Int. J. Mendel, Vol. 30 (1-4), 13-15, 2014 STATISTICS A Comprehensive Review Paper ISSN 0970-9649

RANKED SET SAMPLING METHODS FOR VEGETATION RESEARCH

Arun Kumar Sinha

Key words and Phrases : Cost-effectiveness, Equal and Unequal Allocation, Ranking Information and Error, Relative Precision, Simple Random Sampling, Stratification.

Usually simple random sampling is used in vegetation research which may not provide a representative sample under a budgetary constraint. Also, one needs a large sample size to represent a population. Because of these reasons McIntyre (1952) proposed a cost-effective survey sampling method, which is currently known as ranked set sampling in the literature. In this method a fairly large number of randomly identified sampling units are partitioned into small subsets of the same size. The units of each subset are ranked separately with respect to the characteristic of interest without using their actual measurements. As the ranking induces stratification on the population through rank orders, the method provides a more structured sample than a simple random sample does with the same size. The paper discusses its theory, methods, and some reported and suggested applications in vegetation research. Besides, this explores some future directions that could save enormous resources. This work could be of particular interest to those who look for a cost-effective survey sampling and monitoring technique in vegetation research.

INTRODUCTION

For adequately representing heterogeneous populations large sample sizes are required. But a budgetary constraint does not allow to obtain the desired number of measurements. The "best of both worlds" scenario is referred to as the observational economy. For observational economy to be feasible, identification and acquisition of sampling units should be inexpensive as compared with their quantification. Ranked set sampling (RSS) is one such method for achieving observational economy. McIntyre (1952) proposed a method of sampling for estimating pasture yields and referred to it as "a method of unbiased selective sampling using ranked sets" in this direction. The present name, "ranked set sampling" (RSS), was proposed by Halls and Dell (1966). RSS is useful in situations in which measurements are difficult or costly to obtain, but ranking of a subset of units with respect to the characteristics of interest is relatively economical and convenient. In fact, RSS combines the convenience of purposive sampling and the control of simple random sampling (SRS). Unlike stratified sampling that presumes a stratification of the population, it accomplishes stratification through samples using easily available and cheap outside information. The ranking is carried out with respect to the characteristic of interest by means of some crude method like visual perception that does not require the exact measurements. It, in turn, exploits the experience and expertise of the field personnel for improving upon the efficacy of SRS. One of its strengths is flexibility and model robustness regarding the nature of the auxiliary information needed for ranking. For achieving the cost-effectiveness, the method presumes that the quantification rather than identification, acquisition and ranking of sampling units is the main component of the total sampling cost. Though it was initially proposed for estimating the mean pasture yields more efficiently, it has been used in other sampling situations advantageously. The rigorous mathematical support to this technique, without referring to the McIntyre's work, was given by Takahasi and Wakimoto (1968) and, independently, by Dell (1969) while investigating the McIntyre's contribution. The McIntyre's method appears to have lain dormant for almost fourteen years until Halls and Dell (1966) conducted a field survey to examine its effectiveness for estimating weights of browse and of herbage in a pine hardwood forest in the USA. In fact, they established empirically that it was more efficient than SRS. As errors could get involved while ordering due to dependence on the ranker's judgment, Dell and Clutter (1972) showed that the RSS estimator of a population mean remains unbiased, and is at least as efficient as the SRS estimator with the same number of quantifications. They pointed out that its performance would depend upon the characteristics of the population and also on the magnitude of the errors in ranking. Patil, Sinha and Taillie (1994b) and Sinha (2005) presented a comprehensive review of the literature with some new results of immediate interest. In a recent paper Muttlak and Al-Saleh (2000) also discussed some recent developments in RSS. Johnson, Patil and Sinha (1993) described its applications in vegetation research and Mode, Conquest and Marker (1999) discussed its relevance for ecological research. As it presumes sampling from an infinite population Patil, Sinha and Taillie (1995), and Takahasi and Futatsuya (1988) investigated its characteristics under a finite population scenario. Apart from visual perceptions, ranking may be carried out on the basis of remotely sensed information, prior information, results of earlier sampling episodes, rank correlated covariates, expert-opinion/ expert systems, etc., or some combinations of these methods. Contrary to this procedure other techniques like ratio and regression methods that utilize the auxiliary information for improving upon the SRS estimator of a population mean require very detailed specifications for the outside information. Besides the McIntyre's RSS (MRSS) method, Takahasi (1970) developed an RSS method that uses the ranking information at the estimation stage unlike MRSS that needs the information at the selection stage. But it is slightly less efficient than the former. On ignoring the ranking information one gets a simple random sample from a Takahasi's RSS (TRSS). Further, multiple characteristics may be estimated more efficiently

Department of Statistics, Central University of Bihar
B I T Patna Campus, PO: B V College, Patna-800 014, Bihar, INDIA
E-mail : arunkrsinha@yahoo.com

following these RSS methods. (Norris, Patil and Sinha, 1995 and Patil, Sinha and Taillie, 1994c). Moreover, Norris, Patil and Sinha (1995) proposed some modifications to TRSS, and named it as a TNPS (Takahasi, Norris, Patil, and Sinha) RSS (TNPSRSS). It is probably the first paper to apply TRSS and MTRSS in real life situations. These, in turn, have enhanced its applicability tremendously to deal with real-life sampling and monitoring situations more effectively. Ridout (2002) extended the use of the sampling technique in this direction further. Sinha *et al.* (2001) compared the performances of TRSS and MTRSS for estimating a population mean as compared with the linear regression (LR) estimator with and without double sampling under perfect and concomitant ranking scenarios. This investigation finds the RSS methods better than the LR estimators to estimate a population mean when the correlation between the main variable and the concomitant variable is not very high. Though the MRSS estimator is slightly more efficient than the MTRSS estimator, other advantages of the latter offset this loss. Also, like the former this method reveals its maximum potentiality under perfect ranking scenario. In this paper an attempt is made to present a review of its theory, methods and applications in vegetation research. This work may be of some particular interest to those who look for a cost-effective sampling method in this area.

