Hacettepe Journal of Mathematics and Statistics
Volume 36 (1) (2007), 89 – 99

A GENERALIZED FORMULA FOR INCLUSION PROBABILITIES IN RANKED SET SAMPLING

Yaprak Arzu Özdemir* and Fikri Gökpınar*

Received 09 : 03 : 2006 : Accepted 13 : 03 : 2007

Abstract

In probability sampling, the inclusion probability of any element in the population is the probability of the element which will be chosen in the sample. Al-Saleh and Samawi (*A note on inclusion probability in ranked set sampling and some of its variations*, Test., in press) introduced inclusion probabilities in ranked set sampling for sample sizes 2 and 3. In this paper we gave a generalized formula of these inclusion probabilities for any sample size. Also we compare these probabilities with simple random samplings for various given samples and population sizes.

Keywords: Ranked Set sampling, Simple random sampling, Inclusion probability, Finite population.

2000 AMS Classification: Primary 62 D 05. Secondary 65 C 60.

1. Introduction

The most basic probability sampling technique is Simple Random Sampling (SRS). In this technique, all elements of the population have an equal inclusion probability. If the population consists of N distinct elements and the sample size is n, then each element of the population has an inclusion probability of n/N. Let $\pi_N(k)$ be the probability that element u_k of a population of size N will be chosen in a sample of size n. Then $\sum_{k=1}^{N} \pi_N(k) = n$ for all probability sampling techniques.

In SRS, because of the equal inclusion probability, there is no control over which element enters the sample. On the other hand, Ranked Set Sampling (RSS) introduced by McIntre [3], is a more controlled sampling technique than SRS. RSS is a common sampling technique that has been used recently in some areas such as; environment,

*Gazi University, Faculty of Art and Science, Department of Statistics, 06500 Teknikokullar, Ankara, Turkey.
E-mail: (Y. A. Özdemir) yaprak@gazi.edu.tr (F. Gökpınar) fikri@gazi.edu.tr

ecology and agriculture. In these areas, the measurements of the units according to variables of interest can be quite difficult in some cases, in terms of cost, time and other factors. In such conditions, by using RSS, the sample selection process is done with less cost and time, than with the SRS technique.

RSS was first used by McIntyre [3] to estimate the mean a pasture yields. But this first study was not based on a mature mathematical theory. The mathematical theory of RSS was developed later by Takahasi and Wokimoto [8]. They showed that the sample mean obtained by RSS is an unbiased estimator for the population mean. Moreover, the variance of this estimator is smaller than the variance of the sample mean obtained from SRS with the same sample size. Dell and Clutter [2] indicated that the RSS estimator was also an unbiased estimator for the population mean under the presence of ranking error. Even if there are ranking errors, the RSS design is at least as efficient as SRS for the same sample size. Most of these studies are based on the assumption of an infinite population. In recent years, RSS has also been investigated under a finite population assumption. Takahasi and Futatsuya [6, 7] were the first to give a finite population theory in RSS. Patil *et al.* [4] generalized the results of Takahasi and Futatsuya [6, 7] for a larger set size. Ozturk *et al.* [5] demonstrated the practical use of RSS relative to SRS for the estimation of the population mean and variance in a finite population. In these studies, for obtaining the RSS sample, they used the same selection procedure as for a infinite population. This procedure may cause some problems for small population sizes. For this reason, Al-Saleh and Samawi [1] gave an adjusted selection procedure for RSS. In this study, we have adopted the same adjusted procedure as follows;

1) A SRS of size m is selected (without replacement) from the population and the minimum of the sample with respect to the characteristic of interest is identified by judgment. All other elements returned to the population.
2) A second SRS of size m is selected (without replacement) from the population and the second minimum of the sample with respect to the characteristic of interest is identified by judgement. All other elements are returned to the population.
3) In the i^{th} step, i^{th} minimum of the i^{th} chosen SRS is identified by judgment; $i = 1, 2, \ldots, m$.

The m identified elements make up a Ranked Set Sample of size m. The entire cycle may be repeated, if necessary, r times to produce a Ranked Set Sample of size $n = mr$ [1]. In this study we consider only the case $r = 1$. Our main interest is to find a generalized formula for the inclusion probabilities for any sample size in the RSS procedure.

In the second section, we generalize the formula for the inclusion probabilities and gave an illustration of this formula for $m = 3$ and $N = 5$. In the third section, we calculate the inclusion probabilities to compare the RSS and SRS designs under different sample and population size.

