Ranked Set Sampling (RSS)

Aya Amro
College of Applied Science
Palestine Polytechnic University
Hebron, Palestine
Y_amro@yahoo.com

Dr. Monjed Samoh
College of Applied Science
Palestine Polytechnic University
Hebron, Palestine
monjedsamuh@ppu.edu

Abstract- **Ranked set sampling (RSS) is a method of data collection that makes use of the sampler's judgment of relative sizes of potential sample units. In this research I demonstrate how RSS can be utilized in survey sampling settings and how we can construct a standard ranked set sampling with more representative data than the data obtained using Simple Random Sample (SRS). After that, I consider a case study where RSS is applied on a real data. Finally, I mentioned some of the major applications in the medical field that requires using RSS.**

Keywords-component; RSS; SRS ; Cycle; Judjment Rank; U-Test-

I. INTRODUCTION

Over the span of several decades, many scientists and researchers have sought the accuracy of the data being collected of their experiments in the field of statistics, especially those obtained when doing sampling of the population of interest. Hence, variety of mechanisms have been experienced for obtaining such data and some of these mechanisms have been considered to be the most common sampling techniques, such as, Simple Random Samples (SRS) and Stratifying Sampling . These sampling designs provide a good representation of the population through their obtained data.

With any of these techniques, once the sample items are chosen, the desired measurement(s) is collected from each of the selected items. Usually, taking the actual measurements through doing the laboratory analysis of the sample observations could be costly, destructive, or time consuming. That is, when a new data collecting approach has arisen on the hand of McIntyre (1952). Many years later this approach has been adopted and developed by many scientists to be recently known as Ranked Set Sampling (RSS).

Ranked Set Sampling rather than the other sampling techniques, it has the ability to utilize basic intuitive properties associated with SRS but it also take advantage of additional information available in the population to provide an "artificially Stratified" Sample with more structure that enables us to direct our attention toward the actual measurements of more representative units in the population.

II. RANKED SET SAMPLIG

Many developments have taken place recently for the sake of taking the notion of RSS to a higher level since McIntyre has first proposed the original idea of how this sampling approach could be more effective than other sampling techniques. Some of the developments have made this method applicable in a wider range than originally intended.

THE CONCEPT OF RSS

RSS is a method of collecting data that improves estimation by utilizing the sampler's judgment or auxiliary information about the relative sizes of the sampling units. Prior to quantifying the data, the researcher samples from the population then ranks the samples units based on his or her judgment about their relative sizes of the variable of interest.

RSS is considered as cost-effective sampling method since the judgment of the researcher usually depends on his or her speculation according to the visual inspection rather than the actual measurements. However RSS is said to be more applicable whenever ranking of the sampling can be carried out easily through visual inspection, also, when sampling is much cheaper than the measurements. i.e., there exist certain obtainable concomitant variables.

HOW TO CONSTRUCT AN RSS?

To create an RSS, the sampler first selects m independent simple random samples from the population of interest. Each sample is of size m and is drawn without replacement. Thus, the total initial sample size is m^2. We call each SRS a set. Within each of the m sets, the sampled items are ranked based on the researcher's judgment of their relative sizes. This ranking is performed prior to measuring the variable of interest. Therefore, the researcher must have some method for estimating relative sizes of the variable of interest. The researcher could use visual inspection of the items to form rankings or the value of an auxiliary variable correlated with the variable of interest. After ranking the m items in each of the m sets, a subsample is drawn for measurement. This subsample consists of the smallest ranked unit from the first set, the second-smallest ranked unit from the second set, and so forth, so that the subsample contains m items, each representing a different rank from the sets. The variable of interest is then measured for the subsample. Note that, although m^2 units are sampled initially, only m of them are measured with respect to the variable of interest.

The above procedure describes one cycle of the RSS process. The full experiment then consists of k independent cycles, yielding a total sample size of mk measurements on the variable of interest. These measurements are called the judgment order statistics; we denote the rth judgment order statistic obtained in the ith cycle by $X_{[r]i}$, $r = 1,2,\dots,m$, and $i = 1,2,\dots,k$. This form of RSS is considered balanced, since each rank is represented in the sample by the same number of judgment order statistics from that rank.

