Ranked Set Sampling:

an Approach to More Efficient Data Collection

by

Douglas A. Wolfe

Department of Statistics

Ohio State University

Abstract. This paper is intended to provide the reader with an introduction to ranked set sampling, a statistical technique for data collection that generally leads to more efficient estimators than competitors based on simple random samples. Methods for obtaining ranked set samples are described and the structural differences between ranked set samples and simple random samples are discussed. Properties of the sample mean associated with a balanced ranked set sample are developed. A nonparametric ranked set sample estimator of the distribution function is discussed and properties of a ranked set sample analogue of the Mann-Whitney-Wilcoxon statistic are presented.

Key words and phrases: Distribution function estimation, improved precision, Mann-Whitney-Wilcoxon statistic, mean estimation, nonparametric procedures, sampling techniques, structured samples.

1. INTRODUCTION

One of the keys to any statistical inference is that the data involved be obtained via some formal mechanism that enables the experimenter to make valid judgements on the question(s) of interest. One of the most common mechanisms for obtaining such data is that of a simple random sample. Other more structured sampling designs, such as stratified sampling or probability sampling, are also available to help make sure that the obtained data collection provides a good representation of the population of interest. Any such additional structure of this type revolves around how the sample data themselves should be collected in order to provide an informative image of the larger population. With any of these approaches, once the sample items have been chosen the desired measurement(s) is collected from **each** of the selected items.

The concept of ranked set sampling is a recent development that enables one to provide more structure to the collected sample items, although the name is a bit of a misnomer as it is not as much a sampling technique as it is a data measurement technique. This approach to data collection was first proposed by McIntyre (1952) for situations where taking the actual measurements for sample observations is difficult (e.g., costly, destructive, time-consuming), but mechanisms for either informally or

formally ranking a set of sample units is relatively easy and reliable. In particular, McIntyre was interested in improving the precision in estimation of average yield from large plots of arable crops without a substantial increase in the number of fields from which detailed expensive and tedious measurements needed to be collected. For discussions of some of the settings where ranked set sampling techniques have found application, see Patil (1995) and Barnett and Moore (1997).

Since its inception with the paper by McIntyre, a good deal of attention has been devoted to the topic in the statistical literature, particularly over the past fifteen years. Some of this work has been geared toward specific parametric families and some has been developed under minimal nonparametric distributional assumptions. However, many of the important concepts and features of the ranked set sampling methodology transcend the parametric or nonparametric categories. We will structure this paper around these more general features but make a point to illustrate them with nonparametric procedures. We begin with a description of the basic structure leading to collection of a ranked set sample from a single population.

2. OBTAINING A RANKED SET SAMPLE

When we select a simple random sample $X_1,\ldots,X_n$ from a fixed population of interest, what makes resulting statistical inference procedures appropriate is not the fact that each individual measurement in the sample is likely to be representative of the population characteristic, say mean or median, of interest. Rather it is through the concept of sampling distributions of the relevant statistics that we should, "on the

average", obtain a set of sample observations that are truly representative of the entire population. However, in practice we obtain only a single random sample and the "on the average" concept does not help much if the particular population items selected for our sample are, in fact, not really very representative of the entire population. We are simply bound by the statistical inferences for this particular sample that go with the "on the average" concept unless we are willing to increase our sample size and expand the number of sample observations.

There are a number of ways to address the problems associated with obtaining an unrepresentative sample from a population. One method for dealing with this issue is to involve a more structured sampling scheme than simple random sampling. Such approaches include stratified sampling schemes, proportional sampling, and the use of concomitant variables to help in selecting appropriate sampling units for measurement. All of these approaches provide more structured sample data than that resulting from a simple random sample scheme. Note that this additional structure about which items to collect and measure is imposed on our data collection process *prior* to the actual decision, and, as such, is correctly viewed as a sampling technique.

On the other hand, despite the name, ranked set sampling is more a data collection technique rather than simply a more representative sampling scheme. It utilizes the basic intuitive properties associated with simple random samples but it also takes 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 measurement of more representative units in the population. The net

result is a collection of measurements that are more likely to span the range of values in the population than can be guaranteed by virtue of a simple random sample.

