Ranked Set Sampling

RATIONALE AND HISTORICAL DEVELOPMENT

The goal of any scientific inquiry is to obtain enough information from the study to reach informed conclusions about the questions of interest. Sampling schemes designed to elicit this information in a reliable format are critical to the process. Of course, it is also critical to take into account the cost involved in the sampling procedure.

Statistical inference procedures based on Simple Random Samples (SRS) are justified through the concept of sampling distributions of the relevant statistics.

"On average", we should obtain a set of SRS observations that are truly representative of the population and, therefore, the associated statistics are reliable representatives of their population counterparts.

However, nothing in this process requires that the units in any *specific* SRS that we might collect are truly representative of the full population and, thereby, of the parameter of interest. Are there ways to mitigate this possibility?

Avoiding the Possibility of an Unrepresentative SRS

Stratified Sampling

Cluster Sampling

Proportional Sampling

All three of these approaches use additional information about the population to *a priori* partition it into more homogeneous subgroups that are designed to more fully cover the entire population. SRSs are then collected independently from these subgroups to form a more structured overall set of sample data that by design will more likely be representative of the entire population

Ranked set sampling (RSS) also uses additional information from the population. In RSS, however, this additional information is not used to partition the full population prior to the collection of appropriate simple random samples. Rather, in RSS, potential simple random samples are selected directly from the full population and then auxiliary population information is used to impose an "artificially post-stratified" structure that enables us to collect measurements from units that are more representative of the full spectrum of values in the population.

The concept of RSS was first proposed by McIntyre (1952) (reprinted in 2005) 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 with regard to the aspect of interest are relatively easy and reliable. Takahasi and Wakimoto (1968) and Dell and Clutter (1972) were the first to provide some of the basic properties for statistical procedures based on RSS data.

WHAT IS A RSS AND HOW DO WE COLLECT IT?

To obtain a RSS of k observations from a population, we proceed as follows. First, an initial SRS of k units is selected from the population and rank ordered on the attribute of interest. A variety of mechanisms can be used to obtain this ranking, including visual comparisons, expert opinion, or through the use of auxiliary variables, but it cannot involve actual measurements of the attribute of interest on the sample units. The unit that is judged to be the smallest in this ranking is included as the first item in the RSS and the attribute of interest is formally measured for the unit. This initial measurement is called the first judgment order statistic and is denoted by $X_{[1]}$, where a square bracket is used instead of the usual round bracket (1) for the smallest order statistic because $X_{[1]}$ may or may not actually have the smallest attribute measurement among the k units in the SRS, even though our ranking judged it to be the smallest. The remaining k-1 units (other than $X_{[1]}$) in our initial SRS are not considered further in making inferences about the population—their role was solely to assist in the selection of the smallest ranked unit for measurement.

Following the selection of $X_{[1]}$, a second SRS (independent of the first SRS) of size k is selected from the population and ranked in the same manner as the first SRS. From this second SRS we select the item ranked as the second smallest of the k units (i.e., the second judgment order statistic) and add its attribute measurement, $X_{[2]}$, to the RSS. From a third SRS (independent of both previous SRS's) of size k we select the unit ranked to be the third smallest (i.e., the third judgment order statistic) and include its attribute measurement, $X_{[3]}$, in the RSS.

This process continues until we have selected the unit ranked to be the largest of the k units in the k^{th} independent SRS and included its attribute measurement, $X_{[k]}$, in our RSS.

This results in the k measured observations $X_{[1]}, \ldots, X_{[k]}$ and is called a *cycle*. The number of units, k, in each SRS is called the *set size*. Thus to complete a single ranked set cycle, we need to use a total of k^2 units from the population to separately rank k independent simple random samples of size k each. The measured observations $X_{[1]}, \ldots, X_{[k]}$ constitute a *balanced ranked set sample of size k*, where the descriptor "balanced" refers to the fact that we have collected one judgment order statistic for each of the ranks 1, 2, …, k.

To obtain a balanced RSS with a desired total number of measured observations (i. e., sample size) $n = km$, we repeat the entire process for m independent cycles, yielding the following balanced RSS of size n:

Cycle 1	$X_{[1]1}$	$X_{[2]1}$	$X_{[3]1}$	…	$X_{[k]1}$
Cycle 2	$X_{[1]2}$	$X_{[2]2}$	$X_{[3]2}$	…	$X_{[k]2}$
⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮
Cycle m	$X_{[1]m}$	$X_{[2]m}$	$X_{[3]m}$	…	$X_{[k]m}$

Example — Gasoline Reid Vapor Pressure

Unburned hydrocarbons emitted from automobile tailpipes and via evaporation from manifolds are among the primary contributors to ground level ozone and smog levels in large cities. One way to reduce the effect of this factor on air pollution is through the use of reformulated gasoline, designed to reduce its volatility, as measured by the Reid Vapor Pressure (RVP) value. To monitor compliance, regular samples of reformulated gasoline from pumps in metropolitan areas are collected and RVP values are measured.

The RVP value for a sample can either be measured by a crude field technique right after collection at the gasoline pump or via a more sophisticated analysis after the sample has been shipped to a government laboratory. While the laboratory analysis is not overly expensive, it is costly to ship these gasoline samples to the laboratory, since they are flammable and must be packed to prevent gaseous hydrocarbons from escaping en route. It would be beneficial to use these cruder, less expensive, field RVP measurements as reliable surrogates for the more expensive laboratory RVP measurements to reduce the required number of formal laboratory tests without significant loss of accuracy.

Nussbaum and Sinha (1997) suggested the use of RSS as an aid in achieving this goal. Thirty-six of the field RVP measurements (collected at the pumps) considered by Nussbaum and Sinha are given in the following table.

Sample Number	Field RVP Value	Sample Number	Field RVP Value
1	7.60	19	7.85
2	9.25	20	7.86
3	7.73	21	7.92
4	7.88	22	7.95
5	8.89	23	7.85
6	8.88	24	7.95
7	9.14	25	7.98
8	9.15	26	7.80
9	8.25	27	7.80
10	8.98	28	8.01
11	8.63	29	7.96
12	8.62	30	7.86
13	7.90	31	8.89
14	8.01	32	7.89
15	8.28	33	7.73
16	8.25	34	9.21
17	8.17	35	8.01
18	10.72	36	8.32

Source: B. D. Nussbaum and B. K. Sinha (1997).