2. Inclusion probabilities of population elements in RSS

Knowing the inclusion probability of each element in the population is very important for sampling theory. Inclusion probabilities give an insight into how the RSS design has more control over which element enters the sample than SRS has. Also, we can determine the probability distribution of any statistics from the ranked set sample using these inclusion probabilities.

In this section, we give a generalized formula for calculating inclusion probabilities in RSS for any given sample size m and population size N.

Let $u_1 < u_2 < \ldots < u_N$ be the ordered population elements.

Also let the inclusion probability of any element u_k be defined as,

$$\pi_N(k) = \pi_N^{(1)}(k) + \pi_N^{(2)}(k) + \cdots + \pi_N^{(m)}(k), \tag{1}$$

where $\pi_N^{(j)}(k)$ is the inclusion probability of u_k in the j^{th} selection, $(j = 1, 2, \ldots m)$. Using conditional probabilities this can be written as follows :

$$\pi_N^{(j)}(k) = \sum P(A_j / B_{j-1}^{l_{j-1}} \cap B_{j-2}^{l_{j-2}} \cap \ldots \cap B_1^{l_1}) P(B_{j-1}^{l_{j-1}} / B_{j-2}^{l_{j-2}} \cap \ldots \cap B_1^{l_1}) \ldots P(B_1^{l_1}). \tag{2}$$

In (2) the summation is over all the 2^{j-1} possible permutations of the values l_j. Some necessary definitions for these inclusion probabilities are as follows:

A_j: the event of choosing u_k in the j^{th} selection

y_j: the element selected in the j^{th} selection

$$l_j = \begin{cases} 0 & y_j > u_k \\ 1 & y_j < u_k \end{cases}$$

$B_j^{l_j} = B_j^0$: the event $\{y_j > u_k\}$ when $l_j = 0$,

$B_j^{l_j} = B_j^1$; the event $\{y_j < u_k\}$ when $l_j = 1$,

and $a = j - (1 + l_1 + l_2 + \cdots + l_{j-1})$ denotes the number of elements greater than u_k selected in previous selections.

When $l_j = 0$, that is $y_j > u_k$ in the j^{th} selection, then all possible selection cases of y_j in the j^{th} selection where $y_j > u_k$ are as given in Table 1.

In Table 1, the number of elements smaller than u_k is $k-j+a$, the number of elements greater than u_k is $N-k-a$ and because of the distinctness of the population elements there is only one u_k. Thus, the population size in the j^{th} selection is $N-j+1$. We want to choose y_j to be greater than u_k. One of the possible cases is when all m elements are chosen from elements greater than u_k, another possible case is where $m-1$ elements are chosen from elements greater than u_k and the other is u_k, and so on. These are summarized in Table 1.

Table 1. Possible selection cases for $y_j > u_k$.

The number of elements smaller than u_k in the j^{th} selection	u_k	The number of elements greater than u_k in the j^{th} selection
$\mathbf{k-j+a}$	**1**	$\mathbf{N-k-a}$
-	-	m
-	1	$m-1$
1	-	$m-1$
$\vdots$	$\vdots$	$\vdots$
$j-2$	1	$m-(j-1)$
$j-1$	-	$m-(j-1)$

Using Table 1, the probability of $P(B_j^0 / B_{j-1}^{l_{j-1}} \cap B_{j-2}^{l_{j-2}} \ldots \cap B_1^{l_1})$ can be written as,

$$\begin{aligned} P(B_j^0 / B_{j-1}^{l_{j-1}} \cap B_{j-2}^{l_{j-2}} \cap \ldots \cap B_1^{l_1}) &= \frac{\binom{k-j+a}{0}\binom{1}{0}\binom{N-k-a}{m} + \binom{k-j+a}{0}\binom{1}{1}\binom{N-k-a}{m-1}}{\binom{N-j+1}{m}} \\ &+ \cdots + \frac{\binom{k-j+a}{j-2}\binom{1}{1}\binom{N-k-a}{m-(j-1)} + \binom{k-j+a}{j-1}\binom{1}{0}\binom{N-k-a}{m-(j-1)}}{\binom{N-j+1}{m}} \end{aligned} \tag{3}$$

Using Pascal's rule, this equation can be simplified as follows:

$$
\begin{aligned}
P(B_j^0/B_{j-1}^{l_{j-1}} \cap B_{j-2}^{l_{j-2}} \cap \ldots \cap B_1^{l_1}) &= \frac{\binom{N-k-a}{m} + \binom{k-j+a+1}{1}\binom{N-k-a}{m-1}}{\binom{N-j+1}{m}} \\
&\quad + \cdots + \frac{\binom{k-j+a+1}{j-1}\binom{N-k-a}{m-(j-1)}}{\binom{N-j+1}{m}} \\
&= \frac{\sum_{i=0}^{j-1} \binom{k-j+a+1}{i}\binom{N-k-a}{m-i}}{\binom{N-j+1}{m}}
\end{aligned} \tag{4}
$$

When $l_j = 1$, that is $y_j < u_k$ in the j^{th} selection, then all possible selection cases of y_j where $y_j < u_k$ are as given in Table 2.

