The Performance of "The Five-Number Summary" Obtained Using Different Sampling Techniques

By

Amin Khaled Mohammad Na'amnih

Supervisor

Prof. Mohammad Fraiwan Al-Saleh

Program: Statistics

December 28, 2011

The Performance of "The Five-Number Summary" Obtained Using Different Sampling Techniques

By

Amin Khaled Mohammad Na'amnih

B.Sc. (Mathematics), Irbid National University, 2004

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in the Department of Statistics, Yarmouk University, Irbid, Jordan.

Approved by

Prof. Mohammad Fraiwan Al-Saleh -- **Chairman**

Professor of Statistics, Yarmouk University

Dr. Mohammed H. Baker AL-Haj Ebrahem -- **Member**

Associate Professor of Statistics, Yarmouk University

Dr. Amer Ibrahim Al-Omari -- **Member**

Assistant Professor of Statistics, Al al-Bayt University

December 28, 2011

Acknowledgment

At the beginning, I want to thank Allah, the most merciful and most gracious.

I would like to express my sincere thanks to my supervisor, Prof. Dr. Mohammad Fraiwan Al-Saleh for his encouragement and guidance during the course of this research and the members of the defense exam committee for their suggestions.

Special thanks to my parents (father and mother), my brothers and sisters for their support and encouragement.

Finally, thanks to all friends, faculty members and students in the Department of Statistics at Yarmouk University.

Table of Contents

Chapter Three : Coverage Probability of the "Five-Number Summary" Based on Different Sampling Techniques

Chapter Four : Conclusions and Suggestion for Further Research

List of Tables

Page

Terminology

SRS	Simple random sampling
RSS	Ranked set sampling
MERSS	Moving extreme ranked set sampling
CP	Coverage probability
$E(R)$	Expected range
$E(R_{ij})$	Expected range for the (i,j) division
RCP	Relative coverage probability
RCP_{ij}	Relative coverage probability for the (i,j) division
$X_{(i)}$	The i^{th} order statistic
$X_{(1)}$	Minimum order statistic
$X_{(n)}$	Maximum order statistic
Q_2	The median
Q_1	First quartile
Q_3	Third quartile
μ	The mean
σ^2	The variance
$X_{(i:m)}^{k}$	The i^{th} order statistic for a SRS of size m in the k^{th} cycle
iid	Independent and identically distributed
pdf	Probability density function $f(x)$
cdf	Cumulative distribution function $F(x)$
$f_{(i)}(x)$	Probability density function of $X_{(i)}$
RP	Relative precision
RS	Relative saving
Δ	Overlapping coefficient
$X_{(i:m)}^{ORSS}$	Ordered RSS, ORSS, obtained by arranging $X_{(i:m)}$

$f_{i:m}^{ORSS}(x)$	The density function of $X_{(i:m)}^{ORSS}$
WLOG	Without loss of generality
w. r. t.	With respect to
$Y_{(i:n)}$	The i^{th} order statistics of the RSS or MERSS (after ordered)
$U(0,1)$	Uniform distribution
$N(0,1)$	Standard normal distribution
$Cauchy(0,1)$	Cauchy distribution
$Exp(1)$	Exponential distribution
$I(a,b)$	The indicator function
$E(X)$	Expectation of X
$E(X\|Y)$	Conditional mean of X given Y
$BN(n,P)$	Binomial with parameters n and P

Abstract

Na'amnih, Amin khaled. The Performance of "The Five-Number Summary" Obtained Using Different Sampling Techniques. Master of Science Thesis, Department of Statistics, Yarmouk University, 2011. (Supervisor : Prof. Mohammed Fraiwan Al-Saleh)

In this thesis, we focused on comparing the three sampling techniques, SRS, RSS, MERSS, with respect to coverage probability, range and relative coverage probability for the pieces of the "5-number summary" obtained based on a sample taken using one of these procedures. Also, different properties of the elements of the sample with respect to the three sampling techniques are discussed; (Independence, probability of ordering, overlapping coefficient). This, gives us some idea about how well the "5-number summary" represents the different portions of the population of interest. We noted that overall, the RSS technique is more suitable than the SRS and MERSS. Real data that consists of heights and diameters of 1083 trees is used for illustration.

Keywords: Simple Random Sampling, Ranked Set Sampling, Moving Extreme Ranked Set Sampling, "5-number summary", Coverage Probability, Range, Relative Coverage Probability, Probability of Ordering, Overlapping Coefficient.

الملخص بالعربي

نعامنة، أمين خالد. أداء "ملخص الخمسة أرقام" باستخدام تقنيات أخذ العينات المختلفة. أطروحة ماجستير في قسم الإحصاء، جامعة اليرموك، ٢٠١١. (المشرف : الأستاذ الدكتور محمد فريوان الصالح).

في هذه الأطروحة، ركزنا على مقارنة طرق أخذ العينات الثلاث (MERSS, RSS, SRS) فيما يتعلق بنطاق التغطية، والمدى والاحتمال النسبي لتغطية القطع من "ملخص الخمسة أرقام". أيضا، تمت مناقشة خصائص مختلفة لهذه الطرق، (الاستقلال، واحتمال الترتيب، ومعامل التداخل). وهذا يعطينا فكرة عن أداء "ملخص الخمسة أرقام" بالنسبة للأجزاء المختلفة. وقد تبين لنا أن أسلوب RSS كان الأكثر ملائمة. ولمزيد من التوضيح تم استخدام مجموعة بيانات حقيقية تتألف من الطول والقطر لـ ١٠٨٣ شجرة.

الكلمات المفتاحية : المعاينة العشوائية البسيطة، المعاينة المرتبة، المعاينة المرتبة المتحركة، "ملخص الخمسة أرقام"، احتمال التغطية، المدى، الاحتمالية النسبية للتغطية، احتمال الترتيب، معامل التداخل.

Chapter One

Introduction and Literature Review

1.1 Introduction

The main objective of statistics is to make inference about the population based on the information contained in a sample.

A population is a collection of elements about which we wish to make inference and a sample is any non-empty subset of the population. Once a sample is obtained, the information in the sample is summarized using some descriptive statistical techniques. One popular summary is the "5-number summary"; which consists, of 5 order statistics when n is suitably chosen. The amount and accuracy of the information contained in these 5 order statistics depends heavily on the sampling technique used to collect the data. Our main objective here is to study the properties of this summary with respect to some different sampling techniques.

1.2 Five Number Summary Statistic

The "5-number summary" is a method of summarizing a data set. Reporting five numbers avoids the need to decide on the most appropriate summary statistic. Let $X_1, X_2, ..., X_n$ denote a random sample of size n and let $X_{(1)}, X_{(2)}, ..., X_{(n)}$ be the ordered data. The "5-number summary" consists of the following statistics:

1. **Minimum**: The first order statistic (smallest order statistic); it is denoted by $X_{(1)}$.

2. **Maximum**: The highest order statistic (largest order statistic); it is denoted by $X_{(n)}$.

3. **Median**: The median is: $X_{(n+1)/2}$ if n is odd and $\frac{X_{(\frac{n}{2})}+X_{(\frac{n}{2}+1)}}{2}$ if n is even. It can be written also as:

$$Q_2 = \frac{X_{([\frac{n+1}{2}])}+X_{([\frac{n}{2}]+1)}}{2}.$$

4. **First Quartile**: It is the median of the lower half of the data; it is denoted by Q_1.

5. **Third Quartile**: It is the median of the upper half of the data; it is denoted by Q_3.

Therefore, the "5-number summary" statistic is

$$(X_{(1)}, Q_1, Q_2, Q_3, X_{(n)}).$$

The "5-number summary" gives information about the location (from median), spread (from the quartiles) and range (from the minimum and maximum) of the observations. The general shape of the underlying distribution can be inferred from the "5-number summary".

1.3 Sampling Techniques

There are several sampling methods in sampling from a finite population without replacement. The following is a description of the main ones:

1. Simple random sampling (SRS): It is a sampling technique which involves the drawing of n units from a population of size N in such away that every possible sample of size n has the same chance (probability= $\frac{1}{\binom{N}{n}}$) of being selected. As a consequence of this definition, all elements of the population have equal chances, $\frac{n}{N}$, of being selected.
2. Stratified random sampling: It is another sampling technique that can be used when the population can be divided (stratified) into relatively similar within but different among subgroups (strata). A stratified random sample consists of several random samples, one from each stratum. The technique is suitable when there is a relatively high variability among strata and little variability within each stratum.
3. Systematic random sampling: "A random sample obtained by randomly selecting one element from the first k elements in the frame and every k^{th} element thereafter is called a $1-in-k$ systematic sample", [Scheaffer et al.(1986)]. k is a suitable positive

integer chosen so that we end up with a sample of size n; usually k is the greatest integer of N/n; $k = \left[\frac{N}{n}\right]$.

4. Cluster random sampling: This technique is a sampling procedure used when the population consists of groups, (or clusters). The set of elements of a random sample of these clusters is called a cluster random sample. So, the ideal situation occurs when there is a relatively small variability among clusters and large variability within clusters.

