



CDF estimation in multistage pair ranked set sampling

M. Mahdizadeh^{*1} , Ehsan Zamanzade^{2,3} 

¹Department of Statistics, Hakim Sabzevari University, P.O. Box 397, Sabzevar, Iran

²Department of Statistics, Faculty of Mathematics and Statistics, University of Isfahan, Isfahan 81746-73441, Iran

³School of Mathematics, Institute for Research in Fundamental Sciences (IPM), P.O. Box 19395-5746, Tehran, Iran

Abstract

Multistage pair ranked set sampling (MSPRSS) is a rank-based design that improves statistical inference with respect to simple random sampling of the same size. It is applicable when exact measurement is difficult, but judgment raking of the potential sample units can be done fairly accurately and easily. The ranking is usually performed based on personal judgment or a concomitant variable, and need not be totally free of errors. This article deals with estimating the cumulative distribution function in MSPRSS. The proposed estimator is theoretically compared with its contenders in the literature. The findings are supported by numerical evidence from simulation, and real data in the context of body fat analysis. Finally, a cost analysis is performed to show the advantage of the estimator.

Mathematics Subject Classification (2020). 62G05, 62G30

Keywords. CDF estimation, cost efficiency, judgment ranking

1. Introduction

Ranked set sampling (RSS) is a sampling technique that enhances inferential methods by incorporating the experimenter's judgment or additional information on the characteristic of interest. In this design, units sampled from the population are first informally ranked. Utilizing the ranks of the units, a subset of the sample is actually quantified. A reasonable method for ranking the units is to employ a concomitant variable.

For comparable sample sizes, RSS-based methods are generally superior to their counterparts in simple random sampling (SRS). The improvement in precision stems from the structure, in the form of the sampler's ranking, added to the data. Obviously, such a structure is absent in SRS.

McIntyre [16] proposed RSS as a method for obtaining better estimates of crop yield. Many statistical methods have been studied under this scheme. Bouza-Herrera and Al-Omari [3] discuss some new developments in this area. Some applications include auditing [8], environmental studies [9, 18], cluster randomized designs [19], and medicine [15].

*Corresponding Author.

Email addresses: mahdizadeh.m@hsu.ac.ir (M. Mahdizadeh), e.zamanzade@sci.ui.ac.ir (E. Zamanzade)

Received: 01.12.2021; Accepted: 16.07.2022

To draw a ranked set sample of size mn using set size m , the following procedure is replicated for n cycles:

- (1) First, m independent simple random samples of size m are drawn from the population.
- (2) The elements of the i th ($i = 1, \dots, m$) sample are ordered, and the unit with rank i is identified.
- (3) Finally, m units identified in step 2 are measured.

The final sample is given by $\{X_{ij} : i = 1, \dots, m; j = 1, \dots, n\}$, where X_{ij} be the i th judgement order statistic from the j th cycle. If the ranking process is not affected by errors, it is said that perfect ranking holds. Otherwise, imperfect ranking happens.

The basic RSS protocol has been modified, in different ways, to redress its shortcomings for specific situations. Frey and Feeman [6, 7], Frey [5], and Mahdizadeh and Zamanzade [10, 11, 14] are examples of recent works on generalizations of RSS. Muttalak [17] introduced pair ranked set sampling (PRSS) that reduces the number of sampled units from the population to almost half of that in RSS. A pair ranked set sample of size mn using set size m is obtained by replicating the next procedure for n cycles:

- (1) First, k independent simple random samples of size m are drawn from the population, where k is equal to $m/2$ or $(m+1)/2$ if m is even or odd.
- (2) The elements of the i th ($i = 1, \dots, k$) sample are ordered, and the units with ranks i and $m+1-i$ are identified.
- (3) Finally, m units identified in step 2 are measured.

If Y_{ij} denotes the i th judgement order statistic from the j th cycle, then the final sample is given by $\{Y_{ij} : i = 1, \dots, m; j = 1, \dots, n\}$. Under the perfect ranking setup, the units with ranks i and $m+1-i$, where $i = 1, \dots, k$ in each cycle of PRSS have a positive correlation.

To reduce possible errors in the judgment rankings, small values of m are used in the above designs. There exist generalizations of RSS and PRSS which allow to attain higher efficiency with a fixed set size. These are known as multistage ranked set sampling (MSRSS) and multistage pair ranked set sampling (MSPRSS). In this article, we study the cumulative distribution function (CDF) estimation in the latter design.

In Section 2, MSRSS and MSPRSS procedures are delineated. In Section 3, the CDF estimator based on MSPRSS is presented, and its mathematical properties are treated. In Section 4, the estimator and its competitor in MSRSS are compared by means of simulation and a real data set. Section 5 presents a cost analysis to demonstrate merit of the suggested estimator. In Section 6, we conclude with a summary. Figures are gathered in an appendix.