Table 2. Possible selection cases for $y_j < u_k$.

The number of elements smaller than u_k in the j^{th} selection	u_k	The number of elements greater than u_k in the j^{th} selection
$\mathbf{k-j+a}$	**1**	$\mathbf{N-k-a}$
m	-	-
$m-1$	1	-
$m-1$	-	1
$\vdots$	$\vdots$	$\vdots$
$m-(m-j)$	1	$(m-j+1)$
$m-(m-j)$	-	$(m-j)$

Using Table 2, the probability of $P(B_j^1/B_{j-1}^{l_{j-1}} \cap B_{j-2}^{l_{j-2}} \cap \ldots \cap B_1^{l_1})$ can be written as,

$$
\begin{aligned}
P(B_j^1/B_{j-1}^{l_{j-1}} \cap B_{j-2}^{l_{j-2}} \cap \ldots \cap B_1^{l_1}) &= \frac{\binom{k-j+a}{m}\binom{1}{0}\binom{N-k-a}{0} + \binom{k-j+a}{m-1}\binom{1}{1}\binom{N-k-a}{0}}{\binom{N-j+1}{m}} \\
&\quad + \cdots + \frac{\binom{k-j+a}{m-(m-j)}\binom{1}{1}\binom{N-k-a}{m-j-1} + \binom{k-j+a}{m-(m-j)}\binom{1}{0}\binom{N-k-a}{m-j}}{\binom{N-j+1}{m}}
\end{aligned} \tag{5}
$$

Using Pascal rule, this equation can be simplified as follows:

$$
\begin{aligned}
P(B_j^1/B_{j-1}^{l_{j-1}} \cap B_{j-2}^{l_{j-2}} \cap \ldots \cap B_1^{l_1}) &= \frac{\binom{k-j+a}{m} + \binom{k-j+a}{m-1}\binom{N-k-a+1}{1}}{\binom{N-j+1}{m}} \\
&\quad + \cdots + \frac{\binom{k-j+a}{m-(m-j)}\binom{N-k-a+1}{(m-j)}\binom{N-j+1}{m}}{\binom{N-j+1}{m}} \\
&= \frac{\sum_{i=0}^{m-j} \binom{k-j+a}{m-i}\binom{N-k-a+1}{i}}{\binom{N-j+1}{m}},
\end{aligned} \tag{6}
$$

and for choosing u_k in the j^{th} selection, $j-1$ elements must be chosen from those greater than u_k and $m-j$ elements from those less than u_k. So the probability of choosing u_k in the j^{th} selection can be written as

$$
P(A_j/B_{j-1}^{l_{j-1}} \cap B_{j-2}^{l_{j-2}} \cap \ldots \cap B_1^{l_1}) = \frac{\binom{k-j+a}{j-1}\binom{N-k-a}{m-j}}{\binom{N-j+1}{m}} \tag{7}
$$

Using these formulae, the inclusion probabilities for all the elements in the population can be derived easily. For example, when $N = 5$ and $m = 3$, the population consists of

$u_1 < u_2 < u_3 < u_4 < u_5$ elements. The inclusion probability of u_k, $(k = 1, 2, 3, 4, 5)$, can be written using (1) as follows:

$$\pi_5(k) = \pi_5^{(1)}(k) + \pi_5^{(2)}(k) + \pi_5^{(3)}(k). \tag{8}$$

In (8), the probability of choosing u_k in the first selection can be written as

$$\pi_5^{(1)}(k) = P(A_1) = \frac{\binom{k-1+0}{1-1}\binom{5-k-0}{3-1}}{\binom{5-1+1}{3}} = \frac{\binom{5-k}{2}}{\binom{5}{3}}. \tag{9}$$

The inclusion probabilities of all elements in the first selection (i.e. $j = 1$) can now be calculated:

$$\begin{aligned}
&\pi_5^{(1)}(1) = \frac{\binom{5-1}{2}}{\binom{5}{3}} = \frac{6}{10}, \quad \pi_5^{(1)}(2) = \frac{\binom{5-2}{2}}{\binom{5}{3}} = \frac{3}{10},\\
&\pi_5^{(1)}(3) = \frac{\binom{5-3}{2}}{\binom{5}{3}} = \frac{1}{10}, \quad \pi_5^{(1)}(4) = \frac{\binom{5-4}{2}}{\binom{5}{3}} = 0,\\
&\pi_5^{(1)}(5) = \frac{\binom{5-5}{2}}{\binom{5}{3}} = 0.
\end{aligned} \tag{10}$$

