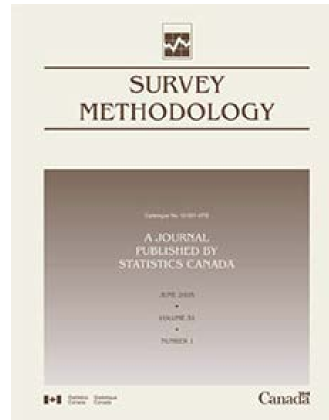
Catalogue no. 12-001-X
ISSN 1492-0921

Survey Methodology

Model based inference using ranked set samples



by Omer Ozturk and Konul Bayramoglu Kavlak

Release date: June 21, 2018

Statistics Canada Statistique Canada

Canada

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

email at STATCAN.infostats-infostats.STATCAN@canada.ca

telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following toll-free numbers:

- Statistical Information Service 1-800-263-1136
- National telecommunications device for the hearing impaired 1-800-363-7629
- Fax line 1-877-287-4369

Depository Services Program

- Inquiries line 1-800-635-7943
- Fax line 1-800-565-7757

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "Contact us" > "Standards of service to the public."

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Standard table symbols

The following symbols are used in Statistics Canada publications:

. not available for any reference period
.. not available for a specific reference period
... not applicable
0 true zero or a value rounded to zero
0^s value rounded to 0 (zero) where there is a meaningful distinction between true zero and the value that was rounded
^{p} preliminary
^{r} revised
x suppressed to meet the confidentiality requirements of the *Statistics Act*
^{E} use with caution
F too unreliable to be published
* significantly different from reference category ($p < 0.05$)

An HTML version is also available.

Cette publication est aussi disponible en français.

Survey Methodology, June 2018
Vol. 44, No. 1, pp. 1-16
Statistics Canada, Catalogue No. 12-001-X

Model based inference using ranked set samples

Omer Ozturk and Konul Bayramoglu Kavlak[1]

Abstract

This paper develops statistical inference based on super population model in a finite population setting using ranked set samples (RSS). The samples are constructed without replacement. It is shown that the sample mean of RSS is model unbiased and has smaller mean square prediction error (MSPE) than the MSPE of a simple random sample mean. Using an unbiased estimator of MSPE, the paper also constructs a prediction confidence interval for the population mean. A small scale simulation study shows that estimator is as good as a simple random sample (SRS) estimator for poor ranking information. On the other hand it has higher efficiency than SRS estimator when the quality of ranking information is good, and the cost ratio of obtaining a single unit in RSS and SRS is not very high. Simulation study also indicates that coverage probabilities of prediction intervals are very close to the nominal coverage probabilities. Proposed inferential procedure is applied to a real data set.

Key Words: Ranked set sampling; Finite population; Mean square prediction error; Sampling cost model; Coherent ranking; Concomitant ranking; Visual ranking.

1 Introduction

In many survey sampling studies, it is very common that the sampling frame has additional auxiliary information in addition to characteristic of interest. Under a fairly strong modeling assumption, this auxiliary information improves the statistical inference. For example, ratio and regression estimators use covariate information under a linearity assumption to estimate the population mean or total. The auxiliary information can also be used under a weaker assumption in a ranked set sample (RSS) and judgment post stratified (JPS) sample. These samples use auxiliary information to increase the information content of each measured unit through a ranking process. The ranking process is performed in a small set of size H formed by combining the measured unit with an additional $H-1$ unmeasured units from the population. Ranking process is performed either before or after measurement and determines the relative position of each measured unit. Ranking information can be obtained from either a visual inspection or some other form of ranking process. A reasonable ranking mechanism requires some sort of monotonic relationship between the ranking variable and response, which is much weaker than the strong linearity assumption of regression and ratio estimators.

A balanced ranked set sample of set size H and cycle size d can be constructed by first selecting $n = Hd$ simple random samples of size H from the population and ranking the units in each sample without measurement from smallest to largest. In these n ranked sets (samples), one then measures the units with rank 1 in the first d sets, the unit with rank 2 in the next d sets and so on. This yields samples of H different sets of judgment order statistics, each of which has d independent and identically distributed judgment order statistics.

A sharp contrast exists between an observation from SRS and RSS, where the observation from an SRS sample provides information only about the unit on which it was measured while the observation from an RSS sample, in addition to the information that the measured unit provides, also provides limited

1. Omer Ozturk, Department of Statistics, The Ohio State University, 1958 Neil Avenue, Columbus, OH, 43210, U.S.A. E-mail: omer@stat.osu.edu; Konul Bayramoglu Kavlak, Department of Actuarial Sciences, Hacettepe University, Ankara, Turkey. E-mail: konul.bayramoglu@hacettepe.edu.tr.

information about the other $(H-1)$ unmeasured units in the set through the relative position (rank) of measured unit. Since ranking process does not require a formal measurement and is usually less expensive in comparison with formal measurement, the RSS sample provides substantial amount of reduction in sampling cost.

A JPS sample differs from an RSS sample in that the ranking step comes after the construction of an SRS sample. Construction of a JPS sample of size n requires a set size H. Once the set size H is determined, one first draws a simple random sample of size n and makes a measurement on each of the n units. For each measured unit in the sample, one then selects additional $H-1$ units to form a set of size H. The units in this set are ranked from smallest to largest without measurement and the rank of the measured unit in the set is recorded. The JPS sample then consists of n measured values, together with their ranks.

Both RSS and JPS samples induces a stochastic structure among measured units in which observations in judgment class h are usually smaller than the observations in judgment class h', $h < h'$. This stochastic ordering feature spreads the measured units in the support of the distribution and creates a better representative sample than a simple random sample. The nature of stochastic ordering in a JPS sample is significantly different from the stochastic ordering in an RSS sample. A JPS sample consists of a simple random sample and an associated rank vector. This rank vector is loosely related to the sample and may be ignored if desired. On the other hand, an RSS sample is measured as judgment order statistics, judgment ranks can not be separated from the observed values. An RSS sample can not be treated as an SRS sample.

Both JPS and RSS sampling designs have generated extensive research interest in a finite population setting. Patil, Sinha and Taillie (1995) used ranked set sample to estimate population mean for a population of size N when the sample is constructed without replacement. Takahasi and Futatsuya (1998) showed that the ranked set sample estimator of the population mean is more precise than the simple random sample estimator when samples are drawn without replacement from a finite population. Deshpande, Frey and Ozturk (2006) described three different sampling designs and constructed nonparametric confidence intervals for population quantiles. Al-Saleh and Samawi (2007), Ozdemir and Gokpinar (2007 and 2008), Gokpinar and Ozdemir (2010), Ozturk and Jozani (2013), Frey (2011) and Ozturk (2014, 2015, 2016a) computed inclusion probabilities and constructed Horwitz-Thompson type estimators for population mean and total based on a ranked set sample. These research papers show that an RSS design yields a substantial amount of improvement in efficiency over the usual simple random sampling design. Ozturk (2016b) developed estimators for population mean based on a JPS sample where he showed that the estimator needs a finite population correction factor similar to the one used in a simple random sample.