2. Multistage schemes

Accuracy of the judgment ranking process has a large impact on the efficiency of RSS. Small choices for m helps to lower the judgmental errors. MSRSS is a modification of RSS capable of attaining higher efficiency, given a fixed m . Al-Saleh and Al-Omari [1] studied estimating the population mean in MSRSS.

To draw an r th stage ranked set sample of size mn using set size m , the following procedure is replicated for n cycles:

- (1) First, m^{r+1} units are randomly identified from the population.
- (2) Next, the m^{r+1} units are randomly divided into m^{r-1} sets of size m^2 .
- (3) Steps 1 and 2 of RSS algorithm are done on each set in step 2 to have a (judgement) ranked set of size m . This yields m^{r-1} (judgement) ranked sets of size m .
- (4) Step 3 is done on the m^{r-1} ranked sets to have m^{r-2} second stage (judgement) ranked sets of size m .

- (5) Step 3 is repeated until ending in an r th stage (judgement) ranked set of size m .
- (6) Finally, m units identified in step 5 are measured.

The final sample is denoted by $\{X_{ij}^{(r)} : i = 1, \dots, m; j = 1, \dots, n\}$, where $X_{ij}^{(r)}$ is the i th judgement order statistic in the j th cycle. We note that the usual RSS is obtained by selecting $r = 1$.

It was mentioned that PRSS decreases the number of sampled units by almost half, as compared with RSS. Sparked by this idea, Mahdizadeh and Zamanzade [12] introduced MSPRSS, and applied it for estimating the population mean. An r th stage pair ranked set sample of size mn using set size m is obtained by replicating the next procedure for n cycles:

- (1) First, $k^r m$ units are randomly identified from the population, where k is equal to $m/2$ or $(m+1)/2$ if m is even or odd.
- (2) Next, the $k^r m$ units are randomly divided into k^{r-1} sets of size km .
- (3) Steps 1 and 2 of PRSS algorithm are done on each set in step 2 to have a (judgement) pair ranked set of size m . This yields k^{r-1} (judgement) pair ranked sets of size m .
- (4) Step 3 is done on the k^{r-1} pair ranked sets to have k^{r-2} second stage (judgement) pair ranked sets of size m .
- (5) Step 3 is repeated until ending in an r th stage (judgement) pair ranked set of size m .
- (6) Finally, m units identified in step 5 are measured.

Similarly, $\{Y_{ij}^{(r)} : i = 1, \dots, m; j = 1, \dots, n\}$ is the final sample, where $Y_{ij}^{(r)}$ is the i th judgement order statistic in the j th cycle. Clearly, PRSS is obtained by selecting $r = 1$.

MSPRSS is a viable alternative to MSRSS as it needs fewer number of sampling units. For example, by setting $m = 3$, $r = 5$ and $n = 1$, MSRSS and MSPRSS would require random samples of sizes 729 and 96, respectively. For common choices of the design parameters, the ratio of number of units used in MSRSS to that in MSPRSS is given by $\tau = (m/k)^r$. Figure 1 shows values of τ for $m \in \{3, 4, 5\}$ and $r \in \{1, 2, 3, 4, 5, 6\}$. It emerges that for fixed m , τ is increasing in r . Surprisingly, the ratio could be larger than 60 in some cases. This provides evidence that using MSPRSS is better suited for practical situations.

3. The CDF estimation

Suppose Z_1, \dots, Z_m is a random sample from a population with distribution function F . The CDF estimator in SRS is given by

$$\hat{F}(t) = \frac{1}{m} \sum_{i=1}^m I(Z_i \leq t), \quad (3.1)$$

with $I(\cdot)$ being the indicator function.

In what follows, we assume that MSRSS and MSPRSS procedures are performed using $n = 1$ since the relative efficiency is invariant when n changes. The notation for a sample of size m drawn from F under the two designs is therefore simplified as $X_1^{(r)}, \dots, X_m^{(r)}$ and $Y_1^{(r)}, \dots, Y_m^{(r)}$. Results of this section may be easily extended for the general case. The corresponding CDF estimators are denoted by $\hat{F}_{\text{RSS}}^{(r)}(t)$ and $\hat{F}_{\text{PRSS}}^{(r)}(t)$, i.e.

$$\hat{F}_{\text{RSS}}^{(r)}(t) = \frac{1}{m} \sum_{i=1}^m I(X_i^{(r)} \leq t) \quad (3.2)$$

and

$$\hat{F}_{\text{PRSS}}^{(r)}(t) = \frac{1}{m} \sum_{i=1}^m I(Y_i^{(r)} \leq t). \quad (3.3)$$

Al-Saleh and Samuh [2] investigated estimator (3.2). In the following, properties of estimator (3.3) is studied in the perfect ranking setup. For a technical reason in some proofs, it is assumed that $F(t)$ cannot be 0 or 1.

If k is the same as given in the MSPRSS, then k^* is defined as

$$k^* = \begin{cases} k & \text{if } k \text{ is even,} \\ k - 1 & \text{if } k \text{ is odd.} \end{cases}$$

This convention will be helpful in proving some results.