We now describe how this additional structure is captured in a single ranked set sample of k measured observations. First, an initial simple random sample of k units from the population is selected and subjected to ordering on the attribute of interest via some ranking process. This judgement ranking can result from a variety of mechanisms, including expert opinion, visual comparisons, or the use of easy-to-obtain auxiliary variables, but it cannot involve actual measurements of the attribute of interest on the sample units. Once this judgement ranking of the k units in our initial random sample has been accomplished, the item judged to be the smallest is included as the first item in our ranked set sample and the attribute of interest will be formally measured on this unit. The remaining k-1 unmeasured units in the first random sample are not considered further. We denote this measurement by $X_{[1]}$, where a square bracket [1] is used instead of the usual round bracket (1) for the smallest order statistic because $X_{[1]}$ is only the smallest judgment ordered item. It may or may not actually have the smallest attribute measurement among our k sampled units. Note that the remaining (other than $X_{[1]}$) units in our first random sample are not considered further in the selection of our ranked set sample or eventual inference about the population. The sole purpose of these other k-1 units is to help select an item for measurement that represents the smaller attribute values in the population.

Following selection of $X_{[1]}$, a second independent random sample of size k is selected from the population and judgement ranked without formal measurement on

the attribute of interest. This time we select the item judged to be the second smallest of the *k* units in this second random sample and include it in our ranked set sample for measurement of the attribute of interest. This second measured observation is denoted by $X_{[2]}$.

From a third independent random sample we select the unit judgement ranked to be the third smallest, $X_{[3]}$, for measurement and inclusion in the ranked set sample. This process is continued until we have selected the unit judgement ranked to be the largest of the k units in the k^{th} random sample, denoted by $X_{[k]}$, for measurement and inclusion in our ranked set sample. This entire process is referred to as a *cycle* and the number of observations in each random sample, *k* in our example, is called the *set size*. Thus to complete a single ranked set cycle, we need to judgment rank k independent random samples of size k involving a total of k^2 sample units in order to obtain k measured observations $X_{[1]}, X_{[2]}, \ldots, X_{[k]}$. These k observations represent a *balanced ranked set sample with set size k*, where the descriptor balance refers to the fact that we have collected one judgement order statistic for each of the ranks i = 1, …, k. In order to obtain a ranked set sample with a desired total number of measured observations km, we repeat the entire cycle process m independent time, yielding the data $X_{[1]j}, \ldots, X_{[k]j}$, for j = 1, …, m.

3. STRUCTURE OF A RANKED SET SAMPLE

To understand what makes the ranked set sample (RSS) different from a simple random sample (SRS) of the same size, we consider the simple case of a single cycle (m

= 1) with set size k and perfect judgement ranking. In this case, the ranked set sample observations are also the respective order statistics. Let $X_1, \ldots, X_k$ denote a simple random sample of size k from a continuous population with p.d.f. f(x) and c.d.f. F(x) and let $X_1^*, \cdots, X_k^*$ be a ranked set sample of size k obtained as described in Section 2 from k independent random samples of k units each.

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. Letting $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(k)}$ be the order statistics associated with these SRS observations, we note that they are dependent random variables with joint p.d.f. given by

$$g_{SRS}(x_{(1)}, \ldots, x_{(k)}) = k! \prod_{i=1}^{k} f(x_{(i)}) I_{\{-\infty < x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(k)} < \infty\}}(x_{(1)}, \ldots, x_{(k)}) \ .$$

For the RSS setting, additional information and structure has been provided through the judgement ranking process involving a total of k^2 sample units. The k measurements $X_{(1)}^*, \ldots, X_{(k)}^*$ 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. The joint p.d.f. for $X_{(1)}^*, \ldots, X_{(k)}^*$ is given by

$$g_{RSS}(x_{(1)}^*, \ldots, x_{(k)}^*) = \prod_{i=1}^{k} f_{(i)}(x_{(i)}^*) \ ,$$

where

$$f_{(i)}(x^*_{(i)}) = \frac{k!}{(i-1)!(k-i)!}[F(x^*_{(i)})]^{i-1}[1-F(x^*_{(i)})]^{k-i}\, f(x^*_{(i)})$$

is the p.d.f. for the i[th] order statistic for a SRS of size k from the population with p.d.f. f(x) and c.d.f. F(x). It is this extra structure provided by the judgement 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. In the next section, we illustrate both of these aspects via comparison of the RSS and SRS sample means.