5. Ranked set sampling (RSS): The idea of ranked set sampling was first proposed by McIntyre (1952), to increase the accuracy of crop yield estimates without increasing the number of observations that need to be quantified. This method of sampling can be described as follows:

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

Step 2: Randomly partition the sample into m sets, each of size m.

Step 3: The units within each set are then ranked w.r.t. the variable of interest; this may be based on personal judgment or any inexpensive method without actual measurements.

Step 4: m measurements are obtained by taking the smallest unit from the first set, the second smallest from the second set; the procedure continues in this manner until the largest unit has been selected from last set.

Step 5: Repeat steps (1- 4) if necessary, r times until the desired sample of size $n = rm$, is obtained.

The following display contains RSS of size $n = rm$,

$$\begin{array}{cccc} X^1_{(1:m)} & X^1_{(2:m)} & \cdots\cdots & X^1_{(m:m)} \\ X^2_{(1:m)} & X^2_{(2:m)} & \cdots\cdots & X^2_{(m:m)} \\ \vdots & \vdots & \cdots\cdots & \vdots \\ X^r_{(1:m)} & X^r_{(2:m)} & \cdots\cdots & X^r_{(m:m)} \end{array}$$

$X^k_{(i:m)}$: is the i^{th} order statistic for a SRS of size m in the k^{th} cycle.

We note that, the elements in each column are independent and identically distributed (iid), while the elements in each row are only independent.

Many extensions of the RSS were developed including double RSS, multistage RSS, bivariate RSS, moving extreme RSS, median RSS,…,etc.

6. Moving extreme ranked set sampling (MERSS): It is another recent modification of ranked set sampling. This method was introduced and investigated by Al-Odat and Al-Saleh (2001) and investigated further by Al-Saleh and Al–Hadrami (2003). The suggested method uses the maximum or the minimum of varied set size to obtain judgment order statistics. The procedure of MERSS is described as follows:

Step 1: Randomly select m simple random samples of sizes $1, 2, 3, \ldots, m$, respectively.

Step 2: Order the elements in each sample w.r.t. the variable of interest, by eye or some other relatively inexpensive method, without doing any actual measurements on the characteristic of interest.

Step 3: Measure accurately, the maximum ordered observation from the first set, the maximum ordered observation from the second set; the procedure continues in this fashion until the maximum ordered observation from the m^{th} set is measured.

Step 4: Repeat steps (1- 3), if necessary r independent cycles until the desired sample of size $n = rm$ is obtained for analysis.

Note: In the above procedure, we may use the minimum instead of the maximum. Also, sometime we may use both.

The following display contains MERSS of size $n = rm$.

$X_{(1:1)}^{1}$	$X_{(2:2)}^{1}$	………	$X_{(m:m)}^{1}$
$X_{(1:1)}^{2}$	$X_{(2:2)}^{2}$	………	$X_{(m:m)}^{2}$
$\vdots$	$\vdots$	………	$\vdots$
$X_{(1:1)}^{r}$	$X_{(2:2)}^{r}$	………	$X_{(m:m)}^{r}$

Again, the elements in each column are independent and identically distributed (iid), while the elements in each row are only independent.

1.4 Literature Review

Ranked Set Sampling was introduced as a cheap and efficient method of estimating mean pasture yield by **McIntyre (1952).** RSS has applications in many areas such as agriculture, ecology, environmental sciences and medical studies.

The first theoretical result about RSS was obtained by **Takahasi and Wakimoto (1968)**. Under perfect ranking, they proved that the sample mean of RSS is an unbiased estimator of the population mean, with smaller variance as compared to the sample mean of simple random sample with the same size. Their main results are summarized as follows:-

Let $f(x)$ be a probability density function with mean μ and variance σ^2; let $X_{(i:m)}$ be the i^{th} ordered statistic of a random sample $X_1, X_2, \dots, X_n$ from $f(x)$. Let $f_{i:m}(x)$ be the density of $X_{(i:m)}$ with mean $\mu_{(i:m)}$ and variance $\sigma^2_{(i:m)}$, then:

$$f(x) = \frac{1}{m}\sum_{i=1}^{m} f_{i:m}(x), \qquad \mu = \frac{1}{m}\sum_{i=1}^{m} \mu_{(i:m)}.$$

Under perfect ranking, let $\{ X^k_{(i:m)}, k = 1, \dots, r, i = 1, \dots, m\}$ be a RSS of size $n=rm$. An unbiased estimator of the population mean μ is defined as

$$\hat{\mu}_{RSS} = \frac{1}{mr}\sum_{k=1}^{r}\sum_{i=1}^{m} X^k_{(i:m)}.$$

The variance is

$$\sigma^2 = \frac{1}{m}\left(\sum_{i=1}^{m} \sigma_{(i:m)}^2 + \sum_{i=1}^{m} (\mu_{(i:m)} - \mu)^2\right).$$

Next, Takahasi and Wakimoto (1968) compared the performance of RSS estimator, $\hat{\mu}_{RSS}$ with that of the SRS estimator based on the same sample size. They defined the relative precision (RP) and relative saving (RS) by $RP = \frac{Var(\hat{\mu}_{SRS})}{Var(\hat{\mu}_{RSS})}$ and $RS = 1 - \frac{1}{RP}$, respectively. They showed that

$$1 \leq RP \leq \frac{m+1}{2} \text{ and } 0 \leq RS \leq \frac{m-1}{m+1}.$$

Dell and Clutter (1972) studied the case in which the ranking may not be perfect. They showed that the mean using the RSS is an unbiased estimator of the population, whether or not there are errors in ranking.

Stokes (1976) used ranked set sampling for the estimation of scale and location parameters. In sampling form $F\left(\frac{X-\mu}{\sigma}\right)$, the RSS estimator of μ and σ^2 are shown to be more efficient than the SRS estimators, at least for large sample size n.

Stokes (1977) studied RSS using a concomitant variable, she assumed that the variable of interest Y has a linear relationship with another variable X that is easy to rank.

Stokes and Sager (1988) used RSS to estimate distribution functions. They showed that the empirical distribution function (cdf) using RSS is an unbiased estimator of the distribution function and has smaller variance than that based on SRS.

Lam et al. (1994) used RSS to estimate the two parameters of exponential distribution:

$$f(x) = \frac{1}{\sigma} e^{-\frac{(x-\theta)}{\sigma}} \quad x \geq \theta, \sigma > 0.$$

They obtained the best linear unbiased estimator of θ and σ.

Samawi et al. (1996) proposed extreme RSS (ERSS) method and showed that the estimator of the mean in ERSS is more efficient than the mean estimator of SRS when the underlying distribution is symmetric.

Muttlak (1997) suggested an alternative RSS method that can be used to estimate the population mean. This method is called Median Ranked Set Sampling (MRSS).

Al-Saleh and Al-Kadiri (2000) suggested Double Ranked Set Sampling method (DRSS). The method increases the efficiency of the RSS estimator without increasing the set size m. It was shown that the DRSS estimator obtained is more efficient than that using RSS. Furthermore, it was shown that the ranking in the second stage is easier than the ranking in first stage. The method was generalized to multistage RSS by **Al-Saleh and Al-Omari (2002).**

Al-Odat and Al-Saleh (2001) introduced the concept of varied set size RSS, which is coined later as Moving Extreme Ranked Set Sampling by **Al-Saleh and Al-Hadrami (2003).**

Al-Saleh and **Al-Shrafat (2001)** provided a real application of RSS. They considered the estimation of average milk yield of Sheep. They conducted a study on a field of (402) sheep in East Jordan. It appears that the relative saving of sampling units when using RSS with set size (3) is between 32% and 44% compared to simple random sampling. The precision of the estimator obtained using RSS with respect to that obtained using SRS is calculated for both perfect and visual ranking.

Bayesian estimation with RSS was considered by **Al-Saleh and Muttlak (2000)** and **Al-Saleh and Abu-Hawwas (2002).**

Al-Saleh and Zheng (2002) Introduced a bivariate ranked set sampling.

Al-Saleh and Diab (2009) estimated the parameters of Downton's bivariate exponential distribution using ranked set sampling.

Ozturk and Deshpande (2004) proposed a confidence interval that has higher coverage probability and shorter expected length than its counterpart using SRS analog. In order to achieve the desired confidence level, a distribution-free confidence interval that interpolates the adjacent order statistics was constructed.

Zhu and Wang (2005) proposed a new quantile estimator and showed that it has a smaller asymptotic variance for all distributions. Both the optimal rank and the relative efficiency gain with respect to simple random sampling are distribution-free and depend on the set size and the given probability only.

For more details about RSS, see also **Kaur et al (1995), Zheng and Al-Saleh (2003), Wolfe (2004), Sinha (2005), Al-Saleh and Endeer (2007), Al-Omari and Al-Saleh (2009), Al-Omari and Al-Saleh (2010), Al-Saleh and Samawi (2010), Samuh and Al-Saleh (2011), Al-Saleh and Ababneh (2011), and Hanandeh (2011).**

In this research, we are going to investigate the properties of the "5-number summary" obtained based on SRS, RSS and MERSS. The following are the main properties that are to be investigated:

(1) The coverage probability: We will compare the coverage probability of each piece of the "5-number summary". *i.e.* $P(X_{(1)} \leq X \leq X_{(n)})$, $P(X_{(1)} \leq X \leq Q_{(1)})$, $P(Q_{(1)} < X \leq Q_{(2)})$, etc. where X is the population value.