The expectation and variance of $\hat{F}_{PRSS}^{(r)}(t)$ are now obtained.

Proposition 3.1. *Let F be the population distribution function, and $F_i^{(r)}$ ($i = 1, \dots, m$) be that of $Y_i^{(r)}$. It holds that $E(\hat{F}_{PRSS}^{(r)}(t)) = F(t)$ and*

$$Var(\hat{F}_{PRSS}^{(r)}(t)) = \frac{1}{m^2} \left[\sum_{i=1}^m F_i^{(r)}(t) (1 - F_i^{(r)}(t)) + 2 \sum_{i=1}^{k^*} Cov(I(Y_i^{(r)} \leq t), I(Y_{m+1-i}^{(r)} \leq t)) \right].$$

Proof. The unbiasedness directly follows from the identity $F(t) = \sum_{i=1}^m F_i^{(r)}(t)/m$, which is given by Proposition 1 in [12]. The variance expression is derived in view of the estimator form. □

The following lemma is helpful in studying effect of increasing r on the variance of $\hat{F}_{PRSS}^{(r)}(t)$.

Lemma 3.2. *Let $\{X_1, \dots, X_n\}$ be jointly distributed binary random variables, with the corresponding order statistics $X_{(1)} \leq \dots \leq X_{(n)}$. Then, $Cov(X_{(i)}, X_{(j)}) \geq 0$, for any $i \neq j$, where $i, j = 1, \dots, n$.*

Proof. Without loss of generality, assume that $i < j$. If $p_i = E(X_{(i)})$ ($i = 1, \dots, n$), then we have

$$\begin{aligned} E(X_{(i)}X_{(j)}) &= P(X_{(i)} = X_{(j)} = 1) \\ &= P(X_{(i)} = 1) \\ &= p_i. \end{aligned}$$

This means that

$$\begin{aligned} Cov(X_{(i)}, X_{(j)}) &= E(X_{(i)}X_{(j)}) - E(X_{(i)})E(X_{(j)}) \\ &= p_i - p_i p_j \\ &= p_i(1 - p_j), \end{aligned}$$

which completes the proof. □

Suppose $Y_1^{(r-1)}, \dots, Y_m^{(r-1)}$ is drawn from F , and the i th order statistic of this sample is denoted by $Y_{(i)}^{(r-1)}$. From MSPRSS algorithm, it follows that $Y_i^{(r)} \stackrel{d}{=} Y_{(i)}^{(r-1)}$. This property plays a central role in proving the next result.

Proposition 3.3. *For fixed m and t , the variance of $\hat{F}_{PRSS}^{(r)}(t)$ is decreasing in r .*

Proof. We have

$$\begin{aligned}
 \text{Var} \left(\hat{F}_{\text{PRSS}}^{(r-1)}(t) \right) &= \frac{1}{m^2} \text{Var} \left(\sum_{i=1}^m I \left(Y_{(i)}^{(r-1)} \leq t \right) \right) \\
 &= \frac{1}{m^2} \left[\sum_{i=1}^m \text{Var} \left(I \left(Y_{(i)}^{(r-1)} \leq t \right) \right) \right. \\
 &\quad \left. + \sum_{i \neq j=1}^m \text{Cov} \left(I \left(Y_{(i)}^{(r-1)} \leq t \right), I \left(Y_{(j)}^{(r-1)} \leq t \right) \right) \right] \\
 &= \frac{1}{m^2} \left[\sum_{i=1}^m \text{Var} \left(I \left(Y_{(i)}^{(r-1)} \leq t \right) \right) \right. \\
 &\quad \left. + 2 \sum_{i=1}^{k^*} \text{Cov} \left(I \left(Y_{(i)}^{(r-1)} \leq t \right), I \left(Y_{(m+1-i)}^{(r-1)} \leq t \right) \right) + R \right],
 \end{aligned}$$

with R being sum of covariances of the form $\text{Cov} \left(I \left(Y_{(i')}^{(r-1)} \leq t \right), I \left(Y_{(j')}^{(r-1)} \leq t \right) \right)$, on condition that i' and j' cannot be written as i and $m+1-i$ for some $i \in \{1, \dots, k^*\}$.

Owing to the previous lemma, $R \geq 0$, implying that

$$\begin{aligned}
 \text{Var} \left(\hat{F}_{\text{PRSS}}^{(r-1)}(t) \right) &\geq \frac{1}{m^2} \left[\sum_{i=1}^m \text{Var} \left(I \left(Y_{(i)}^{(r-1)} \leq t \right) \right) \right. \\
 &\quad \left. + 2 \sum_{i=1}^{k^*} \text{Cov} \left(I \left(Y_{(i)}^{(r-1)} \leq t \right), I \left(Y_{(m+1-i)}^{(r-1)} \leq t \right) \right) \right] \\
 &= \frac{1}{m^2} \left[\sum_{i=1}^m \text{Var} \left(I \left(Y_i^{(r)} \leq t \right) \right) \right. \\
 &\quad \left. + 2 \sum_{i=1}^{k^*} \text{Cov} \left(I \left(Y_i^{(r)} \leq t \right), I \left(Y_{m+1-i}^{(r)} \leq t \right) \right) \right] \\
 &= \text{Var} \left(\hat{F}_{\text{PRSS}}^{(r)}(t) \right),
 \end{aligned}$$

that completes the proof. \square

The following result allows to compare the variances of $\hat{F}(t)$ and $\hat{F}_{\text{PRSS}}^{(r)}(t)$.

