Estimation of Population Mean under Stratified Median Ranked Set Sampling using Air-Quality Data

Assoc. Prof.Dr. Nursel KOYUNCU

ITACOSM 2019

Hacettepe University Department of Statistics, Ankara, Turkey
Email: nkoyuncu@hacettepe.edu.tr

5-7 June 2019

Outline

- The Aim of Study
- Literature Review
- Sampling Designs
- Simulation Study
- Conclusion

The Aim of Study

Ranked set sampling, developed as an alternative to simple random sampling

- Agriculture,
- Forestry,
- Environment,
- In ecology and medicine fields,

when the cost, timing and effort of measuring interested variables are hard to arrange, we would prefer this sampling method.

In this study, we try to see the effect of different sampling designs to estimation of population mean.

Literature Review

- Ranked Set Sampling (RSS) is introduced by McIntyre (1952)
- To reduce the error in ranking several modifications of the RSS method had been suggested. McIntyre (1952) and Samawi et al. (1996) suggested using an extreme ranked set sampling (ERSS) method.
- Muttlak (1997) suggested using median ranked set sampling (MRSS) methods.
- Al-Saleh and Al-Kadiri (2000)introduced Double Ranked Set Sampling for estimating the population mean.
- Al-Saleh and Al-Omari (2002) suggested Multistage Ranked Set Sampling that increases the efficiency of estimating the population mean for specific value of the sample size.

Literature Review

- Jemain and Al-Omari (2006) suggested Multistage Median Ranked Set Sampling (MMRSS) to estimate a population mean.
- Zamanzade and Al-Omari (2016)introduced neoteric ranked set sampling.
- Samawi (1996) introduced the concept of stratified ranked set sampling (SRSS), to improve the precision of estimating the population mean.
- Ibrahim et al. (2010) proposed stratified median ranked set sampling.
- Syam et al. (2013) considered the problem of estimating the population mean using stratified Double Percentile Ranked Set Sample.

Simple Random Sampling

- Let $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ be a bivariate random sample where the random variable X is the concomitant variable and the target variable Y.
- means μ_x, μ_y,
- variances σ_x^2, σ_y^2 and
- correlation coefficient ρ_{xy}.
- Assume that the auxiliary variable X is correlated with the variable of interest Y.

Simple Random Sampling

The unbiased estimators for the sample mean of study and concomitant variable based on SRS are as follows:

$$\bar{y}_{(ySRS)} = \frac{1}{n}\sum_{i=1}^{n} Y_i; \quad (1)$$

$$\bar{x}_{(SRS)} = \frac{1}{n}\sum_{i=1}^{n} X_i. \quad (2)$$

where $i = 1, 2, \ldots, n$

Ranked Set Sampling Design

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

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

Ranked Set Sampling Design

Let

$$\begin{aligned}
(X_{11j}, Y_{11j}) &,\quad (X_{12j}, Y_{12j}), \ldots, (X_{1mj}, Y_{1mj});\\
(X_{21j}, Y_{21j}) &,\quad (X_{22j}, Y_{22j}), \ldots, (X_{2mj}, Y_{2mj});\\
\vdots &;\\
(X_{m1j}, Y_{m1j}) &,\quad (X_{m2j}, Y_{m2j}), \ldots, (X_{mmj}, Y_{mmj})
\end{aligned}$$

be m independent bivariate random samples with pdf $f(x, y)$, each of size m in the jth cycle, $(j = 1, 2, \ldots, k)$. Let

$$(X_{i(1:m)j}, Y_{i[1:m]j}), (X_{i(2:m)j}, Y_{i[2:m]j}), \ldots, (X_{i(m:m)j}, Y_{i[m:m]j})$$

be the order statistics of $X_{i1j}, X_{i2j}, \ldots, X_{imj}$ and the judgement order of $Y_{i1j}, Y_{i2j}, \ldots, Y_{imj}$ $(i = 1, 2, \ldots, m)$, where round parentheses and square brackets indicate that the ranking of X is perfect and ranking of Y has errors.