As mentioned above, to create ranked sets we must partition the selected first phase sample into sets of equal size. In order to plan an RSS design, we must therefore choose a set size that is typically small, around three or four, to minimize ranking error. Call this set size m, where m is the number of sample units allocated to each set. Now proceed as follows:

Step 1: randomly select m^2 sample units from the population.

Step 2: allocate the m^2 selected units as randomly as possible into m sets, each of size m.

Step 3: without yet knowing any values for the variable of interest, rank the units within each set based on a perception of relative values for this variable. This may be based on personal judgment or done with measurements of a covariate that is correlated with the variable of interest.

Step 4: choose a sample for actual analysis by including the smallest ranked unit in the first set, then, the second smallest ranked unit in the second set, continuing in this form until the largest ranked unit is selected in the last set.

Step 5: repeat steps 1 through 4 for k cycles until the desired sample size, $n = mk$, is obtained for analysis. As an illustration, consider the set size $m = 3$ with $k = 4$ cycles. This situation is illustrated in (Figure 1).

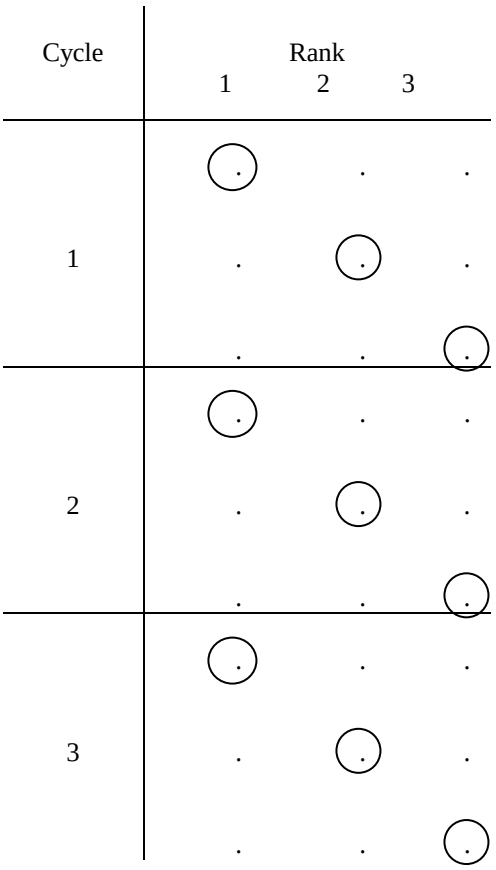

Figure 1: A ranked set sample design with set size $m = 3$ and number of sampling cycles $k = 4$.

(Figure 1) where each row denotes a judgment-ordered sample within a cycle and the units selected for quantitative analysis are circled. Note that 36 units have been randomly selected in four cycles; however, only 12 units are actually analyzed to obtain the ranked set sample of measurements.

Obtaining a sample in this manner maintains the unbiasedness of SRS; however, by incorporating 'outside' information about the sample units, we are able to contribute a structure to the sample that increases its representativeness of the true underlying population. If we quantified the same number of sample units, $mk = 12$ by a simple random sample, then we have no control over which units enter the sample. Perhaps all the 12 units would come from the lower end of the range, or perhaps most would be clustered at the low end while one or two units would come from the middle or upper range. With SRS, the only way to increase the prospect of covering the full range of possible values is to increase the sample size. With RSS, however, we increase the representativeness with a fixed number of sample units, thus saving considerably on quantification costs.

With the ranked set sample thus obtained, it can be shown that unbiased estimators of several important population parameters can be calculated, including the mean and, in case of more than one sampling cycle, the variance.

RSS PROPERTIES (STRUCTURE, MEAN, AND VARIANCE)

In each sampling method there are some of the basic standards such as accuracy, unbiasedness, precision … etc, that defines the method to be effective than other methods. In this section we will have a deeper look to RSS features that distinguish it from other sampling methods.

STRUCTURE OF RSS

Let $X_{[r]i}$ $r = 1, \dots, m, i = 1, \dots, k$ be a ranked set sample of size k obtained as described in Section 2.2 from k independent random samples of k units each. And Let X1, ..., Xk denote a simple random sample of size k from a continuous population with p.d.f. $f(x)$ and c.d.f. $F(x)$.