Proposition 3.4. For fixed m and t , $\hat{F}_{\text{PRSS}}^{(r)}(t)$ is more efficient than $\hat{F}(t)$.

Proof. Thanks to Proposition 3.3, we need to prove the statement for $r = 1$. Let $Z_{(1)}, \dots, Z_{(m)}$ be the order statistics of Z_1, \dots, Z_m . By an argument similar to proof of Proposition 3.3, one can write

$$\begin{aligned}
 \text{Var} \left(\hat{F}(t) \right) &= \frac{1}{m^2} \text{Var} \left(\sum_{i=1}^m I \left(Z_{(i)} \leq t \right) \right) \\
 &= \frac{1}{m^2} \left[\sum_{i=1}^m \text{Var} \left(I \left(Z_{(i)} \leq t \right) \right) + \sum_{i \neq j=1}^m \text{Cov} \left(I \left(Z_{(i)} \leq t \right), I \left(Z_{(j)} \leq t \right) \right) \right] \\
 &\geq \frac{1}{m^2} \left[\sum_{i=1}^m \text{Var} \left(I \left(Z_{(i)} \leq t \right) \right) + 2 \sum_{i=1}^{k^*} \text{Cov} \left(I \left(Z_{(i)} \leq t \right), I \left(Z_{(m+1-i)} \leq t \right) \right) \right] \\
 &= \text{Var} \left(\hat{F}_{\text{PRSS}}^{(1)}(t) \right),
 \end{aligned}$$

as was to be shown. \square

It can be formally shown that the CDF estimator in RSS surpasses its PRSS analog. This is so because

$$\begin{aligned} \text{Var} \left(\hat{F}_{\text{PRSS}}^{(1)}(t) \right) &= \frac{1}{m^2} \text{Var} \left(\sum_{i=1}^m I \left(Y_i^{(1)} \leq t \right) \right) \\ &= \frac{1}{m^2} \left[\sum_{i=1}^m \text{Var} \left(I \left(Y_i^{(1)} \leq t \right) \right) \right. \\ &\quad \left. + 2 \sum_{i=1}^{k^*} \text{Cov} \left(I \left(Y_i^{(1)} \leq t \right), I \left(Y_{m+1-i}^{(1)} \leq t \right) \right) \right] \\ &\geq \frac{1}{m^2} \left[\sum_{i=1}^m \text{Var} \left(I \left(Y_i^{(1)} \leq t \right) \right) \right] \\ &= \text{Var} \left(\hat{F}_{\text{RSS}}^{(1)}(t) \right), \end{aligned}$$

where the inequality is verified as any pair of sample order statistics have a positive covariance. One would expect this property to hold for any $r > 1$. This is supported by simulation results, but a mathematical reasoning does not seem to be an easy job. The difficulty roots in the fact that units resulted from first stage in MSPRSS are dependent. We close this section with a conjecture.

Conjecture 3.5 For fixed m, r and t , $\hat{F}_{\text{RSS}}^{(r)}(t)$ is more efficient than $\hat{F}_{\text{PRSS}}^{(r)}(t)$.

4. Numerical studies

A simulation study was used to assess the performance of $\hat{F}_{\text{PRSS}}^{(r)}(t)$ by comparing it to $\hat{F}(t)$ and $\hat{F}_{\text{RSS}}^{(r)}(t)$ in terms of variance. For each scenario under consideration, we determine the relative precisions (RPs)

$$\text{RP}_1(t) = \frac{\text{Var} \left(\hat{F}(t) \right)}{\text{Var} \left(\hat{F}_{\text{PRSS}}^{(r)}(t) \right)}$$

and

$$\text{RP}_2(t) = \frac{\text{Var} \left(\hat{F}_{\text{RSS}}^{(r)}(t) \right)}{\text{Var} \left(\hat{F}_{\text{PRSS}}^{(r)}(t) \right)},$$

where variance estimation is based on 100,000 samples. Having defined the RPs in this way, values greater than unity signify the preference of $\hat{F}_{\text{PRSS}}^{(r)}(t)$.

We note that above RPs are independent of the parent distribution F . For each combination of $m \in \{3, 4, 5\}$ and $r \in \{1, 3, 5\}$, the RPs were computed at quantiles of order $p \in (0, 1)$. The results from our precision comparison appear in Figures 2-4.