Nussbaum and Sinha recommended using these field RVP values to provide the ranking mechanism for selection of a much smaller subgroup of gasoline samples to submit for formal laboratory analysis. They used a set size of $k = 3$, which leads to a RSS of only $n = 12$ gasoline samples to send for full laboratory RVP measurement.

To select this RSS, using a set size $k = 3$, the first thing we do is to randomly divide the 36 gasoline samples into twelve sets of three each, such as the following:

(10, 13, 23)	(11, 12, 17)	(16, 2, 21)	(15, 18, 36)	(34, 27, 5)	(24, 31, 14)
(22, 35, 19)	(30, 9, 4)	(28, 32, 8)	(33, 26, 29)	(7, 1, 6)	(3, 20, 25)

Next, we decide which four sets will be used to obtain the smallest judgment ordered units, which four will be used to obtain the median judgment ordered units, and which four will be used to obtain the largest judgment ordered units. There is complete flexibility here, but these decisions must be made without knowledge of the actual field RVP values in the twelve sets. For sake of illustration here, we choose to select the minimum judgment ordered unit from the first four sets, the median judgment ordered unit from the second four sets, and the largest judgment ordered unit from the final four sets.

The resulting twelve sets of three RVP values each are given in the following table:

8.98	7.90	7.85	8.63	8.62	8.17	8.25	9.25	7.92
8.28	10.72	8.32	9.21	7.80	8.89	7.95	8.89	8.01
7.95	8.01	7.85	7.86	8.25	7.88	8.01	7.89	9.15
7.73	7.80	7.96	9.14	7.60	8.88	7.73	7.86	7.98

Using our chosen RSS criteria, the units selected for shipment to the laboratory for precise RVP measurements are those gasoline samples corresponding to the bolded, enlarged field RVP values in the following table:

8.98	7.90	**7.85**	8.63	8.62	**8.17**	8.25	9.25	**7.92**
8.28	10.72	8.32	9.21	7.80	**8.89**	7.95	8.89	**8.01**
7.95	8.01	7.85	7.86	8.25	**7.88**	8.01	7.89	**9.15**
7.73	7.80	**7.96**	**9.14**	7.60	8.88	7.73	7.86	**7.98**

Thus, we will send gasoline samples 23, 17, 21, 15, 5, 14, 22, 4, 8, 29, 7, and 25 to the laboratory for more precise RVP determinations to obtain our balanced RSS of 12 more representative RVP values based on a set size $k = 3$.

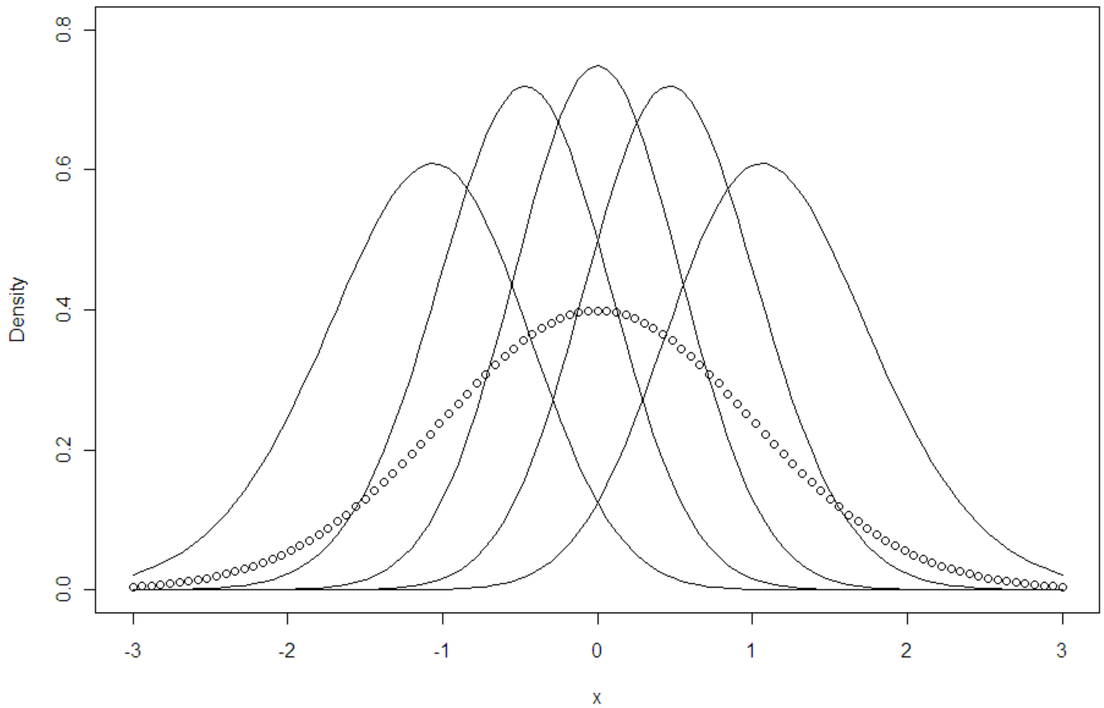
Comparison of RSSs and SRSs

A balanced RSS of size n differs from an SRS of the same size in a number of important ways.

The n observations in an SRS are mutually independent and identically distributed. Probabilistically speaking, each of the individual sample items represents a typical value chosen from the underlying population, and, collectively, we expect them to provide a mirror of the entire population, but we have no way of knowing how well our particular sample does this.

On the other hand, while the individual observations in a balanced RSS remain mutually independent, they are clearly not identically distributed. In fact, the individual judgment order statistics represent very distinctly different portions of the underlying population, as the items in the sample are designed to provide greater assurance that the entire range of population values are represented.

This is best illustrated by considering an example. Suppose that X has a standard normal distribution and let $X_1, X_2, \ldots, X_5$ be a random sample of size five from this distribution. Let $X_{(1)} \leq X_{(2)} \leq X_{(3)} \leq X_{(4)} \leq X_{(5)}$ be the associated order statistics. In the following Figure, we plot the underlying N(0,1) density as well as the **marginal distributions** for the five individual order statistics $X_{(1)}$, $X_{(2)}$, $X_{(3)}$, $X_{(4)}$, and $X_{(5)}$.



