DOKUZ EYLÜL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

STATISTICAL INFERENCE BASED ON RANKED SET SAMPLING

by
Bekir ÇETİNTAV

June, 2018
İZMİR

STATISTICAL INFERENCE BASED ON RANKED SET SAMPLING

A Thesis Submitted to the
Graduate School of Natural And Applied Sciences of Dokuz Eylül University
In Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Mathematics

by
Bekir ÇETİNTAV

June, 2018
İZMİR

Ph.D. THESIS EXAMINATION RESULT FORM

We have read the thesis entitled "**STATISTICAL INFERENCE BASED ON RANKED SET SAMPLING**" completed by **BEKİR ÇETİNTAV** under supervision of **PROF. DR. SELMA GÜRLER** and **PROF. DR. BERNARD DE BAETS** and we certify that in our opinion it is fully adequate, in scope and in quality,as a thesis for the degree of Doctor of Philosophy.

Prof. Dr. Selma GÜRLER
Supervisor

Prof. Dr. Bernard DE BAETS
Supervisor

Prof. Dr. Halil ORUÇ
Thesis Committee Member

Assoc. Prof. Dr. Neslihan DEMİREL
Thesis Committee Member

Doç. Dr. Güvenç ARSLAN
Examining Committee Member

Prof. Dr. G. Mahir TÜRKMEN
Examining Committee Member

Dr. Öğr. Üyesi A. Özgür KINAY
Examining Committee Member

Prof. Dr. Latif SALUM
Director

Graduate School of Natural and Applied Sciences

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my supervisor Prof. Dr. Selma GÜRLER and to my co-supervisor Prof. Dr. Bernard DE BAETS for their continuous support, selfless guidance, generous advice and helpful comments. In particular, I wish to thank Prof. Dr. Selma GÜRLER, for her valuable insights, patience and encouragement during the difficult times throughout my research. I am truly grateful to my co-supervisor, for his never ending guidance and support both during and after my scientific visit to Ghent University, Belgium. Their academic perspectives created a very positive impact on my studies.

Besides my advisors, I would like to thank Assoc. Prof. Dr. Neslihan DEMİREL and Assoc. Prof. Dr. Gözde ULUTAGAY for their contributions and invaluable assistance. I'm further grateful for the insights and efforts put forth by Prof. Dr. Özlem EGE ORUÇ, Prof. Dr. Halil ORUÇ, Prof. Dr. Pembe KESKİNOĞLU and Prof. Dr. Serdar KURT during my PhD study.

This thesis is dedicated to my parents Bahriye and Savaş ÇETİNTAV and my lovely wife Elif ÇETİNTAV because of their endless love, confidence, encouragement and continued support in my life. Words cannot accurately express my thanks to my lovely family.

Finally, it is noted that this study is supported by the Scientific and Technological Research Council of Turkey (TUBITAK-COST Grant No. 115F300) under ISCH COST Action IS1304 and partly carried out while I made a visit to Ghent University supported by TUBITAK-BIDEB Grant No. 1059B141501212.

Bekir ÇETİNTAV

STATISTICAL INFERENCE BASED ON RANKED SET SAMPLING

ABSTRACT

In a scientific research, the data collection method is substantial for statistical and methodological analysis. Ranked Set Sampling (RSS) is an advanced and effective method for collecting data and making inferences about the population. The main impact of RSS is to use the ranking information of the units in the sampling mechanism. When the ranking is done properly, the inference based on RSS generally gives better results compared with simple random sampling (SRS) for both parametric and non-parametric cases. Because of the uncertain nature of the ranking process without actual measurement, inaccuracy among judgment ranks is unavoidable when using the RSS procedure in a real life application. In this thesis, we propose a new approach for modeling uncertainty in the ranking process and combining the information coming from multiple rankers. A new sampling procedure, Fuzzy-weighted Ranked Set Sampling (FRSS), is introduced for dealing with uncertainty in the ranking mechanism of RSS with fuzzy sets perspective. New methods are introduced to combine the information coming from multiple rankers in FRSS procedure. An estimator for the population mean is defined using the measurements of the sampled units and their membership degrees obtained using the new sampling procedure. We use real data sets from biometric researches and comparative simulation studies to show that our new method results in a considerable improvement on the estimation of the population mean over the counterparts in the literature.

Keywords: Ranked set sampling, mean estimation, fuzzy sets approach, imperfect ranking.

SIRALI KÜME ÖRNEKLEMESİNE DAYALI İSTATİSTİKSEL ÇIKARSAMA

ÖZ

Bilimsel bir araştırmada, veri toplama yöntemi istatistiksel ve metodolojik analiz için oldukça önemlidir. Sıralı Küme Örneklemesi (SKÖ) verileri elde etmek ve kitle hakkında çıkarımlar yapmak için geliştirilmiş etkili bir yöntemdir. SKÖ'nün en temel etkisi, örnekleme mekanizmasındaki birimlerin sıra bilgilerini kullanmasıdır. Sıralama işlemi olması gerektiği gibi yapıldığında, SKÖ'ye dayalı çıkarımlar Basit Rasgele Örneklemeye (BRÖ) dayalı çıkarımlara kıyasla, hem parametrik hem de parametrik olmayan yöntemler için, daha iyi sonuçlar vermektedir. SKÖ uygulamalarında, birimlerin gerçek ölçüm değerleri bilinmeden gerçekleştirilen sıralama işleminin belirsiz yapısından ötürü, sıralamalara dair verilen kararlarda hata olması kaçınılmazdır.

Bu tezde, sıralama işlemindeki belirsizliğin modellenmesi ve birden fazla sıralayıcıdan gelen bilginin birleştirilmesi için kullanılabilecek yeni bir yaklaşım önerilmektedir. Sıralama işlemindeki belirsizliğin modellenmesinde bulanık kümeler yaklaşımının kullanıldığı yeni örnekleme yöntemi Bulanık-ağırlıklı Sıralı Küme Örneklemesi (BSKÖ) tanıtılmaktadır. Birden çok sıralayıcıdan gelen bilginin birleştirilmesi için yeni yöntemler önerilmektedir. Örneklem birimleri ve onların üyelik derecelerinin birlikte kullanıldığı yeni bir ortalama kestiricisi tanımlanmıştır. Biyometrik araştırmalardan elde edilmiş gerçek veri setleri ve karşılaştırmalı benzetim çalışmaları kullanarak yeni örnekleme metodumuzun literatürdeki benzerlerine göre kitle ortalaması kestiriminde kayda değer bir üstünlük ortaya koyduğunu gösterilmiştir.

Anahtar kelimeler: Sıralı küme örneklemesi, ortalama kestirimi, bulanık kümeler yaklaşımı, hatalı sıralama

CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER ONE
INTRODUCTION

Data collection is a very important part of any type of scientific research. The quantitative methods of data collection build upon sampling which can be defined as the process of selecting units from a population of interest to generalize the results back to the population. Sampling methods can be grouped into two categories, non-random and random sampling methods. In a non-random sampling method, the sampling units are chosen by the researcher without using a probabilistic model. Some of the famous non-random sampling methods are the Quota Sampling, Judgmental Sampling and Snowball Sampling methods. Random sampling methods are considered more preferable and reliable methods when compared with non-random methods because the sampling units are chosen randomly. Simple random sampling (SRS) is the most commonly used random sampling method (Wolfe, 2012). In settings where the concerned population is infinite (or finite but large enough to approximate), a simple random sample is defined as follows.

Definition 1.0.1. *Wolfe (2012), Let the concerned population has a probability distribution function (p.d.f) f and cumulative distribution function (c.d.f) F. The set of n collected random variables $X_1, X_2, \cdots, X_n$ is said to be a simple random sample if each of n variables are mutually independent and have the same probability distribution as the concerned population with p.d.f. f and c.d.f. F.*

In the finite population settings, a simple random sample is defined using a different set of conditions.

Definition 1.0.2. *Wolfe (2012), A set of sample observations with size n from a finite population with size N, $X_1, X_2, \cdots, X_n$, can be defined as a simple random sample if each of the possible subsets of n observations has the same chance of being selected.*

The most important advantages of SRS are ease of procedure and existence of the wide range of statistical methods defined based on SRS. Also it is known that the

average of the variable/attribute of interest based on SRS would provide a good estimator for the population parameter of the variable/attribute if we take enough multiple random samples. However, SRS procedure does not guarantee that a single random sample provide such a good estimate (Wolfe, 2012). Advanced sampling methods such as Stratified Simple Random Sampling, Systematical Sampling and Cluster Sampling focus on this situation in SRS and try to minimize it using an admissible amount of cost, time and labor. Ranked Set Sampling (RSS) is another advanced sampling method developed for this purpose.

1.1 Ranked Set Sampling in the Literature

Ranked set sampling (RSS) was first introduced by McIntyre (1952) as part of a biometric research. In simple terms, random sets are drawn from a population, the units in the sets are ranked with a ranking mechanism, and one of these ranked units is sampled from each set with a specific scheme in RSS procedure. The ranking mechanism could be the visual inspection of an human expert or a highly-correlated concomitant variable. For cases where the actual measurement of the variable of interest is more difficult or expensive than obtaining the rank information from a human expert or a variable used for ranking, RSS can be introduced as an easier and alternative way to obtain a more representative sample.

The terminology of *ranked set sampling* was first introduced in the study of Halls & Dell (1966) regarding the estimation of forage yields in a pine hardwood forest. The first theoretical results on RSS were given by Takahasi & Wakimoto (1968). They proved that, the mean estimator based on RSS is an unbiased estimator of the population mean if ranking is perfect and the variance of the ranked set sample mean is always smaller than the variance of the mean estimator of a SRS with the same sample size. Dell & Clutter (1972) and David & Levine (1972) were the first to give some theoretical treatments on imperfect ranking. Stokes (1976, 1977) considered the use of the concomitant variables in RSS. Later Stokes (1980a,b) introduced the estimation of population variance and the estimation of correlation coefficient of a

bivariate normal population based on RSS. The estimation of distribution function and quantiles with various settings of RSS was considered by Stokes & Sager (1988), Kvam & Samaniego (1993) and Chen (2000). The RSS counterpart of ratio estimate was considered by Samawi & Muttlak (1996).

Based on RSS, various statistical procedures, parametric and non-parametric, have been investigated. For instance, Bohn & Wolfe (1992, 1994), and Hettmansperger (1995) investigated the RSS version of distribution-free test procedures such as sign test, signed rank test and Mann-Whitney-Wilcoxon test. The regression estimate based RSS was addressed by Yu & Lam (1997) and Chen (2001). Different quality control charts for the sample mean were introduced using RSS by Muttlak & Al-Sabah (2003). The studies to date show that RSS based estimators give better results for both parametric and non-parametric inference than traditional sampling methods in nearly all cases. More detailed information on developments in the area of RSS can be found in the review paper of Wolfe (2012), the book of Chen et al. (2004) and the technical report of Mode et al. (1999).

Some researches have focused on modifying the ranking scheme of RSS. Muttlak (1997) first introduced Median RSS where only medians of the random sets are chosen for the sample for estimation of population mean. Ibrahim et al. (2010) modified Median RSS for stratified cases. Samawi et al. (1996) proposed the Extreme RSS (ERSS) model in which they suggested using only the lowest and largest values in the random sets for estimation. Subsequently, Al-Nasser & Mustafa (2009) modified ERSS and called the new model Robust Extreme RSS. Muttlak (2003a) suggested another modification to the RSS, Quartile RSS in which the quartiles of the random sets are chosen for the sample. Al-Nasser (2007) proposed a robust method L Ranked Set Sampling as a generalization of the many types of modified methods mentioned above. Another modification on the ranking scheme of RSS is defining multiple stages. Al-Saleh & Al-Kadiri (2000) introduced Double RSS. Al-Saleh & Al-Omari (2002) suggested Multistage RSS in order to increase the efficiency of the estimator of the population mean.

The ranking mechanism is one of the major parts of RSS procedure. Since there is an ambiguity in discriminating the rank of one unit with another without actual measurement, the ranking of the units could not be perfect in practice. Clearly, there is uncertainty in the ranking mechanism of RSS in cases of imperfect ranking. In the literature, there are several studies focused on the case of imperfect ranking and approaches were introduced for modeling the uncertainty in ranking with a probabilistic perspective. These studies can be grouped into two categories.

The first category features those studies using the model to search for the impact of the imperfect ranking on the inferences. Dell & Clutter (1972) introduce a pioneering model for imperfect ranking using an additive model of concerning and concomitant variable and investigate the effect of the imperfect ranking on the estimation of population mean. Bohn & Wolfe (1994) proposed an expected spacings model for the probabilities of imperfect judgment rankings based on the expected differences of units in the set. They suggested constructing a single stochastic matrix consisting of specified probabilities of units which are inversely proportional to the expected differences between the order statistics. The properties of tests based on Mann-Whitney U statistics are studied based on their model. Aragon et al. (1999) further developed a probability matrix for imperfect ranking. Presnell & Bohn (1999) used this model to show that RSS procedure is asymptotically at least as efficient as the SRS, even if the ranking is imperfect. Frey (2007) described a new class of models for imperfect rankings, and argued whether it contains essentially any reasonable imperfect rankings model. Ozturk (2008) proposed inference techniques for RSS data in the case of imperfect ranking using the models of Bohn & Wolfe (1994) and Frey (2007). Park & Lim (2012) and Frey (2014) also studied the impact of ranking error on the Fisher information in RSS.

In the second category, there are RSS inspired studies where new sampling procedures are defined using the main idea of ranking in RSS and also the uncertainty in the ranking mechanism is studied. These studies are defined as *the RSS related approaches* or the extensions of RSS in the literature. MacEachern et al. (2004) introduced the Judgment Post-Stratified (JPS) sampling method for a specific case

where a simple random sample is already chosen. In the JPS method, a random set is taken for each unit in the simple random sample. The units in these random sets are used for determining the rank of the units in the simple random sample. Then the rank information of the units is used as additional information for inference. MacEachern et al. (2004) also give a formula for the estimation of population mean when the ranking is imprecise/imperfect. Ozturk (2011c) introduced the partially ranked ordered set (PROS) sampling design for a specific case of imperfect ranking. Rankers are allowed to declare any two or more units are tied in rank whenever the units cannot be ranked with complete confidence. The fully measured units are then selected from partially ordered judgment subsets where the units are replaced. Arslan & Ozturk (2013) develop parametric inference for the location and scale parameters of a location scale family of distributions based on a PROS sample design. In that and subsequent studies, such as Ozturk & Jozani (2014), Nazari et al. (2016), ties are the cause of imprecise rank decisions in sampling design and the tied units split the probability evenly between the assigned ranks.

1.2 The scope of the thesis

The studies in the literature show that the inference based on RSS generally gives better results compared with simple random sampling (SRS) for both parametric and non-parametric cases. This advantage comes from the additional information of rank decisions. In simple terms, the ranker determines the most possible rank for each unit in ranked sets and this information is used in inference. However, the ranking is done without actual measurement in practice and it causes uncertainty. In other words, there could be other possible ranks for a unit. In our study, we introduce a new approach for this situation and propose a fuzzy set perspective for dealing with the uncertainty occurring in the ranking process of RSS. Fuzzy set theory allows the units to belong to different sets with different membership degrees. If we define fuzzy sets for each rank in the ranking process of the random sets, the units could belong to not only the most possible rank but also other possible ranks. In this way, we can use not only the

information coming from most possible ranks but also the information coming from other possible ranks.

With this perspective, we first set out the RSS procedure, the mean estimator based on RSS and its statistical properties in Chapter 2. Modified RSS methods and the studies focused on modeling the uncertainty in the ranking process are given in the same chapter. A brief introduction to fuzzy set theory, the motivation of the study and the new sampling method Fuzzy-weighted Ranked Set Sampling (FRSS) are introduced in Chapter 3. The sampling procedure of FRSS which works for both single and multiple ranker cases, an estimator of population mean with some theoretical remarks are given in the same chapter. In Chapter 4, the performance of our new method is investigated with simulation studies for both single and the multiple ranker cases. The brief applications of FRSS and a comprehensive numerical study based on real data sets are conducted in Chapter 5. Finally, the conclusions are given in Chapter 6.

In this thesis, the concerned population is assumed infinite or quite large relative to the size of the sample (being collected with replacement) for ease of discussion. It is assumed that the random sets in the procedure satisfy the following property.

- The collected random variables X_1, X_2, $\cdots$, X_k have the same probability distribution as the concerned population with p.d.f. f and c.d.f. F and are mutually independent.

For a general information about the settings of RSS without replacement from a finite population see Patil et al. (1995), Ozturk et al. (2005) and Jafari Jozani & Johnson (2011).

CHAPTER TWO
RANKED SET SAMPLING

Advanced sampling methods provide increased assurances on the representativeness of the sample data by using additional information. For example, Stratified SRS uses the information coming from homogenous strata; Cluster Sampling uses the information coming from heterogeneous subsets. When viewed from this aspect, RSS can be defined as the sampling method where the units are sampled from ranked sets which are randomly-chosen and the rank information of the sample units is used as additional information. Thereby, a ranked set sample is more likely to span the full range of values in the population than a simple random sample of the same size.

2.1 RSS procedure

The RSS has an easy-to-use sampling procedure. First, the size of the sets (k) and the sample size should be decided. After that, there is a six-step procedure to obtain the ranked set sample.

1. Select k units at random from a specified population.
2. Rank these k units without measuring them.
3. Keep the smallest judged unit.
4. Select the second set of k units randomly from a specified population, rank these units without measuring them, keep the second smallest judged unit
5. Continue the process until k ordered units are measured. (The first five steps are called a cycle.)
6. The first five steps are repeated n times to get n cycles and $k * n$ units.

The procedure above is known as the Balanced RSS procedure. The term *balanced* means an equal number of units are chosen for each rank. The RSS procedure also allows the choice of different numbers of units for each of the ranks so that the researcher is able to give *more weight* to the center or tails. Our study focused on the balanced form of RSS which is the most widely-used form RSS. For detailed information about the unbalanced forms of RSS, one can see Chen et al. (2004) pages 73-101. An illustration of the RSS procedure (balanced form) as follows. Let X be the variable of interest and Y be the concomitant variable which is used to rank the random sets. Figure 2.1 shows the RSS procedure with a concomitant variable Y and set size $k = 3$.

Cycle 1

$$\begin{bmatrix} \mathbf{Y_{(1)1}} \leq Y_{(2)1} \leq Y_{(3)1} \\ Y_{(1)1} \leq \mathbf{Y_{(2)1}} \leq Y_{(3)1} \\ Y_{(1)1} \leq Y_{(2)1} \leq \mathbf{Y_{(3)1}} \end{bmatrix} \rightarrow O_1 = \begin{bmatrix} \mathbf{X_{[1]1}} \\ \mathbf{X_{[2]1}} \\ \mathbf{X_{[3]1}} \end{bmatrix}$$

Cycle 2

$$\begin{bmatrix} \mathbf{Y_{(1)2}} \leq Y_{(2)2} \leq Y_{(3)2} \\ Y_{(1)2} \leq \mathbf{Y_{(2)2}} \leq Y_{(3)2} \\ Y_{(1)2} \leq Y_{(2)2} \leq \mathbf{Y_{(3)2}} \end{bmatrix} \rightarrow O_2 = \begin{bmatrix} \mathbf{X_{[1]2}} \\ \mathbf{X_{[2]2}} \\ \mathbf{X_{[3]2}} \end{bmatrix}$$

...

