Evaluating New Estimators in Ranked Set Sampling Using Auxiliary Variable

Ruchi Gupta[1], Sangeeta Malik[2]

[1]Research Scholar, Department of Mathematics, Baba Mastnath University, Rohtak, Haryana, India
Email: *ruchigupta5989[at]gmail.com*

[2]Professor, Department of Mathematics, Baba Mastnath University, Rohtak, Haryana, India
Corresponding Author Email: *sangeetastat[at]gmail.com*

Abstract: ***This paper explores the application of Ranked Set Sampling (RSS) for estimating population means using auxiliary variables. We introduce two new estimators within the RSS framework and compare their performance to existing estimators in terms of mean square error and efficiency. Our findings indicate that the proposed estimators provide more efficient estimates under certain conditions, particularly when sample sizes are small and rankings are either perfect or imperfect. Empirical studies using simulations further validate the effectiveness of these new estimators, suggesting their potential for broader application in statistical analyses.***

Keywords: Estimators, Population, Auxiliary Variable, Ranked Set Sampling, Efficiency

1. Introduction

Ranked set sampling was introduced by McIntyre (1952) in the field of agriculture. He gave a concise and enlightening overview of the fundamentals of ranked set sampling and explains why it can produce better estimates than standard random sampling. Without using a rigorous mathematical background, he pointed out that the sample mean provides an unbiased estimator for population mean. Takashi and Wakimoto (1968) provided the notion of RSS with a crucial statistical basis. They demonstrated that, when compared to the sample mean of an SRS with a sample of the same size, the mean per unit estimator in RSS is an unbiased estimator of the population mean with a variance less than that of SRS. Dell and Clutter (1972) showed that the mean per unit estimator in RSS is an unbiased estimator of the population mean and having more efficiency than the mean of SRS in case the rankings are perfect or imperfect. Stokes (1980) demonstrated that the variance of mean per unit estimator in RSS is an asymptotically unbiased estimator of the population variance and more competent than the mean per unit estimator in SRS data with sample of same size.

The empirical distribution function based on RSS was investigated and demonstrated to be an unbiased estimator of the underlying distribution function by Stokes and Sager (1988). A recent summaries of RSS literature appear in two survey articles by Wolfe (2012) and a monograph by Chen et al (1980). Many scientific domains, including as medicine, environmental studies, and agriculture, use a variety of sampling strategies to collect data for inference. Many biological and environmental constraints disrupt the data collection technique during research investigations, such as sample size, cost per sample, and the study variable's destructible sample units. These limits have a significant impact on the statistical analysis and conclusions of the study. RSS, on the other hand, performs better in such cases.

2. Literature Review

Swami and Muttlak (1996) were the first to suggest the estimator for population ratio (R) under RSS which can also be used to estimate population mean after making some modification. But in case of bivariate population perfect ranking of units is possible with respect to study variable (y) or auxiliary variable(x). Further, Samawi and Muttlak (1996) recommended to rank process with respect to the variable that is used in denominator of the estimator. Generally, auxiliary variable is used in the denominator of the ratio type estimator. So, we consider the ranking process that is performed with respect to the auxiliary variable(x) which is also the more practical due to easily availability of the information regarding the auxiliary variable. Let ($y_{j[1]}$,$x_{j[1]}$), ($y_{j[2]}$,$x_{j[2]}$),….., ($y_{j[m]}$,$x_{j[m]}$); j = 1, 2, ..., r be the ranked set sample obtained from the j^{th} replication process, where $y_{j[i]}$ denotes the i^{th} judgment order statistics for the variable y.

Some notations and basic results are as follows: Assume that $\mu_x$and μ_y be the population means of study and auxiliary variables respectively and correspondingly σ_x and σ_y denotes the population variances.