All available research in literature in JPS and RSS sampling designs in a finite population setting considers design-based approach. To our knowledge, super population model has not been used. In this paper, we develop a model-based statistical inference using RSS sampling design for population mean and total in a finite population setting. Similar results, with some additional variation due to random judgment class samples sizes, can also be established for a JPS sampling design. Because of the random judgment class sample sizes, the estimators based on a JPS sample are less efficient than the estimators based on an RSS sample. For this reason, the JPS sample is not considered further in this paper. Section 2 clearly defines the model and describes the sampling designs for RSS under super population model. We show that estimators of population mean and total are model-unbiased and their mean square prediction errors (MSPE) are smaller than the MSPE of the same estimators of an SRS sample. Section 3 constructs unbiased estimators for the MSPE and provides approximate confidence intervals for the population mean and total. Section 4 introduces cost models to account the effect of additional cost (excess of the cost of construction of SRS sample) in construction of RSS sample. Section 5 provides empirical evidence about the performance of the estimators. Section 6 applies the proposed estimators to an example in a finite population setting. Section 7 provides some concluding remarks.

2 Sampling designs

We consider RSS sampling designs from a super population model to draw statistical inference in a finite population setting. Let Y be the characteristic of interest. The copies of Y, $Y_1, \ldots, Y_N$, are considered as independent identically distributed (iid) random variables from a super population. Basic assumption for this super population model can be stated as

$$\text{Model: } Y_1, \ldots, Y_N \text{ independent identically distributed with } E_M(Y_i) = \mu, V_M(Y_i) = \sigma^2. \quad (2.1)$$

The subscript M in model (2.1) is used to highlight that the mean and variance are computed based on a super population model, not the randomization distribution as in Ozturk (2016b). In this super population model, μ and σ^2 represent unknown infinite population parameters.

In super population model, a particular realization, $y_1, \ldots, y_N$, of random variables $Y_1, \ldots, Y_N$ from model (2.1), is considered as a finite population. Let $P^N = \{y_1, \ldots, y_N\}$ denotes this finite population. Ranked set sample is constructed from P^N. Without loss of generality, we assume that $y_{(1)} < y_{(2)} < \ldots < y_{(N)}$ are ordered values of $y_1, \ldots, y_N$ where $y_{(i)}$ is the i^{th} largest value of Y in P^N. Throughout the paper, H and d are used to denote the set and cycle sizes, respectively.

To construct a ranked set sample, one selects a set of H experimental units, $y_{s_1}, \ldots, y_{s_H}$, at random from P^N and ranks them based on their Y values in an increasing magnitude without actual measurement. Ranking process can be performed either using visual inspection or some auxiliary variables and hence subjected to ranking error. The unit that corresponds to the smallest Y, $y_{[1]}$, is identified and measured where the square bracket in the subscript, [1], denotes the rank of the smallest unit (rank 1) in the set $\left\{y_{[1]}, y^*_{[2]}, \ldots, y^*_{[H]}\right\}$. The remaining unmeasured units are denoted with $\left\{y^*_{[2]}, \ldots, y^*_{[H]}\right\}$. After $y_{[1]}$ is measured, none of the H units in the set $\left\{y_{[1]}, y^*_{[2]}, \ldots, y^*_{[H]}\right\}$ are returned to the population. One then selects another set of H experimental units at random from the remaining population P^{N-H} and ranks them without measurement. This time, the unit that corresponds to the second smallest Y, $y_{[2]}$, is identified and measured in $\left\{y^*_{[1]}, y_{[2]}, y^*_{[3]}, \ldots, y^*_{[H]}\right\}$. This process is continued until a simple random sample of size H is taken from the reduced population $P^{N-H(H-1)}$ and the H^{th} smallest unit is identified and measured in the set $\left\{y^*_{[1]}, y^*_{[2]}, \ldots, y^*_{[H-1]}, y_{[H]}\right\}$. This is called a cycle. A cycle selects H disjoint sets, each of size H and only measures H units. The remaining $H(H-1)$ units are used only for ranking purposes. The cycles are repeated d times to yield a ranked set sample of size $n = dH$ units. A ranked set sample can then be represented as

$$W_{h,i,H} \;=\; \left\{y^*_{[1]i}, \ldots, y^*_{[h-1]i}, y_{[h]i}, y^*_{[h+1]i}, \ldots, y^*_{[H]i}\right\}, \quad h = 1, \ldots, H, \quad i = 1, \ldots, d, \tag{2.2}$$

where only $y_{[h]i}$, $h = 1, \ldots, H$, $i = 1, \ldots, d$, are measured. The other values are used to obtain the rank of the measured values. Units in sets $W_{h,i,H}$ and $W_{h',i',H}$ are all independent if either $h \neq h'$ or $i \neq i'$, but the units in $W_{h,i,H}$ are all correlated since they are ranked in the same set. Under model (2.1), means, variances and covariances of judgment order statistics are given by

$$\begin{aligned} E_M(Y_{[h]i}) \;&=\; \mu_{[h]}, \mathrm{Var}_M(Y_{[h]i}) = \sigma^2_{[h]}, \\ \mathrm{Cov}_M(Y_{[h]i}, Y_{[h']i}) \;&=\; \begin{cases} \sigma_{[h,h']} & \text{if } Y_{[h]i}, Y_{[h']i} \text{ are from the same set} \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

It should be noted that since all sets are disjoint no units can be used more than once in any one of the sets. Hence all sample units are distinct. Since the sets are independently ranked $Y_{[h]i}$'s are mutually independent. Observations having the same rank h, $Y_{[h]i}$, $i = 1, \ldots, d$ are identically distributed.

Estimator of the population mean μ based on RSS data in equation (2.2) can be defined as follows.

$$\overline{Y}_R \;=\; \frac{1}{dH}\sum_{h=1}^{H}\sum_{i=1}^{d} Y_{[h]i}. \tag{2.3}$$

It can be immediately observed that the estimator $\overline{Y}_R$ is model unbiased. In other words, under the model (2.1), $E_M\left(\overline{Y}_R - \overline{Y}_N\right) = 0$, where $\overline{Y}_N = \frac{1}{N}\sum_{i=1}^{N} Y_i$.