Cycle n

$$\begin{bmatrix} \mathbf{Y_{(1)n}} \leq Y_{(2)n} \leq Y_{(3)n} \\ Y_{(1)n} \leq \mathbf{Y_{(2)n}} \leq Y_{(3)n} \\ Y_{(1)n} \leq Y_{(2)n} \leq \mathbf{Y_{(3)n}} \end{bmatrix} \rightarrow O_n = \begin{bmatrix} \mathbf{X_{[1]n}} \\ \mathbf{X_{[2]n}} \\ \mathbf{X_{[3]n}} \end{bmatrix}$$

Figure 2.1 Illustration of RSS procedure with a concomitant variable

In the procedure $Y_{(h)j}$ means the value of the concomitant variable Y of the hth unit in a specific set in the jth cycle. $X_{[h]j}$ means the hth ranked unit which is chosen for the sample from the jth cycle for $h = 1, 2, ..., k$ and $j = 1, 2, ..., n$. We use square brackets [] instead of the usual round brackets () for the value of the variable of interest, X, while using round bracket () for the values of the concomitant variable, Y. The reason is that

$X_{[h]j}$ may or may not actually be the hth ordered measurement (in terms of X) among the k units in the random set, because the ranking is done without actual measurement of X in practice of RSS.

We can give another illustration of the application of RSS for a single cycle where the ranking is done by visual inspection of a human ranker. Haq (2011) conducted a study to estimate the mean height of Arabidopsis Thaliana (AT) plants. Balanced RSS procedure was applied and ranking is done visually, as in Figure 2.2.

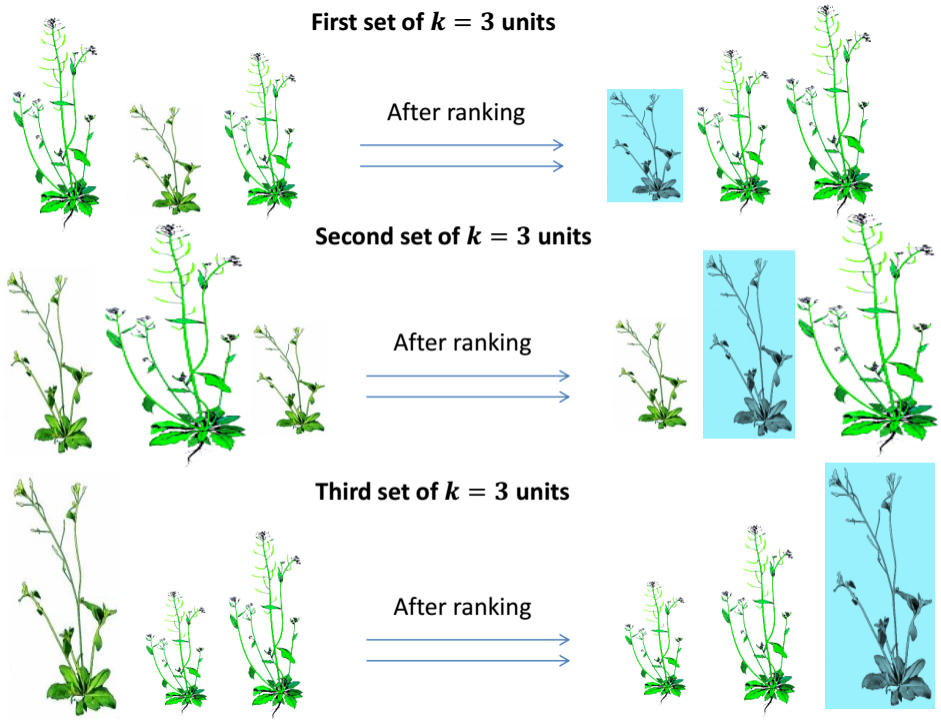

Figure 2.2 Example of ranking with visual inspection for a single cycle (Haq, 2011)

The main aim in RSS is to sample the units which are more likely to span the full range of values in the population, Ozturk (2011b). The ranking part can be defined as the main tool to achieve this aim. The population is *post-stratified* by ranking the random sets. Thus RSS becomes more representative than SRS for the same sample sizes. We can give an illustrative example as follows. Let X have a normal distribution, $N(0, \sigma^2)$. Lets also take a ranked set sample for set size $k = 5$ and cycle size $n = 1$. If the ranking of the units is considered perfect, then we have five units, $X_{[1]}$, $X_{[2]}$, $X_{[3]}$, $X_{[4]}$ and $X_{[5]}$, which behave as order statistics coming from the same population. Figure 2.3 illustrates the marginal densities of the order statistics (denoted

by $jgmt$ 1,2,3,4 and 5 in the figure) and the entire population (denoted by $SimpleR$ in the figure). It can be seen in Figure 2.3 how RSS is more likely to span the full range.

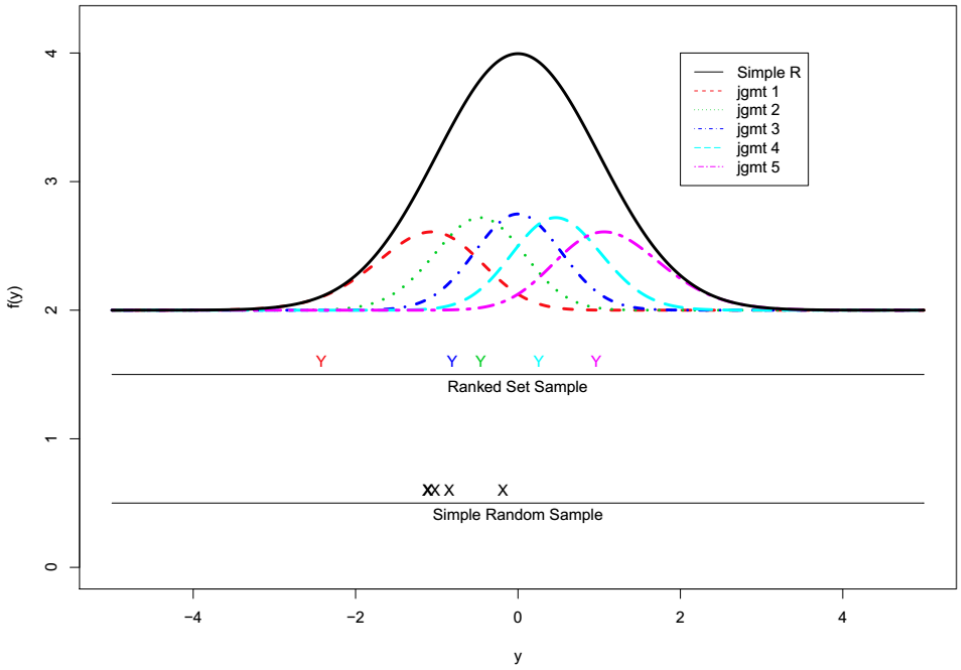

Figure 2.3 Spaning the full range of values in the population with RSS (Ozturk, 2011b)

Since the study of Takahasi & Wakimoto (1968), numerous researchers have shown the advantage of an estimation based on RSS over an estimation based on SRS in terms of efficiency. However, this advantage brings an incremental cost of sampling. Suppose that we take samples from RSS and SRS with equal sizes, $k \times n$, where k is the set size and n is the cycle size in RSS procedure. Let the cost of measuring of the variable of interest X for a single unit be $C(X)$. Let also a concomitant variable, Y, be used for ranking in RSS and the cost of measuring the concomitant variable, Y for a single unit be $C(Y)$. Then the sampling costs ($SCost$) of SRS, RSS and the relative cost (RC) are given by Dell & Clutter (1972)as;

$$SCost(SRS) = (k \times n)C(X), \tag{2.1}$$

$$SCost(RSS) = (k \times n)C(X) + (k^2 \times n)C(Y), \tag{2.2}$$

$$RC = \frac{(k \times n)C(X) + (k^2 \times n)C(Y)}{(k \times n)C(X)} = 1 + \frac{(k^2 \times n)C(Y)}{(k \times n)C(X)}. \tag{2.3}$$

When there is a cost of ranking, the RSS method is more costly than SRS in any case when the sample sizes are equal. Thus, the researcher must chose the optimal set size, k and relative cost of ranking $\frac{C(Y)}{C(X)}$ to attain an acceptable cost of sampling. We can define when RSS is a cost-efficient method compared with SRS. Let $MSE()$ denotes, for example, the mean square error of the estimator based on a specific sampling method. Then the relative efficiency (RE) of RSS is

$$RE = \frac{MSE(SRS)}{MSE(RSS)}. \tag{2.4}$$

When we combine the relative cost (RC) in eq.2.3 and relative efficiency (RE) in Eq. 2.4 to write the formula of the relative cost-efficiency (RCE), it becomes

$$RCE = \frac{SCost(SRS)}{SCost(RSS)} \times \frac{MSE(SRS)}{MSE(RSS)}. \tag{2.5}$$

Finally, we can say that the RSS method can be defined as a cost-efficient sampling method for cases where $RCE > 1$ (Dell & Clutter, 1972) . This definition may be the numeric answer to the questions *"When and why should a researcher use RSS?"*. In other words, RSS is a cost-efficient way to obtain a more representative sample when the actual measurement of the variable of interest is more difficult or expensive than obtaining the rank information from a human expert or a variable used for ranking.

2.2 Mean estimator based on RSS

We should first introduce the mean estimator based on SRS for a clear comparison with RSS. Let N observations, $X_1, X_2, \ldots, X_N$, be chosen as a simple random sample from a continuous population with distribution function F, density function f and finite mean and variance (μ, σ^2). The estimator of μ is;

$$\bar{X}_{SRS} = \frac{1}{N}\sum_{i=1}^{N} X_i. \tag{2.6}$$

We know that $\bar{X}_{SRS}$ is an unbiased estimator of μ with variance $Var(\bar{X}_{SRS}) = \frac{\sigma^2}{N}$. Let another set of N observations be chosen as a ranked set sample from the same population with set size k, cycle size n and $N = k \times n$. The mean estimator of the ranked set sample observations, $X_{[1]1}, X_{[2]1}, \cdots, X_{[k]1}, X_{[1]2}, \cdots, X_{[k]2}, \cdots, X_{[k]n}$, is formulated as follows.

$$\bar{X}_{RSS} = \frac{1}{k}\frac{1}{n}\sum_{h=1}^{k}\sum_{j=1}^{n} X_{[h]j} \tag{2.7}$$

Under the assumption of perfect rankings, we can give the well known theorems of the mean estimator based on RSS given in Eq. 2.7.

Theorem 2.2.1. *Wolfe (2012). $\bar{X}_{RSS}$ is an unbiased estimator for μ, i.e $E(\bar{X}_{RSS}) = \mu$.*

Proof. With the assumption of perfect rankings, the ranked set sample observations, $X_{[1]1}, X_{[2]1}, \cdots, X_{[k]1}, X_{[1]2}, \cdots, X_{[k]2}, \cdots, X_{[k]n}$ can be considered as order statistics from the underlying continuous population, $X_{(1)1}, X_{(2)1}, \cdots, X_{(k)1}, X_{(1)2}, \cdots, X_{(k)2}$, $\cdots, X_{(k)n}$. Then the expectation of $\bar{X}_{RSS}$ is written as follows.

$$E(\bar{X}_{RSS}) = \frac{1}{k}\frac{1}{n}\sum_{h=1}^{k}\sum_{j=1}^{n} E(X_{(h)j}) \tag{2.8}$$

We remember from the order statistics background that expected values of hth ordered observations of random sets of n cycle chosen from the same population are equal.

$$E(X_{(h)1}) = E(X_{(h)2}) = \cdots = E(X_{(h)n}) = E(X_{(h)}) \tag{2.9}$$

Then 2.8 becomes

$$E(\bar{X}_{RSS}) = \frac{1}{k}\sum_{h=1}^{k} E(X_{(h)}) \tag{2.10}$$

We can write the expectation of hth order statistics, $E(X_{(h)})$ as

$$E(\bar{X}_{RSS}) = \frac{1}{k}\sum_{h=1}^{k}\int_{-\infty}^{\infty} x\frac{k!}{(h-1)!(k-h)!}[F(x)]^{h-1}[1-F(x)]^{k-h}f(x)dx,$$

$$= \frac{1}{k}\sum_{h=1}^{k}\int_{-\infty}^{\infty} kx\binom{k-1}{h-1}[F(x)]^{h-1}[1-F(x)]^{k-h}f(x)dx,$$

$$= \int_{-\infty}^{\infty} xf(x)\left[\sum_{h=1}^{k}\binom{k-1}{h-1}[F(x)]^{h-1}[1-F(x)]^{k-h}\right]dx, \tag{2.11}$$

It follows from the binomial distribution that $\sum_{h=1}^{k}\binom{k-1}{h-1}[F(x)]^{h-1}[1-F(x)]^{k-h} = 1$. Therefore,

$$E(\bar{X}_{RSS}) = \int_{-\infty}^{\infty} xf(x)dx = E[X] = \mu. \tag{2.12}$$

□

Theorem 2.2.2. *Wolfe (2012). The variance of the mean estimator based on RSS is always less than or equal to the variance of the mean estimator based on SRS,* $Var(\bar{X}_{RSS}) \leq Var(\bar{X}_{SRS})$.

Proof. We first write an equation which we use in the proof. Let μ_h denote the mean of hth order statistics, $E(X_h) = \mu_h$.

$$E\Big[(X_h - \mu)^2\Big] = E\Big[(X_h - \mu_h + \mu_h - \mu)^2\Big] = E\Big[(X_h - \mu_h)^2\Big] + (\mu_h - \mu)^2.$$

As we know that $E\Big[(X_h - \mu_h)^2\Big] = Var(X_h)$, then we can write

$$Var(X_h) = E\Big[(X_h - \mu)^2\Big] - (\mu_h - \mu)^2. \tag{2.13}$$

We also remember from the order statistics background that variances of hth ordered observations of random sets of n cycle chosen from the same population are equal.

$$Var(X_{(h)1}) = Var(X_{(h)2}) = \cdots = Var(X_{(h)n}) = Var(X_{(h)}) \tag{2.14}$$

Now, we can write the variance of $\bar{X}_{RSS}$ using eq. 2.14 and eq. 2.13 as follows (remembering that the observations are mutually independent).

$$Var(\bar{X}_{RSS}) = \frac{1}{k^2n^2}\sum_{h=1}^{k} nVar(X_{(h)}) = \frac{1}{k^2n}\sum_{h=1}^{k}\left(E\Big[(X_h - \mu)^2\Big] - (\mu_h - \mu)^2\right),$$

$$= \frac{1}{k^2n}\sum_{h=1}^{k} E\Big[(X_h - \mu)^2\Big] - \frac{1}{k^2n}\sum_{h=1}^{k}(\mu_h - \mu)^2. \tag{2.15}$$

We can rewrite the first part of the eq. 2.15 in a similar way as the previous proof,

$$\frac{1}{k^2n}\sum_{h=1}^{k} E\Big[(X_h - \mu)^2\Big] = \frac{1}{k^2n}\sum_{h=1}^{k}\int_{-\infty}^{\infty} k(x-\mu)^2\binom{k-1}{h-1}[F(x)]^{h-1}[1-F(x)]^{k-h}f(x)dx,$$

$$= \frac{1}{kn} \int_{-\infty}^{\infty} (x-\mu)^2 f(x) \left[\sum_{h=1}^{k} \binom{k-1}{h-1} [F(x)]^{h-1} [1-F(x)]^{k-h} \right] dx,$$

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

$$= \frac{1}{kn} \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx = \frac{\sigma^2}{kn} = Var(\bar{X}_{SRS}). \tag{2.16}$$

Finally the proof is completed as follows.

$$Var(\bar{X}_{RSS}) = Var(\bar{X}_{SRS}) - \frac{1}{k^2 n} \sum_{h=1}^{k} (\mu_h - \mu)^2. \tag{2.17}$$

□

Although the theorems are generally demonstrated under the restrictive assumption of perfect ranking, Dell & Clutter (1972) show that the mean estimator based on the (balanced) RSS is an unbiased estimator for the population mean even though rankings are imperfect.

2.3 Modified RSS methods

Although the main point in this thesis concerns the balanced form, to a lesser degree we also give some other forms of RSS, in the literature. Several researcher proposed modifications on the sampling scheme of RSS procedure. The common point of all modified methods is collecting unequal number of units for each of the ranks. For those cases where the underlying distribution of the population is heavy-tailed, unimodal or similar, it may be more useful to use a modified RSS method rather than the balanced one.

2.3.1 Median RSS

The median ranked set sampling (Median RSS) is suggested by Muttlak (1997). In this method, only medians of the random sets are chosen to the sample for estimation of population mean. For the odd set sizes, it is quite easy to chose the median of the each set. For even set sizes, the $m/2$ th ranked units are chosen from the first $m/2$ sets and the $(m+2)/2$ ranked units are selected from the remaining $m/2$ sets. Others are discarded and only selected units are measured. If necessary, procedure can be repeated r times as a cycle and we have $n = mr$ sample of size The procedure for one cycle and the case of $m = 3$ can be shown as below:

$$\begin{bmatrix} Y_{(1)1} \leq & \mathbf{Y_{(2)1}} \leq & Y_{(3)1} & \rightarrow & \mathbf{X_{[2]1}} \\ Y_{(1)2} \leq & \mathbf{Y_{(2)2}} \leq & Y_{(3)2} & \rightarrow & \mathbf{X_{[2]2}} \\ Y_{(1)3} \leq & \mathbf{Y_{(2)3}} \leq & Y_{(3)3} & \rightarrow & \mathbf{X_{[2]3}} \end{bmatrix}$$

Muttlak (1998) used Median RSS to estimate the population mean and regression coefficients for the case for which the ranking is done based on a concomitant variable. Hossain & Muttlak (2008) introduced some test procedures for scale parameters of exponential and rectangular distributions using the Median RSS method. Al-Omari (2012) define two modified ratio estimators for the population mean using auxiliary information in simple random sampling and Median RSS. Ibrahim et al. (2010) also extended Median RSS for stratified sampling.

2.3.2 Percentile RSS

Muttlak (2003b) suggested another modification for the RSS, Percentile RSS, where only the upper and lower percentiles of the random sets are chosen to the sample for determined value of percentile (p). The procedure of the Percentile RSS is similar to the even-set-size procedure of the MRSS. Suppose that m random sets with size m are

chosen from a specific population to sample m units and ordered visually or with an auxiliary variable. In PRSS procedure, the units which represent the lower percentile are chosen from the first $(m/2)$ sets and the units which represent the upper percentile are chosen from the other $(m/2)$ sets. If m is an odd number, the average of the first and third quartiles can be chosen as the final unit from the final set. Example for one cycle and the case of $m = 5$ and $p = 0.3$ can be represented as below for odd set size.

$$\begin{bmatrix} Y_{(1)1} \leq & \mathbf{Y_{(2)1}} \leq & Y_{(3)1} \leq & Y_{(4)1} \leq & Y_{(5)1} & \rightarrow & \mathbf{X_{[2]1}} \\ Y_{(1)2} \leq & \mathbf{Y_{(2)2}} \leq & Y_{(3)2} \leq & Y_{(4)2} \leq & Y_{(5)2} & \rightarrow & \mathbf{X_{[2]2}} \\ Y_{(1)3} \leq & Y_{(2)3} \leq & Y_{(3)3} \leq & \mathbf{Y_{(4)3}} \leq & Y_{(5)3} & \rightarrow & \mathbf{X_{[4]3}} \\ Y_{(1)4} \leq & Y_{(2)4} \leq & Y_{(3)4} \leq & \mathbf{Y_{(4)4}} \leq & Y_{(5)4} & \rightarrow & \mathbf{X_{[4]4}} \\ Y_{(1)5} \leq & Y_{(2)5} \leq & \mathbf{Y_{(3)5}} \leq & Y_{(4)5} \leq & Y_{(5)5} & \rightarrow & \mathbf{X_{[3]5}} \end{bmatrix}.$$

Muttlak (2003a) also suggested another modification for the RSS, Quartile RSS, where only the upper and lower quartiles of the random sets are chosen for the sample. Simply, Percentile RSS method can be considered the generalized version of Quartile RSS. Muttlak & Abu-Dayyeh (2004) introduced weighted version of Percentile RSS. Quartile RSS procedure was also considered to estimate the distribution function of a random variable by Al-Omari (2016).