Standard normal density (dotted curve) and the individual marginal densities of the five order statistics $X_{(1)}$, $X_{(2)}$, $X_{(3)}$, $X_{(4)}$, and $X_{(5)}$ (solid curves, in order of peaks, from the minimum, $X_{(1)}$, on the left to the maximum, $X_{(5)}$, on the right) for a random sample of size five from the standard normal distribution.

If we use perfect rankings to collect a RSS of size five from the standard normal distribution, then these five RSS observations behave like **mutually independent** order statistics from the standard normal and their densities are represented by the five individual marginal density curves in the previous Figure. While these five densities certainly overlap, they assign the bulk of their individual marginal probabilities to five sub-regions of the standard normal domain. As a result, the five RSS observations are much more likely to represent the full range of values for the standard normal distribution than would a SRS of size five; This feature enables RSS to be more effective than SRS in estimation of a population mean. For ease of discussion, we assume throughout this discussion that all sampling is from an infinite population or with replacement from a finite population.

RSS ESTIMATION OF A POPULATION MEAN

Consider two sets of n observations each from a population. One set of n observations, $X_1, \ldots, X_n$, is collected as an SRS and the second independent set of n observations is collected as a balanced RSS, corresponding to set size k and m cycles, with $n = km$. The RSS observations from cycle 1 are denoted by $X_{[1]1}, \ldots, X_{[k]1}$, the RSS observations from cycle 2 are denoted by $X_{[1]2}, \ldots, X_{[k]2}$, and the RSS observations from the final cycle m are denoted by $X_{[1]m}, \ldots, X_{[k]m}$.

The natural ranked set sample estimator, $\hat{\mu}_{RSS}$, for the population mean μ is simply the average of the sample observations, namely,

$$\hat{\mu}_{RSS} = \overline{X}_{RSS} = \sum_{j=1}^{m} \sum_{i=1}^{k} X_{[i]j} / km .$$

Properties of $\hat{\mu}_{RSS}$

For simplicity in our discussion, we consider the case of only a single cycle (m = 1), so that the sample size n is equal to the set size k. A similar comparison holds for an arbitrary number of cycles. Dell and Clutter (1972) showed that $\hat{\mu}_{RSS}$ is an unbiased estimator for the population mean μ regardless of whether the judgment rankings are perfect or imperfect. In addition, the variance of $\hat{\mu}_{RSS}$ is given by

$$\begin{aligned} Var(\bar{X}_{RSS}) &= \frac{1}{k^2}\{k\sigma^2 - \sum_{i=1}^{k}(\mu_{[i]}^* - \mu)^2\} \\ &= \frac{\sigma^2}{k} - \frac{1}{k^2}\sum_{i=1}^{k}(\mu_{[i]}^* - \mu)^2, \end{aligned}$$

where $\mu_{[i]}^* = E[X_{[i]}]$.

Comparison of SRS and RSS Mean Estimators

The SRS estimator for the population mean μ is just the sample mean $\hat{\mu}_{SRS} = \bar{X} = \frac{1}{k}\sum_{j=1}^{k} X_j$ and it is well known that $E[\hat{\mu}_{SRS}] = \mu$ and $Var(\hat{\mu}_{SRS}) = \frac{\sigma^2}{k}$. Thus, both $\hat{\mu}_{SRS}$ and $\hat{\mu}_{RSS}$ are unbiased estimators for the population mean. Moreover, from the previous expression, it follows that

$$Var(\bar{X}_{RSS}) = \frac{\sigma^2}{k} - \frac{1}{k^2}\sum_{i=1}^{k}(\mu_{[i]}^* - \mu)^2 = Var(\bar{X}) - \frac{1}{k^2}\sum_{i=1}^{k}(\mu_{[i]}^* - \mu)^2 \le Var(\bar{X}),$$

since $\sum_{i=1}^{k}(\mu_{[i]}^* - \mu)^2 \ge 0$. Note that greater differences between the $\mu_{[i]}^*{}'s$ and the overall mean μ lead to improved precision from using RSS.

Hence, not only is $\bar{X}_{RSS}$ an unbiased estimator, its variance is always no larger than the variance of the SRS estimator $\bar{X}$ 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 judgment rankings are purely random.

RSS ESTIMATION OF OTHER PARAMETERS

Estimation of the Population Variance

Stokes (1980) and MacEachern et al. (2002) and, more recently, Perron and Sinha (2004) and Zamanzade and Vock (2015), among others, showed that RSS estimators for the population variance are more efficient than their SRS counterpart, although the gain in efficiency is not as great as in the case of estimation of the population mean.

Estimation of the Population Distribution Function

Stokes and Sager (1988) initially studied the use of RSS data to estimate the distribution function $F(\mathrm{t})$ of a population and showed that it was more efficient than its SRS counterpart. Kvam and Samaniego (1993, 1994) considered competitor estimators that allowed for differential weightings of the RSS observations, which automatically accommodated unbalanced RSS as well.

Estimation of a Population Proportion

For populations consisting of binary data corresponding to 'success' or 'failure', the feature of interest is the proportion, *p*, of 'successes' in the population. One natural RSS estimator for *p* is simply the sample percentage of successes, $\hat{p}_{RSS}$, in the RSS. However, this naive estimator does not fully utilize the additional information incorporated in the RSS via the ranking process; that is, unlike a SRS, not all 'successes' in a RSS should be treated equally.

Taking this into account, Terpstra (2004) developed the RSS maximum likelihood estimator, $\hat{p}_{RSS.MLE}$, for *p*. He showed that $\hat{p}_{RSS.MLE}$ is slightly more efficient than $\hat{p}_{RSS}$ and uniformly more efficient than the standard sample percentage of 'successes', $\hat{p}_{SRS}$, for a SRS of the same size.