(2) The range of the pieces: $X_{(n)} - X_{(1)}$, $Q_{(2)} - Q_{(1)}$, etc. These ranges have a close connection with coverage probability described in (1). *i.e.* The comparison should be based on the coverage probability per unit range.

(3) The relative coverage probability, which is the coverage per unit range.

(4) The overlapping between the densities of the elements of the obtained sample.

1.5 Organization of the Thesis

In this thesis, we are going to compare the three sampling techniques, SRS, RSS and MERSS with respect to coverage probability, range and relative coverage probability for the pieces of the "5-number summary". In chapter (2), the basic terminology is introduced. Also, different properties of the elements of the sample with respect to the three sampling techniques are discussed (independence, probability of ordering, overlapping coefficient are among the properties that are to be compared). In chapter (3), the coverage probability, range and relative coverage are obtained and compared for the different pieces of the "5-number summary". Illustration using a real data set is done also in this chapter. General concluding remarks and suggested future work are given in chapter (4).

Chapter Two

Properties of the Elements of SRS, RSS, MERSS

2.1 Introduction

Let $X_1, X_2, \dots, X_m$ be a SRS from a population with pdf $f(x)$ and cdf $F(x)$. Assume that $F(x)$ is absolutely continuous. Let $X_{(1:m)}$, $X_{(2:m)}, \dots,$ $X_{(m:m)}$ be a RSS of size m from the same population. Let $Y_{(1:1)}, Y_{(2:2)}, \dots,$ $Y_{(m:m)}$ be a MERSS of size m from the same population.

In this chapter, we will study some properties of the elements of these samples. Some of these properties may not be new. In section 2, we will discuss the joint densities of the elements of each sample and the probability of different ordering. In section 3, we will investigate the overlapping coefficient for each pair of the elements in each sampling plan. This, gives us some idea about how well the "5-number summary" separates the different portions of the population of interest.

2.2 Joint Densities and Probability of Ordering

The joint density of $X_1, X_2, \dots, X_m$ is

$$f(x_1, x_2, \dots, x_m) = \prod_{i=1}^{m} f(x_i) .$$

The joint density of $X_{(1:m)}$, $X_{(2:m)}, \dots, X_{(m:m)}$ is

$$f(x_1, x_2, \dots, x_m) = \prod_{i=1}^{m} f_{(i:m)}(x_i),$$

$$= \left(\prod_{i=1}^{m} f(x_i)\right) \prod_{i=1}^{m} m\binom{m-1}{i-1}(F(x_i))^{i-1}(1 - F(x_i))^{m-i},$$

while the joint density of $Y_{(1:1)}, Y_{(2:2)}, \dots, Y_{(m:m)}$ is

$$f(y_1, y_2, \dots, y_m) = \prod_{i=1}^{m} f_{(i:i)}(y_i) = \prod_{i=1}^{m} i(F(y_i))^{i-1} f(y_i)$$

$$= m! \left(\prod_{i=1}^{m} f(y_i)\right) \prod_{i=1}^{m} (F(y_i))^{i-1}.$$

Note that, SRS elements may be thought of as dependent order statistics : $(X_{(1)}, X_{(2)}, \dots, X_{(m)})$. Also, $X_{(i:m)} \stackrel{d}{=} X_{(i)}$, but $(X_{(1:m)}, X_{(2:m)}, \dots, X_{(m:m)})$ are independent while $(X_{(1)}, X_{(2)}, \dots, X_{(m)})$ are positively correlated. Also, clearly $Y_{(1:1)}, Y_{(2:2)}, \dots, Y_{(m:m)}$ are independent.

For any permutation $(i_1, i_2, \dots, i_m)$ of $(1, 2, \dots, m)$, we have

$$\mathrm{P}\left(X_{i_1} < X_{i_2} < \cdots < X_{i_m}\right) = \frac{1}{m!}. \tag{2.1}$$

It was shown by **Al-Saleh and Al-Kadiri (2000)** that

$$\pi_{RSS}(i_1, i_2, \dots \dots, i_m) = \mathrm{P}\left(X_{(i_1:m)} < X_{(i_2:m)} < \cdots < X_{(i_m:m)}\right) =$$

$$\frac{(m)^m}{\prod_{j=1}^{m}(i_j-1)!} \sum_{j_1=0}^{m-i_1} \sum_{j_2=0}^{m-i_2} \dots \sum_{j_m=0}^{m-i_m} \frac{(-1)^{\sum_{k=1}^{m} j_k}}{\prod_{k=1}^{m}\left[\sum_{r=1}^{k}(j_r-i_r)\right] \prod_{k=1}^{m}\left[j_k!(m-i_k-j_k)!\right]}. \tag{2.2}$$

Also, the maximum probability occurs when $(i_1, i_2, \dots, i_m) = (1, 2, \dots, m)$. It was shown by **Al-Saleh and Ababneh (2011)** that

$$\pi_{MERSS}(i_1, i_2, \dots, i_m) = \mathrm{P}\left(y_{(i_1:i_1)} < y_{(i_2:i_2)} < \cdots < y_{(i_m:i_m)}\right) = \frac{m!}{\prod_{k=1}^{m} \sum_{j=1}^{k} i_j} \tag{2.3}$$

for any permutation $(i_1, i_2, \dots, i_m)$ of $(1, 2, \dots, m)$. Also, the maximum of (2.3) occurs when $(i_1, i_2, \dots \dots, i_m) = (1, 2, \dots, m)$.

Note: It was shown by the above authors that the minimum probability of ordering occurs when $(i_1, i_2, \dots, i_m) = (m, m-1, \dots, 1)$.

Table (2.1) contains the maximum probability of ordering (*i.e.* when $(i_1, i_2, \dots, i_m) = (1, 2, \dots, m)$), for the three sampling techniques:

Table (2.1): Maximum probability of ordering for $m = 2, 3, 4$, using SRS, RSS and MERSS.

m	SRS	RSS	MERSS
2	0.50	0.8333	0.66667
3	0.16667	0.6095	0.33333
4	0.041667	0.4028	0.13333

Clearly, RSS produces the highest probability of ordering; *i.e.* the spread out in the data is highest in the case of RSS.

2.3 The Overlapping Coefficient or the Size of the "Indifference Zone"

In this section, we study another property which is called the overlapping coefficient or the size "Indifference Zone" between the densities of the elements of a sample. For more details about overlapping coefficient see **Al-Saleh (2007).** The following figure gives an illustration for the overlapping.

Based on (2.5), the overlapping coefficient between the densities of the elements of *iid* sample (SRS) is 1.

2.3.1 Overlapping in RSS

For $i < j$ we want to find the overlapping coefficient between any pair $X_{(i:m)}$ and $X_{(j:m)}$. Denote this coefficient by $\Delta(X_{(i:m)}, X_{(j:m)})$. Now, the pdf of $X_{(i:m)}$ is

$$f_{(i:m)}(x) = a_i\big(F(x)\big)^{i-1}\big(1 - F(x)\big)^{m-i} f(x),$$

where $a_i = m\binom{m-1}{i-1}$.

The following theorem is given by **Al-Saleh (2007).**

Theorem 2.1: Assume that $X_1, X_2, \dots, X_m$ are an *iid* with common absolutely continuous cdf F(x). For i < j, $i, j = 1,2, \dots, m$, the overlapping coefficient between $(X_{(i:m)}, X_{(j:m)})$ is

$$\Delta(X_{(i:m)}, X_{(j:m)}) = 1 - \mathrm{P}\big(i \le W_{ij} \le j - 1\big), \qquad (2.6)$$

where, W_{ij} is $BN(m, P_{ij})$ and $P_{ij} = \frac{\sqrt[j-i]{a_i}}{\sqrt[j-i]{a_i} + \sqrt[j-i]{a_j}}$, $a_i = m\binom{m-1}{i-1}$.

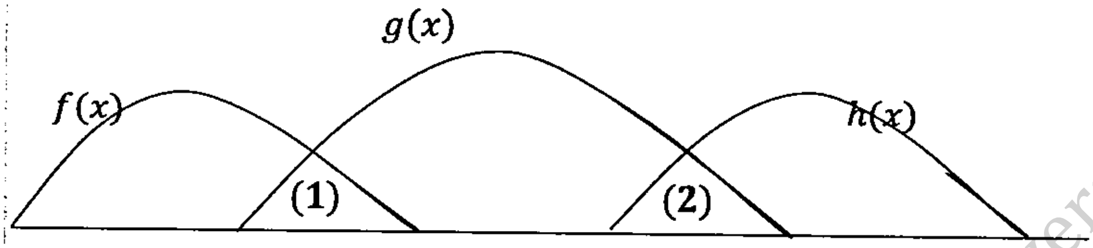
Using this result, we can obtain the overlapping coefficient for the densities of any pair of elements in RSS. For example; W_{1m} is $BN(m, 0.5)$; thus $\Delta(X_{(1:m)}, X_{(m:m)}) = \left(\frac{1}{2}\right)^{m-1}$.