2.3.3 Extreme RSS

Extreme RSS is another modification of RSS suggested by Samawi et al. (1996) to estimate the population mean only using maximum or minimum ranked unit from each set. For estimation based on Extreme RSS, the following procedure can be used. The selected m random sets each of size m units from the population can be ranked within each set with respect to a variable of interest by human expert or any inexpensive method. If the set size m is even, the lowest ranked unit of each set is selected from the first $m/2$ sets and the largest ranked unit of each set is chosen from the other $m/2$ sets. If the set size is odd, select the lowest ranked unit from the first $(m-1)/2$ sets,

select the largest ranked unit from the other $(m-1)/2$ sets and median is taken from the remaining last set. When the size of ERSS is increased by cycling the procedure r times, then we have $n = mr$ sample size. The procedure for one cycle and case of $m = 4$ can be shown as below:

$$\begin{bmatrix} \mathbf{Y_{(1)1}} \leq & Y_{(2)1} \leq & Y_{(3)1} \leq & Y_{(4)1} & \rightarrow & \mathbf{X_{[1]1}} \\ \mathbf{Y_{(1)2}} \leq & Y_{(2)2} \leq & Y_{(3)2} \leq & Y_{(4)2} & \rightarrow & \mathbf{X_{[1]2}} \\ Y_{(1)3} \leq & Y_{(2)3} \leq & Y_{(3)3} \leq & \mathbf{Y_{(4)3}} & \rightarrow & \mathbf{X_{[4]3}} \\ Y_{(1)4} \leq & Y_{(2)4} \leq & Y_{(3)4} \leq & \mathbf{Y_{(4)4}} & \rightarrow & \mathbf{X_{[4]4}} \end{bmatrix}$$

Al-Saleh & Samawi (2010) introduced Moving Extreme RSS on estimating the odds. Al-Nasser & Mustafa (2009) also extended Extreme RSS for robust case.

2.3.4 Balanced groups RSS

Balanced groups ranked set samples method (BGRSS) is can be defined as the combination of Extreme RSS and Median RSS. Jemain et al. (2008) suggested to use BGRSS for estimating the population mean with a special sample size $m = 3k$. In their study, BGRSS procedure is described as follows: $m = 3k$ (where $k = 1, 2, 3, \dots$) sets each of size m are select randomly from a specific population. Units in each set are ranked visually or using another inexpensive method. Sets are randomly allocating into three groups. The k smallest units from the first group, k median units from the second group and k largest units from the third group of ranked sets are chosen. In the second group, when the set size is odd, the median unit is defined as the $((m+1)/2)$th ranked unit in the set and when the set size is even, the median unit is defined as the mean of the $(m/2)$th and the $((m+2)/2)$th ranked units. If k is an even number, the researcher takes the averages of the two units as median unit. BGRSS process for $k = 2$ can be described as follows:

$$\begin{bmatrix} \mathbf{Y_{(1)1}} \leq & Y_{(2)1} \leq & Y_{(3)1} \leq & Y_{(4)1} \leq & Y_{(5)1} \leq & Y_{(6)1} & \rightarrow & \mathbf{X_{[1]1}} \\ \mathbf{Y_{(1)2}} \leq & Y_{(2)2} \leq & Y_{(3)2} \leq & Y_{(4)2} \leq & Y_{(5)2} \leq & Y_{(6)2} & \rightarrow & \mathbf{X_{[1]2}} \\ Y_{(1)3} \leq & Y_{(2)3} \leq & \mathbf{Y_{(3)3}} \leq & \mathbf{Y_{(4)3}} \leq & Y_{(5)3} \leq & Y_{(6)3} & \rightarrow & \frac{\mathbf{X_{[3]3}}+\mathbf{X_{[4]3}}}{2} \\ Y_{(1)4} \leq & Y_{(2)4} \leq & \mathbf{Y_{(3)4}} \leq & \mathbf{Y_{(4)4}} \leq & Y_{(5)4} \leq & Y_{(6)4} & \rightarrow & \frac{\mathbf{X_{[3]4}}+\mathbf{X_{[4]4}}}{2} \\ Y_{(1)5} \leq & Y_{(2)5} \leq & Y_{(3)5} \leq & Y_{(4)5} \leq & Y_{(5)5} \leq & \mathbf{Y_{(6)5}} & \rightarrow & \mathbf{X_{[6]5}} \\ Y_{(1)6} \leq & Y_{(2)6} \leq & Y_{(3)6} \leq & Y_{(4)6} \leq & Y_{(5)6} \leq & \mathbf{Y_{(6)6}} & \rightarrow & \mathbf{X_{[6]6}} \end{bmatrix}.$$

2.4 Studies focused on imperfect ranking and RSS related approaches

In applications of RSS, the efficiency of the estimation is directly related to how well the rankings of the units are done. Although perfect rankings are the goal of any RSS procedure, it is generally not feasible in practice. It means an unavoidable uncertainty in the proper applications of RSS. In the literature, there are studies focused on the case of imperfect ranking and modeling this uncertainty. In this thesis, we categorize these studies with respect to the usage of the model of imperfect ranking. In the first category, there are some studies which develop statistical models to capture the uncertainty of the ranking procedure and use this model to research the impact of imperfect ranking on inferences. The studies in this category are widely-examined under the topic *imperfect ranking*. We also define a second category including some RSS inspired studies because the studies in this category introduce methods to model the uncertainty and to use the model in inference for special cases of imperfect ranking.

2.4.1 Studies focused on imperfect ranking

Dell & Clutter (1972) introduced a pioneering approach for the case of imperfect ranking. The units in the random sets are ranked based on a concomitant variable which

is constructed through an additive model of the actual measurement. The values of the concomitant variable are generated by adding the values of the interested variable and random error is drawn from a parent distribution (generally $N(0, \sigma^2)$). The level of *imperfectness* is controlled by the variance of the error term. Let X be the variable of interest and Y be the concomitant variable using for ranking in the study. Let also x_1, $x_2, \cdots, x_w$ be the true values of the units which are obtainable by accurate measurement in a real life application. Then the x's and y's may be regarded as realizations of linked random variables X_i , Y_i

$$Y_i = g(X_i, Z_i) \text{ for } i = 1, 2, ..., w,$$

where Z_i is defined as judgement error term for a function of g. Then the concomitant variable is computed with a simple formula

$$Y_i = X_i + Z_i \text{ for } i = 1, 2, ..., w, \tag{2.18}$$

where X_1, X_2, $\cdots$, X_w, Z_1, Z_2, $\cdots$, Z_w are mutually independent. Dell & Clutter (1972) also mentioned a theoretical approach to the calculation of the relative efficiency (of RSS) and applications for the case $Z \sim N(0, \sigma^2)$ and $X \sim N(0, 1)$). Dell & Clutter (1972) and David & Levine (1972) used this approach to construct a simulation study to see the impact of imperfect ranking on mean estimation. Because it is simple and easy to apply, Dell & Clutter (1972) approach is widely-used in the simulation part of RSS studies.

Bohn & Wolfe (1994) introduced another approach to model imperfect ranking where probability matrices are used to estimate the quality of judgment ranking information. The distributions of the judgment order statistics is considered mixtures of distributions of the true order statistics and the probabilities of imperfect judgment rankings is determined based on the expected spacings between order statistics in this approach. Simply, the expected difference, $E(X_i - X_j)$, between the order statistics X_i and X_j is called expected spacing. The main idea is that the absolute value of

expected spacing will increase and the probability of mis-specifying decreases as the quantity $|i-j|$ increases. Bohn & Wolfe (1994) preferred to use a matrix notation to show the probabilities as follows,

$$P = \begin{bmatrix} p_{1,1} & p_{1,2} & p_{1,3} & p_{1,4} \\ p_{2,1} & p_{2,2} & p_{2,3} & p_{2,4} \\ p_{3,1} & p_{3,2} & p_{3,3} & p_{3,4} \\ p_{4,1} & p_{4,2} & p_{4,3} & p_{4,4} \end{bmatrix}$$

where $p_{i,j} \geq 0$ is the probability that the jth ordered unit in a set is assigned a judgment rank i. The Bohn & Wolfe (1994) approach indicates that the distribution of each judgment class population ($F_{[i]}(t)$, $i = 1, 2, \ldots, k$) can be expressed as a linear combination of the order statistics in a set of size k,

$$F_{[i]}(t) = \sum_{j=1}^{k} p_{i,j} F_{(j)}(t) \text{ for } i = 1, 2, ..., k. \tag{2.20}$$

The assigning probabilities, $p_{i,j}$'s, have some features. First, the sum of the probabilities of a judgment rank i must be equal to 1, $p_{i,1} + p_{i,2} + ... + p_{i,k} = 1$. Second, the sum of the probabilities of the jth ordered unit must also be equal to 1, $p_{1,j} + p_{2,j} + \cdots + p_{k,j} = 1$. Clearly, these two features make the probability matrix P a double-stochastic matrix.

The Bohn & Wolfe (1994) approach draws considerable interest to the topic of imperfect ranking. Aragon et al. (1999) introduced an advanced approach for the construction of the probability matrix P. Presnell & Bohn (1999) used this model to show that the RSS procedure is asymptotically at least as efficient as the SRS, even if the ranking is imperfect. Frey (2007) described a new class of models for imperfect rankings as an extension of the Bohn & Wolfe (1994) approach and argues whether it contains essentially any reasonable imperfect rankings model. Ozturk (2008) proposed inference techniques for RSS data in the case of imperfect ranking using the

models of Bohn & Wolfe (1994) and Frey (2007) and gave an application for constructing a confidence interval for the population median. There is also a less-known alternative to the Bohn & Wolfe (1994) approach in modelling the imperfect ranking. Fligner & MacEachern (2006) introduced a new model for ranking information through the monotone likelihood ratio principle.

Finally we can say that the common underlying structure in all of these approaches is stochastic ordering of judgment ranking classes that leads to a reasonable ranking mechanism (Ozturk, 2008). They also share a common purpose, modelling the uncertainty in the case of imperfect ranking and using the model to investigate the impact of imperfect ranking on estimation.

2.4.2 Judgment Post Stratification(JPS) Method

Judgment Post Stratification (JPS) is the first method in the second category, RSS related approaches, of the studies focusing on imperfect ranking. MacEachern et al. (2004) introduced the JPS sampling method as a combination of SRS and RSS for a specific case. Briefly, there is a simple random sample which is already chosen and the rank information of each sampled unit is obtained individually by way of random sets using the same method as in RSS (see Figure 2.4). Finally, the simple random sample and the rank information of the units in the sample are used together in estimation.

$$\boldsymbol{X_1} \leftarrow X_{1[1]}, X_{2[1]}, \dots, X_{k-1[1]}$$
$$\boldsymbol{X_2} \leftarrow X_{1[2]}, X_{2[2]}, \dots, X_{k-1[2]}$$
$$\dots$$
$$\boldsymbol{X_n} \leftarrow X_{1[n]}, X_{2[n]}, \dots, X_{k-1[n]}$$

Figure 2.4 Illustration of the JPS method for sample size n

Let $X_1, X_2, \cdots, X_n$ be a simple random sample with size n. Let us also take n random sets (1 set for each unit) including $k-1$ auxiliary units (which is not fully

measured) from the same population to get rank information, R_i for $i = 1, 2, ..., n$. If we consolidate the sampled units with the random sets of auxiliary unit, we will have n sets with size k (and thus k ranks or post-strata). Finally, we obtain the judgment post-stratified sample consists of pairs (X_i, R_i) through ranking the sets. For cases where the rankings are perfect, the estimator of population mean based on JPS is defined as follows.

$$\bar{X}_{JPS-P} = \frac{1}{k}\sum_{h=1}^{k}\frac{\sum_{i=1}^{n} X_i R_{ih}}{\sum_{i=1}^{n} R_{ih}} \text{ for } i = 1, 2, ..., n \text{ and } h = 1, 2, ..., k, \tag{2.21}$$

where R_{ih} is an indicator variable as $R_{ih} = I(R_i = h)$. It can be seen that $\bar{X}_{JPS-P}$ formula is actually the same with RSS estimator for the case the rankings are perfect. However, as in the application of RSS, the ranking of the units should be done without actual measurement in the application of JPS. Therefore MacEachern et al. (2004) proposed another estimator (indeed, three estimators but they don't recommend the other two) for the case of imperfect ranking.

$$\bar{X}_{JPS-I} = \frac{1}{k}\sum_{h=1}^{k}\frac{\sum_{i=1}^{n} p_{ih} X_i}{\sum_{i=1}^{n} p_{ih}} \text{ for } i = 1, 2, ..., n \text{ and } h = 1, 2, ..., k, \tag{2.22}$$

where p_{ih} is an auxiliary variable as $0 \leq p_{ih} \leq 1$ and $\sum_{h=1}^{k} p_{ih} = 1$ for all i and h. In this case, p_{ih} represents the ranker's strength of belief that the unit X_i has rank h. Also $p_{ih}X_i$ assigns a proportion ($\frac{p_{ih}}{\sum_{i=1}^{n} p_{ih}}$) of X_i to post-stratum h in the estimation.

The main advantage of JPS is that it uses simple random samples with additional information of ranking and increase the efficiency of the estimation with this additional information. It is a real problem that some researchers are worried whether subject

matter journals would consider the sampling method to be nonstandard when they use RSS. Therefore, using a simple random sample may be useful for researchers who are strictly loyal to SRS in their researches. For further studies on the topic JPS, see Wang et al. (2006), Ozturk (2011a) for the multiple ranker case and Wang et al. (2008), Frey & Ozturk (2011), Frey & Feeman (2012) for estimations of different parameters.

2.4.3 Partially ranked ordered sets (PROS) method

Another RSS related approach, the Partially Ranked Ordered Sets (PROS) sampling method was introduced by Ozturk (2011c). Rankers are allowed to declare any two or more units are tied in rank whenever the units cannot be ranked with total confidence. The fully measured units are then selected from partially ordered judgment subsets where the units are replaced. In the PROS method, ties are the main cause of imprecise ranking decisions in sampling design and probability is prorated proportionally to the ranks which subsets hold. This approach adds flexibility to improved precision for RSS estimation procedures in settings where it is difficult to distinguish one unit from one or more others in a set for a rank decision.

The PROS method has a relatively complex procedure and notation. It starts with the decision of the sample size N and the set size M. Then N sets of M units are randomly chosen from a population having the distribution F, mean μ and variance σ^2. Suppose that the researcher has a problem to distinguish units for a rank decision in each of N sets. Therefore, he/she makes subsets for each set and ranks them. Ozturk (2011c) also defines a *cycle* structure where N sets are divided into K cycles. The notation of sets and subsets will be denoted as $S_{r,i} = s_{r[1]i}, s_{r[2]i}, ..., s_{r[k_i]i}$, where i shows the cycle number and k_i is the number of subset in the set $S_{r,i}$. The value of k_i varies from set to set and depends on the *ranking ability* of the researcher. The structure of the PROS method is ready for sampling (see Figure 2.5, Ozturk (2011c)).

Group (i)	$S_{r,i}$	Judgment subsets	m_{hi}	Obs
a) Design G, $N = 5$, $K = 2$, $M = 10$				
1	$\{\mathbf{s_{1[1]1}}, s_{1[2]1}\}$	{**1**, **2**, **3**, **4**, **5**}, {6, 7, 8, 9, 10}	5	$X_{s_{1[1]1}}$
	$\{s_{2[1]1}, \mathbf{s_{2[2]1}}, s_{2[3]1}\}$	{1, 2, 3}, {**4**, **5**, **6**, **7**}, {8, 9, 10}	4	$X_{s_{2[2]1}}$
	$\{s_{3[1]1}, s_{3[2]1}, \mathbf{s_{3[3]1}}\}$	{1, 2, 3}, {4, 5}, {**6**, **7**, **8**, **9**, **10**}	5	$X_{s_{3[3]1}}$
2	$\{\mathbf{s_{1[1]2}}, s_{1[2]2}, s_{1[3]2}\}$	{**1**, **2**, **3**}, {4, 5, 6, 7}, {8, 9, 10}	3	$X_{s_{1[1]2}}$
	$\{s_{2[1]2}, \mathbf{s_{2[2]2}}\}$	{1, 2, 3, 4, 5, 6, 7}, {**8**, **9**, **10**}	3	$X_{s_{2[2]2}}$
*b) Design G***, $N = 5$, $K = 2$, $M = 6$				
1	$\{\mathbf{s_{1[1]1}}, s_{1[2]1}, s_{1[3]1}\}$	{**1**, **2**}, {3, 4}, {5, 6}	2	$X_{s_{1[1]1}}$
	$\{s_{2[1]1}, \mathbf{s_{2[2]1}}, s_{2[3]1}\}$	{1, 2}, {**3**, **4**}, {5, 6}	2	$X_{s_{2[2]1}}$
	$\{s_{3[1]1}, s_{3[2]1}, \mathbf{s_{3[3]1}}\}$	{1, 2}, {3, 4}, {**5**, **6**}	2	$X_{s_{3[3]1}}$
2	$\{\mathbf{s_{1[1]2}}, s_{1[2]2}\}$	{**1**, **2**, **3**}, {4, 5, 6}	3	$X_{s_{1[1]2}}$
	$\{s_{2[1]2}, \mathbf{s_{2[2]2}}\}$	{1, 2, 3}, {**4**, **5**, **6**}	3	$X_{s_{2[2]2}}$
*c) Design G****, $N = 4$, $K = 2$, $M = 6$				
1	$\{\mathbf{s_{1[1]1}}, s_{1[2]1}\}$	{**1**, **2**, **3**}, {4, 5, 6}	3	$X_{s_{1[1]1}}$
	$\{s_{2[1]1}, \mathbf{s_{2[2]1}}\}$	{1, 2, 3}, {**4**, **5**, **6**}	3	$X_{s_{2[2]1}}$
2	$\{\mathbf{s_{1[1]2}}, s_{1[2]2}\}$	{**1**, **2**, **3**}, {4, 5, 6}	3	$X_{s_{1[1]2}}$
	$\{s_{2[1]2}, \mathbf{s_{2[2]2}}\}$	{1, 2, 3}, {**4**, **5**, **6**}	3	$X_{s_{2[2]2}}$

Figure 2.5 Illustration of the PROS method for three different designs (Ozturk, 2011c)

$X_{s_{r[h]i}}$ represents the sampled unit which is chosen randomly from hth ordered subset of rth set in cycle i. (Because of the design of the method, a unit is sampled when $r = h$.) There is one additional notation, m_{hi}, in Figure 2.5. m_{hi} is the size of the subset in the hth set of cycle i from where the sampled unit is chosen from. Ozturk (2011c) also gave a definition for a "balanced design" of PROS method as follows.