It is observed that our estimator beats its SRS rival in all situations. For each m and p , values of RP_1 are increasing in r . The CDF estimator in MSPRSS is outperformed by that in MSRSS as RP_2 values are generally smaller than unity. The efficiency loss from using MSPRSS instead of MSRSS is more pronounced with even m , especially when p approaches 0.5 (see Figure 3). All these trends are in agreement with the theoretical results in Section 3.

The above simulation experiment is based on the perfect ranking assumption. We now use a real data set to see how judgment ranking errors affect the suggested estimator. Mahdizadeh and Zamanzade [13] used this data set[†], that includes measurements of 15 variables for 252 men. Body fat percentage (X) and abdomen circumference (Y) are

[†]It is accessible at <http://lib.stat.cmu.edu/datasets/bodyfat>

the two variables considered for any man. Thanks to its high degree of accuracy, dual-energy X-ray absorptiometry is now a popular method for body fat analysis. The use of this technology is, however, hampered by the implementation cost. Here, Y serves as a readily accessible measure of obesity. Exact quantification of X is expensive with respect to the judgment ranking using Y . In this situation, employing the RSS-based sampling techniques is fitting.

Suppose we aim to estimate the CDF of X in the population consisting of the 252 men. The two RPs defined above were determined using 100,000 samples drawn with replacement from the population. In so doing, $m \in \{3, 5\}$ and $r \in \{1, 2, 3\}$ were chosen. Also, the ranking phase was based on Y . It is worth noting that X and Y have a correlation coefficient of 0.81. Figures 5 and 6 display the results that reveal properties similar to those in Figures 2-4.

5. Comparison with cost consideration

In Section 4, the CDF estimator in MSPRSS was compared with its analogs in SRS and MSRSS, given a fixed sample size. This is not a fair method because the cost involved in the judgment ranking are not incorporated appropriately. To resolve this problem, we employ a cost model in RSS due to [4]. It allows to compare different estimators based on fixed cost.

Assume that the stratification cost for each quantified unit in MSPRSS is equal to c_s . Also, let c_q be the cost of drawing and quantifying a unit, regardless of the ranking. Then the relative efficiency (RE) of MSPRSS to SRS in estimating the population CDF is

$$\begin{aligned} \text{RE}_1(t) &= \frac{N_P}{N_S} \text{RP}_1(t) \\ &= \frac{c_q}{c_q + c_s} \text{RP}_1(t), \end{aligned}$$

where N_P and N_S are the total number of measured units in MSPRSS and SRS, respectively. The number of sample units required in MSRSS is τ times of that in MSPRSS, where τ was given in Section 2. This yields a similar relationship between the stratification cost for each quantified unit in the two designs. Thus, the RE of MSPRSS to MSRSS in estimating the population CDF is

$$\begin{aligned} \text{RE}_2(t) &= \frac{N_P}{N_R} \text{RP}_2(t) \\ &= \frac{c_q + \tau c_s}{c_q + c_s} \text{RP}_2(t), \end{aligned}$$

where N_R indicates the total number of quantified units in MSRSS.

To get insight into effect of c_s on the above REs, a numerical study was performed in the perfect ranking setup. It was assumed that $m = 3$ and $r \in \{1, 3, 5\}$. Next, values of $\text{RP}_1(t)$ and $\text{RP}_2(t)$ were determined using 100,000 samples at $t = F^{-1}(0.5)$, i.e. the population median. Using these estimates, the REs were plotted in Figure 7 as a function of c_s when $c_q = \$50$.

It transpires that the CDF estimator in MSPRSS may be less efficient than its SRS version if the stratification cost is too high (see the upper panel in Figure 7). This is consistent with a point emphasized in the literature about necessary condition for implementing rank-based designs like MSPRSS. It states that a small set of sample units can be ranked fairly accurately and inexpensively. Finally, the trend observed in the lower panel of Figure 7 is very interesting. It signifies that the CDF estimator in MSPRSS surpasses its competitor in MSRSS as the stratification cost goes higher. This is more pronounced with larger stage numbers.

6. Conclusion

RSS utilizes a judgment ranking mechanism to put additional structure in data. This structure resembles stratifying the population prior to drawing a simple random sample. While stratified SRS incorporates auxiliary information from the entire population, RSS uses that from only the units in an initial sample. For comparable sample sizes, RSS-based methods are generally more efficient than their SRS counterparts.

MSRSS is a modification of RSS aimed at reducing effect of the judgment ranking errors. The required number of units in MSRSS becomes excessive by increasing r , especially for large m . This feature may hinder the use of MSRSS in practice. MSPRSS is a newly suggested design that needs less sampling process than MSRSS.

In this article, a CDF estimator under MSPRSS is developed. The suggested estimator is theoretically compared with its contenders in MSRSS and SRS. Numerical studies are also conducted to get insight of the finite sample behavior of the suggested estimator. The advantage of our estimator over its MSRSS analog is demonstrated using a cost analysis. In a subsequent work, we plan to study the proportion estimation in MSPRSS design.