Example 2.1: Suppose that $m = 5$, the overlapping coefficient between pairs of elements in RSS is given in Table (2.2).

Table (2.2): Overlapping coefficient for some pairs of elements of RSS, $m = 5$.

Pairs $(X_{(i:m)}, X_{(j:m)})$	P_{ij}	$\Delta(X_{(i)}, X_{(j)})$
$(X_{(1:5)}, X_{(5:5)})$	0.5	0.0625
$(X_{(1:5)}, X_{(2:5)})$	0.2	0.5904
$(X_{(2:5)}, X_{(3:5)})$	0.4	0.6544
$(X_{(3:5)}, X_{(4:5)})$	0.6	0.6544
$(X_{(4:5)}, X_{(5:5)})$	0.8	0.5904

We notice from the previous table, that $\Delta(X_{(1:5)}, X_{(2:5)}) = \Delta(X_{(4:5)}, X_{(5:5)})$, and $\Delta(X_{(2:5)}, X_{(3:5)} = \Delta(X_{(3:5)}, X_{(4:5)})$. It can be verified easily that $\Delta(X_{(i:m)}, X_{(i+1:m)}) = \Delta(X_{(m-i:m)}, X_{(m-i+1:m)})$. See **(Al-Saleh, 2007)**.



Region (1) represents the overlapping between $f(x)$ and $g(x)$, while region (2) represents the overlapping between $g(x)$ and $h(x)$.

To measure the similarity of two probability distributions, the overlapping coefficient (Δ) of Weitzman (1970) can be used. This coefficient measures the similarity, agreement, or closeness of two probability distributions. This measure has been considered by several authors such as **Inman and Bradley (1989), Reiser and Faraggi (1999), Clemons and Bradley (2000), and Mulekar and Mishra (2000).**

Let $f(x)$ and $g(x)$ be two pdfs, then (Δ) was defined by Weitzman (1970) as

$$\Delta = \int_{-\infty}^{\infty} min(f(x), g(x))dx. \qquad (2.4)$$

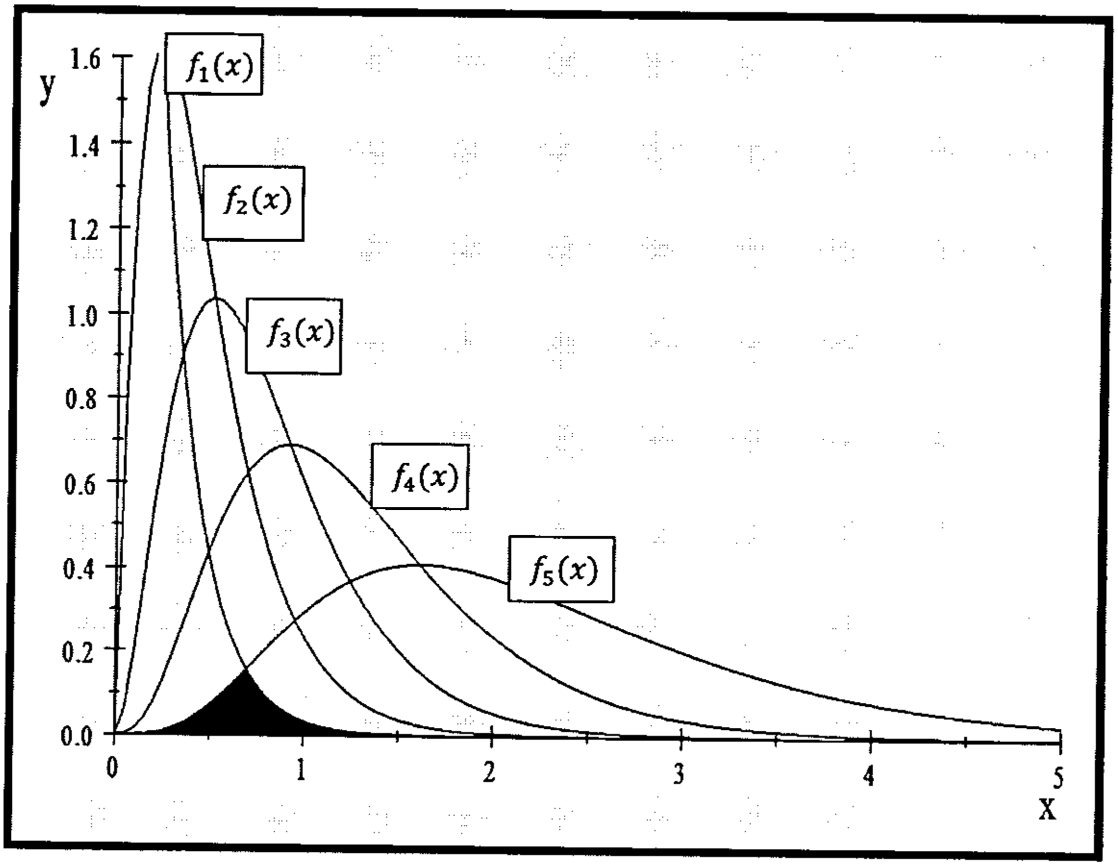
Using the identity $min(a, b) = \frac{a+b}{2} - \frac{|a-b|}{2}$, formula (2.4) can be rewritten as

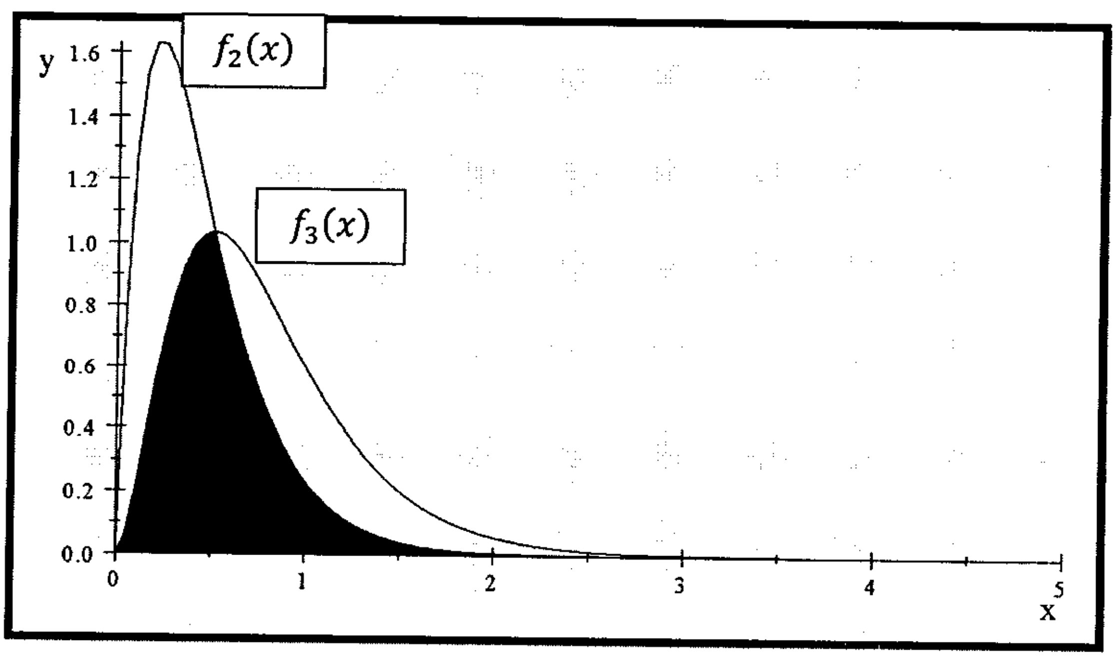
$$\Delta = 1 - \frac{1}{2}\int_{-\infty}^{\infty} |f(x) - g(x)|dx. \qquad (2.5)$$