Definition 2.4.1. *(Ozturk (2011c).) A PROS sampling design is called balanced if*

$$\bigcup_{i=1}^{K} \left(\bigcup_{h=1}^{k_i} s_{h[h]i} \right) = KM, \tag{2.23}$$

where K is the cycle size and M is the set size. According to this definition, all of the three designs in Figure 2.5 are balanced designs of PROS. The definition 2.4.1 can also be written with an easier way.

Definition 2.4.2. *A PROS sampling design is called balanced if*

$$\sum_{i=1}^{K} \sum_{h=1}^{k_i} m_{hi} = KM. \tag{2.24}$$

Based on a balanced design of the PROS method, the unbiased estimator of the population mean is given as,

$$\bar{X}_{PROS} = \frac{1}{KM} \sum_{i=1}^{K} \sum_{h=1}^{k_i} m_{hi} X_{s_{h[h]i}}. \tag{2.25}$$

When $\left(\frac{m_{hi}}{KM}\right)$ is defined as the weight of $X_{s_{h[h]i}}$, this weight assigns a proportion of sampled unit $X_{s_{h[h]i}}$ to stratum h proportional to size of the subset, m_{hi}. Clearly, a balanced design of PROS becomes balanced RSS for $m_{hi} = 1$ and SRS for $m_{hi} = M$.

2.4.4 Combining the information coming from multiple rankers

In the literature, some studies focused on using multiple rankers as another way of reducing ranking error. Ridout (2003) counted on a multiple attributes (rankers) case for the selection of sample units in RSS. He proposed a new method to sample the units in the sets according to the rank decisions of all attributes. Ridout (2003) introduced a new method that sought to achieve samples that are nearly balanced with respect to the rank decisions of all rankers. However this new method based on multiple rankers is shown to result in loss of precision compared to single ranker case. Ozturk (2011a) proposed a more powerful and easier way to combine the ranking information of multiple observers with a *strength of agreement measure* for RSS and JPS. Simply, each ranker makes a decision about the rank of the units and gives 1 to a suitable rank and 0 to the others. Then the average of the rank decisions from multiple rankers is taken as the *strength of agreement measure* for each unit. Let X_1, X_2, $\cdots$, X_k be a random set with size k chosen for RSS. Let the rankers decide the rank of the units in the set and $w_h^c(r)$ denotes the rank decision of ranker c for unit X_h to rank r, where $c = 1, 2, ..., C$, $h = 1, 2, ..., k$ and $r = 1, 2, ..., k$. Then we have C rank decision matrices for C ranker for this single set. For $k = 4$ it will be as follows.

$$W^c = \begin{bmatrix} w_1^c(1) & w_1^c(2) & w_1^c(3) & w_1^c(4) \\ w_2^c(1) & w_2^c(2) & w_2^c(3) & w_2^c(4) \\ w_3^c(1) & w_3^c(2) & w_3^c(3) & w_3^c(4) \\ w_4^c(1) & w_4^c(2) & w_4^c(3) & w_4^c(4) \end{bmatrix} \text{ for instance } \begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Let $\bar{w}_h(r)$ shows the *strength of agreement measure* for unit X_h to rank r in the set. It is defined as

$$\bar{w}_h(r) = \frac{1}{C} \sum_{c=1}^{C} w_h^c(r). \tag{2.27}$$

Finally we have a *strength of agreement measure* matrix, $\bar{W}$, for a single set and the unit which has the biggest $\bar{w}$ value for rank r is chosen for the sample for rank r. (For example the unit which has $max(\bar{w}_1(1), \bar{w}_2(1), ..., \bar{w}_k(1))$ is chosen for rank r).

$$\bar{W} = \begin{bmatrix} \bar{w}_1(1) & \bar{w}_1(2) & \bar{w}_1(3) & \bar{w}_1(4) \\ \bar{w}_2(1) & \bar{w}_2(2) & \bar{w}_2(3) & \bar{w}_2(4) \\ \bar{w}_3(1) & \bar{w}_3(2) & \bar{w}_3(3) & \bar{w}_3(4) \\ \bar{w}_4(1) & \bar{w}_4(2) & \bar{w}_4(3) & \bar{w}_4(4) \end{bmatrix}$$

When we repeat the process for k times for each of the ranks, one unit will be sampled for each rank. It is called a cycle and we can repeat the cycle for n times to get a ranked set sample with size $k \times n$. We note that Ozturk (2011a) preferred not to use cycle structure. Instead, he suggested repeating the single set process N times (where N is the sample size and it is equal to $k \times n$ in cycle notation) and cautioned that researcher must chose equal number of units for each rank to have a balanced RSS. With the notation we prefer in this study (cycle notation), the estimator of the population mean is given by

$$\bar{X}_{RSS-M} = \frac{1}{k}\sum_{r=1}^{k} \frac{\sum_{j=1}^{n}\sum_{h=1}^{k} \bar{w}_{h,j}(r) X_{[h]j}}{\sum_{j=1}^{n}\sum_{h=1}^{k} \bar{w}_{h,j}(r)}. \tag{2.29}$$

The main idea of the estimator $\bar{X}_{RSS-M}$ came from the study of MacEachern et al. (2004). The eq. (2.29) is a transformed version of eq. (2.22). Ozturk (2011a) also adapted the strength of agreement measure to JPS in the same study and PROS design in Ozturk (2013) for multiple ranker case. For further studies on multiple ranker case, see Murray et al. (2000), Wang et al. (2006).

CHAPTER THREE
FUZZY-WEIGHTED RANKED SET SAMPLING

In the application of RSS, ranking is performed via two main ways: i) One (or more) human ranker(s) can decide the rank of a unit in a set by visual inspection taking the other units in the set into account, or ii) one (or more) concomitant variable(s) can be used for ranking. In both ways, ranking must be done without actual measurement of the variable of interest and ranking cost must be smaller than the cost to measure the variable of interest. Since the ranker makes the decision of ranks without actual measurement, these decisions could not always be perfect even if ranking is done using powerful criteria with a high relationship. Clearly, there is an uncertainty in the ranking mechanism and it causes a loss of information. In previous chapter, we have given information about the approaches focusing on this uncertainty from a probabilistic perspective. In this chapter, we introduce a different perspective using fuzzy sets approach.

3.1 A brief introduction to fuzzy set theory

Zadeh (1965) introduced fuzzy sets for representing and manipulating data when they are not precise. With a simple definition, Zadeh (1965) transformed the structure of membership in classical set theory from a binary categorization (0-1) into a more flexible one which allowed partial membership with a degree between 0 and 1. While a classical set has a black or white form, a fuzzy set can include the grey zones. Fuzzy set theory has transformed not only set theory but also logic, reasoning and inference by taking into account the human subjectivity and imprecision. Thus there are a group of famous topics which are developed out of fuzzy set theory such as *fuzzy logic*, *approximate reasoning*, *fuzzy inference*, *possibility theory*, *soft computing*, etc. For detailed information, see Zimmermann (2001), the literature review of Kahraman et al. (2016) and the theoretical review of Zimmermann (2010) for general fuzzy set theory; Ross (2016) and the review of Mohd Adnan et al. (2015) for fuzzy logic and

applications; Dubois & Prade (1988) for possibility theory; Chaturvedi (2008) for soft computing methods and applications.

3.1.1 Fundamentals of fuzzy set theory

A fuzzy set is a natural extension of a classical set. Thus it needs a reference set (called universe of discourse) which always has a classical-crisp form. Let $X = \{x_1, x_2, ..., x_n\}$ be a classical set of units x_i. Then the fuzzy set is defined as follows.

Definition 3.1.1. *Zimmermann (2010). If X is a collection of objects denoted generically by x, then a fuzzy set $\tilde{A}$ in X is a set of ordered pairs:*

$$\tilde{A} = (x, \mu_{\tilde{A}}(x))|x_i \in X,$$

where $\mu_{\tilde{A}}$ is is called the membership function (generalized characteristic function) of $\tilde{A}$ and maps X to the membership space M.

The membership function plays an important role in the definition of fuzzy sets which is stated in the next definition.

Definition 3.1.2. *Bilgiç & Türkşen (2000). A fuzzy (sub)set, say $\tilde{A}$, has a membership function $\mu_{\tilde{A}}$ defined as a function from a well defined universe (the referential set), X, into the unit interval as:*

$$\mu_{\tilde{A}} : X \to [0, 1].$$

The degree of membership in a fuzzy set is represented by a number between 0 and 1. 0 shows the unit is entirely not in the set, 1 shows the unit completely in the set and there are an infinite number of membership degrees between 0 and 1. The main role of the membership function is to represent human perceptions and decisions, which are usually individual and subjective, as a member of a fuzzy set. The most commonly used membership functions in practice are triangular, trapezoidal and Gaussian type membership functions because of their ease of use for arithmetic and fuzzy operations.

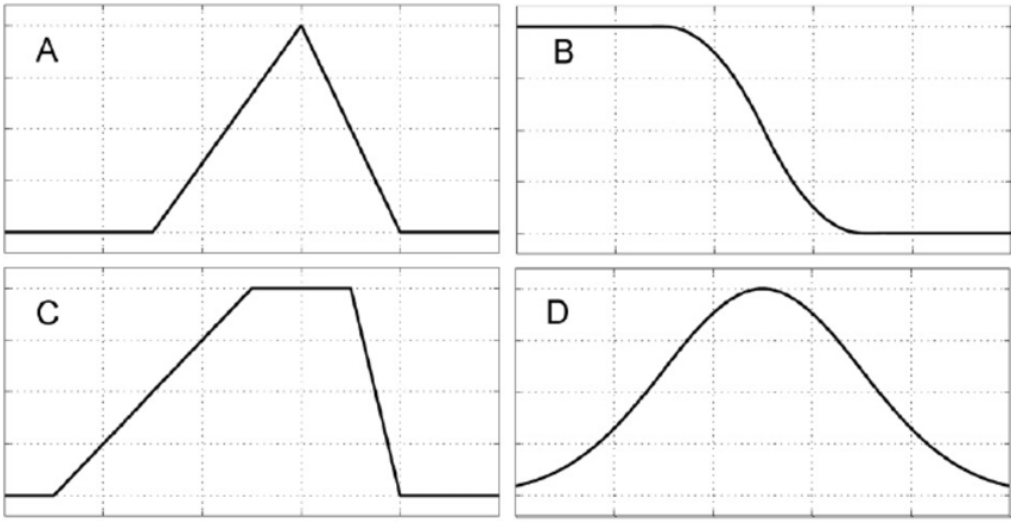

Figure 3.1 Different types of membership functions. (A) Triangular, (B) Z-shaped, (C) Trapezoidal, (D) Gaussian

The existing properties of the classical sets, such as emptiness, equality, inclusion, cardinality and so on are also extended to the fuzzy set concept.

Definition 3.1.3. *Werro (2015). Defined over the classical set* $X = x_1, x_2, ..., x_n$*, a fuzzy set* $\tilde{A}$ *is empty if*

$$\mu_{\tilde{A}}(x_i) = 0, \forall x_i \in X.$$

Definition 3.1.4. *Werro (2015). Two fuzzy sets* $\tilde{A}$ *and* $\tilde{B}$ *defined over the same reference set* X*, are equal if*

$$\mu_{\tilde{A}}(x_i) = \mu_{\tilde{B}}(x_i), \forall x_i \in X.$$

Definition 3.1.5. *Werro (2015). Two fuzzy sets* $\tilde{A}$ *and* $\tilde{B}$ *are defined over the same reference set* X*.* $\tilde{A}$ *can be said to be included in* $\tilde{B}$ *if*

$$\mu_{\tilde{A}}(x_i) \leq \mu_{\tilde{B}}(x_i), \forall x_i \in X.$$

The cardinality of a classical set is defined as the total number of units in the set. A unit is a member of a fuzzy set with a membership degree. Thus, cardinality is defined as the sum of the membership degrees of the units in the set.

Definition 3.1.6. *Werro (2015). Cardinality of fuzzy sets $\tilde{A}$ defined over set X is*

$$Card(\tilde{A}) = \sum_{x_i \in X} \mu_{\tilde{A}}(x_i).$$

Fuzzy set theory allows us to define both convex and non-convex fuzzy sets. But we note that convex fuzzy sets are used in the great majority of the analysis.

Definition 3.1.7. *Zimmermann (2010). A fuzzy set $\tilde{A}$ which is defined over set X is convex if*

$$\mu_{\tilde{A}}(\lambda x_i + (1-\lambda)x_j) \geq min(\mu_{\tilde{A}}(x_i), \mu_{\tilde{A}}(x_j)), \forall x_i, x_j \in X, \lambda \in [0,1].$$

The properties, *support*, α-*cut* and *kernel* are widely used in fuzzy set operations. In particular, α-*cut*s are important for application of fuzzy sets because any fuzzy set can also be described by using its α-*cut*s.

Definition 3.1.8. *Werro (2015). The support of a fuzzy set $\tilde{A}$ defined over a crisp set X is a crisp subset of X where the membership degree of the units is positive. That is*

$$Supp(\tilde{A}) = \{x_i \in X, \mu_{\tilde{A}}(x_i) > 0\}$$

Definition 3.1.9. *Werro (2015). The kernel of a fuzzy set $\tilde{A}$ defined over the crisp set X is a crisp subset of X where the membership degree of the units is equal to* 1.

$$Kern(\tilde{A}) = \{x_i \in X, \mu_{\tilde{A}}(x_i) = 1\}$$

Definition 3.1.10. *Werro (2015). The α-cut of a fuzzy set $\tilde{A}$ defined over the crisp set X is a crisp subset of X where the membership degree of the units are equal or greater then α.*

$$\tilde{A}_\alpha = \{x_i \in X, \mu_{\tilde{A}}(x_i) \geq \alpha\}$$

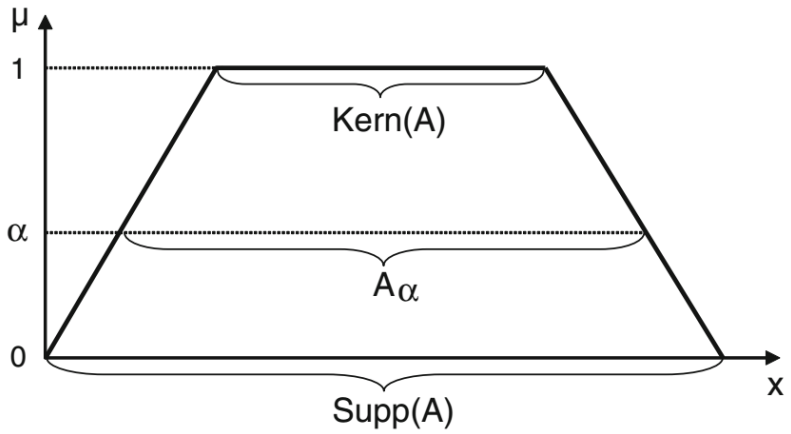

Figure 3.2 Illustration of support, kernel and α-cut of a fuzzy set, Werro (2015)

3.1.2 Operations on fuzzy sets

Complement, *intersection* and *union* are operations of classical set theory. Each of the operation is extended to a large class of operations in the flexible structure of fuzzy set theory. We have provided some basic definitions and notable operations in this section. One can see Zimmermann (2010) and Chaturvedi (2008) for a detailed information about fuzzy set operations.

Definition 3.1.11. *The complement of a fuzzy set $\tilde{A}$ defined over a crisp set X is denoted as $\tilde{A}^c$ and defined as follows.*

$$\tilde{A}^c : \mu_{\tilde{A}^c}(x_i) = 1 - \mu_{\tilde{A}}(x_i), \forall x_i \in X.$$

In fuzzy set concept, the class of intersection operators is called *T-norm* (Triangular Norm). Zadeh (1965)'s *min* operation is the pioneer and the most famous of this class.

Definition 3.1.12. *Zimmermann (2010). The membership function of the Zadeh's intersection of two fuzzy sets $\tilde{A}$ and $\tilde{B}$ defined over a the crisp set X is defined as*

$$\tilde{A} \cap \tilde{B} : \mu_{\tilde{A} \cap \tilde{B}}(x_i) = min(\mu_{\tilde{A}}(x_i), \mu_{\tilde{B}}(x_i)), \forall x_i \in X. \tag{3.1}$$

Some of the typical T-norms are listed below to give an idea of the class of intersection operations. For two fuzzy sets $\tilde{A}$ and $\tilde{B}$ defined over a crisp set X, there are several ways to define intersection ($\tilde{A} \cap \tilde{B}$),

- **Drastic product**

$$\mu_{\tilde{A}\cap\tilde{B}}(x_i) = \begin{cases} min(\mu_{\tilde{A}}(x_i), \mu_{\tilde{B}}(x_i)), & if\ \ max(\mu_{\tilde{A}}(x_i), \mu_{\tilde{B}}(x_i)) = 1 \\ 0, & otherwise \end{cases}$$

- **Bounded difference**

$$\mu_{\tilde{A}\cap\tilde{B}}(x_i) = max[0, (\mu_{\tilde{A}}(x_i) + \mu_{\tilde{B}}(x_i) - 1)]$$

- **Einstein product**

$$\mu_{\tilde{A}\cap\tilde{B}}(x_i) = \frac{\mu_{\tilde{A}}(x_i)\mu_{\tilde{B}}(x_i)}{2 - [\mu_{\tilde{A}}(x_i) + \mu_{\tilde{B}}(x_i) - \mu_{\tilde{A}}(x_i) \times \mu_{\tilde{B}}(x_i)]}$$

In relation to *T-norm* operators, the class of union operators is called *T-conorm* (Triangular Conorm) or *s-norm*.

Definition 3.1.13. *Zimmermann (2010). The membership function of the Zadeh's union of two fuzzy sets $\tilde{A}$ and $\tilde{B}$ defined over a crisp set X is*

$$\tilde{A} \cup \tilde{B} : \mu_{\tilde{A}\cup\tilde{B}}(x_i) = max(\mu_{\tilde{A}}(x_i), \mu_{\tilde{B}}(x_i)), \forall x_i \in X.$$

Some of the typical T-conorms are also listed below. For two fuzzy sets $\tilde{A}$ and $\tilde{B}$ defined over a the crisp set X, there are several ways to define union ($\tilde{A} \cup \tilde{B}$).

- **Drastic sum**

$$\mu_{\tilde{A}\cup\tilde{B}}(x_i) = \begin{cases} max(\mu_{\tilde{A}}(x_i), \mu_{\tilde{B}}(x_i)) & if\ \ min(\mu_{\tilde{A}}(x_i), \mu_{\tilde{B}}(x_i)) = 0 \\ 1 & otherwise \end{cases}$$

- **Bounded sum**

$$\mu_{\tilde{A}\cup\tilde{B}}(x_i) = min[1, (\mu_{\tilde{A}}(x_i) + \mu_{\tilde{B}}(x_i))]$$

- **Einstein sum**

$$\mu_{\tilde{A}\cup\tilde{B}}(x_i) = \frac{\mu_{\tilde{A}}(x_i) + \mu_{\tilde{B}}(x_i)}{1 - \mu_{\tilde{A}}(x_i)\mu_{\tilde{B}}(x_i)}$$

T-norm and T-conorm operators are also called nonparameterized operators because there is no additional parameter in the formulation of the membership function, $\mu_{\tilde{A}\cup\tilde{B}}(x_i)$ or $\mu_{\tilde{A}\cap\tilde{B}}(x_i)$. There are also some *parameterized* operators, such as Hamacher union, Yager intersection, Dubois and Prade intersection and so on. (See Zimmermann (2010) for a brief information.) In any case, an operation should satisfy the properties listed below to be called as a T-norm or T-conorm operator.