We now consider the mean square prediction error (MSPE) of the estimator $\overline{Y}_R$ under model (2.1)

$$\mathrm{MSPE}_M\left(\overline{Y}_R\right) \;=\; E_M\left(\overline{Y}_R - \frac{1}{N}\sum_{i=1}^{N} Y_i\right)^2 = E_M\left(\overline{Y}_R - \overline{Y}_N\right)^2.$$

Since the predictor $\overline{Y}_R$ is model unbiased for $\overline{Y}_N$, $E_M\left(\overline{Y}_R - \overline{Y}_N\right) = 0$, the mean square prediction error (MSPE) of $\overline{Y}_R$ is the same as $\mathrm{Var}_M\left(\overline{Y}_R - \overline{Y}_N\right)$.

Theorem 1: *Let $Y_{[h]i}, h = 1, \ldots, H, i = 1, \ldots, d$, be a ranked set sample from a finite population P^N. Under a super population model in equation (2.1), the mean square prediction error of the estimator $\overline{Y}_R$ is given by*

$$\sigma^2_{\mathrm{RSS}} \;=\; \mathrm{MSPE}_M\left(\overline{Y}_R\right) = \frac{N-n}{Nn}\sigma^2 - \frac{1}{nH}\sum_{h=1}^{H}\left(\mu_{[h]} - \mu\right)^2. \tag{2.4}$$

We note that expression on equation (2.4) is very similar to the sample variance of an infinite population RSS sample. Only difference is due to the coefficient $\frac{N-n}{Nn}$. In infinite population setting the fraction $\frac{N-n}{Nn}$ in equation (2.4) becomes $\frac{1}{n}$. Hence, $\left(1 - \frac{n}{N}\right)$ is the finite population correction (fpc) factor for the variance of RSS sample mean. If the sample size is not small in comparison with the population size N, the fpc, $\frac{N-n}{Nn}$, makes a correction on the variance of an RSS sample mean. This correction would be substantial if n is relatively large with respect to N. If n is small, fpc is close to 1 and the impact of finite population correction factor is minimal.

Corollary 1: *Assume that* n *and* N *increase in such a way that the ratio* $\frac{n}{N}$ *approaches to a limit at* a, $\lim_{n\to\infty} \frac{n}{N} = a$.

(i) *If* $a > 0$, σ^2_{RSS} *converges to a simple form*

$$\lim_{n\to\infty} n\sigma^2_{RSS} = (1-a)\sigma^2 - \frac{1}{H}\sum_{h=1}^{H}\left(\mu_{[h]} - \mu\right)^2,$$

(ii) *if* $a = 0$, $\lim_{n\to\infty} n\sigma^2_{RSS} = \frac{1}{H}\sigma^2_{[h]}$, *which is the same as the variance of the sample mean of a balanced ranked set sample in an infinite population setting,*

(iii) *if* a *is strictly positive, then* $\lim_{n\to\infty} n\sigma^2_{RSS} < \frac{1}{H}\sigma^2_{[h]}$.

The corollary indicates that when sample and population sizes grow at a certain rate, variance of sample mean of an RSS (σ^2_{RSS}) sample in a finite population setting reduces to simple form. If a is strictly positive, variance of an RSS sample mean is smaller than the variance of an RSS sample mean in an infinite population setting.

3 Unbiased estimators

In this section, we construct an unbiased estimator for σ^2_{RSS}. By rewriting the estimator for σ^2_{RSS} in a slightly different form, we obtain

$$\begin{aligned}
\sigma^2_{\text{RSS}} &= \left(\frac{N-n}{Nn}\right)\sigma^2 - \frac{1}{nH}\sum_{h=1}^{H}\left(\mu_{[h]} - \mu\right)^2 \\
&= \left(\frac{1}{n} - \frac{1}{N}\right)\sigma^2 - \frac{1}{nH}\left(H\sigma^2 - \sum_{h=1}^{H}\sigma^2_{[h]}\right) \\
&= \left(\frac{-1}{N}\right)\sigma^2 + \frac{1}{nH}\sum_{h=1}^{H}\sigma^2_{[h]}.
\end{aligned}$$

Let

$$\begin{aligned}
T_1^* &= \frac{1}{2d^2H^2}\sum_{h=1}^{H}\sum_{h\neq h'}^{H}\sum_{i=1}^{d}\sum_{j=1}^{d}\left(Y_{[h]i} - Y_{[h']j}\right)^2 \\
T_2^* &= \frac{1}{2d(d-1)H^2}\sum_{h=1}^{H}\sum_{i=1}^{d}\sum_{j\neq i}^{d}\left(Y_{[h]i} - Y_{[h]j}\right)^2.
\end{aligned}$$

Using these definitions, one can easily establish the following result.

Theorem 2: *Let* $Y_{[h]i}$, $i = 1, \ldots, n, h = 1, \ldots, H$ *be an RSS sample of set size* H *from a finite population. An unbiased estimator of* σ^2_{RSS} *is given by*

$$\hat{\sigma}^2_{RSS} = T_2^*\left(\frac{H}{n}\right) - \left(T_1^* + T_2^*\right)\frac{1}{N}. \tag{3.1}$$

Theorem 2 indicates that the variance estimator is unbiased for any sample and set sizes regardless of the quality of ranking information. Unbiased estimator of the variance of $\overline{Y}_R$ allows us to construct confidence interval for population mean and total. Using normal approximation, $(1-\alpha)100\%$ confidence interval for the population mean is given by

$$\overline{Y}_R \pm t_{n-H,\alpha/2}\hat{\sigma}^2_{\mathrm{RSS}},$$

where $t_{df,a}$ is the a^{th} upper quantile of $t-$distribution with degrees of freedom df. The degrees of freedom $n-H$ is suggested to account the heterogeneity among H judgment classes. The choice of $df = n-H$ is also suggested in Ahn, Lim and Wang (2014) in infinite population setting.

4 Cost model

Efficiency improvement of the RSS estimator results from the relative position (rank) information of the measured observation among unmeasured $H-1$ units in a set. This extra information comes at the cost of sampling a set of size H and obtaining the subsequent ranking. Ranking can be performed either using concomitant (auxiliary) variable or visual inspection of the physical units in each set. Hence, these two approaches, visual and concomitant ranking models, may lead to different cost structures. In either case, there needs to be some sort of consistency in ranking process to develop a meaningful cost function. Patil, Sinha and Taillie (1997) defined a coherent ranking process in which ranking of a set is consistent for all subsets and supersets. Under a coherent ranking scheme, the rank order of H units would remain identical when ranking any of their subsets or supersets containing them. For further detail in coherent ranking, readers are referred to Patil et al. (1997) or Nahhas, Wolfe and Chen (2002).

Concomitant ranking uses an auxiliary variable to rank H units in a set. The quality of ranking depends on monotonic (not necessarily to be linear) relationship between the variable of interest and auxiliary variable. On the other hand, visual inspection can be performed in different ways. One of the strategy is to use pairwise comparison. Under coherent ranking, not all $\binom{H}{2}$ pairwise comparisons are necessary for a visual ranking. For example, in a set of size $H = 3,$ if unit 1 is judged to be smaller than unit 2 and unit 2 is smaller than unit 3, we reasonably assume unit 1 is less than unit 3 without a comparison. In order to differentiate the impact of the cost structures of the concomitant and visual ranking schemes, we denote the estimator in equation (2.3) with $\overline{Y}_{\mathrm{RC}}$ for concomitant ranking and $\overline{Y}_{\mathrm{RV}}$ for visual ranking.