4. PROPERTIES OF THE SAMPLE MEAN

Let $\overline{X} = \sum_{i=1}^{k} X_i / k$ and $\overline{X}^* = \sum_{i=1}^{k} X^*_{(i)} / k$ be the SRS and RSS sample mean, respectively, for common measured number of observations, k. It is well known that $\overline{X}$ is an unbiased estimator of the population mean, μ, and that it has variance σ^2/k, where σ^2 is the population variance. How does $\overline{X}^*$ compare? First, we note that the mutual independence of the $X^*_{(i)}$'s, i = 1, …, k, enables us to write

$$\mathrm{E}[\overline{X}^*] = \frac{1}{k}\sum_{i=1}^{k} E[X^*_{(i)}] \quad \text{and} \quad Var(\overline{X}^*) = \frac{1}{k^2}\sum_{i=1}^{k} Var(X^*_{(i)}). \qquad (1)$$

Moreover, since we have assumed perfect rankings, $X^*_{(i)}$ is distributed like the i[th] order statistic from a continuous distribution with p.d.f. f(x) and c.d.f. F(x). Hence, we have

$$E[X^*_{(i)}] = \int_{-\infty}^{\infty} x \frac{k!}{(i-1)!(k-i)!} [F(x)]^{i-1} [1-F(x)]^{k-i} f(x)dx\,, \qquad (2)$$

for i = 1, …, k. Combining equations (1) and (2), we obtain

$$E[\overline{X}^*] = \frac{1}{k} \sum_{i=1}^{k} \{ \int_{-\infty}^{\infty} kx \binom{k-1}{i-1} [F(x)]^{i-1} [1-F(x)]^{k-i} f(x)dx \}$$

$$= \int_{-\infty}^{\infty} x\, f(x) \{ \sum_{i=1}^{k} \binom{k-1}{i-1} [F(x)]^{i-1} [1-F(x)]^{k-i} \} dx\,. \qquad (3)$$

Letting q = i - 1 in the summation in equation (3) we see that

$$\sum_{i=1}^{k} \binom{k-1}{i-1} [F(x)]^{i-1} [1-F(x)]^{k-i} = \sum_{q=0}^{k-1} \binom{k-1}{q} [F(x)]^{q} [1-F(x)]^{(k-1)-q} = 1\,,$$

since the latter sum is just the sum over the entire sample space of the probabilities for a binomial random variable with parameters k – 1 and p = F(x).

Using this fact in equation (3) we obtain

$$E[\overline{X}^*] = \int_{-\infty}^{\infty} x f(x)dx = \mu\,.$$

Thus, $\overline{X}^*$ is also an unbiased estimator for μ.

Of course, there is certainly a difference between these unbiased estimators $\overline{X}$ and $\overline{X}^*$. The k components of the SRS average $\overline{X}$ are mutually independent and identically distributed and each is itself an unbiased estimator for μ. While the k components of the RSS average $\overline{X}^*$ are also mutually independent, they are not identically distributed and none of them (except for the middle order statistic when k is odd and the underlying distribution is symmetric about μ) are individually unbiased for μ. Yet the averaging process leaves $\overline{X}^*$ unbiased. Interestingly, it is the additional

structure associated with the non-identical nature of the distributions for the terms in $\bar{X}^*$ that leads to the improvement in precision for $\bar{X}^*$ relative to $\bar{X}$, as we now show.

Letting $\mu^*_{(i)} = E[X^*_{(i)}]$, for i = 1, …, k, we note that

$$E[(X^*_{(i)} - \mu)^2] = E[(X^*_{(i)} - \mu^*_{(i)} + \mu^*_{(i)} - \mu)^2]$$

$$= E[(X^*_{(i)} - \mu^*_{(i)})^2] + (\mu^*_{(i)} - \mu)^2, \qquad (4)$$

since the cross product terms are zero. Combining equations (1) and (4) yields the expression

$$Var(\bar{X}^*) = \frac{1}{k^2}\sum_{i=1}^{k} E[(X^*_{(i)} - \mu)^2] - \frac{1}{k^2}\sum_{i=1}^{k}(\mu^*_{(i)} - \mu)^2. \qquad (5)$$

Now, proceeding as we did with $E[\bar{X}^*]$, we see that

$$\sum_{i=1}^{k} E[(X^*_{(i)} - \mu)^2] = \sum_{i=1}^{k}\int_{-\infty}^{\infty} k(x-\mu)^2 \binom{k-1}{i-1}[F(x)]^{i-1}[1-F(x)]^{k-i} f(x)dx$$

$$= k\int_{-\infty}^{\infty}(x-\mu)^2 f(x)\{\sum_{i=1}^{k}\binom{k-1}{i-1}[F(x)]^{i-1}[1-F(x)]^{k-i}\}dx.$$

Once again using the binomial distribution, the interior sum is equal to 1 and we obtain

$$\sum_{i=1}^{k} E[(X^*_{(i)} - \mu)^2] = k\int_{-\infty}^{\infty}(x-\mu)^2 f(x)dx = k\sigma^2. \qquad (6)$$