Ranked Set Sampling Design

Assume measured units using RSS are

$$(X_{1(1:m)j}, Y_{1[1:m]j}), (X_{2(2:m)j}, Y_{2[2:m]j}), \ldots, (X_{m(m:m)j}, Y_{m[m:m]j}).$$

Then the RSS estimators of population mean for study variable and population mean of auxiliary variable can be written as

$$\bar{y}_{(yRSS)} = \frac{1}{mk}\sum_{j=1}^{k}\sum_{i=1}^{m} Y_{i[i:m]j}; \quad (3)$$

$$\bar{x}_{(RSS)} = \frac{1}{mk}\sum_{j=1}^{k}\sum_{i=1}^{m} X_{i(i:m)j}. \quad (4)$$

Median Ranked Set Sampling Design

Zamanzade and Al-Omari(2016)' sampling design and notations. Median ranked set sampling design can be described as in the following steps:

1. Select m random samples each of size m bivariate units from the 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 m is odd, select the $((m+1)/2)$ th-smallest ranked unit X together with the associated Y from each set, i.e., the median of each set. If m is even, from the first $m/2$ sets select the $(m/2)$th ranked unit X together with the associated Y and from the other sets select the $((m+2)/2)$ the ranked unit X together with the associated Y .
4. The whole process can be repeated k times if needed to obtain a sample of size $n = mk$ units.

Median Ranked Set Sampling Design

Let

$$(X_{i(1:m)j}, Y_{i[1:m]j}), (X_{i(2:m)j}, Y_{i[2:m]j}), \ldots, (X_{i(m:m)j}, Y_{i[m:m]j})$$

be the order statistics of $X_{i1j}, X_{i2j}, \ldots, X_{imj}$ and the judgement order of $Y_{i1j}, Y_{i2j}, \ldots, Y_{imj}$ $(i = 1, 2, \ldots, m; j = 1, 2, \ldots, k)$,where () and [] indicate that the ranking of X is perfect and ranking of Y has errors. For odd and even sample sizes the units measured using MRSS are denoted by MRSSO and MRSSE, respectively. For odd sample size let

$$(X_{1(\frac{m+1}{2}:m)j}, Y_{1[\frac{m+1}{2}:m]j}), (X_{2(\frac{m+1}{2}:j)}, Y_{2[\frac{m+1}{2}:j]}), \ldots, (X_{m(\frac{m+1}{2}:j)}, Y_{m[\frac{m+1}{2}]j})$$

denote the observed units by MRSSO.

Median Ranked Set Sampling Design

The estimators MRSS are given,

$$\bar{x}_{(MRSSO)} = \frac{1}{mk}\sum_{j=1}^{k}\sum_{i=1}^{m} X_{i(\frac{m+1}{2}:m)j} \tag{5}$$

$$\bar{y}_{(yMRSSO)} = \frac{1}{mk}\sum_{j=1}^{k}\sum_{i=1}^{m} Y_{i[\frac{m+1}{2}:m]j}. \tag{6}$$

Median Ranked Set Sampling Design

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

Median Ranked Set Sampling Design

The estimators of X and Y are given respectively

$$\bar{x}_{(MRSSE)} = \frac{1}{mk}\sum_{j=1}^{k}\left(\sum_{i=1}^{\frac{m}{2}} X_{i(\frac{m}{2}:m)j} + \sum_{i=\frac{m+2}{2}}^{m} X_{i(\frac{m+2}{2}:m)j}\right) \quad (7)$$

$$\bar{y}_{(yMRSSE)} = \frac{1}{mk}\sum_{j=1}^{k}\left(\sum_{i=1}^{\frac{m}{2}} Y_{i(\frac{m}{2}:m)j} + \sum_{i=\frac{m+2}{2}}^{m} Y_{i(\frac{m+2}{2}:m)j}\right) \quad (8)$$