For visual ranking, we use visual inspection model of Nahhas et al. (2002). This model always compares the selected unit with the largest element previously ranked. It chooses a unit at random and compares it with the unit previously judged to be largest. If it is judged to be larger, then it becomes the largest among all judged units. Otherwise, it is compared with the next largest previously judged unit until it is assigned a rank. The number of required pairwise comparisons under this ranking strategy with a coherent ranking scheme is an integer valued random variable having the support $H-1, H, H+1, \ldots, \binom{H}{2}.$ The expected number of pairwise comparison for this ranking scheme is approximately equal to $f(H) = (H+2)(H-1)/4.$ The reader is referred to Nahhas et al. (2002) for further development on expected number of pairwise comparisons.

We now introduce cost definitions for three models; concomitant, visual ranking and simple random sampling models: C_C = total cost for concomitant ranking, C_V = total cost for visual ranking, C_S = total cost for simple random sampling, c_i = cost of sampling a single unit, c_{qy} = cost of quantification of the variable of interest (Y) for one unit, c_{qx} = cost of quantification of concomitant (auxiliary X) variable for one unit, c_r = cost of one pairwise comparison. We assume that overhead cost in SRS model to be zero, but the overhead cost (in excess of the overhead cost of SRS) of RSS concomitant (visual) ranking model is absorbed in $c_{qx}(c_r)$. Total cost for these three models are then given by

$$C_S = n_s\,(c_i + c_{qy}),\ C_C = n_c\,(Hc_i + Hc_{qx} + c_{qy}),\ C_V = n_V(Hc_i + f(H) + c_{qy}),$$

where n_S, n_C and n_V are the total (measured) observations in SRS, RSS concomitant and RSS visual ranking models. Readers are referred to Nahhas et al. (2002) for further details on these cost functions.

We now fix the total cost on these three models $C_S = C_C = C_V = C$. Under this fixed cost, we look at the relative efficiency of $\bar{Y}_{\mathrm{RC}}$ and $\bar{Y}_{\mathrm{RV}}$ with respect to SRS sample mean $\bar{Y}_{\mathrm{SRS}}$. Let

$$\mathrm{RP} = \frac{1}{1-D}, \quad D = 1 - \frac{1}{H\sigma^2}\sum_{h=1}^{H}\left(\mu_{[h]} - \mu\right)^2,$$

where RP is the relative precision of RSS sample mean with respect to SRS sample mean in an infinite population setting. Under super population model, we can establish the following efficiency result.

Theorem 3: *Let $Y_{[h]i}$, $h = 1, \ldots, H$, $i = 1, \ldots, d$, be a ranked set sample from a finite population P^N. For a fixed cost, under super population model and coherent ranking scheme, the following efficiency results are established.*

$$\mathrm{RE}\left(\bar{Y}_{\mathrm{RC}}, \bar{Y}_{\mathrm{SRS}}\right) = \frac{\mathrm{Var}\left(\bar{Y}_{\mathrm{SRS}}\right)}{\mathrm{Var}\left(\bar{Y}_{\mathrm{RC}}\right)} \geq 1, \quad \text{if } \mathrm{RP} \geq \frac{Hc_i + Hc_{qx} + c_{qy}}{c_i + c_{qy}}$$

$$\mathrm{RE}\left(\bar{Y}_{\mathrm{RV}}, \bar{Y}_{\mathrm{SRS}}\right) = \frac{\mathrm{Var}\left(\bar{Y}_{\mathrm{SRS}}\right)}{\mathrm{Var}\left(\bar{Y}_{\mathrm{RV}}\right)} \geq 1, \quad \text{if } \mathrm{RP} \geq \frac{Hc_i + f(H)\,c_r + c_{qy}}{c_i + c_{qy}}.$$

The fractions on the right hand side of the inequalities in the above theorem is the ratio of the cost of selecting and measuring a single unit in RSS and SRS, respectively. If the cost of sampling a unit and cost of ranking a set are negligible (free), the cost ratio becomes 1. One of the basic assumptions, in settings where RSS is used, is that ranking cost of units is relatively cheap with respect to the cost of measurement. Hence, it is not unreasonable to assume that cost ratio will be very close to 1 for settings where use of RSS is appropriate. It is established in the literature that RP is always greater than or equal to 1 (see Dell and Clutter (1972), Patil et al. (1997), Nahhas et al. (2002)). It is equal to 1 only under random ranking. The values of RP for normal population for different values of ρ (correlation coefficient between response Y and auxiliary variable X) and set sizes are given in Table 4.1. It is now reasonable to say that RSS estimator under super population model is more efficient if the cost of sampling and ranking a unit is relatively cheap in comparison with measurement cost.

Table 4.1
Relative precision (RP) of RSS sample mean with respect to SRS sample mean under infinite population setting for normal distribution $N(0,1)$. ρ is the correlation coefficient between response and auxiliary variable, and H is the set size

ρ	$H=2$	$H=3$	$H=4$	$H=5$	$H=6$
1.00	1.467	1.914	2.347	2.770	3.186
0.90	1.347	1.631	1.869	2.073	2.251
0.75	1.218	1.367	1.477	1.561	1.628
0.50	1.086	1.136	1.168	1.190	1.207

5 Empirical results

In this section, we conduct a simulation study to check the finite sample properties of the estimator for different values of simulation parameters. Data sets are generated from normal $(\mu = 10, \sigma = 4)$ and log normal $(\mu = 0, \sigma = 1)$ super populations. We consider two different finite populations with population sizes $N = 150$ and $N = 1{,}000$ to see the impact of population sizes on the estimators. Sample and set size combinations (n, H) are selected to be (10, 2), (15, 3), (20, 4), (25, 5). The quality of ranking information is modeled through a perceptual error model in Dell and Clutter (1972). The Dell and Clutter model considers two variables, the variable of interest Y and a correlated ranking variable X. The ranking variable is modeled through an additive model $X = Y + \epsilon$, where ϵ is a random noise generated independently with respect to Y. To implement the perceptual error model, we generate a set (size H) of simple random sample, $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_H)$, from the true population of interest with mean μ and variance σ^2. Another set (size H) of random numbers are generated from a normal distribution with mean zero and variance, σ_ϵ^2, $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \ldots, \epsilon_H)$. The perceptual error model is then defined by $X_i = Y_i + \epsilon_i$ $i = 1, 2, \ldots, H$. The random numbers (X_i, Y_i) are ranked with respect to the first components $(X_{(i)})$ and the second components are taken to be the judgment ranked order statistics $(Y_{[i]})$. The quality of the ranking information is controlled by the correlation coefficient between Y and X, $\rho = \mathrm{corr}(Y, X) = \left(\frac{\sigma^2}{\sigma^2 + \sigma_\epsilon^2}\right)^{1/2}$. Since the units are ranked based on concomitant variable X, the ranking model is equivalent to concomitant ranking in Section 3. In the simulation study, we used $\rho = 1$ for perfect ranking and $\rho = 0.75, 0.50$ for imperfect ranking.