Acknowledgment. The authors wish to thank the reviewers for careful reading of the manuscript, and providing many constructive comments. Ehsan Zamanzade's research was carried out in IPM Isfahan branch and was in part supported by a grant from IPM, Iran (No. 1401620420).

References

- [1] M.F. Al-Saleh and A.I. Al-Omari, *Multistage ranked set sampling*, J. Statist. Plann. Inference **102** (2), 273-286, 2002.
- [2] M.F. Al-Saleh and M.J. Samuh, *On multistage ranked set sampling for distribution and median estimation*, Comput. Statist. Data Anal. **52** (4), 2066-2078, 2008.
- [3] C.N. Bouza-Herrera and A.I. Al-Omari, *Ranked Set Sampling, 65 Years Improving the Accuracy in Data Gathering*, Academic Press, 2019.
- [4] T.R. Dell and J.L. Clutter, *Ranked set sampling theory with order statistics background*, Biometrics **28** (2), 545-555, 1972.
- [5] J. Frey, *Bounds on the efficiency of unbalanced ranked-set sampling*, Comm. Statist. Theory Methods **49** (1), 243-256, 2020.
- [6] J. Frey and T.G. Feeman, *Efficiency comparisons for partially rank-ordered set sampling*, Statist. Papers **58** (4), 1149-1163, 2017.
- [7] J. Frey and T.G. Feeman, *Finding the maximum efficiency for multistage ranked-set sampling*, Comm. Statist. Theory Methods **47** (17), 4131-4141, 2018.
- [8] N.M. Gemayel, E.A. Stasny, J.A. Tackett and D.A. Wolfe, *Ranked set sampling: An auditing application*, Rev. Quant. Finance Account. **39** (4), 413-422, 2012.
- [9] M. Mahdizadeh and E. Strzalkowska-Kominiak, *Resampling based inference for a distribution function using censored ranked set samples*, Comput. Statist. **32** (4), 1285-1308, 2017.
- [10] M. Mahdizadeh and E. Zamanzade, *Reliability estimation in multistage ranked set sampling*, REVSTAT **15** (4), 565-581, 2017.
- [11] M. Mahdizadeh and E. Zamanzade, *Dynamic reliability estimation in a rank-based design*, Probab. Math. Statist. **39** (1), 1-18, 2019.
- [12] M. Mahdizadeh and E. Zamanzade, *Efficient body fat estimation using multistage pair ranked set sampling*, Stat. Methods Med. Res. **28** (1), 223-234, 2019.
- [13] M. Mahdizadeh and E. Zamanzade, *Estimating asymptotic variance of M-estimators in ranked set sampling*, Comput. Statist. **35** (4), 1785-1803, 2020.
- [14] M. Mahdizadeh and E. Zamanzade, *Estimation of a symmetric distribution function in multistage ranked set sampling*, Statist. Papers **61** (2), 851-867, 2020.

- [15] M. Mahdizadeh and E. Zamanzade, *Smooth estimation of the area under the ROC curve in multistage ranked set sampling*, *Statist. Papers* **62** (4), 1753-1776, 2021.
- [16] G.A. McIntyre, *A method of unbiased selective sampling using ranked sets*, *Aust. J. Agric. Res.* **3** (4), 385-390, 1952.
- [17] H.A. Muttalak, *Pair rank set sampling*, *Biom. J.* **38** (7), 879-885, 1996.
- [18] E. Strzalkowska-Kominiak and M. Mahdizadeh, *On the Kaplan-Meier estimator based on ranked set samples*, *J. Stat. Comput. Simul.* **84** (12), 2577-2591, 2014.
- [19] X. Wang, J. Lim and L. Stokes, *Using ranked set sampling with cluster randomized designs for improved inference on treatment effects*, *J. Amer. Statist. Assoc.* **111** (516), 1576-1590, 2016.

Appendix A.

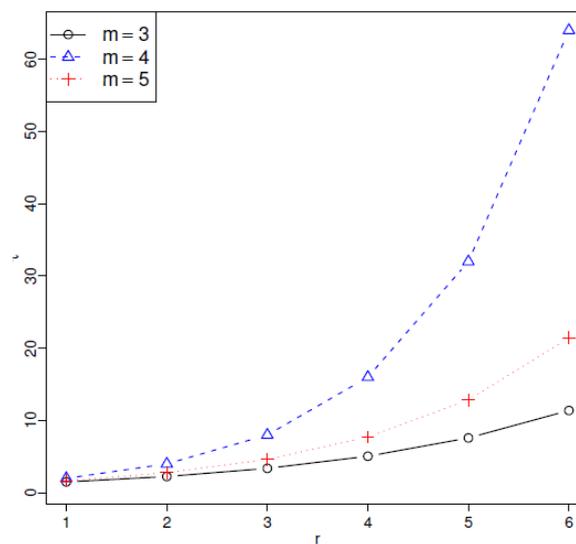


Figure 1. Values of τ for some choices of m and r .