As seen in (10) the smaller values (i.e. u_1, u_2) have greater inclusion probabilities $(6/10, 3/10)$ in the first selection. If we compare RSS with SRS in terms of inclusion probabilities of elements in the first selection, all elements have the same inclusion probability in SRS (i.e. $1/5$), however in RSS the smaller elements have a greater chance of being selected. On the contrary the greater elements have no chance of being selected (i.e. u_4, u_5).

The inclusion probability of all elements in the second selection, can be written as:

$$\pi_5^{(2)}(k) = P(A_2/B_1^0)P(B_1^0) + P(A_2/B_1^1)P(B_1^1), \tag{11}$$

where, using (4)-(6),

$$P(A_2/B_1^1) = \frac{\binom{k-2+0}{2-1}\binom{5-k-0}{3-2}}{\binom{5-2+1}{3}} = \frac{\binom{k-2}{1}\binom{5-k}{3-2}}{\binom{5-1}{3}}, \tag{12}$$

$$P(A_2/B_1^0) = \frac{\binom{k-2+1}{2-1}\binom{5-k-1}{3-2}}{\binom{5-2+1}{3}} = \frac{\binom{k-1}{1}\binom{5-k-1}{3-2}}{\binom{5-1}{3}}, \tag{13}$$

$$P(B_1^1) = \frac{\sum_{i=0}^{3-1}\binom{k-1+0}{3-i}\binom{5-k+1-0}{i}}{\binom{5-1+1}{3}} = \frac{\binom{k-1}{3} + \binom{k-1}{2}\binom{5-k+1}{1} + \binom{k-1}{1}\binom{5-k+1}{2}}{\binom{5}{3}}, \tag{14}$$

$$P(B_1^0) = \frac{\sum_{i=0}^{1-1}\binom{k-1+1+0}{i}\binom{5-k-0}{3-i}}{\binom{5-1+1}{3}} = \frac{\binom{5-k}{3}}{\binom{5}{3}}. \tag{15}$$

Thus, the inclusion probability of u_k in the second selection is given by (16):

$$\begin{aligned}
\pi_5^{(2)}(k) = {} & \frac{\binom{k-2}{1}\binom{5-k}{3-2}}{\binom{5-1}{3}} \frac{\binom{k-1}{3} + \binom{k-1}{2}\binom{5-k+1}{1} + \binom{k-1}{1}\binom{5-k+1}{2}}{\binom{5}{3}}\\
& + \frac{\binom{k-1}{1}\binom{5-k-1}{3-2}}{\binom{5-1}{3}} \frac{\binom{5-k}{3}}{\binom{5}{3}}.
\end{aligned} \tag{16}$$

For $k = 1, 2, \ldots, 5$, respectively, the inclusion probability of u_k can be written as follows:

$$\pi_5^{(2)}(1) = 0 + 0 = 0$$

$$\pi_5^{(2)}(2) = 0 + \frac{\binom{2-1}{1}\binom{5-2-1}{3-2}}{\binom{5-1}{3}} \frac{\binom{5-2}{3}}{\binom{5}{3}} = \frac{2}{40}$$

$$\pi_5^{(2)}(3) = \frac{\binom{3-2}{1}\binom{5-3}{3-2}}{\binom{5-1}{3}} \frac{0 + \binom{3-1}{2}\binom{5-3+1}{1} + \binom{3-1}{1}\binom{5-3+1}{2}}{\binom{5}{3}} + 0 = \frac{18}{40}$$

$$\pi_5^{(2)}(4) = \frac{\binom{4-2}{1}\binom{4-1}{3-2}}{\binom{5-1}{3}} \frac{\binom{4-1}{3} + \binom{4-1}{2}\binom{5-4+1}{1} + \binom{4-1}{1}\binom{5-4+1}{2}}{\binom{5}{3}} + 0 = \frac{20}{40}$$

$$\pi_5^{(2)}(5) = 0 + 0 = 0$$

Hence we have,

$$\sum_{k=1}^{N=5} \pi_5^{(2)}(k) = 0 + \frac{2}{40} + \frac{18}{40} + \frac{20}{40} + 0 = 1.00.$$