$\mu_{x(i)}$ = E ($x_{j[i]}$) $\qquad$ $\mu_{y(i)}$ = E ($y_{j[i]}$)
$\sigma^2_{x(i)}$ = V ($x_{j[i]}$) = E $(x_{j[i]} - \mu_{x(i)})^2$ $\qquad$ $\sigma^2_{y(i)}$ = V ($y_{j[i]}$) = E $(y_{j[i]} - \mu_{y(i)})^2$
$\tau_{x(i)} = \mu_{x(i)} - \mu_x$ $\qquad$ $\tau_{y(i)} = \mu_{y(i)} - \mu_y$
$\tau_{xy(i)} = \tau_{x(i)}\,\tau_{y(i)}$ $\qquad$ $\sigma_{xy[i]}$ = E $(x_{j[i]} - \mu_{x(i)})\,(y_{j[i]} - \mu_{y(i)})$
Also, we can easily verify the following results
$\sum_{i=1}^{m}\mu_{x(i)} = m\mu_x$ $\qquad$ $\sum_{i=1}^{m}\mu_{y(i)} = m\mu_y$
$\sum_{i=1}^{m}\tau_{x(i)} = 0$ $\qquad$ $\sum_{i=1}^{m}\tau_{y(i)} = 0$
$\sum_{i=1}^{m}\sigma^2_{x(i)} = m\sigma^2_x - \sum_{i=1}^{m}\tau^2_{x(i)}$ $\qquad$ $\sum_{i=1}^{m}\sigma^2_{y(i)} = m\sigma^2_y - \sum_{i=1}^{m}\tau^2_{y(i)}$
$\sum_{i=1}^{m}\sigma^2_{xy(i)} = m\sigma^2_{xy} - \sum_{i=1}^{m}\tau^2_{xy(i)}$

In this section two different estimators for population mean have been proposed under RSS

3. Proposed Estimator – I

$$\bar{y}_{prop1} = \frac{\bar{y}_{rss}}{2}\left[\left(\frac{\bar{x}_{rss}}{\mu_x}\right)^{1+\alpha_0} + \left(\frac{\mu_x}{\bar{x}_{rss}}\right)^{-1+\alpha_0}\right]$$

Where α_0 is a real constant, $\bar{y}_{rss}$ and $\bar{x}_{rss}$ are the sample means of study and auxiliary variables respectively.

To obtain the approximate expressions for the bias and MSE of the proposed estimator, we express $\bar{y}_{rss}$ and $\bar{x}_{rss}$ in terms of δ's as

$\bar{y}_{rss} = \mu_y(1+\delta_1)$ $\bar{x}_{rss} = \mu_x(1+\delta_2)$

$$\bar{y}_{prop1} = \frac{\mu_y(1+\delta_1)}{2}\left[\left(\frac{\mu_x(1+\delta_2)}{\mu_x}\right)^{1+\alpha_0} + \left(\frac{\mu_x}{\mu_x(1+\delta_2)}\right)^{\alpha_0-1}\right]$$
$$= \frac{\mu_y(1+\delta_1)}{2}\left[(1+\delta_2)^{1+\alpha_0} + ((1+\delta_2)^{(-1)(\alpha_0-1)}\right]$$
$$= \frac{\mu_y(1+\delta_1)}{2}\left[(1+\delta_2)^{1+\alpha_0} + (1+\delta_2)^{1-\alpha_0}\right]$$

After solving and including the terms upto degree two in δ_1 and δ_2, we have

$$\bar{y}_{prop1} - \mu_y = \mu_y\left(\delta_1 + \delta_2 + \frac{(\alpha_0\delta_2)^2}{2} + \delta_1\delta_2\right) \quad (1)$$

After taking the expectation on both sides, we can obtain the expression for bias upto the first order approximation as

$$\text{Bias}(\bar{y}_{prop1}) = \frac{1}{\mu_y}\left[\frac{\alpha_0^{\,2}}{2} R^2\, V(\bar{x}_{rss}) + R\, Cov(\bar{y}_{rss}, \bar{x}_{rss})\right]$$