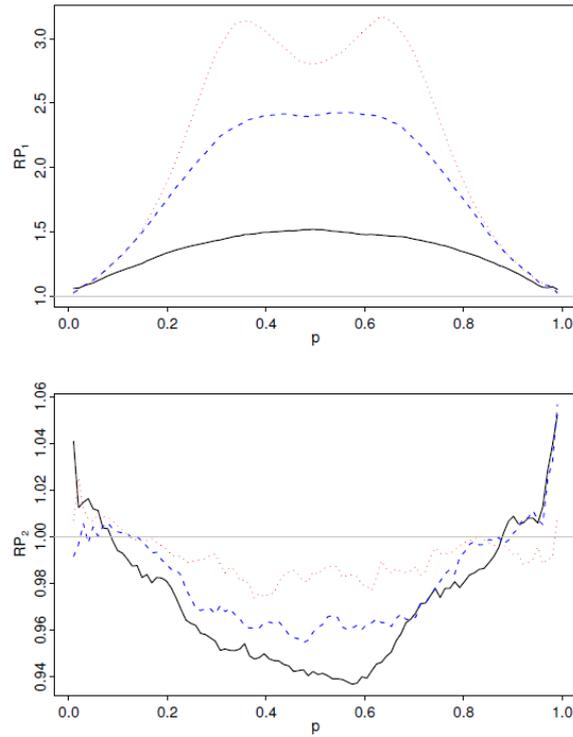


Figure 2. Estimated RPs for $m = 3$, where black/solid, blue/dashed and red/dotted curves relate to $r = 1$, $r = 3$ and $r = 5$, respectively.

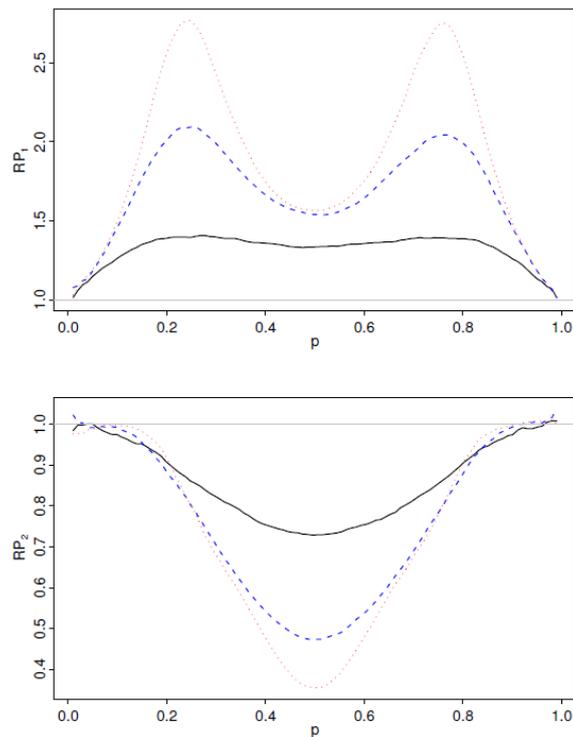


Figure 3. Estimated RPs for $m = 4$, where black/solid, blue/dashed and red/dotted curves relate to $r = 1$, $r = 3$ and $r = 5$, respectively.

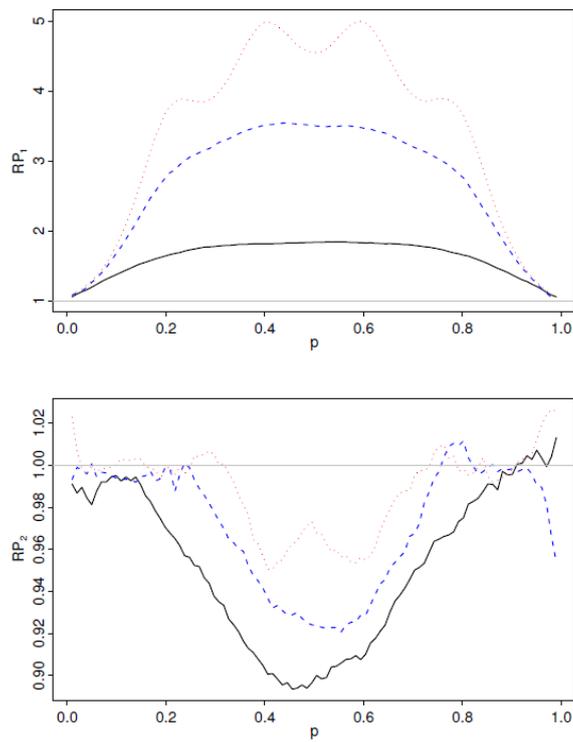


Figure 4. Estimated RPs for $m = 5$, where black/solid, blue/dashed and red/dotted curves relate to $r = 1$, $r = 3$ and $r = 5$, respectively.