In the case of a SRS the k observations are independent and each of them is viewed as representing a typical value from the population. However, there is no additional structure imposed on their relationship to one another. While For the RSS setting, additional information and structure is being provided through the judgment ranking process involving a total of k^2 sample units. The k measurements are also order statistics but in this case they are independent observations and each of them provides information about a different aspect of the population.

It is this extra structure provided by the judgment ranking and the independence of the resulting order statistics that enables procedures based on RSS data to be more efficient than comparable procedures based on a SRS with the same number of measured observations. On the other hand, these same features also make the theoretical development of properties for RSS procedures more difficult than for their SRS counterparts.

MEAN AND VARIANCE OF RSS

There are several characteristics of the judgment order statistics that are important to understanding the performance of the estimator of the population mean ($\hat{\mu}$rss). First, the judgment order statistics $X_{[r]i}$, $r = 1,2, \dots, m$, and $i = 1,2, \dots, k$ are mutually independent because they are obtained from independent samples. Second, if we assume that the ranking procedure does not change from cycle to cycle, the judgment order statistics $X_{[r]i}$, $i = 1,2, \dots, k$ are identically distributed for each fixed r. We let f [r], μ [r], and σ 2[r] denote the probability function, mean, and variance, respectively, of $X_{[r]i}$.

Important relationships exist between the distribution of the judgment order statistics and the parent distribution from which they are obtained. Suppose the initial sets used for ranking were sampled from a distribution with probability function f, mean μ, and variance σ 2. Then the ranking partitions the sample space so that

$$\frac{1}{\mathrm{m}} \sum_{r=1}^{m} f_{[r]}(x) = f(x) \qquad (1)$$

by the Law of Total Probability (Dell and Clutter [1]). The ranking procedure partitions the population into m classes, each with probability of selection $1/\mathrm{m}$. The value $f_{[r]}(x)$ is the probability of selecting x given partition r was chosen. This result holds regardless of the accuracy of the rankings. The worst-case scenario is that the rankings are assigned randomly; in this case, or if $m = 1$, $f_{[r]} = f$ for all r.

It follows from Equation (1) that

$$\sum_{r=1}^{m} \mu_{[r]} = m\mu \qquad (2)$$

and

$$\sum_{r=1}^{m} \sigma^2{}_{[r]} = m\sigma^2 - \sum_{r=1}^{m} \left(\mu_{[r]} - \mu\right)^2 \qquad (3)$$

The fact that the $X_{[r]i}$ are identically distributed for fixed r can be used in conjunction with Equation (2) to prove that $\hat{\mu}_{\mathrm{rss}}$ is an unbiased estimator of the population mean μ. This result holds regardless to the accuracy of the rankings.

Independence of the judgment order statistics implies that

$$var(\hat{\mu}_{rss}) = var\left(\left(\frac{1}{mk}\right)\sum_{i=1}^{k}\sum_{r=1}^{m} X_{[r]i}\right) \qquad (4)$$

$$= \frac{1}{(mk)^2}\sum_{i=1}^{k}\sum_{r=1}^{m} var\left(X_{[r]i}\right)$$

$$= \frac{1}{(mk)^2}\sum_{i=1}^{k}\sum_{r=1}^{m} \sigma^2_{[r]}$$

$$= \frac{1}{(mk)^2}\sum_{r=1}^{m} \sigma^2_{[r]}$$

The following inequality (Takahasi and Wakimoto []) holds under perfect ranking

$$\frac{1}{m}\sum_{r=1}^{m} \sigma^2_{[r]} > \frac{1}{m+1}\sum_{r=1}^{m} \sigma^2_{[r]} \qquad (5)$$

As a result,

$$\frac{1}{\mathrm{m}^2}\sum_{\mathrm{r}=1}^{\mathrm{m}} \sigma^2_{[\mathrm{r}]} > \frac{1}{(m+1)^2\mathrm{k}}\sum_{\mathrm{r}=1}^{\mathrm{m}+1} \sigma^2_{[\mathrm{r}]} \qquad (6)$$

Substituting Equation (3) into Equation (4) results in following expression for the variance of $\hat{\mu}_{rss}$:

$$\mathrm{var}(\hat{\mu}_{rss}) = \frac{\sigma^2}{mk} - \frac{\sum_{r=1}^{m}\left(\mu_{[r]} - \mu\right)^2}{m^2 k} \qquad (7)$$