Where $R = \frac{\mu_y}{\mu_x}$

Now substituting the values of $V(\bar{x}_{rss})$ and $Cov(\bar{y}_{rss}, \bar{x}_{rss})$, we get

$$\text{Bias}(\bar{y}_{prop1}) = \frac{1}{rm}\frac{1}{\mu_y}\left[\frac{\alpha_0^{\,2}}{2} R^2\sigma_x^{\,2} + R\sigma_{xy}\right] - \frac{1}{r}\frac{1}{m^2}\frac{1}{\mu_y}\left[\frac{\alpha_0^{\,2}}{2} R^2 \sum_{i=1}^m \tau_{x(i)}^2 + R\sum_{i=1}^m \tau_{xy(i)}\right]$$

To find out the approximate expression for MSE, first take square on the both sides of (1) and consider the terms of δ's upto degree two

$$(\bar{y}_{prop1} - \mu_y)^2 = \mu_y^{\,2}(\delta_1^{\,2} + \delta_2^{\,2} + 2\delta_1\delta_2)$$

Taking expectation on both sides, we get

$$MSE(\bar{y}_{prop1}) = V(\bar{y}_{rss}) + R^2\, V(\bar{x}_{rss}) + 2R\, Cov(\bar{y}_{rss}, \bar{x}_{rss})$$
$$= \frac{1}{rm}(\sigma_y^{\,2} + R^2\sigma_x^{\,2} + 2R\sigma_{xy}) - \frac{1}{r}\frac{1}{m^2}\sum_{i=1}^m(\tau_{y(i)} + R\tau_{x(i)})^2$$

Which is the required approximate expression for MSE

Case 1: When $\alpha_0 = 1$

$$\bar{y}_{prop1} = \frac{\bar{y}_{rss}}{2}\left[\left(\frac{\bar{x}_{rss}}{\mu_x}\right)^2 + 1\right]$$

Which is the average of the mean per unit and quadratic product type estimator in RSS

Case 2: when $\alpha_0 = 0$

$$\bar{y}_{prop1} = \bar{y}_{rss}\left[\frac{\bar{x}_{rss}}{\mu_x}\right]$$
$$= \bar{y}_{Prss}$$

Which is the product estimator in RSS

Case 3: When $\alpha_0 = -1$

$$\bar{y}_{prop1} = \frac{\bar{y}_{rss}}{2}\left[1 + \left(\frac{\mu_x}{\bar{x}_{rss}}\right)^{-2}\right]$$
$$= \frac{\bar{y}_{rss}}{2}\left[1 + \left(\frac{\bar{x}_{rss}}{\mu_x}\right)^2\right]$$

Which is the average of the mean per unit and quadratic product type estimator in RSS

Comparison with $\bar{y}_{rss}$

$MSE(\bar{y}_{prop1}) - V(\bar{y}_{rss}) = R^2\, V(\bar{x}_{rss}) + 2R\, Cov(\bar{y}_{rss}, \bar{x}_{rss})$

$MSE(\bar{y}_{prop1}) - V(\bar{y}_{rss}) < 0$ iff $\rho_{\bar{y}_{rss}, \bar{x}_{rss}} < \frac{1}{2R}\sqrt{\frac{V(\bar{x}_{rss})}{V(\bar{y}_{rss})}} = \frac{1}{2}\frac{C_{\bar{x}_{rss}}}{C_{\bar{y}_{rss}}}$

Where $C_{\bar{x}_{rss}}$ and $C_{\bar{y}_{rss}}$ denotes the coefficient of variation for $\bar{x}_{rss}$ and $\bar{y}_{rss}$ respectively and $\rho_{\bar{y}_{rss}, \bar{x}_{rss}}$ is the correlation coefficient between $\bar{x}_{rss}$ and $\bar{y}_{rss}$.