Combining equations (5) and (6), it follows that

$$Var(\bar{X}^*) = \frac{1}{k^2}\{k\sigma^2 - \sum_{i=1}^{k}(\mu^*_{(i)} - \mu)^2\}$$

$$= \frac{\sigma^2}{k} - \frac{1}{k^2}\sum_{i=1}^{k}(\mu^*_{(i)} - \mu)^2 = Var(\bar{X}) - \frac{1}{k^2}\sum_{i=1}^{k}(\mu^*_{(i)} - \mu)^2$$

$$\leq Var(\bar{X}), \text{ since } \sum_{i=1}^{k}(\mu^*_{(i)} - \mu)^2 \geq 0.$$

Hence, in the case of perfect rankings not only is $\overline{X}^*$ an unbiased estimator, its variance is always no larger than the variance of the SRS estimator $\overline{X}$ based on the same number of measured observations. In fact, this is a strict inequality unless $\mu^*_{(i)} = \mu$ for all i = 1, …, k, which is the case only if the judgement rankings are purely random.

5. OTHER IMPORTANT ISSUES FOR RANKED SET SAMPLES

All of the earlier discussion in this paper involved a balanced ranked set sample with fixed set size k and perfect judgement rankings. Of course, these factors can clearly affect the performance of ranked set sample estimators and hypothesis tests. In particular, they interact with one another in a variety of ways. For example, remember that each measured ranked set sample observation utilizes additional information obtained from its ranking among k – 1 other units from the population. Clearly this additional information is an increasing function of k so long as there are no errors in our judgement rankings. Thus, with perfect judgement rankings we would want to take our set size k to be as large as economically possible within available funds. However, it is also clear that the likelihood of errors in our judgement rankings is an increasing function of the set size as well; that is, the larger k is the more likely we are to experience errors in our judgement rankings. Thus to select the set size k optimally we need to be able to both model the probabilities of imperfect judgement rankings and then to assess their impact on our RSS statistical procedures. Initial work in modeling imperfect judgement rankings was provided by Bohn and Wolfe (1992). For a nice

general discussion of modeling probabilities of imperfect judgement rankings, the interested reader is referred to Presnell and Bohn (1999).

Even under perfect judgement rankings, the costs of the various components of ranked set sampling, namely, identifying sampling units, ranking of sets of sampling units, and eventual measurement of units selected for inclusion in the ranked set sample all affect the choice of optimal set size k. For a basic discussion of these factors and optimal set size selection, the reader is referred to Nahhas, Wolfe, and Chen (2002).

We have thus far discussed only balanced ranked set samples; that is, ranked set samples where each judgement order statistic, ranging from $X^*_{[1]}$ to $X^*_{[k]}$, is represented once in each cycle. However, for some situations it is quite reasonable to consider unbalanced ranked set samples, where the various judgement order statistics have differential representation in a given cycle (but common from cycle to cycle).

For example, consider an underlying distribution that is unimodal and symmetric about its median θ. Suppose we are interested only in making inferences about θ using ranked set sample data based on an odd set size k. Among all the order statistics for a random sample of set size k, we know that the sample median, $X_{(\frac{k+1}{2})}$, contains the most information about θ. Thus, to estimate θ in this setting, is it natural to consider using the drastically unbalanced ranked set sample where only a single judgement order statistic, the judgement median, $X^*_{[\frac{k+1}{2}]}$, is represented in the RSS and it is measured all k times in each of the cycles. For discussion of the pros and cons of

balanced versus unbalanced RSS in this setting as well as others, see Özturk and Wolfe (2000a, 2000b).

Finally, we note that it might be logically appealing to collect more than a single judgement ordered item from each ranked set of size k. However, it is generally not statistically optimal to do so unless the cost of the judgement ranking is quite large relative to the cost of actual unit measurement. For most settings where RSS is appropriate in the first place, the optimal choice is to collect only a single observation from each ranked set of size k. This is true regardless of whether it is better to collect a balanced or an unbalanced ranked set sample. For more details, see Özturk and Wolfe (2000c).

6. NONPARAMETRIC RANKED SET SAMPLE PROCEDURES

The previous discussion in this paper is broadly applicable to both parametric and nonparametric methodologies. For example, the general property of unbiased for the sample mean discussed in Section 4 is not dependent on the assumption of any particular underlying distribution. (The variance of the RSS mean is, of course, dependent on the underlying distribution through μ, σ^2, and the $\mu_{(i)}$'s, i = 1, …, k.) For the remainder of the paper we concentrate solely on a number of important nonparametric RSS procedures.