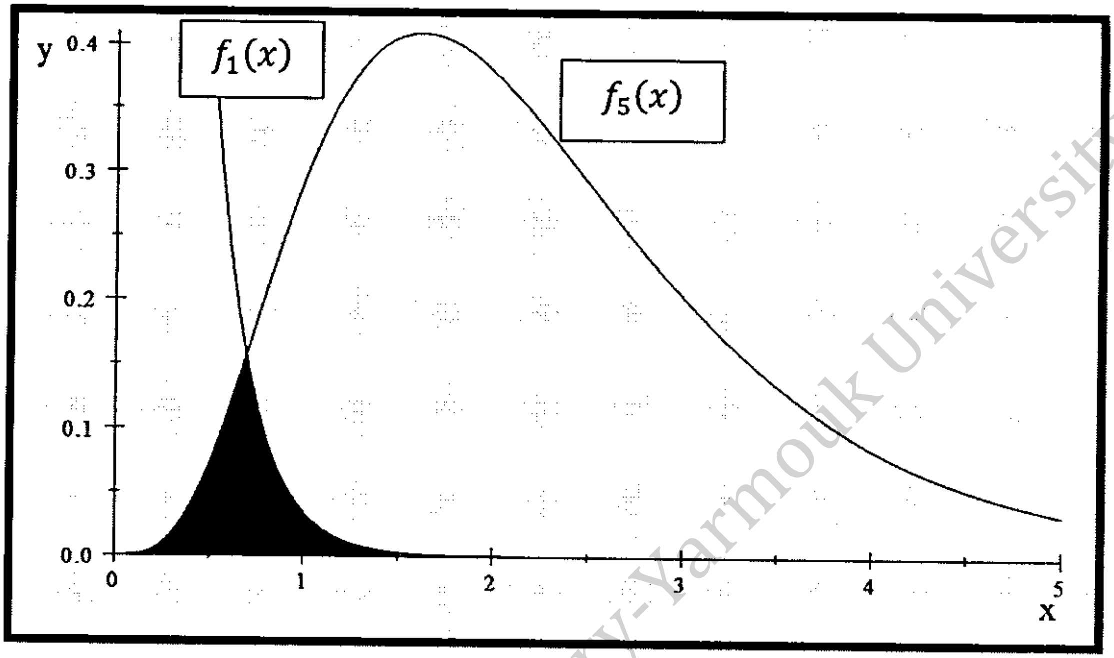
Note that the value of (Δ) is between zero and one. $\Delta = 1$ iff $f(x) = g(x)$; $\Delta = 0$ iff $f(x)g(x) = 0$. (*i.e.* $f(x)$ and $g(x)$ are distinct).

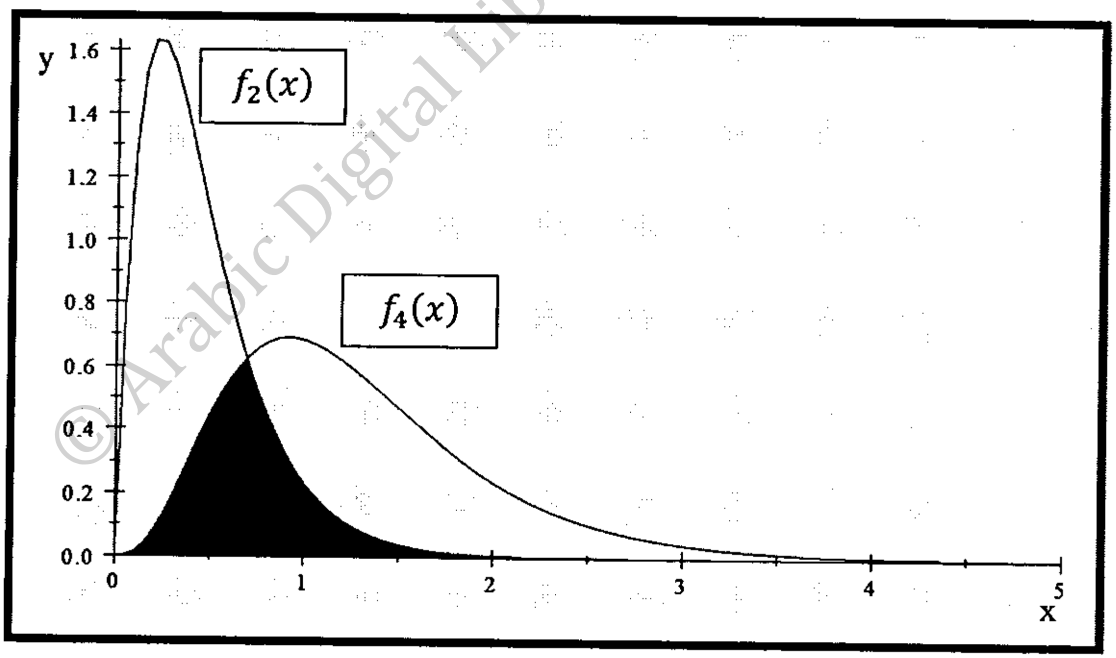
© Arabic Digital Library - Yarmouk University

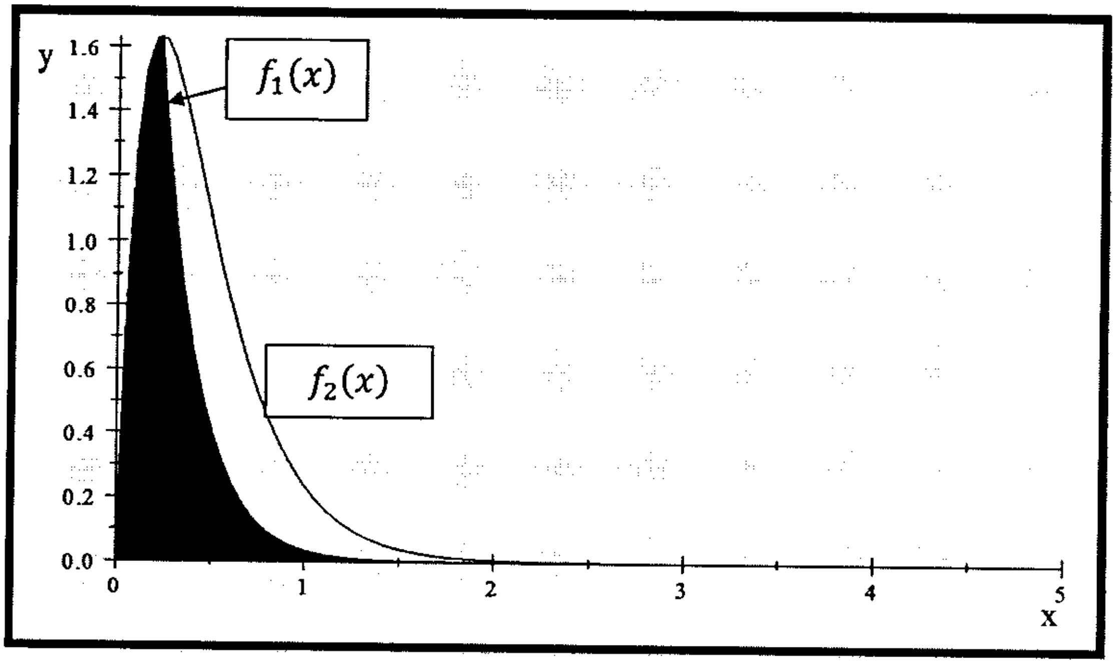
Example 2.2: The following graphs display the overlapping between the densities of elements of RSS, for a sample of size $m = 5$ taken from the exponential distribution with $\theta = 1$.

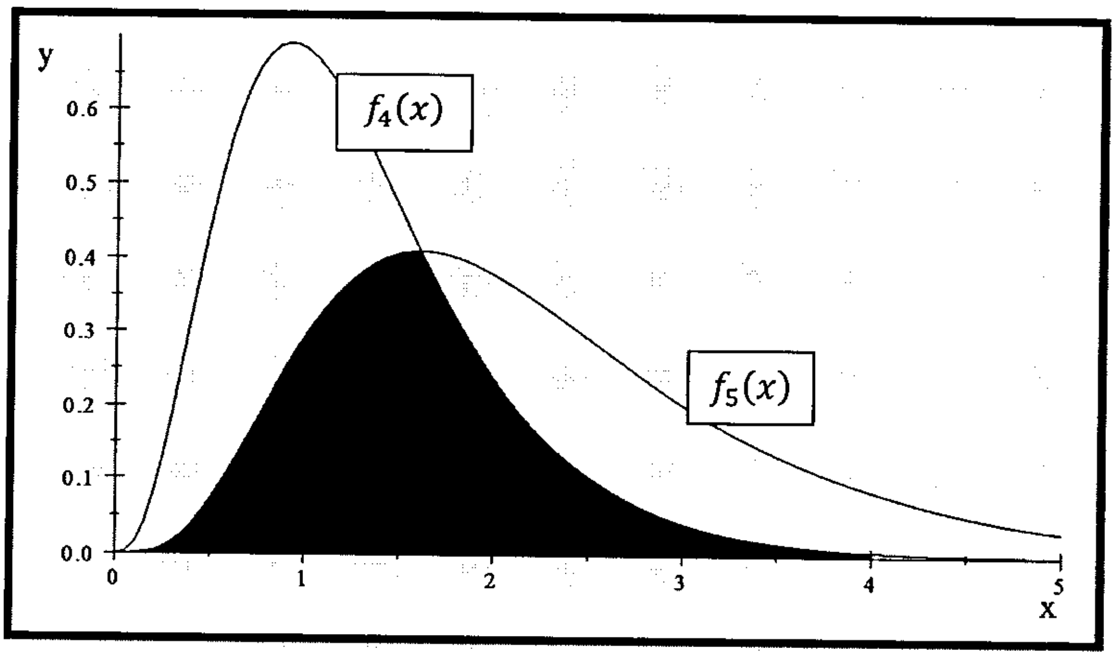





y
0.4
0.3
0.2
0.1
0.0
0
1
2
3
4
5
x
$f_1(x)$
$f_5(x)$


y
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0
1
2
3
4
5
x
$f_2(x)$
$f_4(x)$


y
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0
1
2
3
4
5
x
$f_1(x)$
$f_2(x)$


y
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
1
2
3
4
5
x
$f_4(x)$
$f_5(x)$

2.3.2 Overlapping in MERSS

Let $\{Y_{(i:i)}, i = 1,2,3,\dots,m\}$ be a MERSS of size m, where $Y_{(i:i)}$ is the maximum element of a SRS of size i from a population with pdf $f(y)$, and absolutely continuous cdf $F(y)$.

