n}\sum_{i=1}^{n} X_{(i:n)} \qquad \ldots (4)$$

with variance

$$Var(\overline{X}_{RSS}) = \frac{1}{n^2}\sum_{i=1}^{n} \sigma^2_{(i:n)} \qquad \ldots (5)$$

where $\sigma^2_{(i:n)} = E\left[(X_{(i:n)} - E(X_{(i:n)}))^2\right]$

$$\therefore Var(\overline{X}_{RSS}) = \frac{\sigma^2}{n} - \frac{1}{n}\sum_{i=1}^{n} (\mu_{(i:n)} - \mu)^2$$

2-3 Decile Ranked Set Sampling:

To obtain a sample of size n based on CRSS Method, the following steps are required to be carried out.

Step (1): select n random samples of size n units from the population under studied.

Step (2): rank the units within each sample with respect to the characteristic under investigation. The n measurements are then obtained depend on whether the set size is even or odd.

- If the sample size is even, then select for measurement from the first (n/2) samples the $(d_{L(n+1)})^{th}$ smallest ranked unit and from the second (n/2) samples, the $(d_{u(n+1)})^{th}$ smallest ranked unit where d_L lower decile and d_u upper decile for example, if d_L = 10% then d_u = 90% or d_L = 20% then d_u = 80% etc.

- If the sample size is odd, select from the first $(\frac{n-1}{2})$ sample, the $(d_{L(n+1)})^{th}$ smallest rank unit and from the other $(\frac{n-1}{2})$ samples the $(d_{u(n+1)})^{th}$ smallest rank unit and from one sample the median for that sample for actual measurement.

Step (3): The Cycle may be repeated m times is needed to get nm units. These nm units are from the CRSS data.

- For even sample size:

Let $X_{i(dL(n+1))}$ be the $(d_{L(n+1)})^{th}$ smallest ranked unit of the i^{th} sample (i=1,2,…,L), where L = (n/2) and Let $X_{i(du(n+1))}$ be the $(d_{u(n+1)})^{th}$ smallest ranked unit of the i^{th} sample (i = L+1, L+2, …,n)

In this case, $X_{1\,(dL(n+1))}$, $X_{2(dL(n+1))}$, …, $X_{\frac{n}{2}\,(dL(n+1))}$, $X_{\frac{n+2}{2}(du(n+1))}$, …, $X_{n(du(n+1))}$, denote the measured CRSS.

- For odd sample size:

Let $X_{i(dL(n+1))}$ be the $(d_{L(n+1)})^{th}$ smallest ranked unit of the i^{th} sample (i=1,2,…,L), where $L_1 = (\frac{n-1}{2})$ and $X_{i((n+1)/2)}$ be the median of the i^{th} sample of rank i = (n+1)/2 and Let $X_i\,(d_{u(n+1)})$

be the $(d_{u(n+1)})$ smallest ranked unit of the i^{th} sample, $(i= L_1+2, L_1+3, \ldots, n)$

In this case, $X_{1(dL(n+1))}$, $X_{2(dL(n+1))}$, …, $X_{\frac{n-1}{2}(dL(n+1))}$, $X_{\frac{n+1}{2}((n+1)/2)}$, $X_{\frac{n+3}{2}(du(n+1))}$, $X_{\frac{n+5}{2}(du(n+1))}$, …, $X_{n(du(n+1))}$, denote the measured CRSS.