6.1. Distribution Function Estimation and Mann-Whitney-Wilcoxon Procedures

Utilization of information obtained from rankings is clearly part and parcel of the ranked set sample concept through the judgement ranking process used to select the specific items for measurement. However, it was not until the seminal paper by Stokes and Sager (1988) that a nonparametric approach was considered for analysis of the RSS measurements themselves. In their paper they considered the use of RSS data to estimate the distribution function of a population.

Let $X^*_{[1]j},\cdots,X^*_{[k]j}$, for j = 1, …, m, be the ranked set sample (for set size k and m cycles) from a distribution with c.d.f. F(t). The natural RSS estimator for F(t) considered by Stokes and Sager (1988) is the empirical c.d.f. for the RSS data, namely,

$$F^*(t)=\frac{1}{mk}\sum_{i=1}^{k}\sum_{j=1}^{m} I_{(-\infty,t]}(X^*_{[i]j}).$$

Stokes and Sager show that $F^*(t)$ is an unbiased estimator of F(t) and that

$$Var(F^*(t))\le Var(\hat{F}(t)) \text{ for all t,} \qquad (7)$$

where $\hat{F}(t)$ is the usual empirical c.d.f. for a SRS of equal size mk. They also show how to use the RSS empirical c.d.f. in conjunction with the Kolmogorov-Smirnov statistic to provide simultaneous confidence bands for the distribution function F(t).

Sparked by the Stokes and Sager (1988) paper, Bohn and Wolfe (1992) initiated the development of distribution-free inference procedures based on ranked set samples.

They used the Stokes and Sager RSS estimate of the distribution function to develop RSS analogs of the Mann-Whitney version of the SRS Mann-Whitney-Wilcoxon two-sample test and estimation procedures.

Once again, let $X^*_{[1]j},\cdots,X^*_{[k]j}$, for j = 1, …, 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 $Y^*_{[1]t},\cdots,Y^*_{[q]t}$, for t = 1, …, n, be the ranked set sample (for set size q and n cycles) from a second distribution with c.d.f. G(t) = F(t-Δ), with $-\infty < \Delta < \infty$. Here we assume that both F and G represent continuous distributions. Let $F^*_{m,k}(t)$ and $G^*_{n,q}(t)$ be the empirical distribution functions for the X and Y ranked set samples, respectively, and let Ψ(t) = 1, 0 as t ≥ , < 0. The RSS version of the Mann-Whitney statistic is given by

$$U_{RSS} = mnkq\int_{-\infty}^{\infty} F^*_{m,k}(t)dG^*_{n,q}(t)$$

$$\sum_{s=1}^{q}\sum_{t=1}^{n}\sum_{i=1}^{k}\sum_{j=1}^{m}\Psi(Y^*_{[s]t} - X^*_{[i]j}) = (\,\#X's \le Y's \text{ in the RSS data}). \qquad (8)$$

To conduct hypothesis tests of the null hypothesis H_0: Δ = 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 judgement rankings for both the X and Y ranked set samples. Bohn and Wolfe showed that just as for the SRS setting, the RSS Mann-Whitney statistics U_{RSS} (with perfect rankings) is distribution-free under H_0 over the entire class of continuous distributions F. However, there is a major difference in the null distributions and how critical values are obtained for the two settings. For the SRS setting, the mk + nq combined sample X and Y observations are not only mutually

independent but they are also identically distributed. Thus it suffices to look at each of the $\binom{mk+nq}{mk}$ distinct (i. e., unchanged by permutations within the X's and Y's separately) ordered arrangements of these combined sample observations and, moreover, they are all equally likely. This makes tabulation of the associated null distribution for the SRS Mann-Whitney statistic relatively straightforward. However, the equally likely nature of these arrangements does not carry over to the RSS setting, due to the fact that the ranked set X's and Y's, while still mutually independent, are no longer identically distributed. For example, even in the case of perfect rankings there is nothing to prevent the smallest ordered item from one ranked set from being larger than the largest item from a second ranked set. While this probability will generally be small, it will not be zero as in the case of SRS. This means that for RSS data it is no longer sufficient to look at the $\binom{mk+nq}{mk}$ distinct (i. e., unchanged by permutations within the X's and Y's separately) ordered arrangements of the combined sample observations. Instead we need to calculate the probability of each of the (mk+nq)! permutations separately and then combine them to obtain the null distribution for U_{RSS}. Fortunately the probabilities of these (mk+nq)! permutations under RSS still do not depend on the form of the common, continuous $F \equiv G$ under H_0, although the tabulation can be tedious. We illustrate the necessary computations with a small example.