Another factor that is important to consider when applying RSS methodology to estimation of a population proportion is the curious aspect of initially "ranking" binary data to implement the ranked set sample structure. This is not an issue if individuals are used to subjectively judgment rank the candidates within a set with respect to their relative likelihoods of being 'successes'. However, if we wish to use additional quantitative information from the population to aid in these within-sets binary rankings, then appropriate mechanisms are required to enable that process. Terpstra and Liudahl (2004) suggested the use of a single concomitant to facilitate the ranking of binary data and Chen et al. (2005) expanded on this concept through the use of logistic regression to incorporate multiple concomitants in a formal mechanism for ranking such data. They found that this use of logistic regression substantially improves the accuracy of the preliminary ranking in the RSS process, which, in turn, can lead to considerable gains in precision.

RSS ANALOGUES OF THE MANN-WHITNEY-WILCOXON TWO-SAMPLE PROCEDURES (BOHN-WOLFE)

Let $X_{[1]w}, \ldots, X_{[k]w}$, $w = 1, \ldots, c$, be a balanced RSS of size $m = kc$ from a continuous population with c.d.f. $F(x)$, where k is the set size and m is the number of cycles, and let $Y_{[1]t}, \ldots, Y_{[q]t}$, $t = 1, \ldots, d$, be a balanced RSS of $n = qd$ observations from a second continuous population with c.d.f. $G(x) = F(x\text{-}\Delta)$, where q is the set size and d is the number of cycles. All $N = m + n$ RSS observations are assumed to be mutually independent and we assume (for now) that the ranking processes used to obtain the two RSSs are perfect, so that the RSS observations are true order statistics from their respective populations.

The Bohn-Wolfe (1992) ranked set sample analogue to the two-sample Mann-Whitney U statistic is then

$$BW = \sum_{s=1}^{k}\sum_{t=1}^{c}\sum_{u=1}^{q}\sum_{v=1}^{d}\phi(X_{[s]t}, Y_{[u]v}) = (\#\ of\ X_{[s]t}\text{'}s < Y_{[u]v}\text{'}s\ in\ the\ RSS\ data),$$

where

$$\phi(X_{[s]t}, Y_{[u]v}) = \begin{cases} 1, & \text{if } X_{[s]t} < Y_{[u]v} \\ 0, & \text{otherwise} \end{cases} ,$$

for $s = 1, \ldots, k$; $t = 1, \ldots, c$; $u = 1, \ldots, q$; $v = 1, \ldots, d$.

Analogous to the Mann-Whitney procedure, the Bohn-Wolfe procedure rejects the null hypothesis H_0: $\Delta = 0$ in favor of the alternative H_A: $\Delta > 0$ for large values of BW.

Effect of Imperfect Rankings on BW – Lack of Level Robustness

All of the properties of the test procedures based on *BW* are predicated on the assumption that the judgment rankings are perfect. This assumption leads directly to the distribution-free property for *BW* under the null hypothesis H_0: $\Delta = 0$ and enables one to use standard distributional properties of order statistics to construct the necessary critical values for the test procedures. Under perfect rankings, *BW* is uniformly more powerful than the Mann-Whitney procedure. In practice, of course, it is likely that some ranking errors will occur in obtaining our ranked set sample data. Bohn and Wolfe (1994) used expected spacings to develop an imperfect ranking model to address this issue. They found that small imperfections in the rankings did not seriously affect the overall performance of tests based on *BW*. However, it was also clear from their study that significant ranking error could lead to substantial inflation of the true significance level over the nominal level set by using the null distribution of BW under perfect rankings. Fligner and MacEachern (2006) and Frey (2007) studied this issue under different classes of imperfect ranking models. This level inflation for the *BW* test procedures under imperfect rankings clearly emphasizes the importance of having a reliable ranking process for collecting the ranked set sample data.

Modifications to Accommodate Imperfect Rankings

The Bohn-Wolfe statistic BW is no longer distribution-free under the null hypothesis H_0 in the presence of ranking errors during the collection of the RSS data and this can lead to substantial inflation of the true significance level over the nominal level under the assumption of perfect rankings for the associated test procedures based on BW. While there is little that can be done to ameliorate these concerns in the presence of substantial ranking errors and small sample sizes, Ozturk (2008, 2010) proposed a way to deal with them in the case of larger sample sizes, using a number of techniques, including minimizing a distance measure and nonparametric maximum likelihood, to estimate unknown parameters in the classes of imperfect ranking models developed by Bohn and Wolfe (1994) and Frey (2007).

Comparisons Only Within, but not Across, Judgment Ranks

For the setting where the set size is the same for the X and Y ranked set samples (that is, $k = q$), Fligner and MacEachern

(2006) proposed a competitor statistic T that uses the Mann-Whitney comparisons between X's and Y's only for those (X, Y)'s that have the same within set judgment ranks.

Let

$$T_j = \sum_{r=1}^{c}\sum_{v=1}^{d} \phi(X_{[j]r}, Y_{[j]v}), \text{ for } j = 1, \ldots, k,$$

where, as before,

$$\phi(X_{[j]r}, Y_{[j]v}) = \begin{cases} 1, & \text{if } X_{[j]r} < Y_{[j]v} \\ 0, & \text{otherwise} \end{cases}.$$

Thus, T_j is simply the Bohn-Wolfe statistic utilizing Mann-Whitney counts only between those X and Y observations that have the same within set judgment rank j, for $j = 1, \ldots, k$. The Fligner-MacEachern test statistic is then the sum of these common-rank Mann-Whitney statistics, namely,

$$FM = \sum_{j=1}^{k} T_j \ .$$

The statistic FM is distribution-free under H_0: $\Delta = 0$ for any ranking mechanism (perfect or imperfect) that is the same for both the X and Y populations. The null distribution for

FM can be obtained as the convolution of k independent Mann-Whitney null distributions, each for the same sample sizes of c X's and d Y's, and this null distribution is the same whether the ranking process is perfect or imperfect in any fashion, including completely at random.

OTHER IMPORTANT ISSUES FOR RSS

Ranked set sampling has been an important aspect of statistical research for the past thirty years and continues to attract considerable attention even sixty plus years post-McIntyre. Part of this richness is due to the great flexibility provided by the ranked set paradigm. We briefly discuss some aspects of this flexibility that provide both excellent research opportunities and address complexities in applications.