The overlapping coefficient can be obtained directly from the following lemma.

Lemma 2.1. Assume that Y is a continuous random variable with pdf $f(y)$ and cdf $F(y)$. Let $Y_{(1:1)}, Y_{(2:2)}, \dots, Y_{(n:n)}$ be a MERSS from $f(y)$. Then, for $i < j$,

$$\Delta(Y_{(i:i)}, Y_{(j:j)}) = 1 - \left((P_{ij})^i - (P_{ij})^j\right), \quad i,j = 1,2,\dots,m, \quad (2.7)$$

where

$$P_{ij} = \sqrt[j-i]{\frac{i}{j}} .$$

Proof: The pdf of $Y_{(i:i)}$ is

$$f_{(i:i)}(y) = i\left(F(y)\right)^{i-1} f(y), \quad i = 1,2,\dots,m.$$

$$\Delta(Y_{(i:i)}, Y_{(j:j)}) = \int_{-\infty}^{\infty} min(f_{(i:i)}(y), f_{(j:j)}(y))\, dy,$$

$$= 1 - 0.5 \int_{-\infty}^{\infty} \left|f_{(i:i)}(y) - f_{(j:j)}(y)\right| dy,$$

$$= 1 - 0.5 \int_{-\infty}^{\infty} \left|i\left(F(y)\right)^{i-1} f(y) - j\left(F(y)\right)^{j-1} f(y)\right| dy .$$

Let $F(y) = u$, $\frac{du}{dy} = f(y), -\infty < y < \infty,\ 0 < u < 1.$

Thus,

$$\Delta\left(Y_{(i:i)}, Y_{(j:j)}\right) = 1 - 0.5 \int_0^1 \left| iu^{i-1} - ju^{j-1} \right| du.$$

Now, for $i < j$, $\quad iu^{i-1} - ju^{j-1} \geq 0,$

iff $\quad iu^{i-1} \geq ju^{j-1},$

iff $\quad u \leq \sqrt[j-i]{\frac{i}{j}} = P_{ij}$.

Hence,

$$\left| f_{(i:i)}(y) - f_{(j:j)}(y) \right| = \begin{cases} i(u)^{i-1} - j(u)^{j-1}, & u \leq P_{ij} \\ j(u)^{j-1} - i(u)^{i-1}, & u > P_{ij}. \end{cases}$$

Thus,

$$\Delta\left(Y_{(i:i)}, Y_{(j:j)}\right) = 1 - 0.5 \left[\int_0^{P_{ij}} \left(i(u)^{i-1} - j(u)^{j-1}\right) du + \left(\int_{P_{ij}}^1 j(u)^{j-1} - i(u)^{i-1}\right) du\right].$$

$$\therefore \Delta\left(Y_{(i:i)}, Y_{(j:j)}\right) = 1 - \left((P_{ij})^i - (P_{ij})^j\right) \quad \blacksquare$$

Example 2.2: Suppose that $m = 5$, the overlapping coefficient between pairs of elements in MERSS is given in Table (2.3).

Table (2.3): Overlapping coefficient for some pairs of elements of MERSS with $m = 5$.

Pairs $(Y_{(i:i)}, Y_{(j:j)})$	P_{ij}	$\Delta(Y_{(i:i)}, Y_{(j:j)})$
$(Y_{(1:1)}, Y_{(5:5)})$	0.66874	0.46501
$(Y_{(1:1)}, Y_{(2:2)})$	0.50	0.75
$(Y_{(2:2)}, Y_{(3:3)})$	0.66667	0.85185
$(Y_{(3:3)}, Y_{(4:4)})$	0.75	0.89453
$(Y_{(4:4)}, Y_{(5:5)})$	0.80	0.91808

Based on Tables (2.2) and (2.3), it's clear that the overlapping coefficient for any adjacent pair of RSS is less than the corresponding pair of MERSS, which means that, the method of RSS has a better ability to separates the data than SRS and MERSS.

Chapter Three

Coverage Probability of the "Five-Number Summary" Based on Different Sampling Techniques

3.1 Introduction

Order statistics have applications in many areas of statistical inference, *i.e.* nonparametric statistics, estimation theory, etc. In this research we discuss and investigate some properties of the "5-number summary", which consists of 5 statistics, with respect to the three sampling techniques.

In chapter 1, we have denoted the "5- number summary" statistic as follows:

$$\left(X_{(1)}, Q_1, Q_2, Q_3, X_{(n)}\right).$$

We note here that the results of this chapter may not be all new, but for completeness we derive and discuss them.

In section 3.2, we first study and discuss the coverage probability in SRS. The coverage probability in RSS is discussed in section 3.3. The coverage probability in MERSS is discussed in section 3.4. Comparison of the coverage probability among the three sampling techniques is discussed in section 3.5. Application on real data (trees data) is given in section 3.6. Concluding remarks are given in section 3.7.

3.2.1 The Coverage Probability (CP), in the Case of SRS

Assume that our data consist of a simple random sample $X_1, X_2, \ldots, X_n$. The "5-number summary" divides the sample elements into 4 non overlapping regions $[X_{(1)}, Q_1]$, $(Q_1, Q_2]$, $(Q_2, Q_3]$ and $(Q_3, X_{(n)}]$. About 25% of the sample elements belong to each of the 4 divisions. Our interest is in calculating the percentage of the population elements in each of these regions. In other words, if X is any future value from the population, then what are the values of $P(X_{(1)} \leq X \leq Q_1)$, $P(Q_1 < X \leq Q_2)$, $P(Q_2 < X \leq Q_3)$ and $P(Q_3 < X \leq X_{(n)})$. We may be interested also in $P(X > X_{(n)})$, $P(X < X_{(1)})$, $P(X_{(1)} \leq X \leq X_{(n)})$, etc. The closer the CP to the corresponding sample coverage, the better is the "5-number summary", or the sampling technique that generated them.

Lemma 3.1. Let $X_1, X_2, \ldots, X_n$, be a SRS of size n from an absolutely continuous distribution with cdf $F(x)$ and pdf $f(x)$. Let X be a r.v. from $f(x)$, independent of $X_1, X_2, \ldots, X_n$ (*i.e.* X is a future value). Then

$$P(X_{(i)} < X < X_{(j)}) = \frac{j-i}{n+1},$$

where $1 \leq i < j \leq n$.

Proof: The pdf of $X_{(i)}$ is:

$$f_i(x_{(i)}) = n\binom{n-1}{i-1} (F(x_{(i)}))^{i-1} (1 - F(x_{(i)}))^{n-i} f(x_{(i)}).$$

$$P(X \le X_{(i)}) = n\binom{n-1}{i-1} \int_{-\infty}^{\infty} \int_{-\infty}^{x_{(i)}} f(x) f(x_{(i)})\, dx\, dx_{(i)}$$

$$= n\binom{n-1}{i-1} \int_{-\infty}^{\infty} \int_{-\infty}^{x_{(i)}} f(x) (F(x_{(i)}))^{i-1} (1 - F(x_{(i)}))^{n-i} f(x_{(i)})\, dx\, dx_{(i)},$$

(X and $X_{(i)}$ are independent),

Let $F(x) = w$, $\frac{dw}{dx} = f(x)$, $-\infty < x < x_{(i)}$, $0 < w < u$,

$F(x_{(i)}) = u$, $\frac{du}{dx_{(i)}} = f(x_{(i)})$, $-\infty < x_{(i)} < \infty$, $0 < u < 1$.

$$P(X \le X_{(i)}) = n\binom{n-1}{i-1} \int_0^1 \int_0^u u^{i-1} (1-u)^{n-i} dw\, du$$

$$= n\binom{n-1}{i-1} \int_0^1 u^i (1-u)^{n-i} du$$

$$= n\binom{n-1}{i-1} \frac{\Gamma(i+1)\Gamma(n-i+1)}{\Gamma(n+2)}$$

$$= \frac{n(n-1)!(n-i)!i!}{(i-1)!(n-i)!(n+1)!}$$

$$= \frac{i}{n+1}.$$

$$\therefore P(X \le X_{(i)}) = \frac{i}{n+1}. \qquad (3.1)$$

Thus, the CP between any two order statistics $X_{(i)}, X_{(j)}$, where $i < j$, using (3.1) is

$$P(X_{(i)} < X < X_{(j)}) = P(X \le X_{(j)}) - P(X \le X_{(i)})$$

$$= \frac{j}{n+1} - \frac{i}{n+1}$$

$$= \frac{j-i}{n+1}. \; \blacksquare \qquad (3.2)$$

Now, we can obtain directly using (3.2) the CP of any two order statistics. For the "5-number summary", the CP for different pieces are given in Table (3.1). The sample median may or may not be an order statistic, since there is a single middle value only when n is odd.