Property	T-norm	T-conorm
Identity	$1 \wedge x = x$	$0 \vee x = x$
Commutativity	$x \wedge y = y \wedge x$	$x \vee y = y \vee x$
Associativity	$x \wedge (y \wedge z) = (x \wedge y) \wedge z$	$x \vee (y \vee z) = (x \vee y) \vee z$
Monotonicity	if $v \leq w$ and $x \leq y$ then $v \wedge x \leq w \wedge y$	$v \vee x \leq w \vee y$

Figure 3.3 Properties of *T-norm* and *T-conorm* operators, Werro (2015)

There is also an alternative class of averaging operations for both intersection and union. The averaging operations are generally more optimistic than T-norm operators and more pessimistic than T-conorm operators. Compensatory operators, Yager's OWA and Arithmetic mean-Average are the most famous in this class (Zimmermann (2010)). The averaging operators are also known as to be better suited for modeling aggregation in a human decision environment than T-norm and T-conorm operators. It should be noted that averaging operators do not have the mathematical properties of the T-norms and T-conorms.

3.2 Modelling the uncertainty using a fuzzy set perspective

In Chapter 2, we mentioned the unavoidable uncertainty/impreciseness in the application of RSS and introduced the studies focusing on this problem with probabilistic perspective. In this chapter, we introduce our new approach to deal with this uncertainty, a fuzzy inspired approach. Let us start with an example to provide motivation. Suppose that there is a single human-ranker to decide the rank of a unit in a random set. Based on his/her degree of belief, the ranker assigns the unit to the most likely rank which he/she decides. However, there could be some other possible ranks for that unit because ranking is done without actual measurement. It is known that fuzzy set concept allows the units to belong to different sets with different membership degrees. If we define fuzzy sets for all of the possible ranks, the units in the sets could belong not only to the most likely rank but also to the other possible ranks.

Fuzzy set theory overcomes the limitations of classical set theory by allowing partial membership in a set. It provides a mechanism for measuring the membership degrees as a function represented by a real number in the range [0, 1]. Inspired by fuzzy sets, we introduce a new approach to address uncertainty in ranking. Let $\{X_1, X_2, \ldots, X_k\}$ be a random set of size k from a population. Let also this set be ranked without actual measurement as $\{X_{[1]}, X_{[2]}, \ldots, X_{[k]}\}$. In the proposed FRSS method, k fuzzy sets for k possible rank are defined as follows.

Definition 3.2.1. *Defined over a ranked set $X = \{X_{[1]}, X_{[2]}, \ldots, X_{[k]}\}$, a fuzzy set of rank r, $\tilde{A}_r$, is defined as a set of ordered pairs*

$$\tilde{A}_r = (X_{[h]}, \mu_{\tilde{A}_r}(X_{[h]}))|\forall X_{[h]} \in X, \tag{3.2}$$

where $r = 1, 2, ..., k$ is the rank number and $h = 1, 2, ..., k$ is the order of unit.

$\mu_{\tilde{A}_r}$ is the membership function of $\tilde{A}_r$ and has a key role in the FRSS method through the determination of membership degrees, $\mu_{\tilde{A}_r}(X_{[h]})$. For clearity, we use

$m_h(r)$ instead of $\mu_{\tilde{A}_r}(X_{[h]})$ in the rest of the study to denote the membership degree of hth ordered unit to rank r. In the RSS method, ranking is done with two main tools, with the visual inspection of a human expert or with a concomitant variable. The FRSS method uses the same tools but with an additional task, determination of the membership degrees. For a proper determination of the membership degrees we need some assumptions as follows:

1. The ranker should be a real expert or a highly correlated concomitant variable.
2. A unit belongs to a specific rank with the maximum membership degree, 1, when the ranker considers this rank is most possible rank for this unit.
3. Membership degrees degrade for the other possible ranks.
4. Ties are not allowed if there is one (single) ranker for determination of the memberships.

The first assumption concerns the aim of the FRSS (and also all RSS and related methods) for ranking accurately. The other three assumptions is necessary to make the fuzzy sets of the ranks *convex* and force to have a kernel such as $Kern(\tilde{A}_{r=h}) = X_{[h]}$ for the case of single ranker where $X_{[h]}$ is hth ordered unit in the set.

3.2.1 Determination of the membership degrees by a human-ranker's visual inspection

In the first way of ranking, the ranker can decide the membership degrees of the units by visual inspection, taking the other units in the set into account. In this case, the main role of the membership degrees is to represent the human-ranker's perceptions on rank decisions. Let $r = 1, 2, ..., k$ be the rank number, $h = 1, 2, ..., k$ be the set size and $j = 1, 2, ..., k$ be the cycle number.The membership degree of hth ranked unit in a set of jth cycle to a fuzzy set of rank r determined by the ranker is shown as $m_{h,j}(r)$.

An example of the determined membership degrees by visual inspection is given for a random set of cycle j and set size $k = 3$ as follows.

$$\begin{bmatrix} m_{1,j}(1) & m_{1,j}(2) & m_{1,j}(3) \\ m_{2,j}(1) & m_{2,j}(2) & m_{2,j}(3) \\ m_{3,j}(1) & m_{3,j}(2) & m_{3,j}(3) \end{bmatrix} => \begin{bmatrix} 1 & 0.8 & 0.1 \\ 0.7 & 1 & 0.6 \\ 0.1 & 0.7 & 1 \end{bmatrix}$$

We can give an example to support this case. Suppose that we have a random set from a specific population and it includes 3 units represented by the letters *A, B* and *C*. If we consider it is fully ranked by a human ranker, each unit in the set is ordered 1 to 3. The ranker decides with his/her specific criteria that *unit A* is the 1st, *unit B* is the 2nd and *unit C* is the 3rd. He/she assigns rank 1 to unit A because he/she thinks 1st is the most possible rank for it. However he/she think that 2nd rank is also possible (but not as likely as 1st rank) and that 3rd rank is the least (but not impossible). In the FRSS method, this ranker can determine the membership degrees as $m_{1,j}(1) = 1$, $m_{1,j}(2) = 0.8$ and $m_{1,j}(3) = 0.1$.

We should note that the membership degree matrix of a ranked set is not a stochastic matrix and the matrix notation is only for proper display of the membership degrees of the units in a ranked set. Only one row of the matrix (membership degrees of the unit which is chosen to the sample) is used in the procedure.

Determination of the membership degrees based on the visual inspection of a human expert may look like arbitrary to the researchers because of the uncertain nature of human reasoning. For example, a human ranker may determine different membership degree for the same unit in the same set at various times. Computing with words (CW) methodology in which words are used in place of numbers for computing and reasoning or other similar approaches may help to build a more consistent method for determination of the membership degrees by a human expert. The simulation of the determination of the memberships by human ranker also will be available if a CW or similar method is defined. However, using these kind of methods will also increase the complexity of FRSS method and this could be a disadvantage

for prefer-ability of FRSS in applications. Therefore, we propose to use the simplest method, determination of the membership degrees based on the visual inspection, for human ranker cases.

3.2.2 Determination of the membership degrees with a concomitant variable

Using the information from a concomitant variable is the second way of ranking. We also suggest using the same idea (fuzzy sets and memberships) for this way. However, when a specific concomitant variable is used for ranking the units in the sets, a membership function also can be defined to determine the membership degrees. Let $Y_{(h)j}$ denote the value of the concomitant variable Y for hth ordered unit of the set in jth cycle. When the distance between the units based on the value of concomitant variable Y is large, the ranks of the units are determined more easily and clearly than the case the distance between the units is short. Therefore, we introduce a new membership function where the memberships are determined inversely proportional to the distance between the values of concomitant variables for each sampled unit in the set.

$$m_{h,j}^{RD}(r) = 1 - \frac{|Y_{(r)j} - Y_{(h)j}|}{(\max_q Y_{[q]j} - \min_q Y_{[q]j})} \text{ for } q = 1, 2, ..., k. \tag{3.3}$$

We called the membership function in eq (3.3) the *relative distance (RD)* method because the distances between the units are calculated relatively to the range of the set. For a second way to determine memberships of the units, a Gaussian type formula can be defined, as follows.

$$m_{(h,j)}^{GT}(r) = e^{-\frac{|Y_{[r]j} - Y_{[h]j}|}{\left(\max_q Y_{[q]j} - \min_q Y_{[q]j}\right)/2}} \text{ for } q = 1, 2, ..., k. \tag{3.4}$$

The equations (3.3) and (3.4) have special situations. Eq (3.3) gives the same results with RSS when the set size is 2. Eq (3.4) gives the membership degrees between e^{-2} and e^{0}. We note that, based on the flexible nature of the fuzzy inspired approach, researchers can define new membership functions which are considered suitable for

their studies.

In practice, there may be some criteria in the human ranker's decisions or he/she may rank arbitrarily. In this case, the main role of the membership degrees is to represent the human-ranker's perceptions on rank decisions and it really suits the main idea of fuzzy set theory. However, the fuzzy sets and the memberships in our approach is not totally same with fuzzy sets and memberships in fuzzy set theory because there is also concomitant variable option for determination of the membership degrees. The word of *fuzzy* in the topic is just for representing the main idea in our approach. The similar situation can be remembered in the topic *fuzzy c means* in data mining.

3.2.3 Combining the ranking information from multiple rankers

If we have more than one ranker (human or concomitant variable or both), it could be a good way to combine the ranking information coming from C ($C > 1$) rankers for a more accurate determination of the membership degrees. In our study, we suggest to use min, max and avg operations to obtain the set of combined membership degree decisions of C ($C > 1$) ranker. We prefer to use the most well-known and easy to apply operations in our study. Similar to fuzzy T-norm operations, min can be defined as a conservative combiner. max can be defined as a liberal combiner, like the T-conorm operations in fuzzy set theory. avg is the balanced combiner which can be define as a statistical operation more than a fuzzy one.

Let $c = 1, 2, ..., C$ be the number of ranker and $m^c_{h,j}(r)$ denotes the membership degree which is determined by the ranker c. Then we have C rank decision matrices for C ranker. Considering the set size as $k = 4$, it will be as follows for a single set of cycle j.

$$W^c = \begin{bmatrix} m^c_{1,j}(1) & m^c_{1,j}(2) & m^c_{1,j}(3) & m^c_{1,j}(4) \\ m^c_{2,j}(1) & m^c_{2,j}(2) & m^c_{2,j}(3) & m^c_{2,j}(4) \\ m^c_{3,j}(1) & m^c_{3,j}(2) & m^c_{3,j}(3) & m^c_{3,j}(4) \\ m^c_{4,j}(1) & m^c_{4,j}(2) & m^c_{4,j}(3) & m^c_{4,j}(4) \end{bmatrix} \text{ for instance } \begin{bmatrix} 1 & 0.8 & 0.5 & 0.1 \\ 0.8 & 1 & 0.7 & 0.2 \\ 0.4 & 0.7 & 1 & 0.4 \\ 0.1 & 0.3 & 0.5 & 1 \end{bmatrix}$$

We define the combination of the membership degrees based on chosen operation as follows.

$$m_{h,j}(r) = \begin{cases} min(m^1_{h,j}(r), ..., m^C_{h,j}(r)), & \text{if } min \text{ is chosen} \\ max(m^1_{h,j}(r), ..., m^C_{h,j}(r)), & \text{if } max \text{ is chosen} \\ avg(m^1_{h,j}(r), ..., m^C_{h,j}(r)), & \text{if } avg \text{ is chosen} \end{cases}$$

Finally we have a *combination of the membership degree decisions* matrix for a single set in cycle *j*.

$$M_j = \begin{bmatrix} m_{1,j}(1) & m_{1,j}(2) & m_{1,j}(3) & m_{1,j}(4) \\ m_{2,j}(1) & m_{2,j}(2) & m_{2,j}(3) & m_{2,j}(4) \\ m_{3,j}(1) & m_{3,j}(2) & m_{3,j}(3) & m_{3,j}(4) \\ m_{4,j}(1) & m_{4,j}(2) & m_{4,j}(3) & m_{4,j}(4) \end{bmatrix}$$

3.3 FRSS procedure

Using the formulas to determine membership degrees and the methods to combine the information of multiple rankers given previously, we introduce the sampling

procedure of FRSS in a general form which can be used for both single and multiple ranker cases as follows.

1. Select k units at random from a specified population.
2. The ranker (each ranker if there are multiple rankers) ranks these k units without measuring them and determine the membership degrees of the units for each rank.
3. Combine the membership degree decisions of each ranker. (This step is ignored for the single ranker case)
4. Chose the unit which has the highest membership degree for the rank in the first order and keep its membership degrees.
5. Repeat the first four steps to choose the units which have the highest membership degrees for the ranks in the second, third, ..., kth orders, respectively.
6. The first five step are called a cycle. The first five steps are repeated for n times to sample $k * n$ observations

Figure 3.4 shows the FRSS procedure for the case where the ranking is done with a concomitant variable. In this procedure, $Y_{(h)j}$ means the value of the concomitant variable Y of hth unit in a specific set in the jth cycle and $X_{[h]j}$ means the hth ranked unit which is chosen for the sample from the jth cycle for $h = 1, 2, ..., k$ and $j = 1, 2, ..., n$. We use round brackets while defining the ordered value of the concomitant variable Y because it represents an order statistics in the random set. However, we prefer to use square brackets to show the value of the variable of interest X since we do not know the real order of the X values in the random set as the ranking is done without actual measurement of X.

The first output of the FRSS procedure is the set of $k * n$ observations, called a fuzzy ranked set sample. The second output is the set of membership degrees of the observations which represent the degrees of affiliation to each rank. $m_{h,j}(r)$ means

membership degree of hth chosen unit from the jth cycle for rank r, $r = 1, 2, ..., k$. (The chosen units from jth cycle are represented by a vector $O_j$and their membership degrees are given in a matrix M_j for clarity in the illustration for $j = 1, 2, ..., n$ in Figure 3.4).

Cycle 1

$$\begin{bmatrix} \mathbf{Y_{(1)1}} \le Y_{(2)1} \le Y_{(3)1} \\ Y_{(1)1} \le \mathbf{Y_{(2)1}} \le Y_{(3)1} \\ Y_{(1)1} \le Y_{(2)1} \le \mathbf{Y_{(3)1}} \end{bmatrix} \rightarrow O_1 = \begin{bmatrix} \mathbf{X_{[1]1}} \\ \mathbf{X_{[2]1}} \\ \mathbf{X_{[3]1}} \end{bmatrix} \; and \; M_1 = \begin{bmatrix} m_{1,1}(1) & m_{1,1}(2) & m_{1,1}(3) \\ m_{2,1}(1) & m_{2,1}(2) & m_{2,1}(3) \\ m_{3,1}(1) & m_{3,1}(2) & m_{3,1}(3) \end{bmatrix}$$

Cycle 2

$$\begin{bmatrix} \mathbf{Y_{(1)2}} \le Y_{(2)2} \le Y_{(3)2} \\ Y_{(1)2} \le \mathbf{Y_{(2)2}} \le Y_{(3)2} \\ Y_{(1)2} \le Y_{(2)2} \le \mathbf{Y_{(3)2}} \end{bmatrix} \rightarrow O_2 = \begin{bmatrix} \mathbf{X_{[1]2}} \\ \mathbf{X_{[2]2}} \\ \mathbf{X_{[3]2}} \end{bmatrix} and \; M_2 = \begin{bmatrix} m_{1,2}(1) & m_{1,2}(2) & m_{1,2}(3) \\ m_{2,2}(1) & m_{2,2}(2) & m_{2,2}(3) \\ m_{3,2}(1) & m_{3,2}(2) & m_{3,2}(3) \end{bmatrix}$$

...

Cycle n

$$\begin{bmatrix} \mathbf{Y_{(1)n}} \le Y_{(2)n} \le Y_{(3)n} \\ Y_{(1)n} \le \mathbf{Y_{(2)n}} \le Y_{(3)n} \\ Y_{(1)n} \le Y_{(2)n} \le \mathbf{Y_{(3)n}} \end{bmatrix} \rightarrow O_n = \begin{bmatrix} \mathbf{X_{[1]n}} \\ \mathbf{X_{[2]n}} \\ \mathbf{X_{[3]n}} \end{bmatrix} and \; M_n = \begin{bmatrix} m_{1,n}(1) & m_{1,n}(2) & m_{1,n}(3) \\ m_{2,n}(1) & m_{2,n}(2) & m_{2,n}(3) \\ m_{3,n}(1) & m_{3,n}(2) & m_{3,n}(3) \end{bmatrix}$$

Figure 3.4 FRSS procedure with a single ranker Y and set size $k = 3$

It can be seen in Figure 3.4, the units chosen to sample, $X_{[h]j}$ ($h = 1, 2, ..., k$ and $j = 1, 2, ..., n$), are the same in both the RSS and FRSS procedures when there is a single ranker. However there is additional membership information about the ranks, $m_{h,j}(r)$ ($h = 1, 2, ..., k$, $h = 1, 2, ..., k$ and $j = 1, 2, ..., n$) in the FRSS procedure.

3.4 Estimation of the population mean based on FRSS

From the FRSS procedure, we obtain two outputs: first are the observations and second is the membership degree matrix consisting of the membership degrees of the observations to each rank. In our approach, we propose to use these membership degrees as the weights of the observations for calculation of the average of each rank. Let $X_{[1]1}, X_{[2]1}, X_{[3]1}, \cdots, X_{[h]j}, \cdots, X_{[k]n}$ be the sample obtained by FRSS from the underlying distribution F having finite variance σ^2. Let the membership degrees,

$m_{h,j}(r)$, for $h = 1, 2, \ldots, k$, $j = 1, 2, ..., n$ and $r = 1, 2, ..., k$ be determined considering the assumptions defined in the previous section. Using the membership degrees as weights of the units, the estimator of the population mean based on the FRSS procedure is introduced as follows.

$$\bar{X}_{FRSS} = \frac{1}{k}\sum_{r=1}^{k}\bar{X}^{r}_{FRSS}, \tag{3.8}$$

where

$$\bar{X}^{r}_{FRSS} = \sum_{h=1}^{k}\sum_{j=1}^{n}\frac{m_{h,j}(r)X_{[h]j}}{\sum_{h=1}^{k}\sum_{j=1}^{n}m_{h,j}(r)}.$$

The new estimator $\bar{X}_{FRSS}$ has the same settings as the prorated estimator given in MacEachern et al. (2004) and Ozturk (2011c, 2013) although the determination of weights does differ. For example, similar to the prorated estimator in eq. (2.22), $m_{h,j}(r)X_{[h]j}$ assigns a proportion, $(\frac{m_{h,j}(r)}{\sum_{h=1}^{k}\sum_{j=1}^{n}m_{h,j}(r)})$, of $X_{[h]j}$ to post-stratum h in the estimation. Thus one can easily see that $\bar{X}_{FRSS}$ has the same asymptotic properties, such as asymptotic normality proved by Ozturk (2013), as the estimators in the equations (2.22) and (2.29).

We also give a remark on the unbiasedness of the new estimator with some special assumptions.