Which shows that $\bar{y}_{prop1}$ is more efficient than the mean per unit estimator $\bar{y}_{rss}$ provided $\rho_{\bar{y}_{rss}, \bar{x}_{rss}} < \frac{1}{2}\frac{C_{\bar{x}_{rss}}}{C_{\bar{y}_{rss}}}$

Comparison with $\bar{y}_{Rrss}$

$MSE(\bar{y}_{Rrss}) - MSE(\bar{y}_{prop1}) = \frac{1}{r}\frac{1}{m^2}\sum_{i=1}^m(\tau_{y(i)} + R\tau_{x(i)})^2 \geq 0$

Which shows that $\bar{y}_{prop1}$ is always more efficient than $\bar{y}_{Rrss}$

Comparison with $\bar{y}_{Prss}$

$MSE(\bar{y}_{Prss}) - MSE(\bar{y}_{prop1}) = \left(\rho_{\bar{y}_{rss}, \bar{x}_{rss}}\sqrt{V(\bar{y}_{rss})} + R\sqrt{V(\bar{x}_{rss})}\right)^2 \geq 0$

Which shows that $\bar{y}_{prop1}$ is always more efficient than $\bar{y}_{Prss}$

4. Proposed Estimator – II

$$\bar{y}_{prop2} = \frac{\bar{y}_{rss}}{2}\left[\left(\frac{\bar{x}_{rss}}{\mu_x}\right)^{-1+\alpha_0} + \left(\frac{\mu_x}{\bar{x}_{rss}}\right)^{1+\alpha_0}\right]$$

Where α_0 is a real constant, $\bar{y}_{rss}$ and $\bar{x}_{rss}$ are the sample means of study and auxiliary variables respectively.

To obtain the approximate expressions for the bias and MSE of the proposed estimator, we express $\bar{y}_{rss}$ and $\bar{x}_{rss}$ in terms of δ's as

$\bar{y}_{rss} = \mu_y(1+\delta_1)$ $\bar{x}_{rss} = \mu_x(1+\delta_2)$

$$\bar{y}_{prop2} = \frac{\mu_y(1+\delta_1)}{2}\left[\left(\frac{\mu_x(1+\delta_2)}{\mu_x}\right)^{-1+\alpha_0} + \left(\frac{\mu_x}{\mu_x(1+\delta_2)}\right)^{1+\alpha_0}\right]$$
$$= \frac{\mu_y(1+\delta_1)}{2}\left[(1+\delta_2)^{-1+\alpha_0} + ((1+\delta_2)^{(-1)(1+\alpha_0)}\right]$$
$$= \frac{\mu_y(1+\delta_1)}{2}\left[(1+\delta_2)^{-1+\alpha_0} + (1+\delta_2)^{-1-\alpha_0}\right]$$

After solving and including the terms up to degree two in δ_1 and δ_2, we have

$$\bar{y}_{prop2} - \mu_y = \mu_y\left(\delta_1 - \delta_2 + (\alpha_0^{\,2} + 2)\frac{\delta_2^{\,2}}{2} - \delta_1\delta_2\right) \quad (2)$$

After taking the expectation on both sides, we can obtain the expression for bias up to the first order approximation as

$$\text{Bias}(\bar{y}_{prop2}) = \frac{1}{\mu_y}\left[\frac{(\alpha_0^{\,2}+2)}{2} R^2\, V(\bar{x}_{rss}) - R\, Cov(\bar{y}_{rss}, \bar{x}_{rss})\right]$$

Where $R = \frac{\mu_y}{\mu_x}$

Now substituting the values of $V(\bar{x}_{rss})$ and $Cov(\bar{y}_{rss}, \bar{x}_{rss})$, we get

Bias $(\bar{y}_{prop2}) = \frac{1}{rm}\frac{1}{\mu_y}\left[\frac{(\alpha_0{}^2+2)}{2} R^2\sigma_x{}^2 - R\sigma_{xy}\right] - \frac{1}{r}\frac{1}{m^2}\frac{1}{\mu_y}\left[\frac{\alpha_0{}^2}{2} R^2 \sum_{i=1}^{m} \tau_{x(i)}^2 - R\sum_{i=1}^{m} \tau_{xy(i)}\right]$