2. McIntyre's RSS Method

For obtaining a ranked set sample through the McIntyre's RSS procedure m random samples with m units in each sample are selected from a population with mean, μ and a finite variance, σ^2. This is equivalent to drawing m^2 units randomly and then subdividing them into m samples, each with m units. The m units of each subset are ranked with respect to the variable of interest without using their exact measurements. For this purpose some outside information like visual perception, past experience, etc., are used. Using this ranking information the unit with the smallest rank is quantified from the first subset; the unit with the second smallest rank is measured from the second subset, and this process of quantification is continued until the unit with the mth rank is measured from the mth subset. This yields m measurements with each of the first m ranks, and these constitute an MRSS of size m. For obtaining a larger sample of size mr the whole procedure is repeated r times. Here, m is referred to as the set size while r is called as the number of cycles. In terms of usual notations we get the sample of size, $n = mr$ from the population with its size, $N \geq m^2r$. Further, one could utilize the prior information, if available, about the population for obtaining a more efficient estimator of the population parameter of interest. For a skewed population the number of quantifications of the ith rank, i.e., r_i is taken as proportional to the standard deviation of the rank order, $\sigma_{(i:m)}$. Because of unequal number of quantifications this allocation may be termed as MRSS with unequal allocation (MRSSUA).

As an illustration suppose we need to select three trees using RSS for measuring the average height of tress of an area. For drawing the sample, first of all, nine trees are randomly identified as shown in Figure 1. The nine trees are arranged into three rows (sets) with each row consisting of three trees. The trees in each row are accorded ranks visually (without measuring their actual heights). In the first row the ranks appear as the tallest, the smallest and of middle height. The trees shown in the second row have ranks as the smallest, middle and the tallest. Similarly, the trees of the third row get the ranks as of the middle, the tallest and of the smallest height. Form the first row the tree getting the smallest height is measured, the tree of the middle height is quantified from the second row and finally the height of tree having the tallest height is obtained from the third row.

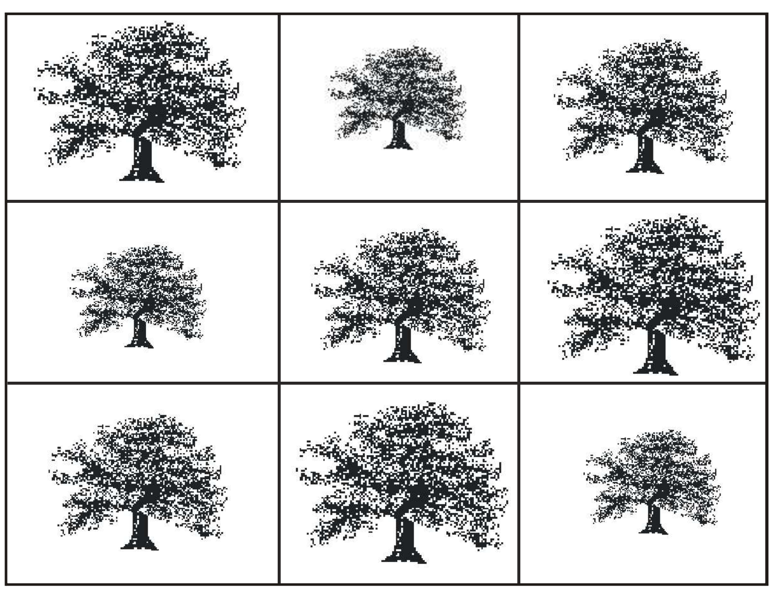

Fig.-1 : Nine randomly selected trees are arranged into three rows (sets) with each row (set) having three trees.

Finally the three selected trees with smallest, middle and tallest heights are shown in Figure 2. Their heights are measured for computing the required statistics.

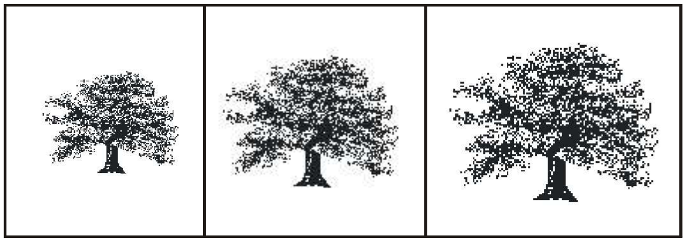

Fig.-2 : A ranked set sample of three selected trees of smallest, middle and highest sizes. Here the set size, m is equal to three.

For getting an RSS of a larger size, the whole procedure is repeated. In the present example the set size is three and it is usually denoted by m. Let us suppose that we need to have a sample of size 12 with the set size three. The whole procedure is to be repeated four times. As the repetition is referred to as a cycle, we need to have four cycles to draw sample of the desired size. As each cycle yields three values, the four cycles will provide 12 measurements with four values of each rank order. One could have larger set sizes, but this depends on the convenience of the ranking of the sampling unit with respect to the characteristics of interest because ranking has to be carried out without measurements. Also it is important to note that the efficiency of RSS increases with the set size. Thus the performance of RSS could be increased by considering a larger set size if it is convenient to the investigator.

3. McIntyre's Estimator

In general suppose that $X_{(i:m)j}$ denotes the *i*th order statistic based on perfect ranking in the *j*th cycle, for *i = 1*, …, *m* and *j = 1*,…, r. Note that these are not independent and identically distributed (*iid*) in general, but for a given value of *i* these are so with $E\left(X_{(i:m)j}\right) = \mu_{(i:m)}$, and $\mathrm{var}(X_{(i:m)j}) = \sigma^2_{(i:m)}$. The McIntyre's estimator, $\hat{\mu}_{MRSS}$, of the population mean, μ, is defined as follows :

$$\hat{\mu}_{MRSS} = \frac{1}{mr}\sum_{i=1}^{m}\sum_{j=1}^{r} X_{(i:m)j} \qquad (1)$$

Also, if $\hat{\mu}_{(i:m)} = \frac{1}{r}\sum_{j=1}^{r} X_{(i:m)j}$ then

$$\hat{\mu}_{MRSS} = \frac{1}{m}\sum_{i=1}^{m}\hat{\mu}_{(i:m)} \quad . \qquad (1a)$$