Note: To guarantee that Q_1 , Q_3 and the median are order statistics of $X_{(1)}, X_{(2)},\dots, X_{(n)}$, we assume that n is odd and also $\frac{n+1}{2}$ is odd, (for example $n = 5, 9, 13, 17, 21, 25, 29, 33,\dots$) *i.e.* $n = 4k+1$ for $k = 1,2,3,\dots$

In this case

$$Q_1 = X_{\left(\frac{n+3}{4}\right)}, \; Q_2 = X_{\left(\frac{n+1}{2}\right)}, \; Q_3 = X_{\left(\frac{3n+1}{4}\right)}.$$

This assumption on n, which is easy to guarantee, greatly simplifies the study of the properties of the "5-number summary". For example, if n=17, then the median is $X_{(9)}$, $Q_1 = X_{(5)}$ and $Q_3 = X_{(13)}$.

It can be seen from Table (3.1), that the CP of each of the 4 regions is $\frac{n-1}{4(n+1)}$, which converges to $\frac{1}{4}$ as $n \to \infty$. A numerical illustration is given in example 3.1.

Table (3.1): The CP for different pieces of the "5-number summary" based on SRS of size n, (n is odd and $\frac{n+1}{2}$ is odd).

Division	CP
$(X_{(1)}, X_{(n)})$	$\frac{n-1}{n+1}$
$(X_{(1)}, Q_1)$	$\frac{n-1}{4(n+1)}$
(Q_1, Q_2)	$\frac{n-1}{4(n+1)}$
(Q_2, Q_3)	$\frac{n-1}{4(n+1)}$
$(Q_3, X_{(n)})$	$\frac{n-1}{4(n+1)}$
$(X_{(1)}, Q_2)$	$\frac{n-1}{2(n+1)}$
(Q_1, Q_3)	$\frac{n-1}{2(n+1)}$
$(-\infty, X_{(1)})$	$\frac{1}{n+1}$
$(X_{(n)}, \infty)$	$\frac{1}{n+1}$

Example 3.1: Coverage probability for $n = 29$.

(1) $P(X_{(1)} < X < X_{(29)}) = \frac{n-1}{n+1} = \frac{28}{30} \cong 0.933,$

(2) $P(X_{(1)} < X < Q_1) = \frac{n-1}{4(n+1)} = \frac{28}{120} \cong 0.233,$

(3) $P(Q_1 < X < Q_2) = \frac{n-1}{4(n+1)} = \frac{28}{120} \cong 0.233,$

(4) $P(Q_2 < X < Q_3) = \frac{n-1}{4(n+1)} = \frac{28}{120} \cong 0.233,$

(5) $P(Q_3 < X < X_{(29)}) = \frac{(n-1}{4(n+1)} = \frac{28}{120} \cong 0.233,$

(6) $P(X < X_{(1)}) = \frac{1}{n+1} = \frac{1}{30} = 0.03,$

(7) $P(X < X_{(29)}) = \frac{1}{n+1} = \frac{1}{30} = 0.03.$

We note that in SRS the CP of the region between the minimum and maximum equal $\frac{n-1}{n+1}$, which goes to 1 as $n \to \infty$.

3.2.2 Expected Range $E(R)$ and Relative Coverage Probability (RCP) in the Case of SRS

We will consider the expected range for any division of the "5-number summary". Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from an absolutely continuous distribution with cdf $F(x)$ and pdf $f(x)$.

Let

$$R_{ij} = X_{(j)} - X_{(i)}.$$

Then

$$E(R_{ij}) = E(X_{(j)}) - E(X_{(i)}),$$

where

$$E(X_{(i)}) = \int_{-\infty}^{\infty} x\, f(x_{(i)})\ dx\,,$$

$$f(x_{(i)}) = n\binom{n-1}{i-1}(F(x))^{i-1}(1-F(x))^{n-i} f(x),$$

Let $u = F(x)$, then $x = F^{-1}(u)$, and $du = f(x)\,dx$. Thus,

$$E(X_{(i)}) = E(F^{-1}(U)), \text{ where } U \text{ is } beta(i\,, n-i+1).$$

The problem is more difficult when we consider RSS and MERSS, therefore, to obtain CP, $E(R_{ij})$, etc we use simulation.

The steps of obtaining the values of CP are as follows :

(1) Simulate independent $Y_i, i = 1, 2, \ldots, n$, where $Y_i \sim$ beta(i , n-i+1).

(2) CP $= \mathrm{P}(X_{(i)} < X < X_{(j)})$; where X is the population value and

$$\text{CP} = E(Y_j) - E(Y_i) = \frac{j-i}{n+1}.$$

(3) We compute the expected range, as

$$E(R_{ij}) = E(X_{(j)}) - E(X_{(i)})$$

$$= E\left(F^{-1}(Y_j)\right) - E(F^{-1}(Y_i)).$$

(4) After that, the $RCP_{ij} = \frac{\textit{The coverage probability}}{\textit{Expected range}} = \frac{CP\ of\ (X_{(i)}, X_{(j)})}{E(R_{ij})}$.

Table (3.2) and Table (3.3) contain the values of the CP, $E(R_{ij})$ and RCP for 4 distributions, U(0,1), N(0,1), Exp(1) and Cauchy(0,1), when $n = 21$ and $n = 45$. For simplicity, we denote the expected range as $E(R)$.

Based on these two tables we can conclude the following:

(1) The CP for $(X_{(1)}, X_{(n)})$, which does not depend on the distribution, increases from 0.91 to 0.96, and CP for $(X_{(1)}, Q_1)$, increases from 0.23 to 0.24. The same can be said about other divisions of the "5-number summary" $\left((Q_1, Q_2), (Q_2, Q_3), (Q_3, X_{(n)})\right)$.

(2) The $E(R)$ of each division also increases. For example for $(X_{(1)}, X_{(n)})$, it increases from 3.78 to 4.42, for N(0,1) and for Exp(1) it increases from 3.60 to 4.38. But the $E(R)$ of the divisions$(Q_1, Q_2)$and(Q_2, Q_3) for the Cauchy distribution tends to decrease.

Also, we note that $E(R)$ of the $(X_{(1)}, X_{(n)})$ for the Cauchy distribution does not exist.

(3) In general, with respect to normal distribution for each division, $(X_{(1)}, X_{(n)})$, $(X_{(1)}, Q_1)$, $(Q_3, X_{(n)})$, it can be seen that the RCP decreases; for example for $(X_{(1)}, X_{(n)})$, it decreases from 0.24 to 0.22. But RCP for the divisions, (Q_1, Q_2) and (Q_2, Q_3), tends to increase. And it is clear that for Cauchy distribution, the RCP tends to increase for the divisions (Q_1, Q_2) and (Q_2, Q_3).

For the exponential distribution, we note that RCP decreases for the divisions, $(X_{(1)}, X_{(n)})$ and $(Q_3, X_{(n)})$, but for the other divisions $(X_{(1)}, Q_1)$, (Q_1, Q_2),(Q_2, Q_3), the RCP tends to increase.

(4) The range of each of the 4 quarters in the population are obtained and given in the following table:

Quarter	U(0,1)	N(0,1)	Exp(1)	Cauchy(0,1)
Lower quarter	0.25	(-)	0.2877	(-)
Second quarter	0.25	0.6745	0.4054	1
Third quarter	0.25	0.6745	0.6932	1
Fourth quarter	0.25	(-)	(-)	(-)

These ranges can be compared with the ranges of the corresponding divisions of the "5-number summary".

- For N(0,1), we can see from the table that for (Q_1, Q_2), the $E(R)$ differs from the range of the population by about 0.04 (bias), for n = 21, and about 0.02 for n = 45. Nearly, the same can be said for (Q_2, Q_3).

- For Exp(1), it is clear that for $(X_{(1)}, Q_1)$, $E(R)$ differs from the range of the population by about 0.01 (bias) for n =21 and n = 45.

- For (Q_1, Q_2) $E(R)$ differs from the range of the population by about 0.02 (bias), for n =21, at about 0.01 for n = 45.

- For (Q_2, Q_3), $E(R)$ differs from the range of the population by about 0.03 (bias), for n = 21 and about 0.02 for n = 45.

- For Cauchy(0,1), for (Q_1, Q_2), $E(R)$ differs for the range of the population by about 0.03 (bias) for n =21, and 0.01 for n = 45. Nearly, the same can be said for (Q_2, Q_3).

- For U(0,1), for each of $(X_{(1)}, Q_1)$, (Q_1, Q_2),(Q_2, Q_3), $(Q_3, X_{(n)})$, the $E(R)$ differs from the range of the population by about 0.02 for n =21 and 0.01 for n = 45.

Table (3.2): The coverage probability, expected range and relative coverage probability for each division of "5-number summary" w. r. t. SRS when n =21.