(2.1)

For a simple random sample (SRS) of size $n = mk$, the variance of $\overline{\mathrm{X}}$ is σ^2/mk. The variance of $\hat{\mu}_{rss}$ therefore, is always less than or equal to the variance of $\overline{\mathrm{X}}$ when both estimators are based on the same number of quantified observations. When $m = 1$, we obtain the same variance as for

a simple random sample of size $n = k$. If the rankings are randomly assigned, then each $\mu_{[r]}$ will be very close to the overall mean because the distribution of each judgment order statistic will approximate the parent distribution. As the rankings become more accurate, the term $\sum_{r=1}^{m}(\mu_{[r]} - \mu)^2$ becomes larger, and the overall variance of $\hat{\mu}_{rss}$ decreases.

MANN-WHITNEY-WILCOXON TEST

The Mann-Whitney-Wilcoxon test or as it also called Rank Sum Test is a non-parametric method which is used as an alternative to the two-sample Student's t-test. Usually this test is used to compare medians of non-normal distributions X and Y. Here, we deal with the RSS version of the Mann-Whitney-Wilcoxon (MWW) test for two independent samples X and Y. Also, we consider an application of this test using real data.

Now, let $X_{(1)j}, \dots, Y_{(k)j}, j = 1, \dots, m$ be the ranked set sample (for set size k and m cycles) from a distribution with c.d.f $F(t)$. In addition, let $Y_{(1)t}, \dots, Y_{(q)t}, t = 1, \dots, n$ be the ranked set sample (for set size q and n cycles) from a distribution with c.d.f $G(t) = F(t - \Delta)$, where $-\infty < \Delta < \infty$. Both F and G represent continuous distributions.

Considering the RSS version of MWW test statistics to be as the following:

$$W_{RSS} = \sum_{s=1}^{q}\sum_{t=1}^{n} R_{st} \qquad (8)$$

Where, R_{st} is the rank of $Y_{(s)t}$ in the pooled sample $\{X_{(i)j}, Y_{(s)t} : i = 1, \dots, k, j = 1, \dots, m, s = 1, \dots, q, t = 1, \dots, n\}$.

It can be shown that

$$W_{RSS} = \frac{1}{2}qn(qn+1) + U_{RSS} \qquad (9)$$

Where,

$$U_{RSS} = \sum_{i=1}^{k}\sum_{j=1}^{m}\sum_{s=1}^{q}\sum_{t=1}^{n} I\left(X_{(i)j} < Y_{(s)t}\right). \qquad (10)$$

The test statistic U_{RSS} measures the degree of separation between the two independent samples, and it indicates that the two samples are well separated with little overlap if it has a large value. Whereas, it's small value indicates that the two samples are not well separated with much overlap. To get the value of U_{RSS} the Mann-Whitney test follows a simple pressure of steps as follow:

Step 1: both samples are combined into one array which is sorted in ascending order. We keep information about which sample the element had come from.

Step 2: after sorting, each element is replaced by its rank (its index in array, from 1 to the sum of both samples set size).

Step 3: the ranks of one of the two samples elements are summarized and the U_{RSS} value is calculated.

Now, to conduct hypothesis tests of the null hypothesis $H_0: \Delta = 0$ against either one- or two-sided alternatives, we need some properties of the null distribution of U_{RSS}. For this purpose, we assume that we have perfect judgment rankings for both the X and Y ranked set samples.

EXPECTATION AND VARIANCE OF TWO-SAMPLE U-STATISTICS

It is easy to deal with the properties of W_{RSS} through those of U_{RSS}. What follow are some properties of U_{RSS}. First we have

$$\mathrm{E}(\mathrm{U_{RSS}}) = \mathrm{E}\left(\sum_{\mathrm{i=1}}^{\mathrm{k}}\sum_{\mathrm{j=1}}^{\mathrm{m}}\sum_{\mathrm{s=1}}^{\mathrm{q}}\sum_{\mathrm{t=1}}^{\mathrm{n}} \mathrm{I}\left(\mathrm{X_{(i)j}} < \mathrm{Y_{(s)t}}\right)\right)$$