Here $\mathrm{E}\left(\hat{\mu}_{(i:m)}\right) = \mu_{(i:m)}$; $\mathrm{E}\left(\hat{\mu}_{MRSS}\right) = \mu$ and $\mathrm{var}\left(\hat{\mu}_{(i:m)}\right) = \frac{\sigma^2_{(i:m)}}{r}$

Thus we get

$$\mathrm{var}\left(\hat{\mu}_{MRSS}\right) = \frac{1}{m^2 r}\sum_{i=1}^{m}\sigma^2_{(i:m)}. \qquad (2)$$

This expression is also expressed as

$$\mathrm{var}\left(\hat{\mu}_{\mathrm{MRSS}}\right) = \frac{1}{mr}\left[\sigma^2 - \frac{1}{m}\sum_{i=1}^{m}\left(\mu_{(i:m)} - \mu\right)^2\right]. \qquad (3)$$

The alternative expression is useful when the variance of the *i*th order statistic is not available. If we denote the SRS estimator of the population mean with the same sample size, *n = mr* by $\hat{\mu}_{SRS}$ and then $\mathrm{var}\left(\hat{\mu}_{SRS}\right)$. Bickel (1967) and Tukey (1958) for a proof of this inequality, which follows from the well known positively associated property of order statistics.

3.1 Relative Precisions of the Estimator

The relative precision (RP) of the MRSS estimator, $\hat{\mu}_{MRSS}$ as compared with simple random sample (SRS) estimator, $\hat{\mu}_{SRS}$ with the same sample size, *n* is computed as follows :

$$\mathrm{RP} = \frac{\mathrm{var}\left(\hat{\mu}_{SRS}\right)}{\mathrm{var}\left(\hat{\mu}_{MRSS}\right)}.$$

As , $\mathrm{var}\left(\hat{\mu}_{SRS}\right) = \frac{\sigma^2}{mr}$ this leads to

$$\mathrm{RP} = \frac{\sigma^2}{\overline{\sigma^2}} \text{ ; where } \overline{\sigma^2} = \frac{\sum_{i=1}^{m}\sigma^2_{(i:m)}}{m}. \qquad (4)$$

The variance $\hat{\mu}_{MRSS}$ given in equation (3) yields the expression for RP as given below :

$$\mathrm{RP} = \frac{1}{1 - \frac{1}{m\sigma^2}\sum_{i=1}^{m}\left(\mu_{(i:m)} - \mu\right)^2} \quad . \qquad (5)$$

An equivalent and useful measure could be relative cost (RC) and relative savings (RS).

These are defined as

$$\text{RC} = \frac{1}{\text{RP}} \text{ and } \text{RS} = 1\text{-RC}.$$

An expression for RS based on equation (4) is given below:

$$\text{RS} = \frac{\sigma^2 - \overline{\sigma^2}}{\sigma^2} \qquad (6)$$

An equivalent expression for RS based on equation (5) is obtained as follows:

$$\text{RS} = \frac{1}{m\,\sigma^2}\sum_{i=1}^{m}\left(\mu_{(i:m)} - \mu\right)^2. \qquad (7)$$

McIntyre (1952), and Takahasi and Wakimoto (1968) showed that $1 \le \text{RP} \le \frac{\text{m}+1}{2}$ and so, $0 \le \text{RS} \le \frac{\text{m-1}}{\text{m}+1}$. The results suggest that there is no loss due to using MRSS instead of SRS. Obviously, there is no gain when ranking is the same as a random ordering, i.e., For obtaining an estimator of RP we use

$$\text{E}\left[\sum_{j=1}^{r}\left(X_{(i:m)j} - \hat{\mu}_{(i:m)}\right)^2 \Big/ (r-1)\right] = \sigma^2_{(i:m)}.$$

An expression for the unbiased estimator of the population variance σ^2 based on a ranked set sample when the number of cycle is more than one is given by

$$\hat{\sigma}^2_{MRSS} = \left[\frac{mr-m+1}{m^2 r(r-1)}\right]\sum_{i=1}^{m}\sum_{j=1}^{r}\left(X_{(i:m)j} - \hat{\mu}_{(i:m)}\right)^2 + \left(\frac{1}{m}\right)\sum_{i=1}^{m}\left(\hat{\mu}_{(i:m)} - \hat{\mu}_{MRSS}\right)^2. \qquad (8)$$

For this result see Stokes (1976), Patil, Sinha and Taillie (1994a and 1993b), Norris, Patil and Sinha (1995) and Yanagawa (2000). Stokes (1980) proposed an expression for the estimator of the population variance that is mentioned below :

$$s^2_{(m)r} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{r}\left(X_{(i:m)j} - \hat{\mu}_{MRSS}\right)^2}{mr-1}.$$

Unlike the unbiased estimator this estimator of the population variance could be used to estimate the population variance even if the number of cycle is one. It is biased but asymptotically unbiased as *m* or *r* increases because

$$E\left(s^2_{(m)r}\right) = \sigma^2 + \frac{\sum_{i=1}^{m}\left(\mu_{(i:m)} - \mu\right)^2}{m(mr-1)}.$$

An estimator of RP, based on unbiased estimators of and , is obtained as follows :

$$R\hat{P} = \frac{mr\text{-}(m\text{-}1)}{mr} + \frac{(r-1)\sum_{i=1}^{m}\left(\hat{\mu}_{(i:m)} - \hat{\mu}_{MRSS}\right)^2}{\sum_{i=1}^{m}\sum_{j=1}^{r}\left(X_{(i:m)j} - \hat{\mu}_{(i:m)}\right)^2}. \qquad (9)$$

The Stoke's biased estimator of σ^2 yields an estimator of RP as given below :

$$R\hat{P} = \left[\frac{m(r-1)}{(mr-1)}\right]\frac{\sum_{i=1}^{m}\left(X_{(i:m)j} - \hat{\mu}_{MRSS}\right)^2}{\sum_{i=1}^{m}\sum_{j=1}^{r}\left(X_{(i:m)j} - \hat{\mu}_{(i:m)}\right)^2}. \qquad (10)$$

But these estimators are useful only when the number of cycles, i.e., *r* is more than one.