In each replication of the simulation, a finite population of size P^N is generated from the normal super population with specified mean μ and standard deviation σ, $P^N = \{y_1, \ldots, y_N\}$. A ranked set sample is then constructed from this finite population, a realization from normal super population, with specified set and cycle sizes. The quality of ranking information in each RSS sample is controlled generating random noise vector $\boldsymbol{\epsilon}$ with specified ρ (or equivalently σ_ϵ) in the perceptual error model. The simulation size is taken to be 50,000.

Simulation results are presented in Tables 5.1, 5.2, 5.3 and 5.4. There are several features that need to be discussed in these tables. For different ρ and sample size combinations (n, H), the relative efficiencies of the RSS estimator with respect to the SRS estimator are given by

$$\mathrm{RE}_{\mathrm{RC}} = \frac{V(\bar{Y}_{\mathrm{SRS}})}{V(\bar{Y}_{\mathrm{RC}})} \tag{5.1}$$

where $V(\bar{Y}_{\text{SRS}})$ and $V(\bar{Y}_{\text{RC}})$ are the MSPE of SRS and RSS sample means from the simulation study under a super population model in equation (2.1), respectively. In equation (5.1), the relative efficiency values (RE_{RC}) greater than one indicate that the RSS estimator is more efficient than the SRS estimator. In all these tables, the RSS sample mean estimator performs better than the SRS estimator. Its efficiency is an increasing function of set size H and correlation coefficient ρ as expected. Under the concomitant cost model, if the cost ratio of obtaining a unit in RSS and a unit in SRS is less than the RP values in Table 4.1, the RSS sample mean has higher efficiency than the SRS sample mean.

The impact of the finite population size N can be observed by comparing the efficiency results in Tables 5.1 and 5.2 for the normal super population and Tables 5.3 and 5.4 for the lognormal super population. When $\rho > 0.50$, relative efficiencies (RE_{RC}) are higher in Table 5.1 $(N = 150)$ than Table 5.2 $(N = 1{,}000)$. In Table 5.1, finite population correction factor is smaller than the finite population correction factor in Table 5.2. Hence, the reduction in MSPE is smaller in RSS estimator. Similar effect is also observed in Tables 5.3 and 5.4.

The simulation study also investigated the properties of the MSPE estimator of RSS sample mean estimator. Theoretical value of the MSPE estimator is given under the heading σ^2_{RSS} when $\rho = 1.0$. The simulated (unbiased) MSPE estimate is given in columns 5 (6) in Tables 5.1-5.4. It is very clear that simulated and unbiased MSPE estimates are almost identical when $\rho \neq 1$ as expected. Under perfect ranking $(\rho = 1)$ theoretical MSPE values, and the simulated and unbiased MSPE estimates are all close to each other within the simulation variation.

The coverage probabilities of the confidence intervals are given under the heading $C(\bar{Y}_{\text{RC}})$ in column 7 in Tables 5.1-5.4. In Tables 5.1 and 5.2, the coverage probabilities of the confidence intervals based on $t-$ approximation are reasonably close to the nominal coverage probability 0.950. On the other hand, the coverage probabilities in Tables 5.3 and 5.4 are smaller than the nominal coverage probability 0.95 for lognormal super population. The coverage probabilities are getting closer to nominal values when the sample size increases. This indicates that for skewed populations, sample sizes should be large enough to have a reasonable coverage probability for the confidence intervals.

Table 5.1
MSPE estimate and relative efficiency of RSS sample estimator, and coverage probability of a 95% confidence interval of population mean. Data sets are generated from a normal super population with $\mu = 10$, $\sigma = 4$ and population size $N = 150$

		Est. from equations		Est. from simu.	UE estimates	Coverage prb.	Relative eff.
H	ρ	σ^2_{RSS}	σ^2_{SRS}	$V(\bar{Y}_{\text{RC}})$	$\hat{\sigma}^2_{\text{RSS}}$	$C(\bar{Y}_{\text{RC}})$	RE_{RC}
2.0	0.50	-	1.493	1.355	1.365	0.949	1.102
3.0	0.50	-	0.960	0.840	0.833	0.947	1.143
4.0	0.50	-	0.693	0.572	0.578	0.948	1.213
5.0	0.50	-	0.533	0.435	0.432	0.948	1.226
2.0	0.75	-	1.493	1.195	1.205	0.949	1.250
3.0	0.75	-	0.960	0.675	0.674	0.947	1.423
4.0	0.75	-	0.693	0.433	0.436	0.946	1.600
5.0	0.75	-	0.533	0.302	0.304	0.945	1.768
2.0	1.00	0.984	1.493	0.974	0.984	0.948	1.534
3.0	1.00	0.451	0.960	0.455	0.451	0.940	2.111
4.0	1.00	0.234	0.693	0.233	0.235	0.936	2.971
5.0	1.00	0.124	0.533	0.125	0.126	0.922	4.273

Table 5.2
MSPE estimate and relative efficiency of RSS sample estimator, and coverage probability of a 95% confidence interval of population mean. Data sets are generated from a normal super population with $\mu = 10$, $\sigma = 4$ and population size $N = 1{,}000$

		Est. from equations		Est. from simu.	UE estimate	Coverage prb.	Relative eff.
H	ρ	σ^2_{RSS}	σ^2_{SRS}	$V(\bar{Y}_{\text{RC}})$	$\hat{\sigma}^2_{\text{RSS}}$	$C(\bar{Y}_{\text{RC}})$	RE_{RC}
2.0	0.50	-	1.584	1.461	1.455	0.950	1.084
3.0	0.50	-	1.051	0.931	0.924	0.949	1.129
4.0	0.50	-	0.784	0.665	0.670	0.950	1.180
5.0	0.50	-	0.624	0.524	0.522	0.950	1.191
2.0	0.75	-	1.584	1.304	1.295	0.949	1.215
3.0	0.75	-	1.051	0.770	0.765	0.948	1.365
4.0	0.75	-	0.784	0.525	0.526	0.951	1.494
5.0	0.75	-	0.624	0.392	0.395	0.951	1.590
2.0	1.00	1.075	1.584	1.075	1.076	0.950	1.473
3.0	1.00	0.541	1.051	0.538	0.541	0.951	1.954
4.0	1.00	0.325	0.784	0.327	0.325	0.949	2.398
5.0	1.00	0.215	0.624	0.217	0.215	0.948	2.877