Stratified Simple Random Sampling Design

Let $(X_1^h, Y_1^h), (X_2^h, Y_2^h), \ldots, (X_n^h, Y_n^h)$ be a bivariate random sample from a population size

$$\bar{y}_{(ySRS,h)} = \frac{1}{n_h}\sum_{i=1}^{n_h} Y_i^h; \quad (9)$$

$$\bar{x}_{(SRS,h)} = \frac{1}{n_h}\sum_{i=1}^{n_h} X_i^h; \quad (10)$$

$$\bar{y}_{(ySSRS)} = \sum_{h=1}^{L} \frac{N_h}{N}\bar{y}_{(ySRS,h)}; \quad (11)$$

$$\bar{x}_{(SSRS)} = \sum_{h=1}^{L} \frac{N_h}{N}\bar{x}_{(SRS,h)}. \quad (12)$$

Stratified Simple Random Sampling Design

$$Var(\bar{y}_{(ySSRS)}) = \sum_{h=1}^{L}(\frac{N_h}{N})^2\frac{\sigma_{yh}^2}{n_h} \qquad (13)$$

$$Var(\bar{x}_{(SSRS)}) = \sum_{h=1}^{L}(\frac{N_h}{N})^2\frac{\sigma_{xh}^2}{n_h} \qquad (14)$$

where variances σ_{xh}^2, σ_{yh}^2 and

Stratified Ranked Set Sampling Design

Let

$$\begin{array}{ccc} (X^h_{11j}, Y^h_{11j}) & , & (X^h_{12j}, Y^h_{12j}), \ldots, (X^h_{1m_hj}, Y^h_{1m_hj}); \\ (X^h_{21j}, Y^h_{21j}) & , & (X^h_{22j}, Y^h_{22j}), \ldots, (X^h_{2m_hj}, Y^h_{2m_hj}); \\ \vdots & ; & \\ (X^h_{m_h1j}, Y^h_{m_h1j}) & , & (X^h_{m_h2j}, Y^h_{m_h2j}), \ldots, (X^h_{m_hm_hj}, Y^h_{m_hm_hj}) \end{array}$$

be m independent bivariate random samples with pdf $f(x, y)$, each of size m_h in the jth cycle and hth stratum , $(j = 1, 2, \ldots, k; h = 1, 2, \ldots, L)$. Let

$$(X^h_{i(1:m_h)j}, Y^h_{i[1:m_h]j}), (X^h_{i(2:m_h)j}, Y^h_{i[2:m_h]j}), \ldots, (X^h_{i(m_h:m_h)j}, Y^h_{i[m_h:m_h]j})$$

be the order statistics of $X^h_{i1j}, X^h_{i2j}, \ldots, X^h_{im_hj}$ and the judgement order of $Y^h_{i1j}, Y_{i2j}, \ldots, Y^h_{im_hj}$ $(i = 1, 2, \ldots, m_h)$, where round parentheses and square brackets indicate that the ranking of X is perfect and ranking of Y has errors.

Stratified Ranked Set Sampling Design

Assume measured units using RSS are

$$(X^h_{i(1:m_h)j}, Y^h_{i[1:m_h]j}), (X^h_{i(2:m_h)j}, Y^h_{i[2:m_h]j}), \ldots, (X^h_{i(m_h:m_h)j}, Y^h_{i[m_h:m_h]j})$$

Then the RSS estimators of mean and mean in the hth stratum for study and auxiliary variable can be written as

$$\bar{y}_{(yRSS,h)} = \frac{1}{m_h k_h} \sum_{j=1}^{k_h} \sum_{i=1}^{m_h} Y^h_{i[i:m_h]j}; \tag{15}$$

$$\bar{x}_{(RSS,h)} = \frac{1}{m_h k_h} \sum_{j=1}^{k_h} \sum_{i=1}^{m_h} X^h_{i(i:m_h)j}. \tag{16}$$

Stratified Ranked Set Sampling Design

Then the stratified RSS estimators of population mean

$$\bar{y}_{(ySRSS)} = \sum_{h=1}^{L} \frac{N_h}{N} \bar{y}_{[yRSS,h]}; \quad (17)$$

$$\bar{x}_{(SRSS)} = \sum_{h=1}^{L} \frac{N_h}{N} \bar{x}_{(RSS,h)}; \quad (18)$$

(19)

Stratified Ranked Set Sampling Design

$$Var(\bar{y}_{(ySRSS)}) = \sum_{h=1}^{L}(\frac{N_h}{Nm_h})^2\frac{1}{k_h}\sum_{i=1}^{m_h}\sigma^2_{y[i]h} \quad (20)$$

$$Var(\bar{x}_{(SRSS)}) = \sum_{h=1}^{L}(\frac{N_h}{Nm_h})^2\frac{1}{k_h}\sum_{i=1}^{m_h}\sigma^2_{x(i)h} \quad (21)$$