3.2 Some More Efficient Estimators

Skewed Distributions

Contrary to the approach of the same sample size for each rank order, McIntyre (1952) suggested to take the sample size of each rank order proportional to its standard deviation while sampling an asymmetrical population. The former may be referred to as equal allocation approach while the latter may be called as unequal allocation method. It means that if *ri* denotes the number of sets having quantified units with rank *i*, then $r_i \propto \sigma_{(i:m)}$ for $i = 1, \cdots, m$. This leads to

$$r_i = \frac{n\sigma_{(i:m)}}{\sum_{i=1}^{m}\sigma_{(i:m)}}; \quad i = 1, \cdots, m \quad . \qquad (11)$$

The RSS estimator, $\hat{\mu}_{MRSS}$, of the population mean, μ, based on the unequal allocation of samples (also called unbalanced allocation) is given by

$$\hat{\mu}_{MRSSUA} = \frac{1}{m}\sum_{i=1}^{m}\frac{T_i}{r_i} \text{ and } \text{var}\left(\hat{\mu}_{MRSSUA}\right) = \frac{1}{m^2}\sum_{i=1}^{m}\frac{\sigma^2_{(i:m)}}{r_i}$$

where *Ti* shows the sum of the quantification of the *ri* units having *i*th rank order. On putting the value of *ri* from equation (11) into the expression for $\text{var}\left(\hat{\mu}_{MRSSUA}\right)$ we have

$$\text{var}\left(\hat{\mu}_{MRSSUA}\right) = \frac{\left(\bar{\sigma}\right)^2}{n} \qquad (12)$$

where $\bar{\sigma} = \frac{\sum_{i=1}^{m}\sigma_{(i:m)}}{m}$.

The relative precision (*RPua*) of $\hat{\mu}_{MRSS}$ relative to $\hat{\mu}_{SRS}$ with the same number of quantifications is given below :

$$RP_{ua} = \frac{\operatorname{var}(\hat{\mu}_{SRS})}{\operatorname{var}(\hat{\mu}_{\text{MRSSUA}})}.$$

This yields that

$$RP_{ua} = \frac{\sigma^2/n}{\sum_{i=1}^{m} \frac{\sigma^2_{(i:m)}}{r_i} \Big/ m^2}. \quad (13)$$

This could also be expressed as

$$RP_{ua} = \left(\frac{\sigma}{\bar{\sigma}}\right)^2. \quad (14)$$

See Patil, Sinha and Taillie (1993b) for these results. Further, it is interesting to note that

$$\operatorname{var}(\hat{\mu}_{MRSS}) - \operatorname{var}(\hat{\mu}_{MRSSUA}) = \frac{\sum_{i=1}^{m}\left(\sigma_{(i:m)} - \bar{\sigma}\right)^2}{mn}.$$

This proves that $RP_{ua} \geq RP.$ Takahasi and Wakimoto (1968) show that $0 \leq RP_{ua} \leq m.$

3.3 Estimator of RP$_{ua}$

For obtaining an estimator of the RP$_{ua}$ we use the estimator of the population variance and that of the population variance of the ith rank order based on unequal sample sizes as given below.

$$\hat{\sigma}^2_{MRSS} = \sum_{i=1}^{m}\left[\frac{m(r_i-1)+1}{m^2 r_i (r_i-1)}\right]\sum_{j=1}^{r_i}\left(X_{(i:m)j} - \bar{X}_{(i:m)}\right)^2 + \left(\frac{1}{m}\right)\sum_{i=1}^{m}\left(\bar{X}_{(i:m)} - \bar{X}_{(m)r}\right)^2. \quad (15)$$

$$\hat{\sigma}^2_{(i:m)UA} = \frac{\sum_{j=1}^{r_i}\left(X_{(i:m)j} - \hat{\mu}_{(i:m)UA}\right)^2}{r_i - 1} \text{ when}$$

$$\hat{\mu}_{(i:m)UA} = \frac{1}{m}\sum_{i=1}^{m}\frac{T_i}{r_i}. \quad (16)$$

See Norris, Patil and Sinha (1995).

4. Takahasi's RSS Method and the Estimator

The randomly selected *m²r* units from an infinite population are arranged into r sets, and each set consists of *m²* units located into *m* rows and *m* columns. A randomly selected unit from each row is measured, and a rank between 1 and *m* (both inclusive) is accorded to the quantification using the experience and expertise of the field investigator / investigators. Evidently, it may not be feasible to get the same frequency for each rank order as in the case of the MRSS method and also, there is a possibility of zero frequency for a rank order. In this situation Takahasi (1970) suggested the McIntyre's method for collecting samples in one cycle. This, in turn, ensures that every rank order gets at least one quantification. This works well to estimate the population means, but this does not help out while estimating the variance of the estimator. In view of these predicaments Norris, Patil and Sinha (1995) suggested to use McIntyre's method in two cycles while using TRSS, and referred to it as a modified TRSS or TNPSRSS (Takahasi, Norris, Patil and Sinha RSS) method.

For the estimator, $\hat{\mu}_{MTRSS}$ we draw an RSS of size (n-2m) based on the Takahasi's method and two cycles of data are obtained following McIntyre's method. The expression, when $r_i \geq 2$, is given as:

$$\hat{\mu}_{MTRSS} = \frac{1}{m}\sum_{i=1}^{m}\frac{S_i}{r_i}. \quad (17)$$

The variance of the estimator $\hat{\mu}_{MTRSS}$ is mentioned below:

$$\operatorname{var}(\hat{\mu}_{MTRSS}) = \frac{1}{m(n-2m+1)}\left[1 - \frac{m}{(n-2m+2)}\left\{1 - \left(1 - \frac{1}{m}\right)^{n-2m+2}\right\}\right]\sum_{i=1}^{m}\sigma^2_{(i:m)}$$

Norris, Patil and Sinha (1995) may be seen for this result.