To find out the approximate expression for MSE, first take square on the both sides of (2) and consider the terms of δ's upto degree two

$(\bar{y}_{prop2} - \mu_y)^2 = \mu_y{}^2(\delta_1{}^2 + \delta_2{}^2 - 2\delta_1\delta_2)$

Taking expectation on both sides, we get

MSE $(\bar{y}_{prop2}) = V(\bar{y}_{rss}) + R^2\, V(\bar{x}_{rss}) - 2R\, Cov(\bar{y}_{rss}, \bar{x}_{rss})$

$= \frac{1}{rm}(\sigma_y{}^2 + R^2\sigma_x{}^2 - 2R\sigma_{xy}) - \frac{1}{r}\frac{1}{m^2}\sum_{i=1}^{m}(\tau_{y(i)} - R\tau_{x(i)})^2$

Which is the required approximate expression for MSE

Case 1: When $\alpha_0 = 1$

$\bar{y}_{prop2} = \frac{\bar{y}_{rss}}{2}\left[\left(\frac{\mu_x}{\bar{x}_{rss}}\right)^2 + 1\right]$

Which is the average of the mean per unit and quadratic ratio type estimator in RSS

Case 2: when $\alpha_0 = 0$

$\bar{y}_{prop2} = \bar{y}_{rss}\left[\frac{\mu_x}{\bar{x}_{rss}}\right]$

$= \bar{y}_{Rrss}$

Which is the ratio estimator in RSS

Case 3: When $\alpha_0 = -1$

$\bar{y}_{prop2} = \frac{\bar{y}_{rss}}{2}\left[1 + \left(\frac{\bar{x}_{rss}}{\mu_x}\right)^{-2}\right]$

$= \frac{\bar{y}_{rss}}{2}\left[1 + \left(\frac{\mu_x}{\bar{x}_{rss}}\right)^2\right]$

Which is the average of the mean per unit and quadratic ratio type estimator in RSS

Comparison with $\bar{y}_{rss}$

MSE $(\bar{y}_{prop2}) - V(\bar{y}_{rss}) = R^2\, V(x_{\bar{r}ss}) - 2R\, Cov(\bar{y}_{rss}, \bar{x}_{rss})$

MSE $(\bar{y}_{prop2}) - V(\bar{y}_{rss}) < 0$ iff $\rho_{\bar{y}_{rss}, \bar{x}_{rss}} > \frac{1}{2R}\sqrt{\frac{V(\bar{x}_{rss})}{V(\bar{y}_{rss})}} = \frac{1}{2}\frac{C_{\bar{x}_{rss}}}{C_{\bar{y}_{rss}}}$

Where $C_{\bar{x}_{rss}}$ and $C_{\bar{y}_{rss}}$ denotes the coefficient of variation for $\bar{x}_{rss}$ and $\bar{y}_{rss}$ respectively and $\rho_{\bar{y}_{rss}, \bar{x}_{rss}}$ is the correlation coefficient between $\bar{x}_{rss}$ and $\bar{y}_{rss}$.

Which shows that $\bar{y}_{prop2}$ is more efficient than the mean per unit estimator $\bar{y}_{rss}$ provided

$\rho_{\bar{y}_{rss}, \bar{x}_{rss}} > \frac{1}{2}\frac{C_{\bar{x}_{rss}}}{C_{\bar{y}_{rss}}}$

Comparison with $\bar{y}_{Rrss}$

MSE $(\bar{y}_{Rrss})$ = MSE $(\bar{y}_{prop2})$

Which shows that $\bar{y}_{prop2}$ is always equally efficient with $\bar{y}_{Rrss}$