Stratified Ranked Set Sampling Design

$$\sum_{i=1}^{m_h} \sigma^2_{y[i]h} = m_h \sigma^2_{yh} - \sum_{i=1}^{m_h} (\mu_{y[i]h} - \mu_{yh})^2 \qquad (22)$$

$$\sum_{i=1}^{m_h} \sigma^2_{x(i)h} = m_h \sigma^2_{xh} - \sum_{i=1}^{m_h} (\mu_{x(i)h} - \mu_{xh})^2 \qquad (23)$$

Stratified Median Ranked Set Sampling Design

Let

$$(X^h_{i(1:m_h)j}, Y^h_{i[1:m_h]j}), (X^h_{i(2:m_h)j}, Y^h_{i[2:m_h]j}), \ldots, (X^h_{i(m_h:m_h)j}, Y^h_{i[m_h:m_h]j})$$

be the order statistics of $X^h_{i1j}, X^h_{i2j}, \ldots, X^h_{im_hj}$ and the judgement order of $Y^h_{i1j}, Y^h_{i2j}, \ldots, Y^h_{im_hj}$ $(i = 1, 2, \ldots, m_h; j = 1, 2, \ldots, k_h)$,where () and [] indicate that the ranking of X is perfect and ranking of Y has errors. For odd and even sample sizes the units measured using MRSS are denoted by MRSSO and MRSSE, respectively. For odd sample size let

$$(X^h_{1(\frac{m_h+1}{2}:m_h)j}, Y^h_{1[\frac{m_h+1}{2}:m_h]j}), (X^h_{2(\frac{m_h+1}{2}:j)}, Y^h_{2[\frac{m_h+1}{2}:j]}), \ldots, (X_{m_h(\frac{m_h+1}{2}:j)^h}, Y^h_{m_h[\frac{m_h+1}{2}]}$$

denote the observed units by MRSSO.

Stratified Median Ranked Set Sampling Design

The estimators MRSS are given,

$$\bar{x}_{(MRSSO,h)} = \frac{1}{m_h k_h} \sum_{j=1}^{k_h} \sum_{i=1}^{m_h} X^h_{i(\frac{m_h+1}{2}:m_h)j} \quad (24)$$

$$\bar{y}_{(yMRSSO,h)} = \frac{1}{m_h k_h} \sum_{j=1}^{k_h} \sum_{i=1}^{m_h} Y^h_{i[\frac{m_h+1}{2}:m_h]j}. \quad (25)$$

$$\bar{y}_{(ySMRSSO)} = \sum_{h=1}^{L} \frac{N_h}{N} \bar{y}_{[yMRSSO,h]}, \quad (26)$$

$$\bar{x}_{(SMRSSO)} = \sum_{h=1}^{L} \frac{N_h}{N} \bar{x}_{(MRSSO,h)}. \quad (27)$$

Stratified Median Ranked Set Sampling Design

For even sample size let

$$\left(X^h_{1(\frac{m_h}{2}:m_h)j}, Y^h_{1[\frac{m_h}{2}:m_h]j}\right), \left(X^h_{2(\frac{m_h}{2}:m_h)j}, Y^h_{2[\frac{m_h}{2}:m_h]j}\right), \ldots, \left(X^h_{m_h(\frac{m_h}{2}:m_h)j}, Y^h_{m_h[\frac{m_h}{2}:m_h]j}\right)$$
$$\left(X^h_{\frac{m_h+2}{2}(\frac{m_h+2}{2}:m_h)j}, Y^h_{\frac{m_h+2}{2}[\frac{m_h+2}{2}:m_h]j}\right), \left(X^h_{\frac{m_h+4}{2}(\frac{m_h+4}{2}:m_h)j}, Y^h_{\frac{m_h+4}{2}[\frac{m_h+4}{2}:m_h]j}\right), \ldots,$$
$$\left(X^h_{m_h(\frac{m_h}{2}:m_h)j}, Y^h_{m_h[\frac{m_h}{2}:m_h]j}\right)$$

denote the observed units by MRSSE.