5. RSS Methods with Concomitant Ranking

RSS presumes that the sampling units are correctly ranked with respect to the variable of interest. This is a perfect ranking (PR) scenario, but this may not always be a rational approach in real life occasions. In these situations one could take help of some other characteristic for ranking, which is inexpensive, easily available and highly correlated with the main characteristic of interest. Unlike perfect ranking scenario the ranking so obtained may be referred to as concomitant ranking (CR) because of its dependence on a concomitant variable. In this situation the relative precision of the estimator will be less than that of the estimator under perfect ranking scenario. But the reduction depends on the magnitude of the correlation between the main variable and the concomitant variable. For a more detailed discussion on concomitant ranking Sinha *et al.* (2001), and Patil, Sinha and Taillie (1993a, 1994a and b) may be looked into.

6. Some Reported and Suggested Applications of RSS in Vegetation Research

6.1. Forage Yields

Mc Intyre's method of sampling was employed first time by Halls and Dell (1966) in a real-life problem. They used it for estimating the weights of browse and herbage in a pine-hardwood forest of east Texas, USA. They found the Mc Intyre's methods of sampling noticeably more efficient than simple random sampling. The variance of mean was found smaller under Mc Intyre's sampling method with perfect ranking under proportional allocation as compared with the method with the same ranking scenario under near equal allocation, which also provided smaller variance of the mean than simple random sampling for both browse and herbage.

For implementing the RSS method for estimating browse 126 random points were selected on a 300-acre tract. A circle of 13 feet radius was placed at each such point, and then metal frames of 3.1 square feet were placed at three randomly selected points within each circle (Figure 3). Quadrats were then ranked as highest, intermediate and lowest according to the apparent weight (without measuring it) of browse. Then, after clipping and drying, the separate weight of browse for each rank order was determined. The same procedure was adopted for estimating herbage at 124 randomly located points with three ranks. For unequal allocation they selected 72 quadrats in the high group, 40 in the intermediate group and 14 in the low group using the standard deviations found in the high, intermediate and low yield group as 27.7, 13 and 7 respectively. For obtaining the SRS estimator for the mean weight of browse one quadrat was randomly selected from each circle without using its rank with respect to its weight. It is the scenario of perfect ranking because the exact weight of each quadrat was known. The results are given in Table 1.

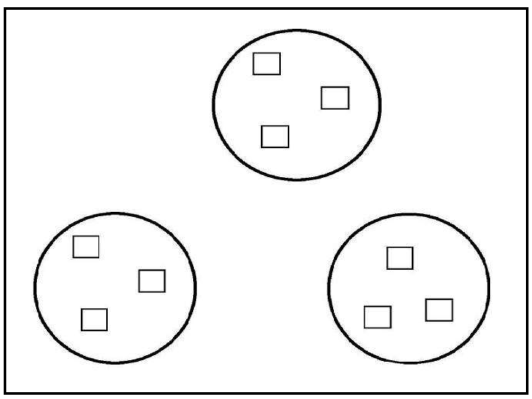

Figure 3 : Quadrats within circles at randomly selected locations

[Source: Patil, G. P., Sinha, A. K. and Taillie, C. (1994b]

For this investigation the ranks to the yields were accorded by two independent observers, one a professional range man and the other a woods worker. There was almost no difference between the results obtained under the two settings.

TABLE-1 : Mean and variance of mean under simple random sampling and Mc Intyre's methods of sampling

Sampling methods	Browse		Herbage	
	Mean	Variance of mean	Mean	Variance of mean
Simple random sampling	14.9	4.55	7.33	1.00
Mc Intyre's sampling method with perfect ranking under <u>near equal allocation</u>	13.2	2.18	7.0	0.73
Mc Intyre's sampling method with perfect ranking under proportional (unequal) allocation	12.9	1.91	7.2	0.58

[Source: Halls and Dell (1966)]

6.2. Seedling Counts

The second field experiment to examine the capability of RSS was conducted by Evans (1967). The author compared the performance of RSS with SRS while using seedling counts in an area of Central Louisiana that was seeded to Longleaf Pine. For this investigation the target area was divided into 24 blocks, and then each block was further partitioned into 25 one-milacre plots. The seedling count was carried out in all 600 the plots. The result is presented in Table 2.

TABLE-2 : The frequency distribution of seedling counts in the 600 milacre plots

Seedling Count	0	1	2	3	4	5	6	7	8	9
Frequency	110	201	157	75	33	17	3	3	0	1

The population mean and standard deviation were obtained as 1.675 and 1.36 respectively. For implementing the RSS procedure three plots were selected randomly from each of the 24 blocks (sets) and the plots were visually ranked with respect to the seedling counts. The lowest ranked plot (L) was selected from the first set, the middle ranked plot (M) was identified from the second set, and the highest ranked plot (L) was chosen from the third set. This process was carried out in the remaining 21 blocks. Finally, the process provided 24 measurements consisting of eight measurements for each rank order. Thus, the author got a ranked set sample of size (n) 24, which is based on the three ranked orders (m = 3) and the eight cycles (r = 8). In other words, m (=3) times r (=8) is n (=24). Next, the three field trials were performed and the results are summarized in Table 3.

TABLE-3 : Means and variances of three ranked set sample trials each with sample size 24

Trial	Mean	Variance
1	1.49	0.043
2	1.62	0.056
3	1.71	0.024

The author also obtained the mean and standard deviations of all the three ranks using all 72 identified plots for each of the three field trials. The results so computed are reproduced in Table 4. One could compare the result of Table 3 with that of Table 4.

TABLE-4 : Means and standard deviations of all 72 (24x3) measurements under all three ranks for each of the three field trials of ranked set sampling

Trial	Means			Mean	Standard Deviations		
	H	M	L		H	M	L
1	2.625	1.500	0.750	1.625	1.173	0.750	0.532
2	2.833	1.625	0.917	1.792	1.880	1.013	0.881
3	3.125	1.708	0.750	1.861	0.927	0.955	0.520

For comparing RSS with SRS Evans resampled the 24 blocks (sets) 80 times to obtain two empirical distributions of the means, which were based on the former and the latter. The result is summarized in Table 5. We observe a significant reduction in the variance due to RSS.