Set Size

The set size plays a critical role in the performance of any RSS procedure. For given set size k, each measured ranked set sample observation utilizes additional information obtained from its ranking relative to $k - 1$ other units from the population. With perfect rankings, this additional information is clearly an increasing function of k. Thus, with perfect rankings, we would want to take our set size k

to be as large as economically possible within available resources. However, it is also clear that the likelihood of errors in our 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 rankings. Therefore, to select the set size k optimally, we need to be able to both model the probabilities for imperfect rankings and to assess their impact on our RSS statistical procedures.

Imperfect Rankings

The effectiveness of RSS procedures depends directly on how well the within-set rankings to select the units for measurement can be accomplished. While perfect rankings are surely the goal of any RSS protocol, it is just as likely not to be feasible. Thus, it is imperative in practice that we be able to assess the effect of imperfect rankings on our procedures and the most appropriate way to do this is to develop statistical models to capture the uncertainty of the ranking process.

Dell and Clutter (1972) proposed the first class of models for this purpose. They viewed the ranks of the experimental units as being based on perceived values that are associated with the true measured values through an additive model. Taking a much different approach, Bohn and Wolfe (1994) considered the distributions of the judgment order statistics to be mixtures of distributions of the true order statistics and based their model on the expected spacings between order statistics. Frey (2007) proposed a much larger class of models through a clever scheme of subsampling order statistics from the basic Bohn-Wolfe model. Fligner and MacEachern (2006) used the monotone likelihood ratio principle to develop a competing class of imperfect ranking models.

Given the importance of reliable ranking protocols in RSS, it is also natural that procedures would be developed to statistically assess whether a given ranking protocol does, in fact, provide reliably "perfect" rankings. Frey et al. (2007) initiated the work on this issue, and many other authors have contributed to our understanding of what constitutes acceptable "perfect rankings", including Li and Balakrishnan (2008), Vock and Balakrishnan (2011, 2013), Frey and Wang (2013), Zhamanzade et al. (2014), and Frey and Zhang (2017).

Unbalanced Ranked Set Sampling

The emphasis in this talk has been entirely on **balanced** ranked set sample data of the form $X_{[i]j}$, $i = 1, \ldots, k$ and $j = 1, \ldots, m$, where k is the common set size and m is the number of cycles. Thus, in the case of balanced ranked set sample data we have the same number, m, of each of the judgment order statistics; that is, we have m mutually independent and identically distributed first judgment order statistics $X_{[1]1}, \ldots, X_{[1]m}$; m mutually independent and identically distributed second judgment order statistics $X_{[2]1}, \ldots, X_{[2]m}$; …; and m mutually independent and identically distributed k^{th} judgment order statistics $X_{[k]1}, \ldots, X_{[k]m}$. While balanced RSS is the most commonly occurring form of ranked set sampling data, there are situations where it is not optimal to collect the same number of measured observations for each of the judgment order statistics.

For example, consider an underlying distribution that is unimodal and symmetric about its median θ and 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 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, it is natural to consider measuring the same judgment order statistic, namely, the judgment median $X_{[\frac{k+1}{2}]}$, in each set, so that it is measured all k times in each of the m cycles. The resulting ranked set sample consists of mk measured observations, each of which is a judgment median from a set of size k. This would be the most efficient ranked set sample for estimating the population median θ for a population that is both unimodal and symmetric about θ, and it is clearly as unbalanced as possible. A similar approach calls for a distinctly different unbalanced ranked set sample for estimating the median of an asymmetric unimodal population. There are, of course, other considerations. While median judgment order statistics do provide an efficient estimator for the median of a symmetric population, they would not be an optimal choice if we also want to estimate the variance of the population—balanced RSS measurements would be preferable for this purpose.

(See Özturk and Wolfe, 2000, for more discussion of the pros and cons of balanced versus unbalanced RSS.)

Chen et al. (2006) and Chen et al. (2009) considered the use of unbalanced ranked set samples in estimation of a population proportion p. They used Neyman allocation to decide on optimal representations of the various judgment order statistics in the formation of a ranked set sample. This approach leads to the preferred use of balanced RSS for values of p near ½, but the unbalanced nature of the optimal allocation grows dramatically as the value of p nears either 0 or 1.

Unequal Set Sizes

Sometimes the sets that arise naturally in RSS applications are of unequal sizes. For instance, commuters on different public buses in a large city or patients in a collection of doctors' waiting rooms represent naturally occurring sets of varying sizes. One alternative in such situations is to pare down the larger sets to agree in size with the smaller sets, but that can lead to a loss of valuable information that could have been obtained from the more comprehensive rankings within the larger sets. Gemayel et al. (2010) proposed an estimator for the median of a symmetric population that combines medians of ranked set samples of varying sizes. While not optimal for any specific symmetric distribution, they showed that the estimator is robust over a wide class of symmetric distributions. Samawi (2011) also considered varied set sizes for RSS mean and ratio estimation.

Cost Considerations

Even under perfect judgment 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 a ranked set sample, all affect the choice of an optimal set size *k*. For a basic discussion of these factors and their effect on optimal set size selection, the interested party is referred to Nahhas et al. (2002).

Multiple Observations Per Set

Thus far, we have only considered measuring a single observation from each set. The rationale behind this approach is the fact that the correlation inherent in measuring more than one observation per set typically leads to a reduction in efficiency for RSS estimation. Wang et al. (2004), however, demonstrated that this is not necessarily the case when the cost involved in the ranking process itself is not small relative to the costs of unit selection and unit measurement. Under such conditions, they found that taking two or more observations from a set can lead to improved RSS estimation.

RSS from Finite Populations

In many situations, we are faced with sampling from a finite population where the size of the population is not large relative to the size of the desired RSS. For such settings, it is important to use appropriate finite population methodology in collecting the RSS sample, as well as in analyzing the collected sample data. Patil et al. 1995) first discussed the necessary finite population corrections for RSS in the setting of mean estimation. Properties of RSS procedures for other finite population settings can be found, for example, in Deshpande et al. (2006) and Jafari Jozani and Johnson (2011).

EXTENSIONS AND RELATED APPROACHES

RSS has also provided a stimulus for the emergence of other important related approaches to statistical inference.