Median Ranked Set Sampling Design

The estimators of X and Y are given respectively

$$\bar{x}_{(MRSSE)} = \frac{1}{m_h k_h}\sum_{j=1}^{k_h}\left(\sum_{i=1}^{\frac{m_h}{2}} X_{i(\frac{m_h}{2}:m_h)j} + \sum_{i=\frac{m_h+2}{2}}^{m_h} X^h_{i(\frac{m_h+2}{2}:m_h)j}\right) \quad (28)$$

$$\bar{y}_{(yMRSSE)} = \frac{1}{m_h k_h}\sum_{j=1}^{k_h}\left(\sum_{i=1}^{\frac{m_h}{2}} Y_{i(\frac{m_h}{2}:m_h)j} + \sum_{i=\frac{m_h+2}{2}}^{m_h} Y^h_{i(\frac{m_h+2}{2}:m_h)j}\right) \quad (29)$$

Stratified Median Ranked Set Sampling Design

$$Var(\bar{y}_{(ySMRSSO)}) = \sum_{h=1}^{L}(\frac{N_h}{Nm_h})^2\frac{1}{k_h}\sum_{i=1}^{m_h}\sigma^2_{y[i:\frac{m_h+1}{2}]h} \quad (30)$$

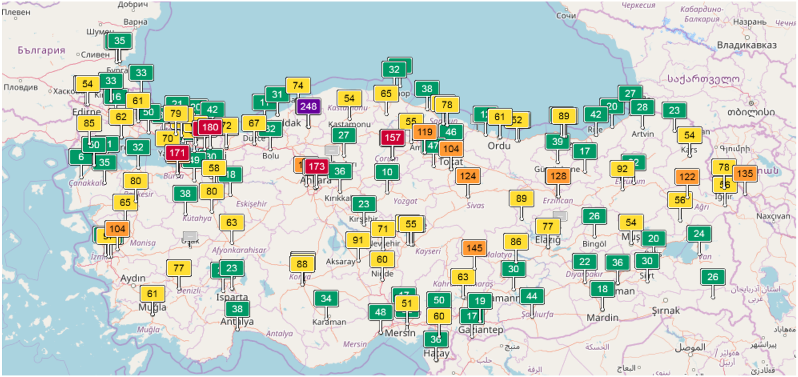
$$Var(\bar{y}_{(ySMRSSE)}) = \sum_{h=1}^{L}(\frac{N_h}{Nm_h})^2\frac{1}{k_h}(\sum_{i=1}^{\frac{m_h}{2}}\sigma^2_{y[i:\frac{m_h}{2}]h} + \sum_{i=\frac{m_h+2}{2}}^{m_h}\sigma^2_{y[i:\frac{m_h+2}{2}]h}) \quad (31)$$

Air Pollution

As a result of urbanization which is conclusion of modern life, air pollution has an impact on global scale as it effects local and regional scale. Because of the significant effects of air pollution to the people, it is important for whole World. Beside of protecting and enhancing the air quality, because it effect health of human beings, it is responsible for authorities , declare current air pollution to the people by communication instruments. When you declare the air airquality to the public, you need to use an easily understandable classification system. There can be calculated 5 basic index of air quality.These are;

- particulated matters (PM10),
- Carbonmonoxide(CO),
- sulphurdioxide(SO2),
- azotedioxide(NO2),
- ozon(O3)

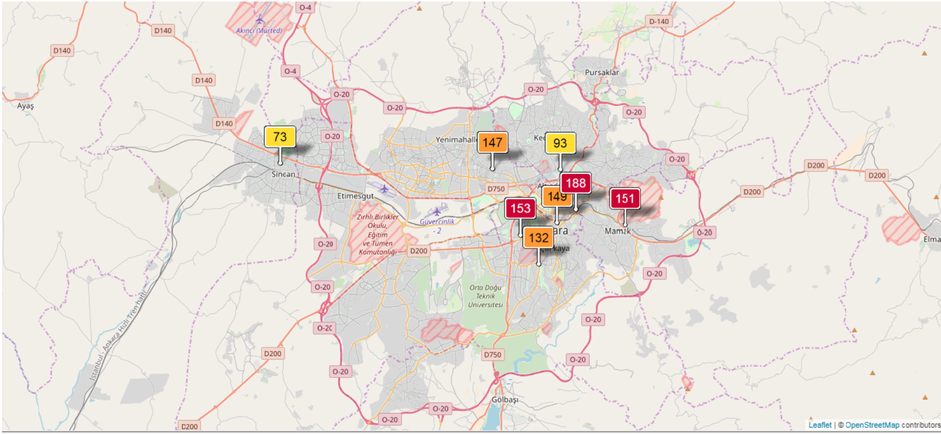
Air Pollution in Turkey: Real-time Air Quality Index Visual Map