Now, by using some of the properties of the expectations, and for more simplicity let $M = km$ and $N = qn$, we get (2.8)

$$E(U_{RSS}) = \frac{1}{2}MN + MN\int_{0}^{\Delta}\int_{-\infty}^{\infty} f(y)\, f(y-x)\,dy\,dx. \qquad (11)$$

For the variance of U_{RSS}, it can be verified that (2.9)

$$\sigma^2(U_{RSS}) = var(U_{RSS}) = MN[N\varsigma_{10} + M\varsigma_{01} + \varsigma_{11}] \qquad (12)$$

Where, (2.10)

$$\varsigma_{10} = var\left(G(X)\right) - \frac{1}{k}\sum_{i=1}^{k}\left[EG\left(X_{(i)}\right) - EG(X)\right]^2 \qquad (13)$$

$$\varsigma_{01} = var\left(F(Y)\right) - \frac{1}{q}\sum_{i=1}^{q}\left[EF\left(Y_{(s)}\right) - EF(Y)\right]^2 \qquad (14)$$

$$\varsigma_{11} = \frac{1}{kq}\sum_{i=1}^{k}\sum_{s=1}^{q} var\left(I\left(X_{(i)} > Y_{(s)}\right)\right) - var\left(G_{(s)}\left(X_{(i)}\right)\right) - var\left(F_{(i)}\left(Y_{(s)}\right)\right). \qquad (15)$$

These properties of U_{RSS} are mainly used to determine the nature of the through the median. We can also note that the distribution of U_{RSS} co distribution only in the neighborhood of the central point (2-4 standard

the values of N and M). Outside of this interval, the tails of the distribution go down more quickly than the tails of normal distribution. This means that if we use normal approximation to make a decision having α less than 0.001, we could accept a wrong null hypothesis. That's why all statistical programs don't use normal approximations for small M and N, instead of that U_{RSS} is tabulated for small M and N (from 5 to 15), and for $M, N > 15$ asymptotic approximation is used. This method allows us to get p-values with satisfactory accuracy.

13	29	16
14	28	30
15	17.5	19

Example 1 In this example we consider two Advanced Calculus classes of two different teachers, the first one with long teaching experience and the second one is with no experienc SRS's of the final's scores from the two classes of size 3 then Construct an RSS fr to be X and another RSS from the second one to be Y using 5 cycles to get a total observations for each sample. The data of the two samples are illustrated below.

Sample X	Sample Y
63	28
70	28
60	28
70	35
63	48
75	60
75	60
70	61
81	72
75	75
80	45
81	80
84	71
83	85
72	73

Now, we are interested in testing the following hypothesis:

$H_0: \text{Med}(X) = \text{Med}(Y)$ VS $H_a: \text{Med}(X) > \text{Med}(Y)$

considering the level of significance to be .05.

- We rank the data of each samples to get

	Ranks for	
Count	Sample X	Sample Y
1	11.5	2
2	14	2
3	8	2
4	14	4
5	11.5	6
6	215	8
7	21.5	8
8	14	10
9	26.5	17.5
10	21.5	21.5
11	24.5	5
12	26.5	24.5

- Applying the Mann-Whitney test we get the results as follows:

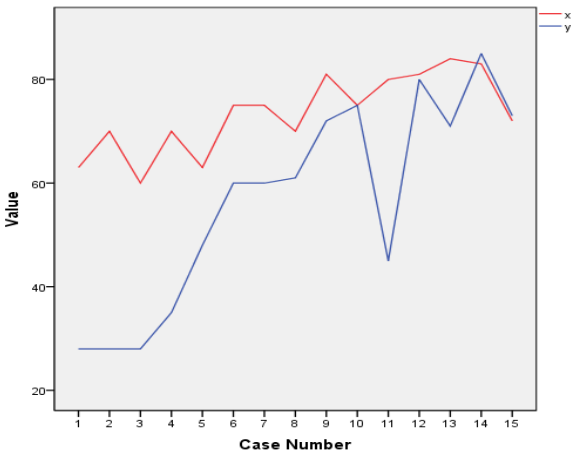

Figure 2: The two samples are well separated with little overlap indi U_{RSS} has a large value.