- If n is even the estimator of the population mean using CRSS with one cycle can be defined as:

$$\overline{X}_{CRSSE} = \frac{1}{n}\left[\sum_{i=1}^{L} X_{i(dL(n+1))} + \sum_{i=L+1}^{n} X_{i(du(n+1))}\right] \qquad \ldots(6)$$

where L = n/2

and variance

$$Var(\overline{X}_{CRSSE}) = \frac{1}{n^2}\left[\sum_{i=1}^{n/2} Var(X_{i(dL(n+1))}) + \sum_{i=\frac{n+2}{2}}^{n} Var(X_{i(du(n+1))})\right] \qquad \ldots(7)$$

- If n is odd the CRSS estimator of the population mean is given by:

$$\overline{X}_{CRSSO} = \frac{1}{n}\left[\sum_{i=1}^{(n-1)/2} X_{i(dL(n+1))} + X_{\frac{(n+1)}{2}((n+1)/2)} + \sum_{i=\frac{n+3}{2}}^{n} X_{i(du(n+1))}\right] \qquad \ldots(8)$$

and variance

$$Var(\overline{X}_{CRSSO}) =$$

$$\frac{1}{n^2}\left[\sum_{i=1}^{(n-1)/2} Var(X_{i(dL(n+1))}) + Var(X_{\frac{(n+1)}{2}((n+1)/2)}) + \sum_{i=\frac{n+3}{2}}^{n} Var(X_{i(du(n+1))})\right] \qquad \ldots(9)$$

Let $\overline{X}_{SRS}$ denote the sample mean of simple random sample of size n. The properties of $\overline{X}_{CRSS}$ are:

1- $\overline{X}_{CRSS}$ is an unbiased estimator of the population mean under the assumption that the population is symmetric about its mean.

2- Var ($\overline{X}_{CRSS}$) is less than Var ($\overline{X}_{SRS}$)

3- If the distribution is asymmetric about μ, then mean square error of $\overline{X}_{CRSS}$ is less than variance of $\overline{X}_{SRS}$

Now, we will prove that $\overline{X}_{CRSS}$ is an unbiased estimator of the population mean and has smaller variance than variance of SRS.

- For sample size is even

$$\overline{X}_{CRSSE} = \frac{1}{n}\left[\sum_{i=1}^{n/2} X_{i(dL(n+1))} + \sum_{i=\frac{n+2}{2}}^{n} X_{i(du(n+1))}\right]$$

$$E(\overline{X}_{CRSSE}) = \frac{1}{n}\left[\sum_{i=1}^{n/2} E(X_{i(dL(n+1))}) + \sum_{i=\frac{n+2}{2}}^{n} E(X_{i(du(n+1))})\right]$$

$$= \frac{1}{n}\left[\frac{n}{2}\mu^{*}_{dL(n+1)} + \frac{n}{2}\mu^{*}_{du(n+1)}\right]$$

If the distribution is symmetric about μ (David and Nagaraja, 2003) then

$$\mu - \mu^{*}_{dL(n+1)} = \mu^{*}_{du(n+1)} - \mu$$

$$\therefore \mu^{*}_{dL(n+1)} + \mu^{*}_{du(n+1)} = 2\mu$$

$$E(\overline{X}_{CRSSE}) = \frac{1}{n}\left[\frac{n}{2}(\mu^{*}_{dL(n+1)}) + (\mu^{*}_{du(n+1)})\right] = \frac{1}{n}\left[\frac{n}{2}(2\mu)\right] = \mu$$

and

$$\operatorname{Var}(\overline{X}_{CRSSE}) = \frac{1}{n^2}\left[\sum_{i=1}^{n/2}\operatorname{Var}(X_{i(dL(n+1))}) + \sum_{i=\frac{n+2}{2}}^{n}\operatorname{Var}(X_{i(du(n+1))})\right]$$

$$= \frac{1}{n^2}\left[\frac{n}{2}\operatorname{Var}(X_{i(dL(n+1))}) + \frac{n}{2}\operatorname{Var}(X_{i(du(n+1))})\right]$$

$$= \frac{1}{2n}\left[\sigma^{*2}_{dL(n+1)} + \sigma^{*2}_{du(n+1)}\right] < \frac{\sigma^2}{n}$$