Judgment Post-Stratification

One of the features of ranked set sampling is that a researcher is required to judgment rank the potential units prior to obtaining any measurements; that is, the researcher must commit to the ranked set sampling approach from the onset of the experiment. MacEachern et al. (2004) introduced a data collection method, called judgment post-stratification (JPS), that enables a researcher to collect an initial simple random sample (SRS) in standard fashion from the population of interest and then to post-stratify the SRS observations by ranking each of them among its own randomly chosen comparison sample. Thus, the variable of interest is first measured on all of the original simple random sample units and only then is relative judgment ranking information obtained from the comparison samples to enable the judgment post-stratification. This approach allows the researcher to utilize the measurements in the full SRS as well as the additional information obtained from the judgment post-stratification process. JPS has been studied for mean estimation and variance estimation by Frey and Feeman (2012, 2013) and for designed experiments by Du and MacEachern (2008).

The JPS approach also provides a mechanism for incorporating both imprecise rankings and information from multiple rankers via the judgment post-stratification process. For additional work on this aspect of JPS, see Wang et al. (2006), Stokes et al. (2007), Wang et al. (2008), and Frey and Özturk (2011).

Order Restricted Randomization

Özturk and MacEachern (2004, 2007) built on the general framework of ranked set sampling to develop order restricted randomized (ORR) designs that utilize subjective judgment ranking to enable restricted randomization in the comparison of two treatments (one of which could be a control). The units within a given set are assigned to different treatments and then instead of the typical RSS approach that selects a single unit from each ranked set for full measurement, the ORR designs allow for all of the units within a set to be fully measured. The positive dependence between the units within sets leads to contrast estimators and confidence intervals with smaller variability than those based on either completely randomized designs or purely ranked set sample designs. An added feature of ORR estimation is that it does not rely on perfect judgment rankings.

Sampling from Partially Rank-Ordered Sets

There are times when it is difficult to rank all of the experimental units in a set with high confidence, particularly when subjective information is utilized in the ranking process. Özturk (2011, 2012) considered a judgment ordering process called *judgment subsetting* that allows a judgment ranker to use tied ranks when it is difficult to fully rank the experimental units in a set. He showed that this added flexibility leads to improved precision for RSS estimation procedures in settings where the full ranking cannot be done with high confidence.

Partially Sequential Ranked Set Sampling

Another method for reducing the sample size necessary to achieve improved statistical inference is the partially sequential sampling (PSS) approach introduced by Wolfe (1977a, 1977b). Recently, this PSS sampling scheme has been combined with ranked set sampling data collection (see, for example, Matthews and Wolfe, 2018a and 2018b, and Matthews et al., 2014) to further improve the efficiency of both estimation and testing procedures in the two-sample location setting.

Applications

Applications of RSS did not begin to appear until nearly fifteen years after the publication of McIntyre's paper. Halls and Dell (1966) discussed its application in a study of forage yields (and they were actually the first to invoke the name ranked set sampling for the methodology). Evans (1967) used this approach in regeneration surveys for long-leaf pine trees. More than ten years later, Martin et al. (1980) employed RSS in the estimation of shrub phytomass in Appalachian oak forests; Nelson et al. (1987) studied the nutrition of *Populus deltoids* plantations in the lower Mississippi River Valley using RSS-collected data; Cobby et al. (1985) utilized RSS in their investigation of grass and grass-clover swards.

More recently, Mode et al. (1999, 2002) investigated the use of RSS in the assessment of stream habitat areas in the Pacific Northwest of the United States in connection with salmon production. Murray et al. (2000) provided an application of RSS in the comparison of different approaches to spraying apple orchards. Al-Saleh and Al-Shrafat (2001) used RSS to estimate average milk yield among sheep and Özturk et al. (2005) used it to estimate the population mean and variance in regard to sheep flock management. Kvam (2003) applied RSS to binary water quality data with covariates. Chen et al. (2004) illustrated the use of an RSS approach in the estimation of tree heights for a set of data collected by Platt et al. (1988) and for estimation of cinchona yield from a previous experiment by Sengupta et al. (1988). Husby et al. (2005) used a crop production data set from the United States Department of Agriculture to demonstrate the practical benefits of RSS in the timely prediction of corn production and corn yield. Muttlak (1996), Özturk (2002), and Özturk et al. (2004) applied an RSS protocol in a one-way analysis of variance setting to assess the relative healthiness of young males raised in different regions of Jordan. Tarr et al. (2005) incorporated RSS in their study of the map accuracy of soil variables using soil electric conductivity as a covariate. Wang et al. (2009) showed how RSS can be used to increase efficiency and reduce costs in fishery research. In a totally different venue, Gemayel et al. (2012) provided an illustration of the cost savings that can result from the application of RSS in the field of auditing.

Additional Discussion and References for RSS

Of course, with the level of RSS research activity in the past thirty years, there are many more published articles in the field than the few that I have briefly discussed here. Chen et al. (2004), Chapter 15 in Hollander et al. (2013) and the recent review articles by Wolfe (2012) and Al-Omari and Bouza (2014) are good sources for additional information and references about RSS methodology.

References

Al-Omari, A. I. and C. N. Bouza. 2014. Review of ranked set sampling: modifications and applications. *Investigación Operacional* **35(3)**: 215-235.

Al-Saleh, M. F. and K. Al-Shrafat. 2001. Estimation of average milk yield using ranked set sampling. *Environmetrics* **12(4)**: 395-399.

Bohn, L. L., and D. A. Wolfe. 1992. Nonparametric two-sample procedures for ranked-set samples data. *J. Amer. Statist. Assoc.* **87**: 552-561.

Bohn, L. L., and D. A. Wolfe. 1994. The effect of imperfect judgment rankings on properties of procedures based on the ranked-set samples analog of the Mann-Whitney-Wilcoxon statistic. *J. Amer. Statist. Assoc.* **89**: 168-176.

Chen, H., E. A. Stasny, and D. A. Wolfe. 2005. Ranked set sampling for efficient estimation of a population proportion. *Stat. Med.* **24**: 3319-3329.

Chen, H., E. A. Stasny, and D. A. Wolfe. 2006. Unbalanced ranked set sampling for estimating a population proportion. *Biometrics* **62**: 150-158.