Table 5.3
MSPE estimate and relative efficiency of RSS sample estimator, and coverage probability of a 95% confidence interval of population mean. Data sets are generated from a log-normal super population with $\mu = 0$, $\sigma = 1$ and population size $N = 150$

		Est. from equations		Est. from simu.	UE estimate	Coverage prb.	Relative eff.
H	ρ	σ^2_{RSS}	σ^2_{SRS}	$V(\bar{Y}_{\text{RC}})$	$\hat{\sigma}^2_{\text{RSS}}$	$C(\bar{Y}_{\text{RC}})$	RE_{RC}
2.0	0.50	-	0.436	0.400	0.400	0.852	1.089
3.0	0.50	-	0.280	0.243	0.242	0.869	1.153
4.0	0.50	-	0.202	0.160	0.162	0.883	1.262
5.0	0.50	-	0.156	0.117	0.116	0.886	1.336
2.0	0.75	-	0.436	0.371	0.372	0.855	1.176
3.0	0.75	-	0.280	0.216	0.217	0.867	1.300
4.0	0.75	-	0.202	0.146	0.146	0.874	1.388
5.0	0.75	-	0.156	0.103	0.103	0.878	1.514
2.0	1.00	0.362	0.436	0.361	0.364	0.839	1.207
3.0	1.00	0.201	0.280	0.197	0.198	0.849	1.423
4.0	1.00	0.128	0.202	0.128	0.127	0.847	1.586
5.0	1.00	0.086	0.156	0.085	0.085	0.845	1.833

Table 5.4
MSPE estimate and relative efficiency of RSS sample estimator, and coverage probability of a 95% confidence interval of population mean. Data sets are generated from a log-normal super population with $\mu = 0$, $\sigma = 1$ and population size $N = 1{,}000$

		Est. from equations		Est. from simu.	UE estimate	Coverage prb.	Relative eff.
H	ρ	σ^2_{RSS}	σ^2_{SRS}	$V(\bar{Y}_{\text{RC}})$	$\hat{\sigma}^2_{\text{RSS}}$	$C(\bar{Y}_{\text{RC}})$	RE_{RC}
2.0	0.50	-	0.462	0.432	0.433	0.851	1.070
3.0	0.50	-	0.307	0.263	0.263	0.868	1.164
4.0	0.50	-	0.229	0.189	0.190	0.882	1.208
5.0	0.50	-	0.182	0.141	0.141	0.889	1.296
2.0	0.75	-	0.462	0.413	0.413	0.852	1.119
3.0	0.75	-	0.307	0.240	0.238	0.868	1.276
4.0	0.75	-	0.229	0.171	0.170	0.878	1.337
5.0	0.75	-	0.182	0.129	0.129	0.884	1.415
2.0	1.00	0.389	0.462	0.387	0.386	0.839	1.195
3.0	1.00	0.228	0.307	0.225	0.227	0.852	1.364
4.0	1.00	0.154	0.229	0.155	0.155	0.857	1.479
5.0	1.00	0.113	0.182	0.113	0.113	0.862	1.614

6 Example

In this section we apply the proposed estimators to a data set which contains a sheep population in a research farm at Ataturk University, Erzurum, Turkey. Data set contains birth weights, mothers' weights at mating and the weights at the 7th month after birth for 224 lambs. The entire data set is given in Hollander, Wolfe and Chicken (2014, page 709). Variable of interest is the weights (Y) at the 7th month after birth for 224 lambs. We use birth weights (X_1) and mothers' weights (X_2) at mating as auxiliary variables to perform ranking process. The ranking variables are positively correlated with the variable of interest Y. The correlation coefficient $(\rho = \text{corr}(X, Y))$ between X_1, Y and X_2, Y are 0.8425 and 0.5941, respectively. The histogram of the variable of interest, Y, is roughly symmetric. Mean and variance of Y are $\bar{Y}_N = 28.125$kg and $S_N^2 = 15.23$kg^2, respectively, where $S_n^2 = \sum_{i=1}^{224} (Y_i - \bar{Y}_N)^2 / 223$. We treated these 224 lambs as a realization from a super population having finite mean μ and variance σ^2. We constructed samples based RSS sampling design using this finite population. Samples are generated for sample and set size combinations, (n, H), (10, 2), (15, 3), (20, 4), (25, 5). Simulation size is taken to be 50,000.

In this example, we incorporate the sampling cost to RSS and SRS sampling designs with concomitant ranking in RSS. We first need to determine reasonable costs associated with various aspects of RSS. Weight measurement is obtained from seven-month-old lambs. These animals are very active and measurement cost is substantial. The measurement process usually require three people for separating the lamb from the flock, bringing it to scale, holding it firm during the measurement. Suppose that the farm employs the workers in an annual salary of \$50,000. This corresponds to a rate of approximatley \$25 per hour per person. Assume that the measurement of a lamb takes about 5 minutes. The measurement cost for a lamb then would be about $c_{qy} = 3(25/12) \approx 6$ for three workers. Ranking will be performed using auxilairy variables X_1 and X_2. These variables are maintained in the data base for some other purposes. Only cost to sampling would be due to personal cost for ranking. Ranking will be performed in the office by selecting sets at random from the data base and ranking them based on auxiliary variables. Suppose that ranking a set of size H takes about 1/2 minute. This leads to ranking cost of $Hc_{qx} = \$0.21$. We may assume that cost related to identification of a unit in the population is negligible $(c_i = 0)$. Under these stipulations, the cost ratio of selecting and measuring a unit in RSS and SRS is given by $\text{ratio} = (Hc_i + Hc_{qx} + c_{qy}) / (c_i + c_{qy}) = (6 + 0.21)/6 = 1.035$. Since this ratio is less than all entries in Table 4.1, we anticipate that $\bar{Y}_{\text{RC}}$ provides higher efficiency than $\bar{Y}_{\text{SRS}}$.

Table 6.1 presents the estimated MSPE and relative efficiency of RSS esimator as well as the coverage probability of the confidence interval of μ for different ρ and sample size combinations. It is clear that the RSS estimator outperform the SRS estimator for all simulation parameter combinations. Estimated MSPEs and coverage probabilities also show similar behaviors as in Section 3. The estimated MSPE values are very close to the simulated MSPE values. The coverage probabilities of the confidence intervals based on $t-$ approximation appear to be very close to the nominal coverage probabiliy, 0.950.