Remark 3.4.1. *Let X be the random variable we want to measure after obtaining the sample via FRSS and the underlying distribution of X is symmetric over 0. If a symmetric matrix containing fixed numbers and satisfying the assumptions;*

- *the diagonals are* 1
- $m_h(r) = m_r(h)$
- $m_h(k - r + 1) = m_r(k - h + 1)$

is used as a common membership matrix for all of the cycles, the proposed estimator

$\bar{X}_{FRSS}$ is an unbiased estimator.

Proof. We obtain the membership matrix with the assumptions defined in the remark for $k = 3, 4$ and 5, for example, as follows.

$$\begin{bmatrix} 1 & a & b \\ a & 1 & a \\ b & a & 1 \end{bmatrix}, \begin{bmatrix} 1 & a & b & c \\ a & 1 & d & b \\ b & d & 1 & a \\ c & b & a & 1 \end{bmatrix} \text{ and } \begin{bmatrix} 1 & a & b & c & d \\ a & 1 & e & f & c \\ b & e & 1 & e & b \\ c & f & e & 1 & a \\ d & c & b & a & 1 \end{bmatrix}$$

First, we can write the expectation of the new estimator as

$$E\left(\bar{X}_{FRSS}\right) = E\left(\frac{1}{k}\sum_{r=1}^{k}\sum_{h=1}^{k}\sum_{j=1}^{n}\frac{m_h(r)X_{[h]j}}{\sum_{h=1}^{k}\sum_{j=1}^{n}m_h(r)}\right). \tag{3.9}$$

The right-side of the eq. (3.9) is equal to

$$\frac{1}{k}E\left(\left[\frac{1}{n\sum_{h=1}^{k}m_h(1)} + ... + \frac{m_1(k)}{n\sum_{h=1}^{k}m_h(k)}\right]\left(X_{[1]1} + ... + X_{[1]n}\right) + \right.$$
$$\left. ... + \left[\frac{m_k(1)}{n\sum_{h=1}^{k}m_h(1)} + ... + \frac{1}{n\sum_{h=1}^{k}m_h(k)}\right]\left(X_{[k]1} + ... + X_{[k]n}\right)\right).$$

Recalling from the order statistics background of RSS, we know that expected values of hth ordered observations of random sets of n cycle chosen from the same population are equal, $E(X_{(1)1}) = E(X_{(1)2}) = \cdots = E(X_{(1)n}) = E(X_{(1)})$. Also the hth ordered units in each cycle have the same membership degree for rank r, $m_{h,1}(r) = m_{h,2}(r) = ... = m_{h,n}(r) = m_h(r)$, when we use a single membership matrix for all of the cycles.

From there, we can rewrite the last equation;

$$\frac{1}{k}\frac{1}{n}\Bigg(\left[\frac{1}{\sum_{h=1}^{k} m_h(1)}+...+\frac{m_1(k)}{\sum_{h=1}^{k} m_h(k)}\right]nE\left(X_{[1]}\right)+... \\ ...+\left[\frac{m_k(1)}{\sum_{h=1}^{k} m_h(1)}+...+\frac{1}{\sum_{h=1}^{k} m_h(k)}\right]nE\left(X_{[k]}\right)\Bigg). \tag{3.10}$$

Describing the multiples of the expectations as the weights of expectations, it can easily be calculated that the weight of $E\left(X_{[h]}\right)$ is equal to weight $E\left(X_{[k-h+1]}\right)$. Therefore eq. (3.10) becomes

$$\frac{1}{k}\Bigg(\left[\frac{1}{\sum_{h=1}^{k} m_h(1)}+...+\frac{m_1(k)}{\sum_{h=1}^{k} m_h(k)}\right]\left[E\left(X_{[1]}\right)+E\left(X_{[k]}\right)\right]+... \\ +\left[\frac{m_{\frac{k}{2}}(1)}{\sum_{h=1}^{k} m_h(1)}+...+\frac{m_{\frac{k}{2}}(k)}{\sum_{h=1}^{k} m_h(k)}\right]\left[E\left(X_{[\frac{k}{2}]}\right)+E\left(X_{[\frac{k}{2}+1]}\right)\right]\Bigg). \tag{3.11}$$

Arnold et al. (1992) (page 27) show that $E\left(X_{(h)}\right) = -E\left(X_{(k-h+1)}\right)$ if the underlying distribution of X is symmetrical over 0. Finally, the expectation becomes

$$E\left(\bar{X}_{FRSS}\right) = 0 = \mu_X. \tag{3.12}$$

□

We note that, the proof is given for even set sizes. It can be easily obtained for odd set sizes in the same way. *Remark 3.4.1* could be useful for the following kind of cases. For a simple application of the FRSS method, the researcher may prefer to use a single membership matrix with fixed values instead of determining the memberships

of each unit individually. This remark also gives us an idea about how we can achieve unbiased results for the new estimator in the cases of symmetric distributions which is now examined in the following chapter.

CHAPTER FOUR
SIMULATION STUDIES

In this section, we examine the performance of the new method with simulation studies for both single and multiple ranker cases. If the performance is examined for only perfect ranking conditions, it is quite easy to generate data sets and rank them with actual measurement. But if imperfect ranking conditions are also considered, we need a method which aims to capture the uncertainty of imperfect ranking in the real life applications. The studies focusing on imperfect ranking and modeling this uncertainty, given in Section 2.4, are also used for the simulation of the imperfect ranking. Since it is the most widely used method in RSS studies, we follow Dell & Clutter (1972)'s method for the generation of data sets from different distributions.

We use Dell & Clutter (1972)'s method in our study as follows. Let X be the variable of interest and Y be the concomitant variable used for ranking in the study. Let also X_1, X_2, $\cdots$, X_W be the true values of the units which are obtainable by accurate measurement in a real life application. W is the population size and chosen as 10000 in our study in order to have a finite but quite large sample. Then the X's and Y's may be regarded as realizations of linked random variables X_i , Y_i

$$Y_i = g(X_i, Z_i), \quad \text{for } i = 1, 2, ..., W$$

where Z_i is defined as judgment error term for the single-valued function of g. Z_i's are generated from normal distributions $Z \sim N(0, \sigma^2)$. Then the concomitant variable is computed with the simple formula devised by Dell & Clutter (1972).

$$Y_i = X_i + Z_i, \quad \text{for } i = 1, 2, ..., W \tag{4.1}$$

where $X_1, X_2, \cdots, X_W, Z_1, Z_2, \cdots, Z_W$ are mutually independent.

We use correlation levels of the variable of interest X_i and concomitant variable

Y_i, $\rho_{X,Y} = \rho$, as the level of *imperfect ranking*. It is quite easy to determine σ^2 of the judgment error term Z_i for different correlation levels when we generate normally distributed X_i. For the other distributions, we determine the value of σ^2 for the judgment error term Z_i by testing.

4.1 Single ranker case

We first generate data sets from different symmetric and asymmetric distributions in order to compare the methods for single ranker case. The ranking mechanism is modeled through the Dell & Clutter (1972) method. The accuracy and efficiency of the proposed sampling design and former ones (SRS and RSS) are addressed. We have studied on the estimators' unbiasedness for their accuracy and obtained at the mean squared errors of the estimators for their efficiency. The two kinds of method to determine membership degrees in the ranking process were introduced previously. Thus we have two mean estimators $\bar{X}_{FRSS-RD}$ and $\bar{X}_{FRSS-GT}$ as the estimator where the memberships are calculated using the relative distance method given in eq. (3.3) and the Gaussian type method given in eq. (3.4), respectively. The relative efficiencies are defined in terms of mean squared errors (MSE) as follows;

$$\begin{aligned} RE1 &= \frac{MSE(\bar{X}_{SRS})}{MSE(\bar{X}_{FRSS-RD})}, RE2 = \frac{MSE(\bar{X}_{RSS})}{MSE(\bar{X}_{FRSS-RD})}, \\ RE3 &= \frac{MSE(\bar{X}_{SRS})}{MSE(\bar{X}_{FRSS-GT})}, RE4 = \frac{MSE(\bar{X}_{RSS})}{MSE(\bar{X}_{FRSS-GT})}, \end{aligned} \tag{4.2}$$

The set sizes are chosen as 2, 3, 4 and 6. The least-common-multiple value of 2, 3, 4 and 6 is 12. Thus we can see the performance of the new estimators in different set sizes for fixed sample sizes, 12 and 24.

According to Table 4.1 and Table 4.2, we can say that the new estimators perform quite well compared to the estimators of the SRS and RSS methods. The relative efficiencies are greater than 1 in the most of the cases. Some of the remarkable results

Table 4.1 Relative efficiency results for FRSS-RD compared with SRS and RSS

				RE1					RE2				
Dist.	S.Size	k	n	$\rho = 1$	$\rho = 0.9$	$\rho = 0.8$	$\rho = 0.7$	$\rho = 0.6$	$\rho = 1$	$\rho = 0.9$	$\rho = 0.8$	$\rho = 0.7$	$\rho = 0.6$
	12	2	6	1.425	1.321	1.242	1.206	1.165	1.000	1.000	1.000	1.000	1.000
		3	4	2.045	1.695	1.475	1.332	1.196	1.103	1.057	1.025	1.011	0.991
N(10,5)		4	3	2.756	1.966	1.641	1.411	1.259	1.180	1.093	1.041	1.017	1.003
		6	2	3.771	2.484	1.801	1.546	1.359	1.238	1.110	1.047	1.024	1.001
	24	2	12	1.487	1.328	1.28	1.167	1.136	1.000	1.000	1.000	1.000	1.000
		3	8	2.119	1.758	1.539	1.349	1.181	1.110	1.052	1.024	1.004	0.992
		4	6	2.676	2.005	1.616	1.380	1.282	1.166	1.083	1.042	1.014	0.998
		6	4	4.042	2.420	1.874	1.476	1.311	1.232	1.112	1.051	1.022	1.001
	12	2	6	1.522	1.387	1.296	1.197	1.123	1.000	1.000	1.000	1.000	1.000
		3	4	2.121	1.776	1.529	1.299	1.176	1.061	1.041	1.015	0.991	0.987
Uni(0,1)		4	3	2.800	1.953	1.717	1.414	1.301	1.120	1.047	1.017	1.000	0.986
		6	2	3.737	2.486	1.907	1.538	1.373	1.053	1.057	1.019	1.000	0.993
	24	2	12	1.478	1.076	1.059	1.055	1.038	1.000	1.000	1.000	1.000	1.000
		3	8	2.188	1.140	1.039	1.026	0.984	1.063	1.040	1.029	1.000	0.987
		4	6	2.692	1.171	1.073	1.022	0.991	1.167	1.048	1.000	1.020	1.000
		6	4	3.889	1.218	1.106	0.998	1.013	1.111	1.000	1.000	1.000	1.000
	12	2	6	1.363	1.274	1.180	1.121	1.077	1.000	1.000	1.000	1.000	1.000
		3	4	1.914	1.588	1.374	1.300	1.184	1.163	1.092	1.050	1.033	1.008
Exp(1)		4	3	2.425	1.971	1.593	1.418	1.198	1.280	1.151	1.090	1.048	1.025
		6	2	3.128	2.289	1.803	1.475	1.339	1.298	1.146	1.086	1.048	1.033
	24	2	12	1.359	1.242	1.169	1.162	1.092	1.000	1.000	1.000	1.000	1.000
		3	8	1.956	1.601	1.421	1.218	1.195	1.161	1.098	1.062	1.026	1.007
		4	6	2.364	1.904	1.524	1.311	1.240	1.205	1.126	1.071	1.034	1.019
		6	4	2.770	2.031	1.682	1.470	1.216	1.122	1.069	1.015	0.997	0.986
	12	2	6	1.147	1.154	1.068	1.169	1.035	1.000	1.000	1.000	1.000	1.000
		3	4	1.697	1.565	1.310	1.300	1.177	1.281	1.194	1.153	1.126	1.072
LogN(1,2)		4	3	2.272	1.872	1.502	1.414	1.201	1.608	1.299	1.205	1.143	1.095
		6	2	2.968	2.051	1.708	1.562	1.301	1.880	1.423	1.285	1.293	1.261
	24	2	12	1.142	1.113	1.122	1.063	1.066	1.000	1.000	1.000	1.000	1.000
		3	8	1.592	1.555	1.275	1.279	1.177	1.355	1.183	1.098	1.100	1.058
		4	6	2.153	1.797	1.535	1.421	1.249	1.395	1.296	1.230	1.169	1.075
		6	4	2.708	1.988	1.642	1.444	1.217	1.759	1.335	1.223	1.125	1.061

in Table 4.1 are given below.

- For $N(10, 5)$ case, relative efficiency rises to 4.042 in comparison to FRSS-SRS and to 1.232 in comparison to FRSS-RSS.
- For $Uni(0, 1)$ case, relative efficiency rises to 3.889 in comparison to FRSS-SRS and to 1.167 in comparison to FRSS-RSS.
- For $Exp(1)$ case, the relative efficiency rises to 3.128 in comparison to FRSS-SRS and to 1.298 in comparison to FRSS-RSS.
- For $LogN(1, 2)$ case, relative efficiency rises to 2.968 in comparison to FRSS-SRS and to 1.880 in comparison to FRSS-RSS.
- For applications of symmetric distributions, relative efficiency generally increases when the sample size increases in the comparisons with FRSS-SRS and FRSS-RSS.
- For applications of asymmetric distributions, the relative efficiency generally decreases when the sample size increases in the comparisons with FRSS-SRS. However the comparison with FRSS-RSS has a reverse situation.

Some of the remarkable results in Table 4.2 are given as follows.

- For $N(10, 5)$ case, relative efficiency rises to 4.716 in comparison to FRSS-SRS and to 1.433 in comparison to FRSS-RSS.
- For $Uni(0, 1)$ case, relative efficiency rises to 3.842 in comparison to FRSS-SRS and to 1.111 in comparison to FRSS-RSS.
- For $Exp(1)$ case, relative efficiency rises to 2.697 in comparison to FRSS-SRS and to 1.138 in comparison to FRSS-RSS.
- For $LogN(1, 2)$ case, relative efficiency rises to 2.366 in comparison to FRSS-SRS and to 1.464 in comparison to FRSS-RSS.

Table 4.2 Relative efficiency results for FRSS-GT compared with SRS and RSS

				RE3					RE4				
Dist.	S.Size	k	n	$\rho=1$	$\rho=0.9$	$\rho=0.8$	$\rho=0.7$	$\rho=0.6$	$\rho=1$	$\rho=0.9$	$\rho=0.8$	$\rho=0.7$	$\rho=0.6$
	12	2	6	1.743	1.445	1.274	1.198	1.109	1.139	1.072	1.025	1.011	0.983
		3	4	2.435	1.792	1.478	1.316	1.125	1.232	1.115	1.036	1.002	0.974
N(10,5)		4	3	3.040	2.139	1.613	1.399	1.218	1.319	1.122	1.030	0.991	0.969
		6	2	4.716	2.636	1.843	1.473	1.296	1.431	1.153	1.048	0.982	0.955
	24	2	12	1.703	1.447	1.348	1.123	1.118	1.147	1.089	1.023	0.995	0.976
		3	8	2.420	1.803	1.512	1.291	1.182	1.231	1.115	1.037	0.991	0.966
		4	6	3.159	2.058	1.637	1.376	1.186	1.311	1.120	1.032	0.979	0.952
		6	4	4.486	2.619	1.828	1.526	1.275	1.433	1.142	1.035	0.986	0.960
	12	2	6	1.468	1.328	1.253	1.175	1.109	1.000	1.000	1.000	1.000	1.000
		3	4	2.000	1.647	1.417	1.245	1.150	1.000	0.980	0.986	0.964	0.950
Uni(0,1)		4	3	2.630	1.913	1.646	1.269	1.266	1.037	0.978	0.969	0.944	0.937
		6	2	3.842	2.324	1.875	1.479	1.231	1.053	1.000	0.982	0.948	0.944
	24	2	12	1.478	1.065	1.042	1.006	1.000	1.000	1.000	1.000	1.000	1.000
		3	8	1.889	1.138	1.046	1.011	1.012	1.059	0.959	0.942	0.950	0.943
		4	6	2.429	1.198	1.107	1.070	1.000	1.077	0.957	0.934	0.941	0.944
		6	4	3.500	1.252	1.073	1.050	1.032	1.111	0.955	0.958	0.944	0.936
	12	2	6	1.329	1.230	1.239	1.169	1.076	1.000	1.000	1.000	1.000	1.000
		3	4	1.314	1.321	1.195	1.167	1.068	1.026	1.016	1.011	1.004	1.001
Exp(1)		4	3	1.318	1.265	1.163	1.171	1.113	1.000	0.999	1.000	1.000	0.999
		6	2	1.288	1.204	1.199	1.115	1.114	1.000	1.000	1.000	1.001	1.000
	24	2	12	1.344	1.257	1.177	1.144	1.052	1.000	1.000	1.000	1.000	1.000
		3	8	1.770	1.536	1.347	1.185	1.095	1.098	1.051	1.022	0.990	0.982
		4	6	2.180	1.750	1.498	1.295	1.181	1.138	1.059	1.026	1.003	0.984
		6	4	2.697	2.012	1.596	1.403	1.262	1.112	1.036	1.005	0.977	0.983
	12	2	6	1.154	1.065	1.057	1.116	1.080	1.000	1.000	1.000	1.000	1.000
		3	4	1.582	1.447	1.295	1.177	1.120	1.208	1.187	1.086	1.060	1.003
LogN(1,2)		4	3	1.871	1.558	1.320	1.227	1.149	1.278	1.245	1.112	1.047	1.038
		6	2	2.366	1.928	1.474	1.401	1.168	1.464	1.288	1.190	1.107	1.012
	24	2	12	1.118	1.129	1.131	1.105	1.046	1.000	1.000	1.000	1.000	1.000
		3	8	1.558	1.413	1.258	1.207	1.064	1.147	1.089	1.057	1.027	1.007
		4	6	1.843	1.532	1.367	1.246	1.103	1.241	1.150	1.088	1.042	1.016
		6	4	2.109	1.823	1.471	1.377	1.229	1.319	1.183	1.117	1.078	1.035

When we compare the new estimators with each other, we can say that $\bar{X}_{FRSS-RD}$ is more efficient than $\bar{X}_{FRSS-GT}$. $\bar{X}_{FRSS-GT}$ is more efficient only in cases where distribution is normal and correlation is very high ($\rho > 0.9$). The performance of $\bar{X}_{FRSS-GT}$ is relatively low in the other cases, especially in the lower correlation levels. The accuracy results are given in Table 4.3. We can say that numerically the new estimators produce unbiased results where X and Y come from symmetric-distributed populations. However there are small biases toward increasing set sizes in asymmetric distributions. For a general discussion, we can indicate the following:

- The new estimators work well for medium-and-higher correlation levels of the response and the concomitant variable. Relative efficiencies increase as the correlation level of the concomitant variable increases. This result indicates that the FRSS method could be a good alternative for researchers having a relatively good ranker (human expert or concomitant variable) in their studies.
- Relative efficiency increases when the set size increases for fix sample size. It means that the proposed estimators constructed based on FRSS can be more useful for cases where the set sizes are relatively big.

As a final comment, we note that the units chosen for sampling are the same in RSS and FRSS if the same sets and ranker are used in the procedures. The only difference is the additional membership information in the procedure of FRSS and its usage in the estimation of the population mean. Therefore, it can be said that increases in the relative efficiencies of the estimators based on the FRSS result from the additional information of memberships. Or, in other words, the single ranker case results indicate that ”defining memberships and using them in the estimation as weights” could be a useful idea to improve efficiency in RSS.