- For the sample size is odd

$$\overline{X}_{CRSSO} = \frac{1}{n}\left[\sum_{i=1}^{(n-1)/2} X_{i(dL(n+1))} + X_{\frac{(n+1)}{2}((n+1)/2)} + \sum_{i=\frac{n+3}{2}}^{n} X_{i(du(n+1))}\right]$$

$$E(\overline{X}_{CRSSO}) = \frac{1}{n}\left[\sum_{i=1}^{(n-1)/2} E(X_{i(dL(n+1))}) + E(X_{\frac{(n+1)}{2}((n+1)/2)}) + \sum_{i=\frac{n+3}{2}}^{n} E(X_{i(du(n+1))})\right]$$

$$= \frac{1}{n}\left[\frac{n-1}{2}(\mu^*_{dL(n+1)} + \mu^*_{du(n+1)}) + \mu\right]$$

because $E(X_{\frac{(n+1)}{2}((n+1)/2)}) = \mu$ If the distribution is symmetric.

$$E(\overline{X}_{CRSSO}) = \frac{1}{n}\left[\frac{n-1}{2}(2\mu) + \mu\right] = \mu$$

and

$$\operatorname{Var}(\overline{X}_{CRSS}) =$$

$$\frac{1}{n^2}\left[\sum_{i=1}^{(n-1)/2}\operatorname{Var}(X_{i(dL(n+1))}) + \operatorname{Var}(X_{\frac{(n+1)}{2}((n+1)/2)}) + \sum_{i=\frac{n+3}{2}}^{n}\operatorname{Var}(X_{i(du(n+1))})\right]$$

$$= \frac{n-1}{2n^2}\left(\sigma^{*2}_{dL(n+1)} + \sigma^{*2}_{du(n+1)}\right) + \frac{1}{n^2}\sigma^{*2}_{(\frac{n+1}{2})((n+1)/2)} < \frac{\sigma^2}{n}$$

3- *Illustrative example*

To compare the proposed estimator for the population mean using the CRSS against the usual estimators using SRS and RSS methods. Two symmetric distributions, namely uniform and normal and two asymmetric distributions, namely, exponential and gamma. The relative efficiency (precision) of estimating the population mean using the RSS with respect to usual estimator using SRS defined by:

$$RP(\bar{X}_{RSS}, \bar{X}_{SRS}) = \frac{Var(\bar{X}_{SRS})}{Var(\bar{X}_{RSS})} \quad \text{...(10)}$$

If the distribution is symmetric the relative precision of the CRSS with respect to SRS is defined by:

$$RP(\bar{X}_{CRSS}, \bar{X}_{SRS}) = \frac{Var(\bar{X}_{SRS})}{Var(\bar{X}_{CRSS})} \quad \text{...(11)}$$

But, If the distribution is asymmetric, the efficiency is defined follow

$$RP(\bar{X}_{CRSS}, \bar{X}_{SRS}) = \frac{Var(\bar{X}_{SRS})}{MSE(\bar{X}_{CRSS})}$$

Where $MSE(\bar{X}_{CRSS}) = Var(\bar{X}_{CRSS}) + (bias)^2$

where $bias = \mu - E(\bar{X}_{CRSS})$

In this study, we assume that the cycle is repeated once, tables 1, 2 and 3 summarize the relative efficiency of the RSS and CRSS estimators with sample size n= 10, 11 and 12 for d_L = 20%, 30% and 40%.

Table (1)

The efficiency for estimating the population mean using RSS and CRSS with sample size n = 10 units for d_L = 20%, 30% and 40%

Distribution	RSS	CRSS		
		20%	30%	40%
Uniform (0,1)	5.451	9.771	9.873	9.632
Normal (0,1)	5.382	10.334	10.334	10.856
Exponential (1)	3.345	4.023	1.817	1.753
Gamma (1,2)	3.271	3.882	1.632	1.234

Table (2)

The efficiency for estimating the population mean using RSS and CRSS with sample size n = 11 units for d_L = 20%, 30% and 40%

Distribution	RSS	CRSS		
		20%	30%	40%
Uniform (0,1)	6.001	10.333	11.314	12.003
Normal (0,1)	5.535	11.112	13.197	13.952
Exponential (1)	3.453	4.561	1.872	1.456
Gamma (1,2)	3.724	4.002	1.931	1.342