TABLE-5 : Test of significance of ranked set sampling versus random sampling

Sampling method	Number of applications	Degrees of freedom	Mean	Sum of squares	Variance	F
Random	80	79	1.079	7.572	0.0958	3.91**
Random Ranked set	80	79	1.647	1.939	0.0245	
** Significant at 0.01 level of probability						

6.3 Shrub Phytomass in Forest Stands

Martin *et al.* (1980) examined the performance of RSS as compared with SRS while estimating shrub phytomass (all vegetation between one and five meters high) at a forested site in Virginia, USA. The vegetation types included (i) mixed hardwood, (ii) mixed oak, (iii) mixed oak and pine and (iv) mixed pine. To begin the experiment an area of 20m by 20m was subjectively identified for each vegetation type, and then each such area was divided into 16 plots, each of size 5m by 5m.

For the computing RSS mean of phytomass (kg/ha) the set size (m) was taken as four. First of all, four sets of four plots were chosen randomly from the 16 plots in each vegetation type. The plots in each set were ranked visually, then the smallest ranked plot was measured from the first set, the second smallest ranked plot was quantified from the second set, the third smallest ranked plot was measured from the third set, and finally the highest ranked plot was quantified from the fourth set. This yielded four the measurements of each rank order. The process was repeated for all the four vegetation types. For SRS four plots were randomly selected without replacement from 16 plots of each vegetation type, and then the selected plots were quantified. This provided 16 measurements based on SRS. In fact, this was a stratified random sample because each vegetation could be considered as a separate stratum. The computation reveals that the RSS mean has a much smaller variance than that of the SRS mean with the same sample size. The complete results are reproduced in Table 6.

TABLE-6 : RSS and SRS results for 16 measured plots across all vegetation types

Sampling Method	Mean Phytomass (kg/ha)	Variance of the Mean (X10^6)	Coefficient of Variation (%)
All 64 Plots	2536	0.15	15
SRS	1936	4.54	108
RSS	2356	2.73	70

[Sources: Martin *et al.* (1980)]

6.4. Herbage Mass

The experiment was conducted at Hurley (UK) by Cobby *et al.* (1985) to compare RSS with SRS for estimating herbage mass in pure grass swards and both herbage mass and clover contents in mixed grass-clover swards. Besides, the authors also examined the impact of (i) imperfect ranking within sets, (ii) greater variation between sets than within sets, and (iii) asymmetric distribution of the quantified values.

The first two experiments were conducted by randomly selecting 15 points within the target area, and then three quadrats were randomly placed at each location. See Figure 4. Several observers accorded ranks to the quadrats within each set. For the last two experiments 45 quadrats were selected randomly from the entire target area. This was meant to get an assessment of the effects of both spatial variation and ranking errors within sets.

The results of the experiments are reproduced in Table 7, which contains RPs of the worst, thee best observers for comparing these values with the RPs under perfect ranking scenario. Besides, between and within sets variances are also included for assessing the spatial variation. The authors suggest the placing of quadrats within sets as apart as possible to minimize the impact of local spatial correlation and recommend RSS over SRS for sampling grass and grass-clover swards.

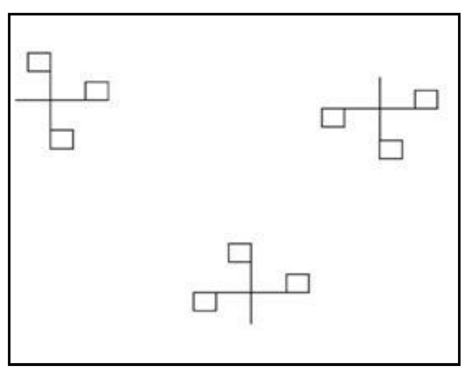

Figure 4. Quadrats at randomly selected locations
[Source: Patil, G. P., Sinha, A. K. and Taillie, C. (1994b)]

TABLE-7 : Relative precisions (RP) ± s.e. of the worst and the beast observers, and under perfect ranking, also the between and the within set variances while estimating herbage mass (grass and mixture) and clover contents

Experiments	Relative Precisions (R P)			Variances	
	Worst	Best	Perfect	Between	Within
1 (Grass)	1.11±0.09	1.23±0.14	1.31±0.17	0.24	0.31
2 (Mixture)	1.11±0.09	1.27±0.10	1.40±0.16	0.07	0.09
3 (Grass)	-	-	1.66±0.17	0.00	1.58
4 (Mixture)	1.36±0.14	1.51±0.15	1.55±0.16	0.11	0.66
2 (Clover)	1.15±0.12	1.34±0.15	1.44±0.16	16.3	34.4
4 (Clover)	1.36±0.19	1.62±0.18	1.72±0.20	16.2	71.6

[Source: Martin *et al.* (1985)]

6. 5. Estimation of Multiple Characteristics

Many times we need to estimate several correlated characteristics economically. For example, one could be interested to investigate not only the characteristics of immediate interest, but also the soil quality under the plantations. See Figure 5. Using the experience and expertise of the field personnel the RSS methods could be employed for estimating multiple characteristics cost-effectively in a single investigation. Of course, the level of cost-effectiveness depends on the magnitude of the correlation between the main variable of interest and other associated variables. Patil, Sinha and Taillie (1994) initiated the work in this direction considering a sampling situation referred to by Sengupta et al. (1951), and discussed by Stokes (1980). For the Government of India Sengupta et al. (1951) conducted a survey of cinchona plants to estimate the yield of dry bark and quinine content, which is used in the treatment of malaria. The yield of dry bark from these plants is obtained after passing through a number of stages like uprooting the plants, stripping the bark and then drying it until its weight gets stabilized. Further, they had observed that the dry bark yield was highly correlated (0.9) with the volume of bark, which could be approximately determined using the height, the girth and thickness of the bark at some specific heights of the plants. As the required data were not available, Stokes generated a bivariate data set (X, Y) of size 150 (and assuming bivariate normality. Finally, she could find a ranked set sample of dry bark weight (X) of size 30 (m = 5 and r = 6). The author obtained the relative precision RP(X:Y) of the RSS estimate of the dry bark weight compared to that of the SRS estimate as 2.07. In real life scenario one might be interested to estimate the average volume of bark as well as its dry weight per plant. This situation could arise for various comparative studies dealing with impact of fertilizers, insecticides, environmental setting, etc.