In the third selection, using (2) the inclusion probabilities are:

$$\begin{aligned}\pi_5^{(3)}(k) &= \sum P(A_3/B_2^{l_2} \cap B_1^{l_1}) P(B_2^{l_2}/B_1^{l_1}) P(B_1^{l_1}) \\ &= P(A_3/B_2^0 \cap B_1^0)P(B_2^0/B_1^0)P(B_1^0) + P(A_3/B_2^1 \cap B_1^0)P(B_2^1/B_1^0)P(B_1^0) \\ &\quad + P(A_3/B_2^0 \cap B_1^1)P(B_2^0/B_1^1)P(B_1^1) + P(A_3/B_2^1 \cap B_1^1)P(B_2^1/B_1^1)P(B_1^1),\end{aligned} \tag{17}$$

$$P(B_2^1/B_1^1) = \frac{\sum_{i=0}^{3-2}\binom{k-2+0}{3-i}\binom{5-k+1-0}{i}}{\binom{5-2+1}{3}} = \frac{\binom{k-2}{3} + \binom{k-2}{2}\binom{5-k+1}{1}}{\binom{5-1}{3}}, \tag{18}$$

$$P(B_2^0/B_1^1) = \frac{\sum_{i=0}^{1}\binom{k-2+1+0}{i}\binom{5-k-0}{3-i}}{\binom{5-2+1}{3}} = \frac{\binom{5-k}{3} + \binom{5-k}{2}\binom{k-1}{1}}{\binom{5-1}{3}}, \tag{19}$$

$$P(B_2^0/B_1^0) = \frac{\sum_{i=0}^{1}\binom{k-2+1+1}{i}\binom{5-k-1}{3-i}}{\binom{5-2+1}{3}} = \frac{\binom{5-k-1}{3} + \binom{k}{1}\binom{5-k-1}{2}}{\binom{5-1}{3}}, \tag{20}$$

$$P(B_2^1/B_1^0) = \frac{\sum_{i=0}^{3-2}\binom{k-2+1}{3-i}\binom{5-k+1-1}{i}}{\binom{5-2+1}{3}} = \frac{\binom{k-1}{3} + \binom{k-1}{2}\binom{5-k}{1}}{\binom{5-1}{3}}. \tag{21}$$

Using (7),

$$P(A_3/B_2^1 \cap B_1^1) = \frac{\binom{k-3+0}{3-1}\binom{5-k-0}{3-3}}{\binom{5-3+1}{3}}, \tag{22}$$

$$P(A_3/B_2^1 \cap B_1^0) = P(A_3/B_2^0 \cap B_1^1) = \frac{\binom{k-3+1}{3-1}\binom{5-k-1}{3-3}}{\binom{5-3+1}{3}}, \tag{23}$$

$$P(A_3/B_2^0 \cap B_1^0) = \frac{\binom{k-3+2}{3-1}\binom{5-k-2}{3-3}}{\binom{5-3+1}{3}}. \tag{24}$$

Now, using (17)–(24), the inclusion probabilities for $k = 1, 2, \ldots, 5$, can be written as:

$$\begin{aligned}&\pi_5^{(3)}(1) = \pi_5^{(3)}(2) = \pi_5^{(3)}(3) = \pi_5^{(3)}(4) = 0, \\ &\pi_5^{(3)}(5) = 1.\end{aligned} \tag{25}$$

Table 3. The inclusion probabilities of u_k $(k = 1, 2, 3, 4, 5)$ in the j^{th} selection, $(j = 1, 2, 3)$, when $m = 3$ and $N = 5$.

	u_1	u_2	u_3	u_4	u_5	Total
$\pi_5^{(1)}(k)$	0.60	0.30	0.10	0	0	1.00
$\pi_5^{(2)}(k)$	0	0.05	0.45	0.50	0	1.00
$\pi_5^{(3)}(k)$	0	0	0	0	1.00	1.00
Total	0.60	0.35	0.55	0.50	1.00	3.00

As seen from Table 3, in the first selection, the smaller elements (i.e. u_1, u_2) have greater inclusion probabilities $(0.6, 0.3)$, on the other hand, greater elements (i.e. u_4, u_5) have zero inclusion probabilities. In the second selection, the extreme values have zero inclusion probability, but the remaining elements, especially u_3 and u_4, have the highest inclusion probabilities $(0.45, 0.5)$. In the third selection, no element except u_5 has non-zero inclusion probabilities, on the other hand u_5 is definitely chosen in the sample. Also, the equality $\sum_{k=1}^{N=5} \pi_5(k) = 3.00$ is satisfied.