Comparison with $\bar{y}_{Prss}$

MSE $(\bar{y}_{Prss})$ – MSE $(\bar{y}_{prop2}) \leq 0$

Which shows that $\bar{y}_{prop2}$ is always less efficient than $\bar{y}_{Prss}$

5. Empirical Study

We performed a simulation study to verify the results of the proposed estimators. A total number of 1,000 samples are drawn from bivariate normal distribution BVN (200, 100, 4, 4, ρ) for ρ = -0.7, -0.5, -0.3, 0.3, 0.5, 0.7 under RSS. The efficiency of an estimator Δ with respect to $\bar{y}_{rss}$ to estimate population mean is

$$\text{Efficiency }(\Delta) = \frac{MSE(\bar{y}_{rss})}{MSE(\Delta)}$$

The MSE values of $\bar{y}_{rss}$, $\bar{y}_{Rrss}$, $\bar{y}_{Prss}$, $\bar{y}_{prop1}$, $\bar{y}_{prop2}$ are obtained for different values of n, m and r in table 1 and its graph in figure 1

The efficiencies of $\bar{y}_{Rrss}$, $\bar{y}_{Prss}$, $\bar{y}_{prop1}$, $\bar{y}_{prop2}$ with respect to $\bar{y}_{rss}$ for different values of ρ and n are shown in table 2 and its graph in figure 2

Table 1: MSE (n = 9, m = 3, r = 3)

ρ	MSE $(\bar{y}_{rss})$	MSE $(\bar{y}_{Rrss})$	MSE $(\bar{y}_{Prss})$	MSE $(\bar{y}_{prop1})$	MSE $(\bar{y}_{prop2})$
-0.7	3.4676	2.2213	1.1237	0.0711	2.2213
-0.5	3.1120	2.2396	1.3192	0.1527	2.2396
-0.3	2.7566	2.1903	1.6370	0.2181	2.1903
0.3	1.6907	1.6382	2.1868	0.2182	1.6382
0.5	1.3352	1.3915	2.2357	0.1955	1.3915
0.7	0.9795	0.9337	2.2175	0.1592	0.9337

Table 2: Efficiencies

ρ	Eff $(\bar{y}_{rss})$	Eff $(\bar{y}_{Rrss})$	Eff $(\bar{y}_{Prss})$	Eff $(\bar{y}_{prop1})$	Eff $(\bar{y}_{prop2})$
-0.7	1	1.5610	3.0858	48.7707	1.5610
-0.5	1	1.3895	2.3590	20.3798	1.3895
-0.3	1	1.2585	1.6839	12.6391	1.2585
0.3	1	1.0320	0.7731	7.7483	1.0320
0.5	1	0.9595	0.5972	6.8296	0.9595
0.7	1	1.0490	0.4417	6.1526	1.0490