Figure 5. A tree with several characteristics of interest

Patil, Sinha and Taillie (1994) obtained the Relative Precision (Y) = 2.76. A higher value for RP(Y) than RP(X:Y) occurs because in the former case the ranking of the plants was assumed to be accomplished on the basis of bark volume of the plant, which was used as a concomitant variable for estimating dry bark yield. Obviously, one could obtain the RSS estimate of the characteristic on which the ranking of the units is based with the maximum RP, and thereafter the RPs of the estimates of other characteristics depend upon the correlation between the characteristic on which the ranking is based, and the characteristic that is being estimated.

Patil, Sinha and Taillie (1994) applied the RSS procedures to estimate the means of more than one characteristics of interest using an observed bivariate data set referred to by Prodan (1968). The paper has shown that the efficiency of the estimation for the other characteristics depends on their correlations with the characteristic on which ranking is based, but is at least as good as SRS with the same number of measurements.

Using the data set of of 399 trees provided by Platt, Evans and Rathbun (1988) the means of diameter, height and age were estimated employing the RSS methods. With the equal allocation of RSS the estimates of the RPs were obtained as

3.40, 2.36 and 1.76 for height, diameter and age respectively. The corresponding values in the case the unequal allocation are 4.21, 3.05 and 2.60 respectively. The TNPS's method gave the estimates of the RP's for height, diameter and age as 3.17, 2.95 and 1.86 respectively. The results appear better than the McIntyre's method based on equal allocation. The results were obtained considering the set size 6 and the number of cycles 10. These results suggest that the unequal allocation is overall the most efficient method of RSS. See Norris, Patil and Sinha (1995) for more details.

6.6. Estimation of Underground Farm Produce

Kumar and Sinha (2012, 2013) have shown that RSS could be employed for estimating potato more efficiently than SRS with the same sample size. This investigation could help minimize the exploitation of farmers because a farmer could estimate the total yield of his underground farm produce more accurately than traditional SRS method.

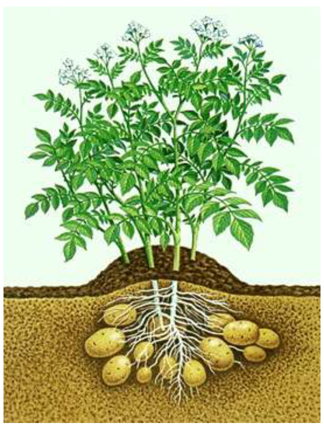

Figure 6. A potato plant

The main results are summarized in Table 8.

TABLE 8 : Relative precision and relative savings under equal and unequal allocation for the set size 4 and the sample size 12 for potato yields

Allocation	Relative Precision	Relative Cost	Relative Savings
Equal Allocation (m=4, r=3, n=12)	1.81	0.55	0.45
Unequal Allocation (m=4, n=12)	1.93	0.52	0.48

The findings show that RSS with unequal allocation performs better than SRS and RSS with equal allocation. But the unequal allocation requires the knowledge of standard deviations of various rank orders for estimating the number of quantifications for each rank order. If this information is not available, then equal allocation needs to be preferred. The illustration recommends that the RSS methods instead of SRS method could be used for estimating farm produces that are grown under the land surface such as potato, ginger, turmeric, garlic, onion, beetroot, peanut, etc.

6.7. Some Briefly Reported Applications

(i) Yangawa and Chen (1980) mentioned that the RSS method was regularly used at the Pastoral Research Laboratory, CSIRO at Armidale, N S W , Australia. Following the RSS procedure a plate with four holes is randomly thrown on a field, the pasture in the four holes is ranked by eye, and a hole is selected for quantification of pasture.

(ii) They also mentioned that the method had been used to estimate rice crops in Okinawa, Japan. They attributed this information to H. Mizuno at the "Mathematical Method in Sampling" symposium held at Chiba University, Japan in 1974.

(iii) Evans (1967) pointed out that the method could save time while determining the cell wall thickness of different species of wood. In the same area of application Dell (1969) mentioned that the RSS procedure could be efficient for estimating average for various properties of cells in a cross section of wood chips.

(iv) The method of sampling could be useful in determining the average length of various kinds of bacterial cells. Also, it could be used to obtain the average number of bacterial cells per unit volume. This is possible because it is convenient to order the tubes containing the cell suspension on the basis of concentration with the help of an optical instrument without knowing the exact number of the bacterial cells. Takahasi and Wakimoto (1968) have suggested these applications. Also, Patil, Gore and Sinha (1992 and 1994) may be consulted for applications in these and some related areas.

(v) The technique may also be used for determining the average height of trees because it is easy to rank the heights of several nearby located trees by a visual perception. The application has been suggesed by Takahasi and Wakimoto (1968).

(vi) For carrying out more rapid assessments of natural resource damage after the occurrence of catastrophic events, Jonson and Myers (1993) proposed ranked set sampling by using the information of remote sensing media for ranking of randomly drawn samples.

(vii) Johnson, Patil and Sinha (1993) presented a review of applications of the RSS method in the area of vegetation as well as they proposed to use indices like greenness, brightness and texture derived from satellite images as concomitant variables for carrying out ranking of vegetation samples.

ACKNOWLEDGEMENT

I would like to thank Prof. M. P Trivedi, Department of Botany, Patna Science College, Patna University, Bihar, India and Prof. M. Roy, Department of Botany, R. N. College, B R A Bihar University, Bihar, India for their comments and suggestions, which improved the contents of the paper considerably.