In addition, as seen from the total inclusion probabilities of u_k, $(k = 1, 2, \ldots, 5)$; u_1 and u_5 have greater inclusion probabilities $(0.6, 1)$ than the others. Thus, this situation agrees with the RSS design which represents the population better than SRS. But these interpretations can vary under different sample and population sizes. For this purpose, we calculate some inclusion probabilities under different sample and population sizes in section 3.

3. Inclusion probabilities for various population and sample sizes

In this section, we investigate the effects of sample size m and population size N on the inclusion probability of elements in the population. The inclusion probabilities are calculated using MATLAB 7.0. The inclusion probabilities calculated are compared with the inclusion probabilities according to SRS with the same sample and population sizes. For this comparison, sample sizes $m = 3, 4, 5$ and population sizes $N = 5(1)12$ are used. These probabilities are given in Table 4–6. In addition, to compare inclusion probabilities for even larger populations, Figure 1-9 are given for $N = 10, 20$ and 500.

Table 4. Inclusion probabilities for $m = 3$

		$m = 3$											
N	Method	1	2	3	4	5	6	7	8	9	10	11	12
5	RSS	0.600	0.350	0.550	0.500	1.000							
	SRS	0.600	0.600	0.600	0.600	0.600							
6	RSS	0.500	0.360	0.414	0.501	0.475	0.750						
	SRS	0.500	0.500	0.500	0.500	0.500	0.500						
7	RSS	0.429	0.343	0.354	0.407	0.427	0.440	0.600					
	SRS	0.429	0.429	0.429	0.429	0.429	0.429	0.429					
8	RSS	0.375	0.319	0.317	0.344	0.368	0.377	0.400	0.500				
	SRS	0.375	0.375	0.375	0.375	0.375	0.375	0.375	0.375				
9	RSS	0.333	0.295	0.289	0.302	0.320	0.331	0.340	0.362	0.428			
	SRS	0.333	0.333	0.333	0.333	0.333	0.333	0.333	0.333	0.333			
10	RSS	0.300	0.272	0.265	0.272	0.283	0.294	0.301	0.309	0.329	0.375		
	SRS	0.300	0.300	0.300	0.300	0.300	0.300	0.300	0.300	0.300	0.300		
11	RSS	0.273	0.252	0.245	0.248	0.256	0.264	0.270	0.276	0.284	0.300	0.333	
	SRS	0.273	0.273	0.273	0.273	0.273	0.273	0.273	0.273	0.273	0.273	0.273	
12	RSS	0.250	0.234	0.228	0.229	0.234	0.240	0.245	0.250	0.254	0.262	0.275	0.300
	SRS	0.250	0.250	0.250	0.250	0.250	0.250	0.250	0.250	0.250	0.250	0.250	0.250

Table 5. Inclusion probabilities for $m = 4$.

		$m=4$											
N	Method	1	2	3	4	5	6	7	8	9	10	11	12
7	RSS	0.571	0.343	0.471	0.495	0.520	0.600	1.000					
	SRS	0.571	0.571	0.571	0.571	0.571	0.571	0.571					
8	RSS	0.500	0.347	0.397	0.451	0.446	0.539	0.520	0.800				
	SRS	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500				
9	RSS	0.444	0.337	0.354	0.397	0.407	0.434	0.484	0.476	0.667			
	SRS	0.444	0.444	0.444	0.444	0.444	0.444	0.444	0.444	0.444			
10	RSS	0.400	0.322	0.325	0.353	0.368	0.380	0.409	0.432	0.439	0.571		
	SRS	0.400	0.400	0.400	0.400	0.400	0.400	0.400	0.400	0.400	0.400		
11	RSS	0.364	0.305	0.301	0.319	0.334	0.343	0.358	0.379	0.392	0.405	0.500	
	SRS	0.364	0.364	0.364	0.364	0.364	0.364	0.364	0.364	0.364	0.364	0.364	
12	RSS	0.333	0.289	0.282	0.293	0.305	0.313	0.322	0.336	0.350	0.359	0.374	0.444
	SRS	0.333	0.333	0.333	0.333	0.333	0.333	0.333	0.333	0.333	0.333	0.333	0.333

Table 6. Inclusion probabilities for $m = 5$.