Table (3)

The efficiency for estimating the population mean using RSS and CRSS with sample size n = 12 units for d_L = 20%, 30% and 40%

Distribution	RSS	CRSS		
		20%	30%	40%
Uniform (0,1)	6.582	11.302	12.103	12.520
Normal (0,1)	6.335	12.141	14.121	14.611
Exponential (1)	4.310	4.661	2.077	1.556
Gamma (1,2)	3.952	4.02	1.765	1.342

Based on Tables 1, 2 and 3 we can conclude the following:

1- In this study, the RSS and CRSS methods are more efficient than SRS with different size of sample and several distributions whether symmetric or asymmetric.

2- In this study, if the underlying distribution is a symmetric, there is decreased in the efficiency if compared with symmetric distribution at the same size of sample and for some d_L values.

For example for the exp(1), with n= 10 and d_L = 20% then the efficiency = 4.023 but the efficiency of normal (0,1) = 10.334.

3- In this study, the efficiency of CRSS is increasing as the sample size increased for symmetric and asymmetric distributions.

For example when d_L = 30% for n=11 and 12 units the efficiency of CRSS is 13.197 and 14.121 respectively for estimating the mean of the standard normal distribution. For the exponential, the efficiency of CRSS when $d_L = 30\%$, for n = 11 and 12 units, is 1.872 and 2.077 respectively for estimating the mean.

4- In this study, the efficiency of CRSS is increased in decile value at the same size of sample when the distributions are symmetric but the efficiency of CRSS is decreased when the distributions are asymmetric.

For example, for $d_L = 20\%$, 30% and 40% with n = 10 units, the efficiency of CRSS is increasing for estimating the mean of the uniform distribution (0, 1), but if the distribution is exponential, the efficiency of CRSS is decreasing, see table (1).

Conclusion:

In this paper, it is observed that the CRSS estimator is unbiased of the population parameters if the distribution is symmetric or asymmetric, and more efficient than the SRS and RSS if the distributions are symmetric and more efficient than the SRS if the distributions are asymmetric, so we suggest using the CRSS for estimating the population mean.

References

1- Al-Saleh, M.F., and Al-Kadiri, M., (2000), "Double Ranked Set Sampling", **Statistical Probability Letters**, Vol.(48), PP.205-212.

2- Cochran, W.G., (1977), "**Sampling Techniques**", 3rd Edition, John Willy and Sons, INC, New York.

3- David, H.A. and Nagaraja, H.N., (2003) "**Order Statistics**", 3rd Edition, John Willy, and Sons, INC, New Jersey.

4- Dell, T.R. and Clutter, J.L., (1972), "Ranked Set Sampling Theory with Order Statistics Background", **Biometrics**, Vol.(28), PP.345-553.

5- McIntyre, G.A., (1952), "A Method of Unbiased Selective Sampling, Using Ranked Sets", **Australian J. Agricultural Research**, Vol.(3), PP.385-390.

6- Muttlak, H.A., (1997), "Median Ranked Set Sampling", **Journal of Applied Statistical Sciences**, Vol.(6), PP.245-255.

7- Muttlak, H.A., (2003), "Investigating The use of Quartile Ranked Set Samples for Estimating The Population Mean", **J. APPL. Math. Comput.**, Vol.(146), PP.437-443.

8- Muttlak, H.A., (2003), "Modified Ranked Set Sampling Methods", **Pak.J. Statist.**, Vol.(19), PP.315-323.

9- Patil, G.P., Sinha, A. K., and Taillie, C., (1994), "Ranked Set Sampling", **A Handbook of Statistics**, Vol.12, PP.167-200.

10- Samawi, H., Abu-Dayyeh, W., and Ahmed, S., (1996), "Extreme Ranked Set Sampling", **The Biometrical Journal**, Vol.(30), PP.577-586.

11- 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**, Vol.(20), PP.1-13.