Chen, H., E. A. Stasny, D. A. Wolfe, and S. N. MacEachern. 2009. Unbalanced ranked set sampling for estimating a population proportion under imperfect rankings. *Comm. Stat. Theory Methods* **38**: 2116-2125.

Chen, Z., Z. D. Bai, and B. K. Sinha. 2004. *Ranked Set Sampling: Theory and Applications*. New York: Springer-Verlag.

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

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

Deshpande, J. V., J. Frey, and Ö. Özturk. 2006. Nonparametric ranked-set sampling confidence intervals for a finite population. *Environ. Ecol. Stat.* **13**: 25-40.

Du, J. and S. N. MacEachern. 2008. Judgment post-stratification for designed experiments. *Biometrics* **64**: 345-354.

Evans, M. J. 1967. Application of ranked set sampling to regeneration surveys in areas direct-seeded to longleaf pine. M. S. Thesis, School of Forestry and Wildlife Management, Louisiana State University, Baton Rouge, LA.

Fligner, M. A. and S. N. MacEachern. 2006. Nonparametric two-sample methods for ranked-set sample data. *J. Amer. Statist. Assoc.* **101**: 1107-1118.

Frey, J. C. 2007. New imperfect rankings models for ranked set sampling. *J. Statist. Plan. Inf.* **137**: 1433-1445.

Frey, J. and T. G. Feeman. 2012. An improved mean estimator for judgment post-stratification. *Comp. Stat. Data Anal.* **56**: 418-426.

Frey, J. and T. G. Feeman. 2013. Variance estimation using judgment post-stratification. *Ann. Inst. Stat. Math.* **65**: 551-569

Frey, J. C. and Ö. Özturk. 2011. Constrained estimation using judgment post-stratification. *Ann. Inst. Stat. Math.* **63(4)**: 769-789.

Frey, J., Ö. Özturk, and J. V. Deshpande. 2007. Nonparametic tests for perfect judgment rankings. *J. Amer. Stat. Assoc.* **102**: 708-717.

Frey, J. and L. Wang. 2013. Most powerful rank tests for perfect rankings. *Comp. Stat. Data Anal.* **60**: 157-168.

Frey, J. and Y. Zhang. 2017. Testing perfect rankings in ranked-set sampling with binary data. *Can. J. Stat.* **45(3)**: 326-339.

Gemayel, N. M., E. A. Stasny, J. A. Tackett, and D. A. Wolfe. 2012. Ranked set sampling: an auditing application. *Rev. Quant. Finance Accounting.* **39**: 413-422.

Gemayel, N. M., E. A. Stasny, and D. A. Wolfe. 2010. Optimal ranked set sampling estimation based on medians from multiple set sizes. *J. Nonpar. Stat.* **22**: 517-527.

Halls, L. K. and T. R. Dell. 1966. Trial of ranked-set sampling for forage yields. *Forest Science* **12**: 22-26.

Hollander, M., D. A. Wolfe, and E. Chicken. 2013. *Nonparametric Statistical Methods*, 3rd edition. John Wiley & Sons, New York, NY.

Husby, C. E., E. A. Stasny, and D. A. Wolfe. 2005. An application of ranked set sampling for mean and median estimation using USDA crop production data. *J. Agr. Biol. and Envir. Stat.* **10(3)**: 354-373.

Jafari Jozani, M. and B. C. Johnson. 2011. Design based estimation for ranked set sampling in finite population. *Environ. Ecol. Stat.* **18**: 663-685.

Kvam, P. H. 2003. Ranked set sampling based on binary water quality data with covariates. *J. Agr. Biol. and Envir. Stat.* **8(3)**: 271-279.

Kvam, P. H. and F. J. Samaniego. 1993. On the inadmissibility of empirical averages as estimators in ranked set sampling. *J. Stat. Plan. Infer.* **36**: 39-55.

Kvam, P. H. and F. J. Samaniego. 1994. Nonparametric maximum likelihood estimation based on ranked set samples. *J. Amer. Stat. Assoc.* **89**: 526-537.

Li, T. and N. Balakrishnan. 2008. Some simple nonparametric methods to test for perfect ranking in ranked set sampling. *J. Stat. Plan. Infer.* **138**: 1325-1338.

MacEachern, S. N., Ö. Özturk, G. V. Stark, and D. A. Wolfe. 2002. A new ranked set sample estimator of variance. *J. Royal Statist. Soc. B* **64**: 177-188.

MacEachern, S. N., E. A. Stasny, and D. A. Wolfe. 2004. Judgement post-stratification with imprecise rankings. *Biometrics* **60**: 207-215.

Martin, W. L., T. L. Sharik, R. G. Oderwald, and D. W. Smith. 1980. Evaluation of ranked set sampling for estimating shrub phytomass in Appalachian oak forests. Publication No. FWS-4-80, School of Forestry and Wildlife Resources, Virginia Polytechnic Institute and State University, Blacksburg, VA.

Matthews, M. J., E. A. Stasny, and D. A. Wolfe. 2014. Two-sample partially sequential median test procedures using ranked set sample data. *J. App. Stat. Sci.* **22(1/2)**: 1-19.

Matthews, M. J. and D. A. Wolfe. 2018a. Partially sequential dual ranked set sample two-sample test procedures. Submitted for publication.

Matthews, M. J. and D. A. Wolfe. 2018b. Partially sequential Fligner-MacEachern test procedure. Submitted for publication.

McIntyre, G. A. 1952. A method for unbiased selective sampling, using ranked sets. *Aust. J. Agric. Res.* **3**: 385-390.

McIntyre, G. A. 2005. A method for unbiased selective sampling, using ranked sets. *Amer. Statistician* **59**: 230-232.

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

Mode, N. A., L. L. Conquest, and D. A. Marker. 2002. Incorporating prior knowledge in environmental sampling: ranked set sampling and other double sampling procedures. *Environmetrics* **13(5-6)**: 513-521.

Murray, R. A., M. S. Ridout, and J. V. Cross. 2000. The use of ranked set sampling in spray deposit assessment. *Aspects of App. Biol.* **57**: 141-146.

Muttlak, H. A. 1996. Estimation of parameters for one-way layout with rank set sampling. *Biom. J.* **38(4)**: 507-515.