		$m=5$											
N	Method	1	2	3	4	5	6	7	8	9	10	11	12
9	RSS	0.556	0.337	0.483	0.461	0.460	0.558	0.524	0.667	1.000			
	SRS	0.556	0.556	0.556	0.556	0.556	0.556	0.556	0.556	0.556			
10	RSS	0.500	0.340	0.387	0.427	0.422	0.475	0.496	0.557	0.563	0.833		
	SRS	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500		
11	RSS	0.455	0.333	0.354	0.389	0.393	0.416	0.450	0.465	0.521	0.510	0.714	
	SRS	0.455	0.455	0.455	0.455	0.455	0.455	0.455	0.455	0.455	0.455	0.455	
12	RSS	0.417	0.323	0.329	0.356	0.365	0.377	0.401	0.419	0.440	0.476	0.471	0.625
	SRS	0.417	0.417	0.417	0.417	0.417	0.417	0.417	0.417	0.417	0.417	0.417	0.417

Figure 1. Inclusion probabilities for $m = 3$, $N = 10$

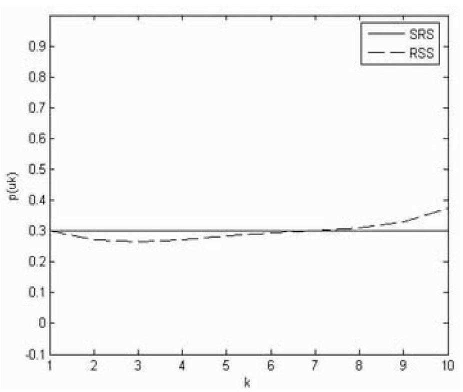


Figure 2. Inclusion probabilities for $m = 4$, $N = 10$

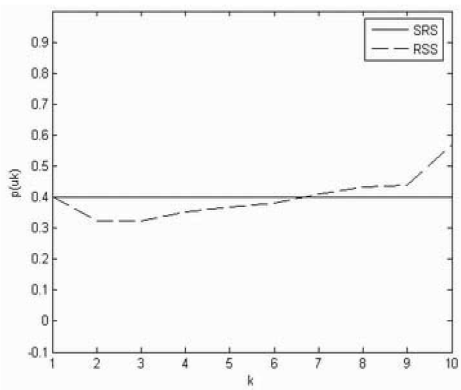


Figure 3. Inclusion probabilities for $m = 5,\ N = 10$

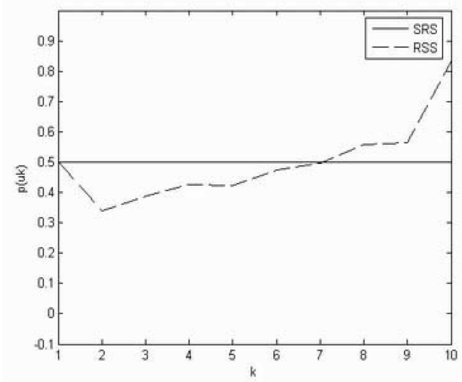


Figure 4. Inclusion probabilities for $m = 3,\ N = 20$

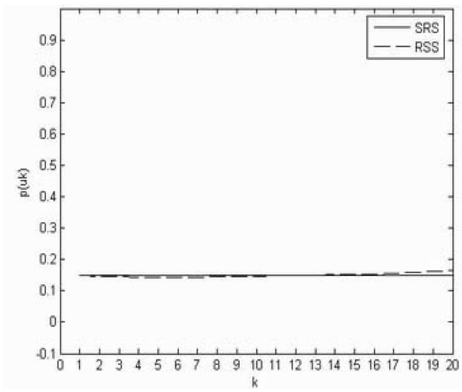


Figure 5. Inclusion probabilities for $m = 4,\ N = 20$

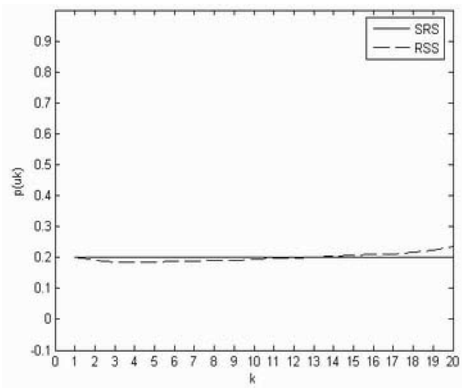


Figure 6. Inclusion probabilities for $m = 5,\ N = 20$

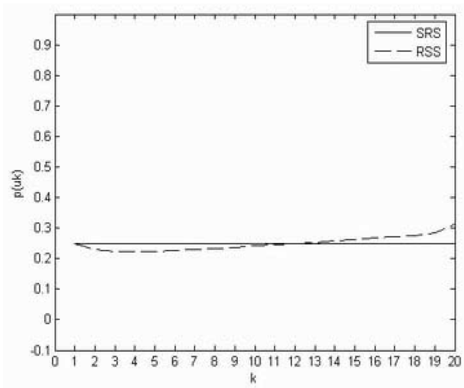


Figure 7. Inclusion probabilities for $m = 3,\ N = 500$

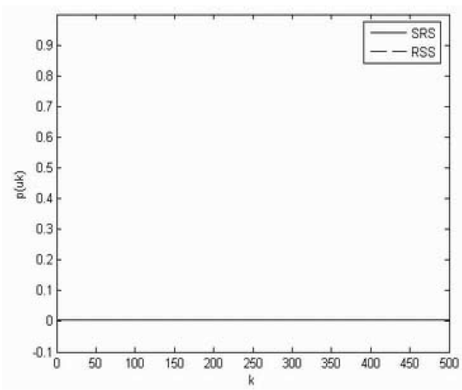


Figure 8. Inclusion probabilities for $m = 4,\ N = 500$

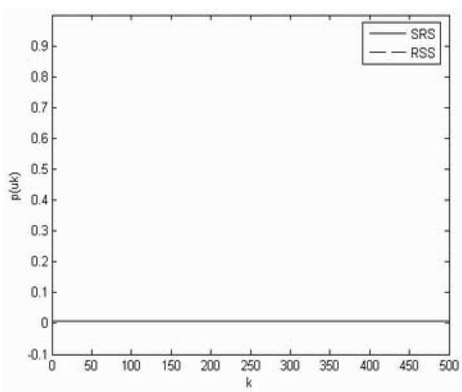
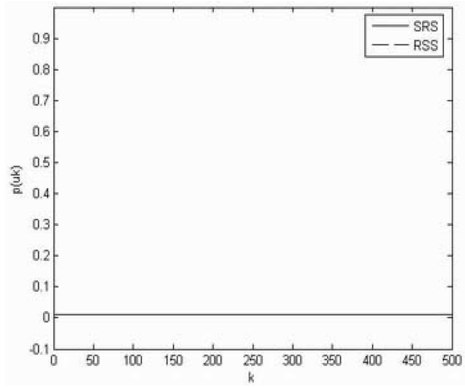


Figure 9. Inclusion probabilities for $m = 5$, $N = 500$



As shown in Tables 4–6 and Figures 1–9, in the RSS design, the elements (u_1, u_N) at the two extremes get higher inclusion probabilities than the others. But u_N takes greater values than u_1. This difference gets smaller as the population size N increases. For example in Table 4, the difference between u_N and u_1 is 0.4 $(1-0.6)$ when the population size $N = 5$, but this difference is 0.05 $(0.3 - 0.25)$ when $N = 12$. Also, the difference in the inclusion probabilities between u_N and u_1 gets smaller when the population size N increases. For example, the difference between the greatest and smallest inclusion probabilities is $0.65(1 - 0.35)$ when $N = 5$, but the difference between the greatest and the smallest inclusion probabilities is 0.072 $(0.3-0.228 = 0.072)$ when $N = 12$. Thus, the difference between SRS and RSS decreases. This decrease can be also seen from Figures 1-9.