References

Bickel, P.J., 1967. Some contributions to the theory of order statistics. In: *Proc. 5th Berkley Symp.*, 1, 575-91.

Cobby, J. M., Ridout, M. S., Bassett, P. J., and Large, R. V., 1985. An investigation into the use of ranked set sampling on grass and grass-clover swards. *Grass and Forage Science* 40, 257-263.

Dell, T. R., 1969. *The theory of some applications of ranked set sampling*. Ph. D. Thesis, University of Georgia, Athens, GA, USA.

Dell, T. R., and Clutter, J. L., 1972. Ranked set sampling theory with order statistics background. *Biometrics*, **28**, 545-553.

Evans, M. J., 1967. Applications of ranked set sampling to regeneration surveys in areas direct-seeded to longleaf pine. Master's Thesis, School of Forestry and Wildlife Management, Louisiana State University, Baton Rouge, USA.

Halls and Dell, 1966. Trial of ranked set sampling for forage yields. *Forest Science* **12** (1), 22-26.

Johnson, G. D., Patil, G. P. and Sinha, A. K., 1993. Ranked set sampling for vegetation research. *Abstarcta Botanica*, **17** (1-2), 87-102.

Johnson, G. D., and Myers, W. L., 1993. Potential of ranked set sampling for disaster Assessment. Presented at IUFROS4.02 Conference on *"Inventory and Mangement Techniques in the Context of Catastrophic Events"* June 1993.

Kumar, Vijay and Sinha, Arun K., 2012. Estimating underground farm produce by ranked set sampling. *Patna University Journal of Science* (ISSN 2249-6165), 3, 83-91.

Kumar, Vijay and Sinha, Arun K. (2013). Estimating underground farm produce by ranked set sampling with unequal allocations. (Manuscript)

Martin, W. L., Sharik, T. L., Oderwald, R. G., and Smith, D. W., 1980. Evaluation of ranked set sampling for estimating shrub phytomass in Appalachian oak forests. *Publication Number FWS 4-80, School of Forestry and Wildlife Resources, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, USA.*

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

Mode, N., Conquest, L. and Marker, D., 1999. Ranked set sampling for ecological research: accounting for the total cost of sampling. *Environmetrics*, **10**, 179-194.

Muttlak, H. A. and Al-Saleh, M. F., 2000. Recent developments in ranked set sampling. *Pakistan Journal of Statistics*,16, 269-290.

Norris, R. C., Patil, G. P. and Sinha, A. K.,1995. Estimation of multiple characteristics by ranked set sampling methods. *Coenoses* (The Journal of the International Center for Theoretical and Applied Ecology, Italy), 10 (2-3), 95-111.

Patil, G. P., Gore, S. D. and Sinha, A. K., 1992. Environmental sampling and statistical modeling with examples. In *Proceedings of the 16th International Biometric Conference*, Hamilton, New Zealand, 148-188.

Patil, G. P., Gore, S. D. and Sinha, A. K., 1994. Environmental chemistry, statistical modeling and observational economy. In *Environmental Statistics, Assessment and Forecasting*. C. R. Cothern and N. P. Ross (eds), 57-97, Lewis Publishers, Ann Arbor, USA.

Patil, G. P., Sinha, A. K. and Taillie, C., 1994a. Ranked set sampling: A novel method to accomplish observational economy in environmental studies. *Center for Statistical Ecology and Environmental Statistics, Department of Statistics, Pennsylvania State University, University Park, PA, USA. Technical Report No. 94-0411*.

Patil, G. P., Sinha, A. K. and Taillie, C., 1994b. Ranked set sampling. In: *Handbook of Statistics*, 12 (G. P. Patil and C. R. Rao eds.), 167-200. North-Holland, Elsevier Science B. V., Amsterdam.

Patil, G. P., Sinha, A. K. and Taillie, C., 1994c. Ranked set sampling for multiple characteristics. *International Journal of Ecology and Environmental Sciences* **20**, 357-373.

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.

Platt, W. J., Evans, G. W., and Rathbun, S. L., 1988. The population dynamics of a long-lived conifer. *The American Naturalist* **131**(4), 491-515.

Prodan, M., 1968. Forest biometrics. Pergamon Press, London.

Ridout, M. S., 2002. A note on ranked set sampling for multiple characteristics. Manuscript.

Sengupta, J. M., Chakravarti, I. M. and Sarkar, D., 1951. Experimental survey for the estimation of cinchona yield. *Bulletin of the International Statistical Institute*, **33**, 313-331.

Sinha, A. K., Perez-Abreu, V., Patil, G. P. and Taillie, C., 2001. On the effectiveness of the Takahasi's ranked set sample estimator as compared with the regression estimator. *American Journal of Mathematical and Management Sciences*, 21, Nos. 1&2,145-163.

Sinha, Arun K., 2005. On some recent developments in ranked set sampling. *Bulletin of Informatics and Cybernetics*, **37**, 137-160.

Stokes, S. L., 1976. *An investigation of the consequences of ranked set sampling*. Ph. D. Thesis. Department of Statistics, University of North Caroliina, Chapel Hill, North Carolina, USA.

Stokes, S. L.,1980. Estimation of variance using judgment ordered ranked set samples. *Biometrics*, **36**, 35-42.

Takahasi, K., 1970. Practical note on estimation of population means based on samples stratified by means of ordering. *Annals of the Institute of Statistical Mathematics*, **22**, 421-428.

Takahasi, K. and Futatsuya, M., 1988. Ranked set sampling from a finite population. *Proceedings of the Ann. Inst. Statist. Math.*, **36**, 55-68. (Japanese with English summary).

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

Tukey, J. W., 1958. A problem of Berkson, and minimum variance orderly estimators. *Annals of Mathematical Statistics*, **29**, 588-92.

Yanagawa, T., 2000. K. Wakimoto's contribution to the ranked set sampling….*The Tenth Japan and Korea Joint Conference of Statistics*, 179-184.

Yanagawa, T. and Chen, S. H., 1980. The MG procedure in ranked set sampling. *Journal of Statistical Planning and Inference*, **4**, 33-44.