Table 4.3 Accuracy results for FRSS estimators

$\bar{X}_{FRSS-RD}$

			N(10,5)					Uni(0,1)				
S.Size	k	n	$\rho = 1$	$\rho = .9$	$\rho = .8$	$\rho = .7$	$\rho = .6$	$\rho = 1$	$\rho = .9$	$\rho = .8$	$\rho = .7$	$\rho = .6$
12	2	6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	3	4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	4	3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	6	2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
24	2	12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	3	8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	4	6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	6	4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

			Exp(1)					LogN(1,2)				
S.Size	k	n	$\rho = 1$	$\rho = .9$	$\rho = .8$	$\rho = .7$	$\rho = .6$	$\rho = 1$	$\rho = .9$	$\rho = .8$	$\rho = .7$	$\rho = .6$
12	2	6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	3	4	0.02	0.02	0.02	0.02	0.02	0.04	0.04	0.04	0.04	0.04
	4	3	0.04	0.04	0.04	0.04	0.04	0.07	0.07	0.07	0.07	0.08
	6	2	0.07	0.07	0.07	0.07	0.07	0.11	0.10	0.11	0.12	0.11
24	2	12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	3	8	0.02	0.02	0.02	0.02	0.02	0.04	0.04	0.04	0.04	0.04
	4	6	0.04	0.04	0.04	0.04	0.04	0.06	0.07	0.07	0.07	0.07
	6	4	0.07	0.07	0.07	0.06	0.06	0.11	0.11	0.11	0.11	0.11

$\bar{X}_{FRSS-GT}$

			N(10,5)					Uni(0,1)				
S.Size	k	n	$\rho = 1$	$\rho = .9$	$\rho = .8$	$\rho = .7$	$\rho = .6$	$\rho = 1$	$\rho = .9$	$\rho = .8$	$\rho = .7$	$\rho = .6$
12	2	6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	3	4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	4	3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	6	2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
24	2	12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	3	8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	4	6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	6	4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

			Exp(1)					LogN(1,2)				
S.Size	k	n	$\rho = 1$	$\rho = .9$	$\rho = .8$	$\rho = .7$	$\rho = .6$	$\rho = 1$	$\rho = .9$	$\rho = .8$	$\rho = .7$	$\rho = .6$
12	2	6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	3	4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	4	3	0.00	0.00	0.00	0.00	0.00	0.10	0.10	0.10	0.10	0.10
	6	2	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10
24	2	12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	3	8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	4	6	0.00	0.00	0.00	0.00	0.00	0.10	0.10	0.10	0.10	0.10
	6	4	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10

4.2 Multiple rankers case

In some of RSS applications, there is more than one ranker (human experts, concomitant variables or both) available and each of these rankers provides judgment for the ranking of units. The experience levels of the human experts or the correlation levels of the concomitant variables are usually different. Furthermore, their rank decisions are usually not the same. For these kind of situations, we referred to a combination method in Section 3.2.3. In this section, a simulation study is conducted to draw conclusions about the performance of FRSS for multiple rankers cases where the variable of interest X and concomitant variables Y_1, Y_2 and Y_3 are generated from the multivariate normal distribution (m, Σ) where

$$m = [10, 8, 5, 3] \text{ and } \Sigma = \begin{bmatrix} 5 & \sigma_{1,2} & \sigma_{1,3} & \sigma_{1,4} \\ \sigma_{2,1} & 5 & \sigma_{2,3} & \sigma_{2,4} \\ \sigma_{3,1} & \sigma_{3,2} & 3 & \sigma_{3,4} \\ \sigma_{4,1} & \sigma_{4,2} & \sigma_{4,3} & 3 \end{bmatrix}.$$

The covariances, $\sigma_{i,j}$'s, are computed from $\rho_{i,j} = \sigma_{i,j}/\sigma_i\sigma_j$ for various correlation levels. Only relative distance method, membership function given in eq. (3.3), is used to determine the membership degrees because of the more efficient results given in the previous section. The estimation of the population mean is obtained individually for each of the combination operations, avg, min and max. $RSS1$, $RSS2$ and $RSS3$ represent the methods of classical RSS where the concomitant variables Y_1, Y_2 and Y_3 are used as a single ranker, respectively. $RSS - M$ represents the combination method and mean estimator defined by Ozturk (2011a). Empirical mean square errors are computed with a repetition rate of 10000 in the simulation study and the results are summarized in Table 4.4 in terms of the relative efficiencies.

$$RE5 = \frac{MSE(\bar{X}_{SRS})}{MSE(\bar{X}_{FRSS})}, RE6 = \frac{MSE(\bar{X}_{RSS1})}{MSE(\bar{X}_{FRSS})},$$

$$RE7 = \frac{MSE(\bar{X}_{RSS2})}{MSE(\bar{X}_{FRSS})}, RE8 = \frac{MSE(\bar{X}_{RSS3})}{MSE(\bar{X}_{FRSS})}, \tag{4.3}$$

$$RE9 = \frac{MSE(\bar{X}_{RSS-M})}{MSE(\bar{X}_{FRSS})}.$$

According to the results in Table 4.4, the idea of combining the decisions of the multiple rankers generally works well. Some of the discussions on combination operations based on the results in Table 4.4 are given bellow.

- The $FRSS$ estimator based on the avg combination method is more efficient than the SRS, RSS and $RSS - M$ methods for estimation of the population mean in all cases.
- The results of the new estimator $\bar{X}_{FRSS-min}$, are not as good as the results of $\bar{X}_{FRSS-avg}$. Nevertheless, its performance is generally close to the performance of $\bar{X}_{FRSS-avg}$.
- The results of the estimator $\bar{X}_{FRSS-max}$ are worse than its counterparts even though they are better than SRS and standard RSS in most cases.
- The min operation does not work for set size 2 because of the formulation in eq. (3.3)

The numerical superiority of the avg combination was not a surprise but the low performance of the other operations, especially the performance of the max operation, was unexpected. We think that better results may be obtained if a more efficient operation is defined for the liberal combination of the rank decisions.

Table 4.4 The relative efficiency results of $\bar{X}_{FRSS}$ with multiple rankers

Corr. Levels	Set	Cycle	$\bar{X}_{FRSS-avg}$					$\bar{X}_{FRSS-min}$					$\bar{X}_{FRSS-max}$				
ρ1-ρ2-ρ3	Size	Size	RE5	RE6	RE7	RE8	RE9	RE5	RE6	RE7	RE8	RE9	RE5	RE6	RE7	RE8	RE9
(95-95-95)	2	6	1.538	1.108	1.111	1.110	1.000	*	*	*	*	*	1.486	1.070	1.074	1.072	0.996
	3	4	2.117	1.241	1.245	1.245	1.056	1.864	1.092	1.096	1.096	0.930	2.056	1.205	1.209	1.209	1.026
	4	3	2.858	1.351	1.349	1.368	1.089	2.598	1.227	1.226	1.244	0.990	2.665	1.259	1.258	1.276	1.016
	6	2	3.980	1.505	1.482	1.506	1.132	3.704	1.401	1.379	1.401	1.053	3.416	1.292	1.272	1.292	0.971
(95-90-30)	2	6	1.443	1.017	1.055	1.367	1.000	*	*	*	*	*	1.095	0.772	0.800	1.037	0.759
	3	4	1.864	1.067	1.148	1.803	1.040	1.522	0.871	0.937	1.472	0.849	1.315	0.752	0.809	1.271	0.734
	4	3	2.275	1.065	1.202	2.146	1.077	2.016	0.944	1.065	1.902	0.954	1.449	0.678	0.766	1.367	0.686
	6	2	2.770	1.057	1.220	2.668	1.158	2.566	0.979	1.130	2.473	1.073	1.548	0.591	0.682	1.491	0.647
(70-70-60)	2	6	1.398	1.170	1.165	1.244	1.000	*	*	*	*	*	1.203	1.007	1.003	1.071	0.861
	3	4	1.706	1.311	1.337	1.418	1.023	1.432	1.078	1.092	1.099	0.794	1.389	1.068	1.089	1.155	0.833
	4	3	1.986	1.436	1.421	1.580	1.090	1.761	1.273	1.260	1.401	0.966	1.496	1.082	1.070	1.190	0.821
	6	2	2.368	1.604	1.612	1.844	1.156	2.230	1.511	1.518	1.737	1.089	1.570	1.064	1.069	1.223	0.766
(95-65-30)	2	6	1.400	0.974	1.182	1.323	1.000	*	*	*	*	*	1.107	0.770	0.935	1.046	0.791
	3	4	1.762	0.975	1.344	1.704	1.058	1.367	0.756	1.043	1.322	0.821	1.309	0.724	0.999	1.266	0.786
	4	3	1.973	0.960	1.534	1.912	1.131	1.802	0.877	1.401	1.746	1.033	1.369	0.666	1.064	1.326	0.784
	6	2	2.456	0.938	1.731	2.279	1.284	2.392	0.914	1.686	2.220	1.251	1.457	0.556	1.027	1.352	0.762
(95-35-30)	2	6	1.261	0.901	1.188	1.233	1.000	*	*	*	*	*	1.065	0.761	1.003	1.041	0.844
	3	4	1.578	0.866	1.457	1.472	1.086	1.311	0.719	1.211	1.223	0.902	1.219	0.669	1.126	1.137	0.839
	4	3	1.730	0.839	1.633	1.646	1.185	1.671	0.810	1.577	1.589	1.145	1.192	0.578	1.125	1.134	0.817
	6	2	2.138	0.795	1.989	1.977	1.366	2.182	0.811	2.030	2.017	1.394	1.298	0.483	1.208	1.200	0.829
(30-30-30)	2	6	1.086	1.027	1.054	1.056	1.000	*	*	*	*	*	1.002	0.948	0.973	0.975	0.923
	3	4	1.086	1.093	1.094	1.047	1.024	0.783	0.788	0.789	0.755	0.738	1.021	1.027	1.028	0.983	0.962
	4	3	1.153	1.083	1.077	1.087	1.032	1.030	0.968	0.963	0.972	0.923	1.079	1.013	1.008	1.017	0.966
	6	2	1.218	1.110	1.117	1.111	1.081	1.124	1.024	1.031	1.025	0.998	1.113	1.013	1.021	1.015	0.988

Some of the other remarkable results in Table 4.4 are as follows.

- For the avg operation, the idea of combining the decisions of multiple rankers works well even if one of them has a really low correlation level ($95-90-30$). The results of the FRSS estimator with multiple ranker are even better than the estimator based on RSS with a highly-correlated ranker.

- For the avg operation, the combination of the rank information coming from 3 moderately-correlated ($70-70-60$) variables increases the efficiency more than 2.3 times compared with the SRS method, 1.6 times compared with the standard RSS methods and 1.1 times compared with the $RSS-M$ method. The combination of the rank information coming from 3 low-correlated ($30-30-30$) variables enhances the efficiency more than 1.2, 1.1 and 1.08 times, respectively. These results encouraged us to promote our new method for cases where there are concomitant variables with only moderate or low correlation levels (this situation is more likely to be faced in real life applications).

- Combining the rank information coming from 3 concomitant variables which have similar correlation levels ($95-95-95$, $70-70-60$ or $30-30-30$) using the avg operation enhances the efficiency of the estimator in comparison with the estimators based on single ranker RSS and RSS-M. It could be another important result because it indicates that our new method with the combination of rank decisions coming from the multiple rankers could be a good alternative for those cases such as when the researcher has more than one concomitant variable having a similar correlation level.

- There is only one case where the avg operation is not the best. In the $95-35-30$ case (combination of rank information coming from 2 low-correlated concomitant variables and a highly-correlated one), the min operation in fact performs best.

CHAPTER FIVE
REAL DATA APPLICATIONS

In this section, we apply the proposed procedure to real data sets coming from two different biometric studies. We pay regard to two points for the selection of the studies. The first one is the convenience for the application of RSS. We know that RSS is a cost-efficient sampling method if and only if the units are ranked without actual measurement of the variable of interest and the ranking cost (the measurement cost of the concomitant variable or the cost of expertise) is lower than the cost of actual measurement of the variable of interest. Thus, the studies in this section include the property that the actual measurement of the variable of interest is relatively time-consuming or expensive than the measurement of the concomitant variable. The second point is the potential for the usage of our new method in the field of the study. Thus, we focus on the biometric studies which the RSS method is first used in and widely applied.

5.1 The blood glucose dataset application

We first addressed the data set from the biometric study of Daramola (2012). Glycosylated hemoglobin (hbA1c) and random blood glucose (rbg) values were derived from 349 patients attending a hospital out-patient department in the original study. The test of hbA1c is a diabetic test to measure glycated hemoglobin which occurs when the hemoglobin protein within the red blood cells combines with the glucose in the blood. Since the red blood cells in our body survive for several weeks (up to 12 weeks) before renewal, medical practitioners can obtain average blood sugar levels over that period.

HbA1c % (old units)	HbA1c mmol/mol Range (new units)
5.8	40
6.7	50
7.6	60
8.6	70
9.5	80
10.4	90
11.3	100
12.2	110
13	120
14	130
15	140
15.9	150

Figure 5.1 hbA1c levels (healthnavigator.org, 2017)

The rbg test checks the glucose level in the blood at the time of measuring. It is easily done by pricking the finger to draw a small drop of blood which is wiped onto a test strip that gives the glucose level results.

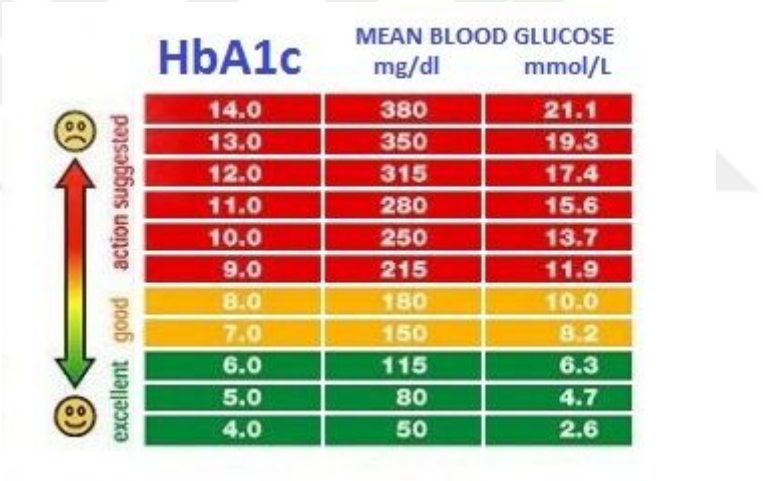

Figure 5.2 Rbg test levels

In this example 349 patients in the dataset are supposed as the population. The concerning feature is the $hbA1c$ values of the patients. We know that the $hba1c$ test is more time consuming and around ten times more expensive than the rbg test. Thus the rbg test can be used as a concomitant variable to obtain a fuzzy-weighted ranked set sample of the patients. The $FRSS$ procedure where the memberships are determined using the relative distance method is applied with settings $setsize = 3$, $cyclesize = 4$ as given in Table 5.1.

Table 5.1 Blood glucose data set application of the $FRSS$ procedure with a single ranker case and the mean estimation

Cycle		Concomitant variable			Membership degrees			Sample
(j)		$Y_{1,j}$	$Y_{2,j}$	$Y_{3,j}$	rank1	rank2	rank3	(X)
	Set1	5.978	7.134	11.179	1.000	0.778	0.000	5.392
1	Set2	6.800	10.401	17.549	0.665	1.000	0.335	11.144
	Set3	6.578	7.428	7.883	0.000	0.651	1.000	9.248
	Set1	6.493	6.505	10.841	1.000	0.997	0.000	11.471
2	Set2	5.978	6.447	9.027	0.846	1.000	0.154	5.131
	Set3	6.727	6.816	11.554	0.000	0.018	1.000	12.974
	Set1	5.978	7.473	9.465	1.000	0.571	0.000	5.850
3	Set2	8.059	8.380	8.469	0.217	1.000	0.783	8.464
	Set3	5.173	5.379	11.016	0.000	0.035	1.000	10.882
	Set1	6.391	6.670	11.895	1.000	0.931	0.000	14.085
4	Set2	9.525	14.912	15.714	0.130	1.000	0.870	18.268
	Set3	6.578	9.422	16.101	0.000	0.299	1.000	26.176
								$\bar{X}_{RSS} = 11.590$
								$\bar{X}_{FRSS} = 11.290$

We can see in Table 5.1 that the units chosen to sample are the same in RSS and FRSS. The slight difference in the estimation comes from the additional membership information in the procedure of FRSS and its usage in the estimation of the population mean.

5.2 The abalone dataset application

In the second application, the real data set (called the abalone data set) obtained from a biometrical study of Nash et al. (1994) about Blacklip Abalones is used to observe the performance of our new method. Haliotis rubra, blacklip abalone, is a

marine gastropod mollusk which has a large and flattened shell. The Haliotis rubra occurs along the south eastern Australian coastline and is one of the most exported marine products of Australia (McIlgorm & Goulstone (2001)).

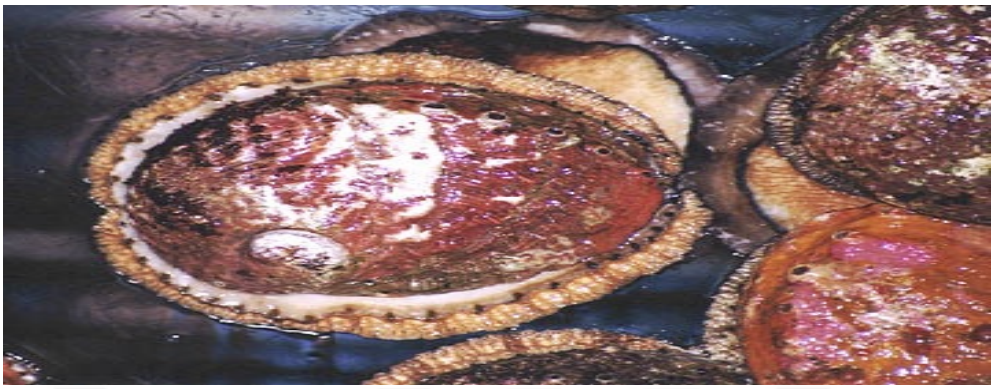

Figure 5.3 Haliotis rubra (McIlgorm and Goulstone, 2001)

The original study was conducted to predict the age of the abalones, an expensive and time-consuming task, from their physical measurements. In their study, Nash et al. (1994) suggested using other physical measurements which are easier to obtain through a regression model to predict the age of the abalones. The data set contains 4177 units with the variables, Sex, $Length$ (mm), $Diameter$ (mm), $Height$ (mm), $Whole\ weight$ $(grams)$, $Shucked\ weight$ $(grams)$, $Viscera\ weight$ $(grams)$, $Shell\ weight$ $(grams)$, $Rings$ (integer $+1.5$ gives the age in years).

Figure 5.4 The habitat and the main trade area of Haliotis rubra

We first apply the proposed $FRSS$ procedure and the mean estimation for a single

ranker case where $diameter$ is chosen as the concomitant variable Y and $number\ of\ rings$ is chosen as the variable of interest, X. $FRSS$ procedure where the memberships are determined using the eq.3.3 is applied with the settings $k = 3$, $n = 4$ as given in Table 5.2.