Air Quality Index

AQI	Air Pollution Level	Health Implications
0 - 50	Good	Air quality is considered satisfactory, and air pollution poses little or no risk
51 -100	Moderate	Air quality is acceptable; however, for some pollutants there may be a moderate health concern for a very small number of people who are unusually sensitive to air pollution.
101-150	Unhealthy for Sensitive Groups	Members of sensitive groups may experience health effects. The general public is not likely to be affected.
151-200	Unhealthy	Everyone may begin to experience health effects; members of sensitive groups may experience more serious health effects
201-300	Very Unhealthy	Health warnings of emergency conditions. The entire population is more likely to be affected.
300+	Hazardous	Health alert: everyone may experience more serious health effects

Air Pollution in Ankara: Real-time Air Quality Index Visual Map



Share: *"Air Pollution in Ankara (Sihhiye): Real-time Air Quality Index Visual Map"*

Air Pollution in Ankara: Stations on the map

There are 8 stations on the map:

- (Sihhiye) (149)
- (Cebeci) (188)
- (Bahcelievler) (153)
- (Dikmen) (132)
- (Kecioren) (93)
- (Kayas) (151)
- (Demetevler) (147)
- (Sincan) (73)

Air Pollution Data

We consider the air quality data since June 2016 to May 2017 in the province of Ankara based on all stations (8 stations).
The population of size ($N = 2562, N1 = 276, N2 = 302, N3 = 358, N4 = 363, N5 = 313, N6 = 341, N7 = 319, N8 = 290$).

Simulation study was carried out by using PM10 (particulate matter) and nitrogen dioxide (NO2) change as concomitant variable.
When we did the simulation work using the air quality data at hand, we had the following results.

The Simulation Study

To see the effect of sampling designs we sample $n = 3, 5, 7, 10$ units from each strata using different sampling schemes with SSRS, SRSS and SMRSS respectively 10000 times.
Then estimate the mean and calculate MSE of estimators.

Results

Table 1: The MSE of mean based on different sampling designs

n	SSRS	SRSS	SMRSS
4	3809.75	3797.09	3436.16
5	3794.20	3788.49	3307.14
6	3801.90	3783.82	3296.11
7	3792.76	3776.83	3231.97

Results

Table 2: The PRE of mean based on different sampling designs

n	SSRS	SRSS	SMRSS
4	100	100.33	110.87
5	100	100.15	114.73
6	100	100.48	115.35
7	100	100.42	117.35

Some References

- McIntyre G.A., (1952). A method for unbiased selective sampling using ranked sets, Australian Journal of Agricultural Research, 3, 385-390.
- Muttlak, H. A., (1997). Median ranked set sampling, Journal of Applied Statistical Sciences, 6, 245-255.
- Samawi, H., Abu-Dayyeh, W., Ahmed, S. (1996) Extreme ranked set sampling, The Biometrical Journal, 30, 577- 586.
- Zamanzade E, Al-Omari AI., (2016). New ranked set sampling for estimating the population mean and variance. *Hacettepe Journal of Mathematics and Statistics* DOI: 10.15672/HJMS.20159213166.
- Samawi, H. M. (1996). Stratified ranked set sample. Pakistan Journal of Statistics, 12, 1, 9-16.
- Ibrahim, K., Syam, M. and Ibrahim Al-omari, A. (2010). Estimating the population mean using stratified median ranked set sampling. Applied Mathematical Science, 47, 2341-2354.
- Zamanzade, E., Mahdizadeh, M., A more efficient proportion estimator in ranked set sampling, Statistics and Probability Letters, 129, 28-33.

Thanks for your attention.