Table 6.1
MSPE estimate and relative efficiency of RSS sample estimator, and coverage probability of a 95% confidence interval of population mean of a sheep population of size $N = 224$

		Est. from equation	Est. from simu.	UE estimate	Coverage prb.	Relative eff.
H	ρ	σ^2_{SRS}	$V(\bar{Y}_{\mathrm{RC}})$	$\hat{\sigma}^2_{\mathrm{RSS}}$	$C(\bar{Y}_{\mathrm{RC}})$	$\mathrm{RE}_{\mathrm{RC}}$
2.0	0.59	1.453	1.279	1.275	0.946	1.136
3.0	0.59	0.946	0.776	0.774	0.948	1.219
4.0	0.59	0.693	0.536	0.537	0.948	1.293
5.0	0.59	0.540	0.399	0.402	0.948	1.353
2.0	0.84	1.453	1.107	1.105	0.945	1.312
3.0	0.84	0.946	0.600	0.602	0.946	1.576
4.0	0.84	0.693	0.377	0.382	0.946	1.839
5.0	0.84	0.540	0.263	0.264	0.944	2.056

7 Concluding remarks

We have developed a model based statistical inference for population mean and total based on RSS samples in a finite population setting where samples are constructed by using a without replacement sampling design. It is shown that the sample mean of RSS samples are model unbiased and they have smaller mean square prediction error (MSPE) than the MSPE of a simple random sample mean. We constructed unbiased estimator for the MSPE and prediction confidence interval for the population mean. A small scale simulation study showed that estimators are as good as or better than SRS estimators when the quality of ranking information in RSS sampling is low or high, respectively, and the cost ratio of obtaining a unit in RSS and a unit in SRS is not too high. The coverage probabilities of the prediction intervals are also very close to the nominal coverage probabilities. Proposed sampling designs and inferential procedures are applied to a data set containing a sheep population in an agricultural research farm.

Acknowledgements

The second author’s research is supported by The Scientific and Technological Research Council of Turkey (TUBITAK). The authors thank the Editor, Associate Editor and two referees for their constructive comments and suggestions.

Appendix

Proof of Theorem 1: We write mean square prediction error (MSPE) as

$$\mathrm{MSPE}_M\left(\bar{Y}_R\right) = E_M\left\{\bar{Y}_R - \frac{1}{N}\sum_{i=1}^{N} Y_i\right\}^2 = E_M\left\{\frac{1}{dH}\sum_{h=1}^{H}\sum_{i=1}^{d} Y_{[h]i} - \frac{1}{N}\sum_{i=1}^{N} Y_i\right\}^2 .$$

Let $Z_i, i = 1, \ldots, N - nH,$ be the responses on $N - nH$ population units that are neither measured nor used in ranking in any one of the randomly selected sets of size H in the construction of the RSS sample. Then the MSPE can be written

$$\text{MSEP}_M\left(\bar{Y}_R\right) = E_M\left\{\frac{1}{dH}\sum_{h=1}^{H}\sum_{i=1}^{d} Y_{[h]i} - \frac{1}{N}\sum_{h=1}^{H}\sum_{i=1}^{d}\left[Y_{[h]i} + \sum_{h \neq h'}^{H} Y^*_{[h']i}\right] - \frac{1}{N}\sum_{i=1}^{N-nH} Z_i\right\}^2$$

where $Y^*_{[h']i}, h' \neq, h,$ are responses on unmeasured units that are used in ranking of units in a set. Hence, $Y^*_{[h']i}$ and $Y_{[h]i}$ are correlated, but they are uncorrelated with Z_i. Let

$$c_{h,h'} = \begin{cases} \dfrac{N-n}{n} & h = h' \\ -1 & h \neq h'. \end{cases}$$

Using the definition of $c_{h,h'},$ we combine $Y^*_{[h']i}$ and $Y_{[h]i}$ under the same summation and write the MSPE as

$$\begin{aligned}
\text{MSPE}_M\left(\bar{Y}_R\right) &= \frac{1}{N^2}\text{var}\left\{\sum_{h=1}^{H}\sum_{i=1}^{d}\sum_{h'=1}^{H} c_{h,h'}Y_{[h]i}\right\} + \text{var}\left\{\frac{1}{N}\sum_{i=1}^{N-nH} Z_i\right\} \\
&= \frac{1}{N^2}\sum_{h=1}^{H}\text{var}\left[\sum_{i=1}^{d}\sum_{h'=1}^{H} c_{h,h'}Y_{[h]i}\right] + \text{var}\left\{\frac{1}{N}\sum_{i=1}^{N-nH} Z_i\right\} \\
&= \frac{d}{N^2}\sum_{h=1}^{H}\sum_{h'=1}^{H}\left(c_{h,h'}\right)^2\sigma^2_{[h']} + \frac{d}{N^2}\sum_{h=1}^{H}\left(\sum_{h'=1}^{H}\sum_{t \neq h'}^{H} c_{h,h'}c_{h,t}\sigma_{[h',t]}\right) \\
&\quad + \text{var}\left\{\frac{1}{N}\sum_{i=1}^{N-nH} Z_i\right\} \\
&= A + B + \frac{(N - nH)\sigma^2}{N^2}.
\end{aligned} \tag{A.1}$$

The expression A reduces to

$$\begin{aligned}
A &= \frac{d}{N^2}\sum_{h=1}^{H}\left(c_{h,h}\right)^2\sigma^2_{[h]} + \frac{d}{N^2}\sum_{h}\sum_{h' \neq h}^{H}\left(c_{h,h'}\right)^2\sigma_{[h,h']} \\
&= \frac{d}{N^2}\left[\left(\frac{N-n}{n}\right)^2 + (H-1)\right]\sum_{h=1}^{H}\sigma^2_{[h]}.
\end{aligned}$$

In a similar fashion, the expression B reduces to

$$\begin{aligned}
B &= \frac{d}{N^2}\sum_{h=1}^{H}\left[\sum_{t \neq h',h}^{H}\left(\sum_{h'=1}^{H} c_{h,h'}c_{h,t}\sigma_{[h',t]} + c_{h,h}c_{h,t}\sigma_{[h,t]}\right) + \sum_{h' \neq h}^{H} c_{h,h'}c_{h,h}\sigma_{[h',h]}\right] \\
&= \frac{d}{N^2}\sum_{h=1}^{H}\left[\sum_{h' \neq h}^{H}\sum_{t \neq h',h}^{H}\sigma_{[h',t]} - 2\left(\frac{N-n}{n}\right)\left(\sum_{t=1}^{H}\sigma_{[h,t]} - \sigma_{[h,h]}\right)\right] \\
&= \frac{d}{N^2}\left[\left(H^2 - 2H\right)\sigma^2 - (H-2)\sum_{h=1}^{H}\sigma^2_{[h]} - 2\left(\frac{N-n}{n}\right)\left(H\sigma^2 - \sum_{h=1}^{H}\sigma^2_{[h]}\right)\right].
\end{aligned}$$