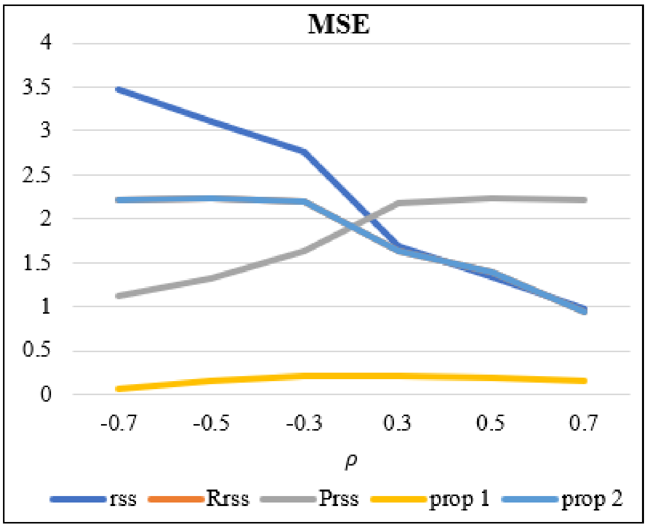

Figure 1: MSE of the various estimators

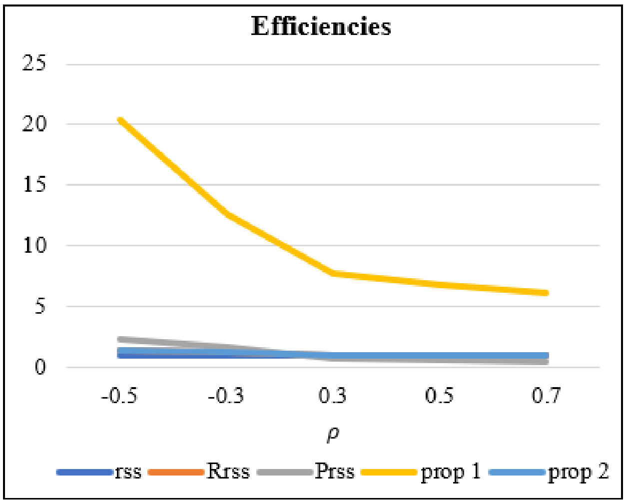

Figure 2: Efficiencies of the various estimators

6. Conclusion

This study demonstrates that the proposed estimators, prop1 and prop2, offer significant improvements in statistical efficiency over traditional estimators like the mean per unit, ratio and product estimators under ranked set sampling. The results from our simulations highlight the advantages of using these new estimators in scenarios with small sample sizes or when rankings are imperfect. This study is significant as it provides a more efficient method for estimating population mean, which can be particularly useful in fields where measurements are costly and difficult to obtain. Future research should explore their applicability in various fields to further validate these findings.

Acknowledgement

There was no specific grant received for this research from a public, private or non-profit funding organization.

References

[1] Bahl S. and Kumar M. Estimation of Finite Population Mean Using Multi-Auxiliary Variables. 2000 3, 67-74.

[2] Chaudhary M.K., Prajapati A. and Singh R. Two-phase sampling in estimation of population mean in the presence of non-response. 2014, 28-41.

[3] Cochran W. G. The estimation of the yields of cereal experiments by sampling for the ratio gain to total produce. Journal of agriculture society. 1940, 30, 262-275.

[4] Cochran, William G. Sampling Techniques. Third Edition, Wiley Eastern Limited, 1977.

[5] Das, A.K. and Tripathi. Use of auxiliary information in estimating the finite population variance. 1978, 139-148.

[6] Dell T.R. and Clutter J.L. Ranked set sampling theory with order statistics background. Biometrics. 1972, 28, 545-555.

[7] Douglas A. Wolfe. Ranked set sampling: Its relevance and impact on statistical inference. 2012.

[8] McIntyre G.A. A method for unbiased selective sampling, using ranked sets. Australian Journal of Agricultural Research. 1952, 3, 385-390.

[9] Stokes S.L. Estimation of variance using judgment ordered ranked set samples. Biometrics. 1980, 36, 35–42.

[10] Stokes S.L. Inferences on the correlation coefficient in bivariate normal populations from ranked set samples. Journal of the American Statistical Association. 1980, 75, 989–995.

[11] Stokes S.L. and Sager T.W. Characterization of a ranked-set sample with application to estimating distribution functions. Journal of the American Statistical Association. 1988, 83, 374–381.

[12] Swami HM, Muttlak HA. Estimation of ratio using rank set sampling. Biometrical Journal. 1996, 38(6), 753-764.

[13] Takahasi K. and Wakimoto K. On unbiased estimates of the population mean based on the sample stratified by means of ordering. Annals of the Institute of Statistical Mathematics. 1968, 20(1), 1-31.

[14] Yanagawa T. and Chen S.H. The mg-procedure in ranked set sampling. Journal of Statistical Planning and Inference. 1980, 4(1), 33–44.

[15] Yanagawa T. and Shirahata S. Ranked set sampling theory with selective probability matrix. The Australian Journal of Statistics. 1976, 18, 45–52.