Moreover, we note that because of the selection process of the RSS design, the minimum population size must satisfy $N \geq 2m - 1$

4. Conclusion

Inclusion probabilities indicate which elements in the population have a greater chance of being selected. In SRS, all elements in the population have an equal probability of being selected. On the other hand, RSS represents the population better than SRS, because it gives a higher probability to extreme values in the population, especially when the population and sample sizes are small. In this study, we generalized the inclusion probabilities in the RSS design for any population and sample size. Using these inclusion probabilities, test statistics of any hypothesis can be obtained theoretically. This study can also be modified for extreme and median RSS designs.

References

[1] Al Saleh, M. F. and Samawi, H. M. *A note on inclusion probability in ranked set sampling and some of its variations.* Test.(In Press)

[2] Dell, D. R. and Clutter, J. L. *Ranked set sampling theory with order statistics background*, Biometrics **28**, 545–555, 1972.

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

[4] Patil, G. P., Sinha, A. K. and Taillie, C. *Finite population correction for ranked set sampling*, Annals of the Institute of Statistical Mathematics **47**, 621–636, 1995.

[5] Öztürk, Ö., Bilgin, Ö.C. and Wolfe, D. A. *Estimation of population mean and variance in flock management: A ranked set sampling approach in a finite population setting*, Journal Of Statistical Computation and Simulation **75**, 905–919, 2005.

[6] Takahasi, K. and Futatsuya, M. *Ranked set sampling from a finite population*, Proceedings of the institute of statistical mathematics **36**, 55–68, 1988.

[7] Takahasi, K. and Futatsuya, M. *Dependence between order statistics in samples from finite population and its application to ranked set sampling*, Proceeding of the institute of statistical mathematics **50**, 49–70, 1998.

[8] 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 **21**, 249–255, 1968.