Example 1. For a single X and Y cycle (i.e., m = n = 1) and common X and Y set size k = q = 2, we must obtain the null probabilities for the 4! = 24 different permutations.

Under the assumption of perfect judgement rankings, the RSS observations $X_{(1)1}$, $X_{(2)1}$, $Y_{(1)1}$, and $Y_{(2)1}$ are independent order statistics with joint p.d.f. given by

$$g_{RSS}(x_{(1)}, x_{(2)}, y_{(1)}, y_{(2)}) = \{\prod_{i=1}^{2} \frac{2!}{(i-1)!(2-i)!}[F(x_{(i)})]^{i-1}[1-F(x_{(i)})]^{2-i} f(x_{(i)})\}$$

$$\times\{\prod_{s=1}^{2} \frac{2!}{(s-1)!(2-s)!}[F(y_{(s)})]^{s-1}[1-F(y_{(s)})]^{2-s} f(y_{(s)})\},$$

which simplifies to

$$g_{RSS}(x_{(1)}, x_{(2)}, y_{(1)}, y_{(2)}) = 16[1-F(x_{(1)})][F(x_{(2)})][1-F(y_{(1)})][F(y_{(2)})]\prod_{i=1}^{2} f(x_{(i)})\prod_{s=1}^{2} f(y_{(s)}) \ .$$

Using this expression for g_{RSS} and straightforward integration, the null probabilities for each of the 4! = 24 permutations of $X_{(1)1}$, $X_{(2)1}$, $Y_{(1)1}$, and $Y_{(2)1}$ can then be computed by integrating over the appropriate region. Thus, for example, the four permutations $\{X_{(1)1} < Y_{(1)1} < X_{(2)1} < Y_{(2)1}\}$, $\{X_{(1)1} < Y_{(1)1} < Y_{(2)1} < X_{(2)1}\}$, $\{Y_{(1)1} < X_{(1)1} < Y_{(2)1} < X_{(2)1}\}$, and $\{Y_{(1)1} < X_{(1)1} < X_{(2)1} < Y_{(2)1}\}$ all have the same null probability of occurrence, p, given by

$$p = \int_{-\infty}^{\infty}\int_{-\infty}^{y_{(2)}}\int_{-\infty}^{x_{(2)}}\int_{-\infty}^{y_{(1)}} g_{RSS}(x_{(1)}, x_{(2)}, y_{(1)}, y_{(2)})\, dx_{(1)} dy_{(1)} dx_{(2)} dy_{(2)} = 41/280.$$

Proceeding in this fashion for all 24 permutations yields the set of null probabilities (independent of the form of the continuous common F) and associated values of U_{RSS} given in Table 1. Combining the null probabilities for the various permutations with the associated values for U_{RSS}, we see that the null distribution of U_{RSS} is given by:

$$P_0(\ U_{RSS} = 0\ ) = P_0(\ URSS = 4\ ) = 1/10,$$

$$P_0(\ U_{RSS} = 1\ ) = P_0(\ U_{RSS} = 3\ ) = 17/90, \text{ and } \ P_0(\ U_{RSS} = 2\ ) = 19/45\ .$$

Note that the null distribution is symmetric about its mean $E_0(U_{RSS}) = mnkq/2 = 2$. This symmetry property holds for the null distribution of U_{RSS} for any (m, n, k, q) configuration.

Just as for the SRS setting, the theoretical properties of the RSS Mann-Whitney statistic U_{RSS} are obtained by using standard results about the general class of U-statistics. (See Randles and Wolfe, 1979, for a discussion of U-statistics.) Let

$$\gamma = \sum_{i=1}^{k}\sum_{s=1}^{q} P(X_{(i)1} < Y_{(s)1}) \ .$$

Then γ is a two-sample, multivariate, estimable parameter of degree (1, 1) and U_{RSS}/mn is the multivariate U-statistic estimator for γ. Standard U-statistic arguments can then be used to establish the following result.

Result 1. Let N = m + n and set $\lambda = \lim_{N\to\infty}(m/N)$. If $0 < \lambda < 1$ and $\lim_{N\to\infty} \frac{N}{m^2 n^2} Var(U_{RSS}) > 0$, then $\frac{\sqrt{N}}{mn}(U_{RSS} - E[U_{RSS}])$ has an asymptotic ($N \to \infty$) normal distribution with mean 0 and finite variance σ_∞^2. An expression for σ_∞^2 can be found in equation (3.3) in Bohn and Wolfe (1992). Under the null hypothesis H_0: $\Delta = 0$ we have $E[U_{RSS}] = mknq/2$ and the asymptotic variance, σ_∞^2, does not depend on the form of the underlying continuous F.