Table 5.2 Abalone data set application of the $FRSS$ procedure with a single ranker and the mean estimation

Cycle		Concomitant variable			Membership degrees			Sample
(j)		$Y_{1,j}$	$Y_{2,j}$	$Y_{3,j}$	rank1	rank2	rank3	(X)
	Set1	0.315	0.335	0.405	1.000	0.778	0.000	7
1	Set2	0.255	0.420	0.540	0.421	1.000	0.579	9
	Set3	0.270	0.440	0.570	0.000	0.568	1.000	9
	Set1	0.185	0.430	0.460	1.000	0.109	0.000	4
2	Set2	0.360	0.405	0.505	0.690	1.000	0.310	9
	Set3	0.360	0.450	0.525	0.000	0.546	1.000	11
	Set1	0.400	0.460	0.495	1.000	0.369	0.000	9
3	Set2	0.355	0.385	0.440	0.647	1.000	0.353	8
	Set3	0.440	0.470	0.495	0.000	0.547	1.000	13
	Set1	0.305	0.435	0.495	1.000	0.316	0.000	10
4	Set2	0.260	0.350	0.440	0.500	1.000	0.500	10
	Set3	0.310	0.420	0.475	0.000	0.546	1.000	18
							$\bar{X}_{RSS} = 9.750$	
							$\bar{X}_{FRSS} = 9.868$	

We can see in Table 5.2 that $\bar{X}_{FRSS}$ is slightly greater than $\bar{X}_{RSS}$. (It was $\bar{X}_{FRSS} < \bar{X}_{RSS}$ in Table 5.1 for the previous application.) This results indicates that the additional membership information can affect the estimation of the population mean bidirectionally.

We also give an application of the combination of the rank information coming from multiple ranker for a single cycle and $setsize = 3$. $Diameter$, $height$ and

$whole\ weight$ are chosen as the concomitant variables, Y_1, Y_2 and Y_3, respectively. The operations avg, min and max are applied separately to determine the membership degrees. After determination of the fuzzy memberships, we can choose the observation which has the highest membership degree for the concerned rank for the sample.

Table 5.3 Applications of combination operations for a single set

Conc. var.	Units		Membership degrees			Comb. method	Combined memberships		
(c)	(i)	Y_i^c	$m_i^c(1)$	$m_i^c(2)$	$m_i^c(3)$		$m_i^c(1)$	$m_i^c(2)$	$m_i^c(3)$
	1	0.390	1.000	0.500	0.000		0.867	0.789	0.133
1	2	0.510	0.000	0.500	1.000	avg	0.000	0.344	1.000
	3	0.450	0.500	1.000	0.500		0.789	0.867	0.211
	1	0.135	0.600	1.000	0.400		0.667	0.333	0.000
2	2	0.150	0.000	0.400	1.000	min	0.000	0.000	1.000
	3	0.125	1.000	0.600	0.000		0.333	0.667	0.000
	1	0.769	1.000	0.868	0.000		1.000	1.000	0.400
3	2	1.456	0.000	0.133	1.000	max	0.000	0.500	1.000
	3	0.860	0.868	1.000	0.133		1.000	1.000	0.500

As can be seen in Table 5.3 different operations can give different membership degrees for the same unit and it sometimes means that different units are chosen to sample.

We conduct a small simulation study to show the performance of the new method using the abalone data set. The settings are determined as $k = 4$, $n = 3$ and $C = 3$. We tried to achieve two different objectives. Our first objective was to obtain the mean estimation of the $number\ of\ rings$, X; $diameter$, $height$ and $whole\ weight$ were chosen as concomitant variables, Y_1, Y_2 and Y_3, respectively. The correlation levels of the concomitants were 0.5747, 0.5575 and 0.5404 for this objective. For the second objective, we wanted to obtain the mean estimation of the $viscera\ weight$, X, and $diameter$, $height$ and $whole\ weight$ were chosen as the concomitant variables. The

correlation levels of the concomitants are 0.9030, 0.8997 and 0.9664 for the second objective. Note that the membership degrees were determined based on the relative distance method given in eq. 3.3 and the combination of the rank information was done by using the avg operation. The relative efficiency results based on the empirical mean square errors computed with a repetition rate of 10000 are summarized for both objectives in Table 5.4.

Table 5.4 The relative efficiency results of the estimator based on FRSS for the abalone dataset

	$\frac{MSE(\bar{X}_{SRS})}{MSE(\bar{X}_{FRSS})}$	$\frac{MSE(\bar{X}_{RSS1})}{MSE(\bar{X}_{FRSS})}$	$\frac{MSE(\bar{X}_{RSS2})}{MSE(\bar{X}_{FRSS})}$	$\frac{MSE(\bar{X}_{RSS-M})}{MSE(\bar{X}_{FRSS})}$
Obj.1	1.1451	1.0129	1.0011	1.0122
Obj.2	2.0517	1.1090	1.0704	1.0486

The relative efficiency results in Table 5.4 are compatible with the simulation study results in the previous section. The relative efficiency of the new estimator $\bar{X}_{FRSS-avg}$ is greater than 1 for both of the objectives. For the second objective in which the highly-correlated concomitant variables are used, it increases to 2.05 and 1.10 in comparison with SRS and RSS . These results encourage us to say that our new method can be a good alternative to the sampling methods using in the biometric studies.

CHAPTER SIX
CONCLUSIONS

Ranked set sampling is an effective method of obtaining data and making inferences about the population. The main impact of RSS is its use of the ranking information of the units in the sampling mechanism. In the practice, the ranking process is done without actual measurement of the variable of interest in RSS. Since, there is an ambiguity in discriminating the rank of one unit with another without actual measurement, this ranking process could not be perfect. This study provides a fuzzy inspired perspective for dealing with the uncertainty occurring in the ranking process of RSS. We constructed fuzzy sets in the ranking process to allow the units to belong not only to the most possible rank but also to other possible ranks. A new sampling procedure, called Fuzzy-weighted Ranked Set Sampling, was introduced to obtain the sample units and their membership degrees. This new procedure can work with both a human ranker and a concomitant variable. We defined the combination methods for the ranking information coming from multiple rankers. Thus, the FRSS procedure can be used in both single and multiple ranker cases. A new estimator for the population mean based on the proposed sampling procedure was introduced.

The simulation studies for single-ranker and multiple-ranker cases were conducted individually. The single-ranker case results show that the proposed sampling method allows us to increase the efficiency of the RSS method with the additional information of membership degrees. There is no additional cost of ranking in FRSS method. The relative efficiency increases with set sizes and the correlation levels and extends to 4.04 in comparison with SRS and 1.88 in comparison with RSS for different cases. The multiple-ranker case results show that the idea of the combination of the membership degrees coming from multiple rankers performs quite well for the combining operations avg and min. In particular, the mean estimator of the $FRSS$ method based on the avg operation provides substantial improvement over the counterparts based on SRS, single rankers of RSS and $RSS - M$. The improvement is generally maintained for different correlation levels of the rankers even if the ranking performance of one of the

rankers is quite low.

The numerical examples and simulation study results based on the biometric data sets indicate that the new sampling method, $FRSS$, could be a useful alternative to SRS and RSS for data collection. The estimation of the population mean based on FRSS is relatively more efficient than its counterparts and additionally its application is not complicated.

Using our approach, further studies can proceed in different ways. Firstly, researchers can improve FRSS by using new functions for the determination of the memberships and/or new operations for the combination of the rank information coming from multiple rankers. In particular, introducing new operations for the *liberal* and the *conservative* combinations could be a good starting point. Based on FRSS, new statistical estimators for different population parameters such as variance and proportion can be defined. The density estimation, regression parameter estimation or other estimations and tests which are studied in the RSS literature can also be addressed. In another way, the FRSS procedure can be modified as in the procedures Median RSS, Extreme RSS, Percentile RSS and so on. Finally, the application of our new method to different real life problems can be beneficial for practitioners.

REFERENCES

Al-Nasser, A. (2007). L ranked set sampling: A generalization procedure for robust visual sampling. *Communications in Statistics - Simulation and Computation*, *36*(1), 33-43.

Al-Nasser, A. D., & Mustafa, A. B. (2009). Robust extreme ranked set sampling. *Journal of Statistical Computation and Simulation*, *79*(7), 859-867.

Al-Omari, A. I. (2012). Ratio estimation of the population mean using auxiliary information in simple random sampling and median ranked set sampling. *Statistics Probability Letters*, *82*(11), 1883 - 1890.

Al-Omari, A. I. (2016). Quartile ranked set sampling for estimating the distribution function. *Journal of the Egyptian Mathematical Society*, *24*(2), 303 - 308.

Al-Saleh, M. F., & Al-Kadiri, M. A. (2000). Double-ranked set sampling. *Statistics and Probability Letters*, *48*(2), 205 - 212.

Al-Saleh, M. F., & Al-Omari, A. I. (2002). Multistage ranked set sampling. *Journal of Statistical Planning and Inference*, *102*(2), 273 - 286.

Al-Saleh, M. F., & Samawi, H. (2010). On estimating the odds using moving extreme ranked set sampling. *Statistical Methodology*, *7*(2), 133 - 140.

Aragon, M. E. D., Patil, G., & Taillie, C. (1999). A performance indicator for ranked set sampling using ranking error probability matrix. *Environmental and Ecological Statistics*, *6:1*, 75-89.

Arnold, B. C., Balakrishnan, N., & Nagaraja, H. N. (1992). *A first course in order statistics*. New York: Wiley.

Arslan, G., & Ozturk, O. (2013). Parametric inference based on partially ordered ranked set samples. *Journal of Indian Statistical Association*, *51*, 1–24.

Bilgiç, T., & Türkşen, I. B. (2000). *Measurement of Membership Functions: Theoretical and Empirical Work*, (195–227). Boston, MA: Springer US.

Bohn, L. L., & Wolfe, D. A. (1992). Non-parametric two sample procedures for ranked set sampling data. *Journal of the American Statistical Association*, *87*, 552-561.

Bohn, L. L., & 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.

Chaturvedi, D. K. (2008). *Soft Computing Techniques and its Applications in Electrical Engineering*. Springer.

Chen, Z. (2000). On ranked-set sample quantiles and their applications. *Journal of Statistical Planning and Inference*, *83*, 125-135.

Chen, Z. (2001). Ranked-set sampling with regression-type estimators. *Journal of Statistical Planning and Inference*, *92*, 181-192.

Chen, Z., Bai, Z., & Sinha, B. K. (2004). *Ranked Set Sampling: Theory and Application*. New York: Springer.

Daramola, O. (2012). *Assessing the validity of random blood glucose testing for monitoring glycemic control and predicting HbA1c values in type 2 diabetics at Karl Bremer hospital*. PhD thesis, Stellenbosch University, South Africa.

David, H. A., & Levine, D. (1972). Ranked set sampling in the presence of judgment error. *Biometrics*, *28*, 553-555.

Dell, T. R., & Clutter, J. L. (1972). Ranked set sampling theory with order statistics background. *Biometrika*, *28*, 545-555.

Dubois, D., & Prade, H. (1988). *Possibility theory: an approach to computerized processing of uncertainty, 1st ed.* Plenum Press, New York .

Fligner, M. A., & MacEachern, S. N. (2006). Nonparametric two-sample methods for ranked-set sample data. *Journal of the American Statistical Association*, *101*(475), 1107-1118.

Frey, J. (2007). New imperfect rankings models for ranked set sampling. *Journal of Statistical Planning and Inference*, *137:4*, 1433-1445.

Frey, J. (2014). A note on fisher information and imperfect ranked-set sampling. *Communications in Statistics - Theory and Methods*, *43:13*, 2726-2733.

Frey, J., & Feeman, T. G. (2012). An improved mean estimator for judgment post-stratification. *Computational Statistics Data Analysis*, *56*(2), 418 - 426.

Frey, J., & Ozturk, O. (2011). Constrained estimation using judgment post-stratification. *Annals of the Institute of Statistical Mathematics*, *63*(4), 769–789.

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

Haq, A. (2011). Partial ranked setsampling design. http://www.math.canterbury.ac.nz/canterbury.../Haq—PhD.pptx, Last accessed on 2018-05-30.

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

Hossain, S., & Muttlak, H. (2008). Hypothesis tests on the scale parameter using median ranked set sampling. *Statistica*, *66*(4), 415–434.

Ibrahim, K., Syam, M., & Al-Omari, A. I. (2010). Estimating the population mean using stratified median ranked set sampling. *Applied Mathematical Sciences*, *4*, 2341-2354.

Jafari Jozani, M., & Johnson, B. C. (2011). Design based estimation for ranked set sampling in finite populations. *Environmental and Ecological Statistics*, *18*(4), 663-685.

Jemain, A. A., Al-Omari, A. I., & Ibrahim, K. (2008). Some variations of ranked set sampling. *Electronic Journal of Applied Statistical Analysis*, *1*, 1–15.

Kahraman, C., Öztayşi, B., & Çevik Onar, S. (2016). A comprehensive literature review of 50 years of fuzzy set theory. *International Journal of Computational Intelligence Systems*, *9*(1), 3-24.

Kvam, P. H., & Samaniego, F. J. (1993). On the inadmissibility of empirical averages as estimators in ranked set sampling. *Journal of Statistical Planning and Inference*, *36*, 39-55.

MacEachern, S. N., Stasny, E. A., & Wolfe, D. A. (2004). Judgement post-stratification with imprecise rankings. *Biometrics*, *60:1*, 207–215.

McIlgorm, A., & Goulstone, A. (2001). Changes in fishing capacity and ownership of harvesting rights in the new south wales abalone fishery. Technical report, Dominion Consulting Pty Ltd and NSW Fisheries, NSW, Australia.

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

Mode, N. A., Conquest, L. L., & Marker, D. A. (1999). Ranked set sampling for ecological research: accounting for the total costs of sampling. Technical report, A National Research Center on Statistics and the Environment, University of Washington , Seattle.

Mohd Adnan, M. R. H., Sarkheyli, A., Mohd Zain, A., & Haron, H. (2015). Fuzzy logic for modeling machining process: a review. *Artificial Intelligence Review*, *43*(3), 345–379.

Murray, R. A., Ridout, M. S., & Cross, J. V. (2000). The use of ranked set sampling in spray deposit assessment. *Aspects of Applied Biology*, *57*, 141-146.

Muttlak, H. A. (1997). Median ranked set sampling. *J. Appl. Statist. Sci*, *6*, 245-255.

Muttlak, H. A. (1998). Median ranked set sampling with concomitant variables and a comparison with ranked set sampling and regression estimators. *Environmetrics*, *9*(3), 255–267.

Muttlak, H. A. (2003a). Investigating the use of quartile ranked set samples for estimating the population mean. *Applied Mathematics and Computation*, *146*, 437-443.

Muttlak, H. A. (2003b). Modified ranked set sampling methods. *Pakistan Journal of Statistics*, *19*(3), 315 - 323.

Muttlak, H. A., & Abu-Dayyeh, W. (2004). Weighted modified ranked set sampling methods. *Applied Mathematics and Computation*, *151*(3), 645 - 657.

Muttlak, H. A., & Al-Sabah, W. (2003). Statistical quality control based on ranked set sampling. *Journal of Applied Statistics*, *30*(9), 1055-1078.

Nash, W. J., Sellers, T. L., Talbot, S. R., Cawthorn, A. J., & Ford, W. B. (1994). The population biology of abalone. Technical report, Sea Fisheries Division, Australia.

Nazari, S., Jozani, M. J., & Kharrati-Kopaei, M. (2016). On distribution function estimation with partially rank-ordered set samples: estimating mercury level in fish using length frequency data. *Statistics : A Journal of Theoretical and Applied Statistics*, *50:6*, 1387-1410.

Ozturk, O. (2008). Inference in the presence of ranking error in ranked set sampling. *Canadian Journal of Statistics*, *36:4*, 577-594.

Ozturk, O. (2011a). Combining ranking information in judgment post stratified and ranked set sampling designs. *Environmental and Ecological Statistics*, *19:1*, 73–93.

Ozturk, O. (2011b). Introduction to ranked set sampling. http://dm.ieu.edu.tr/WorkshopRSS2011/monday1.pdf, Last accessed on 2018-05-30.

Ozturk, O. (2011c). Sampling from partially rank-ordered sets. *Environmental and Ecological Statistics*, *18*, 757–779.

Ozturk, O. (2013). Combining multi-observer information in partially rank-ordered judgment post-stratified and ranked set samples. *Canadian Journal of Statistics*, *41:2*, 304-324.

Ozturk, O., Bilgin, O. C., & Wolfe, D. A. (2005). Estimation of population mean and variance in flock management: a ranked set sampling approach in a finite population setting. *Journal of Statistical Computation and Simulation*, *75*(11), 905-919.

Ozturk, O., & Jozani, M. J. (2014). Inclusion probabilities in partially rank ordered set sampling. *Computational Statistics and Data Analysis*, *69*, 122-132.

Park, S., & Lim, J. (2012). On the effect of imperfect ranking on the amount of fisher information in ranked set samples. *Communications in Statistics-Theory and Methods*, *41*, 3608-3620.

Patil, G. P., Sinha, A. K., & Taillie, C. (1995). Finite population corrections for ranked set sampling. *Annals of the Institute of Statistical Mathematics*, *47*(4), 621-636.

Presnell, B., & Bohn, L. L. (1999). U-statistics and imperfect ranking in ranked set sampling. *Journal of Nonparametric Statistics*, *10:2*, 111–126.

Ridout, M. S. (2003). On ranked set sampling for multiple characteristics. *Environmental and Ecological Statistics*, *10:2*, 255–262.

Ross, T. J. (2016). *Fuzzy Logic with Engineering Applications, 4th Edition*. John Wiley and Sons.

Samawi, H. M., Abu Dayyeh, W., & Ahmed, M. S. (1996). Estimating the population mean using extreme ranked set sampling. *Biometrical Journal*, *38*, 577-568.

Samawi, H. M., & Muttlak, H. A. (1996). Estimation of ratio using rank set sampling. *Biometrical Journal*, *38*, 753-764.

Stokes, S. L. (1976). *An investigation of the consequences of ranked set sampling*. PhD thesis, University of North Carolina.

Stokes, S. L. (1977). Ranked set sampling with concomitant variables. *Communications in Statistics - Theory and Methods*, *12*, 1207-1211.

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

Stokes, S. L. (1980b). Inferences on the correlation coefficient in bivariate normal populations from ranked set samples. *Journal of the American Statistical Association*, *75*, 989-995.

Stokes, S. L., & 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.

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

Wang, X., Lim, J., & Stokes, L. (2008). A nonparametric mean estimator for judgment poststratified data. *Biometrics*, *64*(2), 355–363.

Wang, X., Stokes, L., Lim, J., & Chen, M. (2006). Concomitants of multivariate order statistics with application to judgment poststratification. *Journal of the American Statistical Association*, *101*(476), 1693-1704.

Werro, N. (2015). *Fuzzy Classification of Online Customers*. Springer.

Wolfe, D. A. (2012). Ranked set sampling: Its relevance and impact on statistical inference. *ISRN Probability and Statistics*, *2012*, 1-32.

Yu, P. L. H., & Lam, K. (1997). Regression estimator in ranked set sampling. *Biometrics*, *53*, 1070-1080.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, *8*, 338-353.

Zimmermann, H. (2001). *Fuzzy Set Theory—and Applications, 4th Rev.* Boston: Kluwer Academic Publishers.

Zimmermann, H.-J. (2010). Fuzzy set theory. *Wiley Interdisciplinary Reviews: Computational Statistics*, *2*(3), 317–332.