The value of U_{RSS} is 55.5 since it represents the value of the summatio and in this case it refers to the sample X, Whereas the P_ value of the t less than 0.05, i.e. According to the hypothesis test we accept H_a: Med results lead us to conclude that the class with a long practiced teacher a than the one with a teacher of non experience.

MANN-WHITNEY TEST AND T-TEST

There is a clear similarity between the t-test on independent samples a test as both are testing the identity of two independent populations.

The t test on independent samples is the archetype of the parametric test the two populations are normal with identical variances are essential (s testing the identity of two populations). On the other hand, the M nonparametric: it does not rest on any assumption concerning the underly therefore more widely applicable than the t-test. The price to pay for thi Mann-Whitney test is less powerful than the t-test, because it first conv observations into ranks, and some information is lost in the process. assumptions of the t-test are justified, it can be shown that the Mann-W powerful that the t-test for large samples.
Also, the t-test is sensitive to departures from normality, and to the diffe variances, especially when the sample sizes are different. So whenever th about the validity of these assumptions, the Mann-Whitney test is an exce

III. APPLICATIONS OF RANKED SET SAMPLING

Even though the RSS sampling technique has been neglected for over a decade since it was first proposed by McIntyre in1952, a very wide range of applications have been taken place recently using RSS as one of the most efficient sampling technique in Statistics. In this chapter we discuss two of these applications.

MCINTYRE'S RSS TECHNIQUE

The technique Ranked Set Sampling (RSS) was first introduced by McIntyre as an efficient alternative to simple random sampling for estimating the expected pasture yield. In environmental and ecological sampling, or more generally in special sampling. One may encounter situations where exact measurements of the variable of interest are expensive (in terms of time, money, or other), but where ranking on the basis of visual inspection or on the basis of another highly correlated random variable can be done easily. In McIntyre's case, measuring the plots of pasture yields requires moving and weighting crop yields, which is time consuming. However, a small number of plots can be even though sufficiently well ranked by eye without measurement. McIntyre's goal was to develop a sampling technique to reduce the number of necessary measurements to be made, maintaining the unbiasedness of SRS mean and reducing the variance of the mean estimator by incorporating the outside information provided by visual inspection. Therefore, since the ranking of the plots could be done very cheap, he developed a technique to implement this advantage.

McIntyre's RSS technique did not find further interest for over a decade. Then Halls conducted a field trial by using RSS technique for estimating forage yields in a pine forest. They reported the gain of efficiency by using RSS instead of SRS and mentioned also the practical problems which can arise when using RSS. Halls gave also the name Ranked Set Sampling which is today in use.

THE USAGE OF RSS TECHNIQUE IN THE MEDICAL FIELD

RSS can be used in certain medical studies. For instance, it can be used in the determination of normal ranges of certain medical measures, which usually involves expensive laboratory tests. Samawi considered using RSS for the determination of normal ranges of bilirubin level in blood for new born babies. To establish such ranges, blood sample must be taken from the sampled babies and tested in a laboratory. But, on the other hand, the ranking of the bilirubin levels of a small number of babies can be done by observing whether their face, chest, lower parts of the body and the terminal parts of the whole body are yellowish, since, as the yellowish color goes from face to the terminal parts of the whole body, the level of bilirubin in blood goes higher. RSS also has potential applications in clinical trials. Usually, the cost for a patient to go through a clinical trial is relatively high. However, the patients to be involved in the trial can be selected using the technique of RSS based on their information such as age, weight, height, blood pressure and health history etc., which can be obtained with a relatively negligible cost.

Another example where RSS methodology can be effectively applied in this field, consider the problem of estimation of bone mineral density (BMD) in a human population. Subjects for such a study are plentiful, but measurement of BMD via dual x-ray absorptiometry on the selected subjects is expensive. Thus, it is important to minimize the number of subjects required for such a study without reducing the amount of reliable information obtained about the BMD makeup of the population.

REFERENCES

[1] D. A. Wolfe. Ranked Set Sampling: an Approach to More Efficient Data Collection. Statistical Science, 19:636 – 643, 2004.

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

[3] K. Takahasi and K. Wakimoto. 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, 1968.

[4] M. Samhuh and H. Al-Saleh. The effectiveness of multistage ranked set sampling in stratifying the population. Communication in Statistics – Theory and Methods, 40:1036–1080. 2011.

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

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