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

Nelson, L. E., G. L. Switzer, and B. G. Lockaby. 1987. Nutrition of *Populus deltoids* plantations during maximum production. *Forest Ecol. Manag.* **20(1-2)**: 25-41.

Nussbaum, B. D., and B. K. Sinha. 1997. Cost effective gasoline sampling using ranked set sampling. In *American Statistical Association 1997 Proceedings of the Section on Statistics and the Environment, pages 83-87*. American Statistical Association.

Özturk, Ö. 2002. Rank regression in ranked-set samples. *J. Amer. Stat. Assoc.* **97(460**): 1180-1191.

Özturk, Ö. 2008. Inference in the presence of ranking error in ranked set sampling. *Can. J. Stat.* **36**: 1-18.

Özturk, Ö. 2011. Sampling from partially rank-ordered sets. *Environ. Ecolog. Stat.* **18**: 757-779.

Özturk, Ö. 2012. Quantile inference based on partially rank-ordered set samples. *J. Statist. Plan. Inf.* **142**: 2116-2127.

Özturk, Ö. 2014. Estimation of population mean and total in a finite population setting using multiple auxiliary variables. *J. Agric. Biol. Environ. Stat.* **19(2)**: 161-184.

Özturk, Ö., Ö. C. Bilgin, and D. A. Wolfe. 2005. Estimation of population mean and variance in flock management: a ranked set sampling approach in a finite population setting. *J. Statist. Comp. Sim.* **75**: 905-919.

Özturk, Ö., and S. N. MacEachern. 2004. Order restricted randomized designs for control versus treatment comparison. *Ann. Inst. Statist. Math.* **56**: 701-720.

Özturk, Ö., and S. N. MacEachern. 2007. Order restricted randomized designs and two-sample inference. *Environ. Ecolog. Stat.* **14**: 365-381.

Özturk, Ö., and D. A. Wolfe. 2000. Optimal allocation procedure in ranked set sampling for unimodal and multi-modal distributions. *Environ. Ecolog. Stat.* **7**: 343-356.

Özturk, Ö., D. A. Wolfe, and R. Alexandridis. 2004. Multi-sample inference for simple-tree alternatives with ranked-set samples. *Austral. and New Zeal. J. Stat.* **46(3**): 443-455.

Patil, G. P., A. K. Sinha, and C. Taillie. 1995. Finite population correction for ranked set sampling. *Ann. Inst. Stat. Math.* **47**: 621-636.

Perron, F. and B. K. Sinha. 2004. Estimation of variance based on a ranked set sample. *J. Stat. Plan. Infer.* **120(1)**: 21-28.

Platt, W. J., G. W. Evans, and S. L. Rathbun. 1988. The population dynamics of a long-lived conifer (*Pinus palustris*). *The Amer. Naturalist* **131**: 491-525.

Samawi, H. M. 2011. Varied set size ranked set sampling with applications to mean and ratio estimation. *Int. J. Model. Simul.* **31(1)**: 6-13.

Sengupta, J. M., I. M. Chakravarti, and D. Sarkar. 1951. Experimental survey for the estimation of cinchona yield. *Bull. International Statist. Inst.* **33**: 313-331.

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

Stokes, S. L., and T. W. Sager. 1988. Characterization of a ranked-set sample with application to estimating distribution functions. *J. Amer. Statist. Assoc.* **83**: 374-381.

Stokes, L., X. Wang, and M. Chen. 2007. Judgement post-stratification with multiple rankers. *J. Statist. Theory Appl.* **6**: 344-359.

Takahasi, K., and K. Wakimoto. 1968. On unbiased estimates of the population mean based on the sample stratified by means of ordering. *Ann. Inst. Statist. Math.* **20**: 1-31.

Tarr, A. B., K. J. Moore, C. L. Burras, D. G. Bullock, and P. M. Dixon. 2005. Improving map accuracy of soil variables using soil electrical conductivity as a covariate. *Precision Agr.* **6(3)**: 255-270.

Terpstra, J. T. 2004. On estimating a population proportion via ranked set sampling. *Biometrical J.* **46**: 264-272.

Terpstra, J. T., and L. A. Liudahl. 2004. Concomitant-based rank set sampling proportion estimates. *Stat. Med.* **23**: 2061-2070.

Vock, M. and N. Balakrishnan. 2011. A Jonckheere-Terpstra-type test for perfect ranking in balanced ranked set sampling. *J. Stat. Plan. Infer.* **141**: 624-630.

Vock, M. and N. Balakrishnan. 2013. A connection between two nonparametric tests for perfect ranking in balanced ranked set sampling. *Comm. Stat. Theory Meth.* **42**: 191-193.

Wang, X., J. Lim, and L. Stokes. 2008. A nonparametric mean estimator for judgment poststratified data. *Biometrics* **64**: 355-363.

Wang, X., L. Stokes, J. Lim, and M. Chen. 2006. Concomitants of multivariate order statistics with application to judgment poststratification. *J. Amer. Statist. Assoc.* **101**: 1693-1704.

Wang, Y-G, Z. Chen, and J. Liu. 2004. General ranked set sampling with cost considerations. *Biometrics* **60**: 556-561.

Wang, Y-G, Y. Ye, and D. A. Milton. 2009. Efficient designs for sampling and subsampling in fisheries research based on ranked sets. *ICES J. Marine Sci.* **66(5)**: 928-934.

Wolfe, D. A. 2007a. On a class of partially sequential two-sample test procedures. *J. Amer. Stat. Assoc.* **72(357**): 202-205.

Wolfe, D. A. 2007b. Two-stage two-sample median test. *Technometrics* **19(4)**: 495-501.

Wolfe, D. A. 2012. Ranked set sampling: its relevance and impact on statistical inference. *ISRN Probability and Statistics Spotlight Article*, **Volume 2012, Article ID 568385**, 32 pages.

Zamanzade, E., N. R. Arghami, and M. Vock. 2014. A parametric test of perfect rankings in balanced ranked set sampling. *Comm. Stat. Theory Meth.* **43**: 4589-4611.

Zamanzade, E. and M. Vock. 2015. Variance estimation in ranked set sampling using a concomitant variable. *Stat. Prob. Letters* **105**: 1-5.