Table 1. Null Probabilities and Values of U_{RSS} for the 24 permutations in a RSS with m = n = 1 and k = q = 2.

Permutation	Null Probability	Value of U_{RSS}
$y_{(2)} < y_{(1)} < x_{(2)} < x_{(1)}$	17/2520	0

$y_{(2)} < y_{(1)} < x_{(1)} < x_{(2)}$	7/360	0
$y_{(1)} < y_{(2)} < x_{(1)} < x_{(2)}$	137/2520	0
$y_{(1)} < y_{(2)} < x_{(2)} < x_{(1)}$	7/360	0
$y_{(1)} < x_{(1)} < y_{(2)} < x_{(2)}$	41/280	1
$y_{(1)} < x_{(2)} < y_{(2)} < x_{(1)}$	7/360	1
$y_{(2)} < x_{(1)} < y_{(1)} < x_{(2)}$	7/360	1
$y_{(2)} < x_{(2)} < y_{(1)} < x_{(1)}$	1/280	1
$x_{(1)} < y_{(1)} < y_{(2)} < x_{(2)}$	41/280	2
$x_{(1)} < y_{(2)} < y_{(1)} < x_{(2)}$	137/2520	2
$x_{(2)} < y_{(1)} < y_{(2)} < x_{(1)}$	17/2520	2
$x_{(2)} < y_{(2)} < y_{(1)} < x_{(1)}$	1/280	2
$y_{(1)} < x_{(1)} < x_{(2)} < y_{(2)}$	41/280	2
$y_{(1)} < x_{(2)} < x_{(1)} < y_{(2)}$	137/2520	2
$y_{(2)} < x_{(1)} < x_{(2)} < y_{(1)}$	17/2520	2
$y_{(2)} < x_{(2)} < x_{(1)} < y_{(1)}$	1/280	2
$x_{(1)} < y_{(1)} < x_{(2)} < y_{(2)}$	41/280	3
$x_{(1)} < y_{(2)} < x_{(2)} < y_{(1)}$	7/360	3
$x_{(2)} < y_{(1)} < x_{(1)} < y_{(2)}$	7/360	3
$x_{(2)} < y_{(2)} < x_{(1)} < y_{(1)}$	1/280	3
$x_{(1)} < x_{(2)} < y_{(1)} < y_{(2)}$	137/2520	4
$x_{(1)} < x_{(2)} < y_{(2)} < y_{(1)}$	7/360	4
$x_{(2)} < x_{(1)} < y_{(1)} < y_{(2)}$	7/360	4
$x_{(2)} < x_{(1)} < y_{(2)} < y_{(1)}$	17/2520	4

For given values of k and q, Result 1 can be used to provide approximate critical values for the test of H_0: $\Delta = 0$ based on U_{RSS}. For example, in the special case of m = n

(so that $\lambda = 1/2$) and k = q = 2, it follows from Bohn and Wolfe (1992) that $\sigma_\infty^2 = 16/9$, so that the asymptotic ($N \to \infty$) null distribution of

$\frac{\sqrt{N}}{mn}(U_{RSS} - E_0[U_{RSS}]) = \sqrt{2n}(\frac{U_{RSS}}{n^2} - 2)$ is N(0,16/9). Thus it follows that $P\{\sqrt{2n}(\frac{U_{RSS}}{n^2} - 2) \geq z_{(\alpha)}\} \approx \alpha$, where $z_{(\alpha)}$ is the upper α^{th} percentile for the standard normal distribution. The approximate upper α^{th} percentile for the null distribution of U_{RSS} is then given by $\frac{n^{3/2}}{\sqrt{2}} z_{(\alpha)} + 2n^2$ for the setting k = q = 2.

Bohn and Wolfe (1992) also provided a point estimator and confidence intervals and bounds for Δ associated with the RSS Mann-Whitney statistic U_{RSS}. In addition, they studied the asymptotic ($N \to \infty$) relative efficiency (ARE) of inference procedures based on U_{RSS} relative to the analogous procedures based on the SRS Mann-Whitney statistic U_{SRS}.

In a follow-up paper, Bohn and Wolfe (1994) showed that the statistic U_{RSS} is no longer distribution-free under the null hypothesis H_0: $\Delta = 0$ when the judgement rankings are not perfect. Using an approximate expected spacings model, they studied the effect that imperfect rankings have on the properties of the inferential procedures based on U_{RSS}.