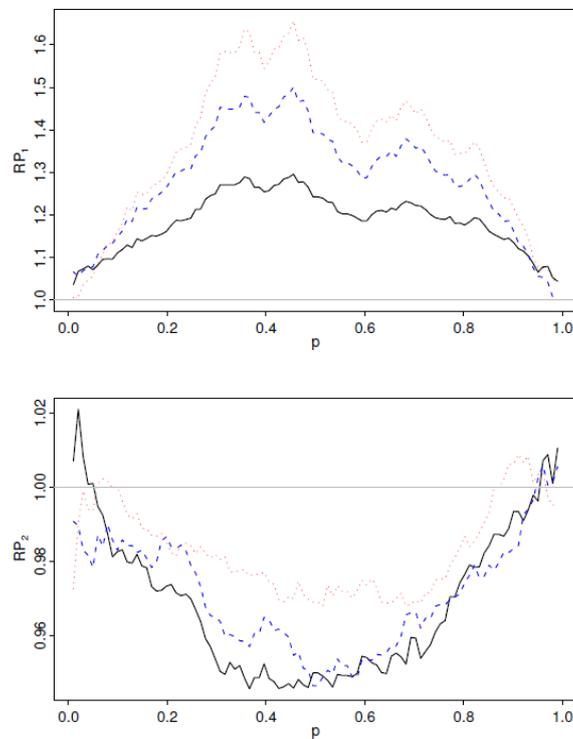


Figure 5. Estimated RPs from the body fat data for $m = 3$, where black/solid, blue/dashed and red/dotted curves relate to $r = 1$, $r = 2$ and $r = 3$, respectively.

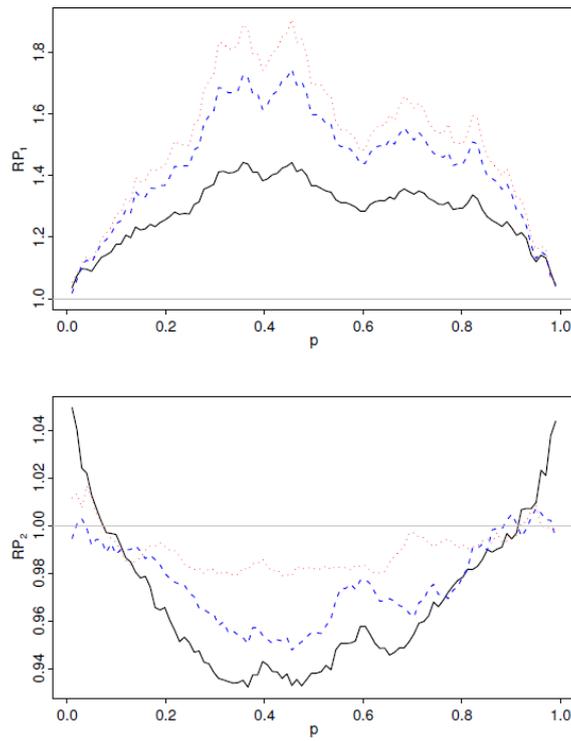


Figure 6. Estimated RPs from the body fat data for $m = 5$, where black/solid, blue/dashed and red/dotted curves relate to $r = 1$, $r = 2$ and $r = 3$, respectively.

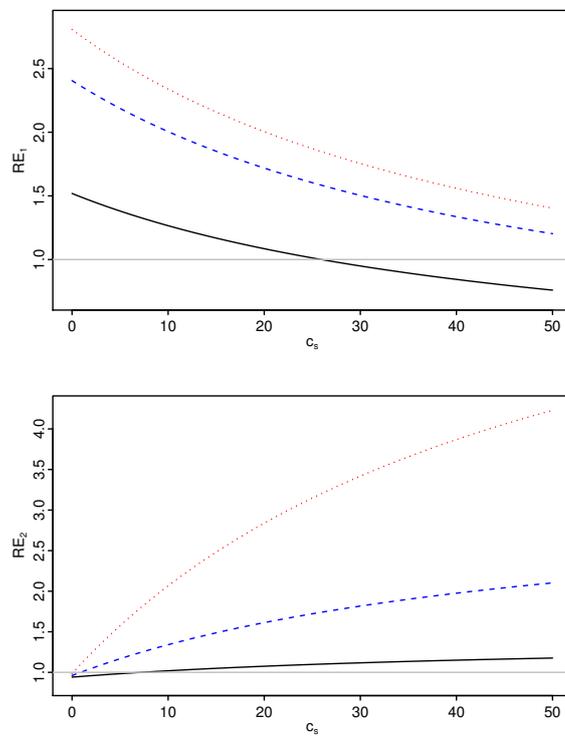


Figure 7. Estimated REs for $m = 3$ at the population median, where black/solid, blue/dashed and red/dotted curves relate to $r = 1$, $r = 3$ and $r = 5$, respectively.