By inserting expressions A and B in equation (A.1), we conclude that

$$\begin{aligned}
\operatorname{MSPE}_M\left(\bar{Y}_R\right) &= \frac{d}{N^2}\left[\left(\frac{N-n}{n}\right)^2+(H-1)\right]\sum_{h=1}^{H}\sigma_{[h]}^2 \\
&+\frac{d}{N^2}\left[(H^2-2H)\sigma^2-(H-2)\sum_{h=1}^{H}\sigma_{[h]}^2-2\left(\frac{N-n}{n}\right)\left(H\sigma^2-\sum_{h=1}^{H}\sigma_{[h]}^2\right)\right] \\
&+\left(\frac{N-nH}{N^2}\right)\sigma^2 \\
&=\left(\frac{N-n}{Nn}\right)\sigma^2-\frac{1}{nH}\sum_{h=1}^{H}\left(\mu_{[h]}-\mu\right)^2
\end{aligned}$$

which completes the proof. Note that to establish the last equality we used the fact that $\sigma^2 = \sum_{h=1}^{H}\sigma_{[h]}^2 \big/ H + \sum_{h=1}^{H}\left(\mu_{[h]}-\mu\right)^2 \big/ H.$

Proof of Theorem 2: We first look at the expected values of T_1^* and T_2^* under the super population model in equation (2.1)

$$\begin{aligned}
E\left(T_1^*\right) &= \frac{1}{H}\sum_{h=1}^{H}\left(\mu_{[h]}-\mu\right)^2+\frac{H-1}{H^2}\sum_{h=1}^{H}\sigma_{[h]}^2=\sigma^2-\frac{1}{H^2}\sum_{h=1}^{H}\sigma_{[h]}^2 \\
E\left(T_2^*\right) &= \frac{1}{H^2}\sum_{h=1}^{H}\sigma_{[h]}^2.
\end{aligned}$$

It is now easy to establish that $E\left(T_1^*+T_2^*\right)=\sigma^2$. The proof is then completed by inserting these expressions in equation (3.1).

Proof of Theorem 3: We sketch the proof for $\operatorname{RE}\left(\bar{Y}_{\mathrm{RC}}, \bar{Y}_{\mathrm{SRS}}\right)$. From the total cost function, we write

$$n_S=\frac{C}{c_i+c_{qy}} \quad \text{and} \quad n_R=\frac{C}{Hc_i+Hc_{qx}+c_{qy}},$$

where C is the fixed total cost. Using these expressions, we have

$$\begin{aligned}
\operatorname{RE}\left(\bar{Y}_{\mathrm{RC}}, \mathrm{SRS}\right) &= \frac{n_R\left(N-n_S\right)}{n_S\left(N-n_r-ND\right)} \\
&= \frac{N\left(c_i+c_{qy}\right)-C}{N\left(Hc_i+Hc_{qx}+c_{qy}\right)-C-ND\left(Hc_i+Hc_{qx}+c_{qy}\right)}.
\end{aligned}$$

We now establish that $\operatorname{RE}\left(\bar{Y}_{\mathrm{RC}}, \mathrm{SRS}\right) \geq 1$ if and only if

$$\begin{aligned}
c_i+c_{qy} &\geq \left(Hc_i+Hc_{qx}+c_{qy}\right)(1-D)=\frac{Hc_i+Hc_{qx}+c_{qy}}{\mathrm{RP}} \\
\mathrm{RP} &\geq \frac{Hc_i+Hc_{qx}+c_{qy}}{c_i+c_{qy}}
\end{aligned}$$

which completes the proof.

References

Ahn, S., Lim, J. and Wang, X. (2014). The students's t approximation to distributions of pivotal statistics from ranked set samples. *Journal of Korean Statistical Society*, 43, 4, 643-652.

Al-Saleh, M.F., and Samawi, H.M. (2007). A note on inclusion probability in ranked set sampling and some of its variations. *Test*, 16, 1, 198-209.

Dell, T.R., and Clutter, J.L. (1972). Ranked-set sampling theory with order statistics background. *Biometrics*, 28, 2, 545-555.

Deshpande, J.V., Frey, J. and Ozturk, O. (2006). Nonparametric ranked set-sampling confidence intervals for a finite population. *Environmental and Ecological Statistics*, 13, 1, 25-40.

Frey, J. (2011). A note on ranked-set sampling using a covariate. *Journal of Statistical Planning and Inference*, 141, 2, 809-816.

Gokpinar, F., and Ozdemir, Y.A. (2010). Generalization of inclusion probabilities in ranked set sampling. *Hacettepe Journal of Mathematics and Statistics*, 39, 1, 89-95.

Hollander, M., Wolfe, D.A. and Chicken, E. (2014). *Nonparametric Statistical Methods, 3rd Edition*. Wiley, 709.

Nahhas, R.W., Wolfe, D.A. and Chen, H. (2002). Ranked set sampling: Cost and optimal set size. *Biometrics*, 58, 964-971.

Ozdemir, Y.A., and Gokpinar, F. (2007). A generalized formula for inclusion probabilities in ranked set sampling. *Hacettepe Journal of Mathematics and Statistics*, 36, 1, 89-99.

Ozdemir, Y.A., and Gokpinar, F. (2008). A new formula for inclusion probabilities in median ranked set sampling. *Communications in Statistics - Theory and Methods*, 37, 13, 2022-2033.

Ozturk, O. (2014). Estimation of population mean and total in finite population setting using multiple auxilary variables. *Journal of Agricultural, Biological and Environmental Statistics*, 19, 2, 161-184.

Ozturk, O. (2015). Distribution free two-sample methods for judgment-post stratified data. *Statistica Sinica*, 25, 1691-1712.

Ozturk, O. (2016a). Estimation of finite population mean and total using population ranks of sample units. *Journal of Agricultural, Biological and Environmental Statistics*, 21, 1, 181-202.

Ozturk, O. (2016b). Statistical inference based on judgment post-stratifed samples in finite population. *Survey Methodology*, 42, 2, 239-262. Paper available at https://www150.statcan.gc.ca/n1/pub/12-001-x/2016002/article/14664-eng.pdf.

Ozturk, O., and Jozani, M.J. (2013). Inclusion probabilities in partially rank ordered set sampling. *Computational Statistics & Data Analysis*, 69, 122-132.

Patil, G.P., Sinha, A.K. and Taillie, C. (1995). Finite population corrections for ranked set sampling. *Annals of the Institute of Statistical Mathematics*, 47, 4, 621-636.

Patil, G.P., Sinha, A.K. and Taillie, C. (1997). Ranked set sampling, coherent rankings and size-biased permutations. *Journal of Statistical Planning and Inference*, 63, 2, 311-324.

Takahasi, K., and Futatsuya, M. (1998). Dependence between order statistics in samples from finite population and its application to ranked set sampling. *Annals of the Institute of Statistical Mathematics*, 50, 1, 49-70.