6.2. Other Nonparametric Procedures

Similar properties have been developed for nonparametric RSS procedures in a number of other settings. Bohn (1996) provides a nice review article that summarizes

the early work on such methodology. Specifically, Hettmansperger (1995) and Koti and Babu (1996) discuss inferences associated with the RSS analog of the sign statistic. Bohn (1998) provides similar results for the RSS version of the signed rank statistic. As with the RSS version of the Mann-Whitney statistic, much of the methodological development for both the RSS sign and signed rank statistics relies on multivariate U-statistic theory. Presnell and Bohn (1999) generalize these results to the entire class of RSS U-statistics.

7. APPLICATIONS OF RSS PROCEDURES

Applications of RSS methodology involve several components. First, there is the initial process of obtaining the sets of SRSs for judgment ranking. Any standard approach for obtaining SRSs can be used for this step. Next there is the process of obtaining the judgment rankings themselves within each of these SRSs. A variety of mechanisms have been proposed for this purpose ranging from totally subjective rankings by experts in the field to the purely objective use of multiple regression or logistic regression based on concomitant variables. Standard software packages can be used for these regressions. Finally, there is the analysis of the RSS data once obtained. At least thus far in its development the statistical analysis of RSS data has been consistently the same as what is standard for analogous SRS data. While this may change as RSS methodology progresses, at this point in time standard software packages are sufficient to analyze RSS data once it has been collected.

As an example where RSS methodology can be effectively applied, 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. Nahhas, Wolfe, and Chen (2002) discuss the selection of an optimal RSS set size for such an application in collaboration with Dr. Velimir Matkovic, a researcher in the Bone and Mineral Metabolism Laboratory at The Ohio State University.

ACKNOWLEDGEMENTS

The author thanks the referees for helpful comments on an earlier draft of this article.

REFERENCES

BARNETT, V. and MOORE, K. (1997). Best linear unbiased estimates in ranked-set sampling with particular reference to imperfect ordering. *Journal of Applied Statistics* **24** 697-710.

BOHN, L. L. (1996). A review of nonparametric ranked-set sampling methodology. *Communications in Statistics Theory and Methods* **25(11)** 2675-2685.

BOHN, L. L. (1998). A ranked-set sample signed-rank statistic. *Journal of Nonparametric Statistics* **9** 295-306.

BOHN, L. L. and WOLFE, D. A. (1992). Nonparametric two-sample procedures for ranked-set samples data. *Journal of the American Statistical Association* **87** 552-561.

BOHN, L. L. and WOLFE, D. A. (1994). The effect of imperfect judgment rankings on properties of procedures based on the ranked-set samples analog of the Mann-Whitney-Wilcoxon statistic. *Journal of the American Statistical Association* **89** 168-176.

HETTMANSPERGER, T. P. (1995). The ranked-set sample sign test. *Journal of Nonparametric Statistics* **4** 263-270.

KOTI, K. M. and BABU, G. J. (1996). Sign test for ranked-set sampling. *Communications in Statistics Theory and Methods* **25(7)** 1617-1630.

MCINTYRE, G. A. (1952). A method for unbiased selective sampling, using ranked sets. *Australian Journal of Agricultural Research* **2** 385-390.

NAHHAS, R. W., WOLFE, D. A. and CHEN, H. (2002). Ranked set sampling: cost and optimal set size. *Biometrics* **58** 964-971.

ÖZTURK, Ö. and WOLFE, D. A. (2000a). Optimal allocation procedure in ranked set sampling for unimodal and multi-modal distributions. *Environmental and Ecological Statistics* **7** 343-356.

ÖZTURK, Ö. and WOLFE, D. A. (2000b). An improved ranked set two-sample Mann-Whitney-Wilcoxon test. *The Canadian Journal of Statistics* **28** 123-135.

ÖZTURK, Ö. and WOLFE, D. A. (2000c). Alternative ranked set sampling protocols for the sign test. *Statistics and Probability Letters* **47** 15-23.

PATIL, G. P. (1995). Editorial: ranked set sampling. *Environmental and Ecological Statistics* **2** 271-285.

PRESNELL, B. and BOHN, L. L. (1999). U-statistics and imperfect ranking in ranked set sampling. *Journal of Nonparametric Statistics* **10** 111-126.

RANDLES, R. H. and WOLFE, D. A. (1979). *Introduction to the Theory of Nonparametric Statistics*. John Wiley and Sons, Inc., New York.

STOKES, S. L. and SAGER, T. W. (1988). Characterization of a ranked-set sample with application to estimating distribution functions. *Journal of the American Statistical Association* **83** 374-381.