Divisions	CP	$E(R)$ and (RCP)			
		U(0,1)	**N(0,1)**	**Exp(1)**	**Cauchy(0,1)**
$(X_{(1)}, X_{(n)})$	0.90909	0.90957 (0.9994)	3.7787 (0.2405)	3.59853 (0.2526)	Does not exist (-)
$(X_{(1)}, Q_1)$	0.22727	0.22732 (0.9997)	1.2611 (0.1802)	0.27861 (0.8157)	Does not exist (-)
(Q_1, Q_2)	0.22727	0.22776 (0.9978)	0.62944 (0.3610)	0.39192 (0.5798)	1.02774 (0.2211)
(Q_2, Q_3)	0.22727	0.22761 (0.9985)	0.62452 (0.3639)	0.66427 (0.3421)	1.0307 (0.2205)
$(Q_3, X_{(n)})$	0.22727	0.22688 (1.0017)	1.2636 (0.1798)	2.2853 (0.0994)	Does not exist (-)

Table (3.3): The coverage probability, expected range and relative coverage probability for each division of "5-number summary" w. r. t. SRS when n =45.

Divisions	CP	$E(R)$ *and* (RCP)			
		U(0,1)	**N(0,1)**	**Exp(1)**	**Cauchy(0,1)**
$(X_{(1)}, X_{(n)})$	0.95652	0.95656 (0.9999)	4.4178 (0.2165)	4.37559 (0.2186)	Does not exist (-)
$(X_{(1)}, Q_1)$	0.23913	0.23867 (1.0019)	1.55907 (0.1533)	0.28318 (0.8444)	Does not exist (-)
(Q_1, Q_2)	0.23913	0.23897 (1.0006)	0.65715 (0.3638)	0.39723 (0.6019)	1.01198 (0.2360)
(Q_2, Q_3)	0.23913	0.23958 (0.9981)	0.65206 (0.3667)	0.66958 (0.35713)	1.01302 (0.2360)
$(Q_3, X_{(n)})$	0.23913	0.23934 (0.9991)	1.54952 (0.1543)	3.0256 (0.0790)	Does not exist (-)

3.3 The Coverage Probability, Expected Range, and Relative Coverage Probability for RSS

In the previous section, we discussed some properties of the "5-number summary" based on SRS. In this section, we will discuss the same properties of the "5-number summary" but based on RSS.

Let $\{X_1, X_2, \dots, X_n\}$ denote a SRS of size n from a continuous population with cdf $F(x)$ and pdf $f(x)$. To obtain a RSS of size n from the same population, "RSS involves selecting one unit among every SRS consisting of m units for quantification. One may select the unit with rank 1 from the first set , and rank 2 from the second set, and so on. The first cycle is completed when the unit with rank m is selected from the m^{th} set. Each cycle involves m^2 units and among which only m units are selected for quantification. This cycle can be repeated for a certain number of times" [**Zhu and Wang (2005)**].

Let $X^1_{(i:m)}$ be the i^{th} minimum of the i^{th} sample in the first cycle $i = 1,2, \dots, m$. Then the RSS is $\{X^1_{(1:m)}, X^1_{(2:m)}, \dots, X^1_{(m:m)}\}$.

For r cycles, let $X^k_{(i:m)}$ be the i^{th} order statistics of the i^{th} sample in k^{th} cycle, $k = 1, 2, \dots, r$.

We have derived in the previous section a formula to find the CP of any two order statistics based on SRS from any continous distribution as

$$P(X_{(i)} < X < X_{(j)}) = \frac{j-i}{n+1},$$

where $1 \le i < j \le n$, and X is the population value.

Now, we will try to find a method for calculating the CP for different divisions, of the "5-number summary" based on RSS.

Lets first consider a 1- cycle RSS. The ordered statistics of RSS was considered by **[Balakrishnan and Li. (2008)]** :

Let $\{X_{(1:m)}^{ORSS} \le X_{(2:m)}^{ORSS} \le \cdots \le X_{(m:m)}^{ORSS}\}$ denote the ordered RSS (ORSS) obtained by arranging $X_{(i:m)}$ in increasing order of magnitude.

The density function of $X_{(i:m)}^{ORSS}$, $(1 \le i \le m)$, is given by

$$f_{(i:m)}^{ORSS}(x) = \frac{1}{(i-1)!\,(m-i)!} \sum_{p[m]} \{ \prod_{k=1}^{i-1} [F_{(j_k:m)}(x) f_{(j_i:m)}(x) \prod_{k=i+1}^{m} [1 - F_{(j_k:m)}(x)]\},$$

where *the* $\sum_{P[m]}$ denotes the summation over all m! permutations $(j_1, j_2, \dots, j_m)$ of (1,2,…, m).

Also, the density function of $X_{(i:m)}^{ORSS}$, can be written as

$$f_{(i:m)}^{ORSS}(x) = \sum_{p[m]} \sum_{c_1=j_1}^{m} \dots \sum_{c_{i-1}=j_{i-1}}^{m} \sum_{c_{i+1}=0}^{c_{i+1}-1} \dots \sum_{c_m=0}^{j_m-1} D_{j.c}^{*}(i) f_{\tilde{i}:m^2}(x), \tag{3.3}$$

where

$$D_{j.c}^{*}(i) = \frac{1}{(i-1)!\,(m-i)!} \left[\prod_{\substack{k=1 \\ k \ne i}}^{m} \binom{m}{c_k} \right] \left[j_i \binom{m}{j_i} \right] \frac{(\tilde{i}-1)!\,(m^2-\tilde{i})!}{(m^2)!},$$

$$i^{\sim} = j_i + \sum_{\substack{k=1 \\ k \neq i}}^{m} c_k.$$

In principle, we can use (3.3) to find P(X < $X_{(i:m)}^{ORSS}$). However, it is very difficult for large m.

Therefore, we will use simulation to obtain the CP. We noted from the previous section that the CP does not depend on the parent distribution $F(x)$. Therefore, without loss of generality (WLOG), we can assume that our parent population is U(0,1). Therefore, the i^{th} order statistics for a SRS of size m is beta(i, m-i+1).

Let m be the set size, r is the number of cycle, $n = rm$, and let $\{X_{(i:m)}^k : i = 1,2, \dots, m, k = 1,2, \dots, r\}$ be the RSS. For fixed i, $X_{(i:m)}^k$, $i = 1,2, \dots, m$ are independent and identically distributed beta(i, m-i +1).

Now, let $Y_{(i:n)}$ be the i^{th} order statistics of RSS of size $n = rm$, for $i < j$. The CP of $(Y_{(i:n)}, Y_{(j:n)})$ for $i < j$ is

$$P(Y_{(i:n)} < X < Y_{(j:n)}) \, ; X \sim U(0,1)$$

$$= E\big(I(Y_{(i:n)} < X < Y_{(j:n)})\big); I(a,b) \text{ is the indicator formula.}$$

Now, using the fact that $E\big(E(X|Y)\big) = E(X)$, we have

$$E\big(I(Y_{(i:n)} < X < Y_{(j:n)})\big) = E\Big(E\big(I(Y_{(i:n)} < X < Y_{(j:n)}) \,|\, y_{(i:n)}, y_{(j:n)}\big)\Big),$$

$$= E\big(P(Y_{(i:n)} < X < Y_{(j:n)}) \,|\, y_{(i:n)}, y_{(j:n)}\big)$$

$$= E\big(Y_{(j:n)} - Y_{(i:n)}\big), \text{ (since } X \text{ is } U(0,1))$$

$$= E\big(Y_{(j:n)}\big) - E\big(Y_{(i:n)}\big). \qquad (3.4)$$

To obtain the $E(R_{ij})$, assume that parent distribution $F(x)$, then

$$E(R_{ij}) = E(Y_{(j:n)}) - E(Y_{(i:n)}).$$

Note: If X is $beta(i\,,j)$, F is a continuous distribution function and $Y = F^{-1}(X)$, then it can be shown

$$f_Y(y) = c\,(F(y))^{i-1}(1 - F(y))^{j-1}, \text{ (c is a normalizing constant).}$$

Thus, if X is an order statistic for a sample from $U(0,1)$, then $Y = F^{-1}(X)$ is the corresponding order statistic from F. This fact can be used to obtain $E(R_{ij})$ by simulation.

To simulate a RSS with m set size, and r cycles we follow the following steps :

(1) Simulate a random sample of size r from $beta(1, m)$.

(2) Simulate a random sample of size r from $beta(2, m-1)$. etc.

(3) Continue in this way, till obtaining a random sample of size r from $beta(m, 1)$.

(4) Denote the sample obtained by $\{X^k_{(1:m)}, X^k_{(2:m)}, \ldots, X^k_{(m:m)}, k = 1, 2, \ldots, r\}$. Denote the elements of RSS from $U(0,1)$, (after ordered), by $\{X^*_{(1:n)}, X^*_{(2:n)}, \ldots, X^*_{(n:n)}\}$ where $X^*_{(i:n)}$ is the i^{th} order statistics of the RSS of size n.

(5) The RSS corresponding to the ordered RSS from F is $F^{-1}(X^*_{(1:n)})$, $F^{-1}(X^*_{(2:n)}), \ldots, F^{-1}(X^*_{(n:n)})$; denote these by $Y_{(1:n)}, Y_{(2:n)}, \ldots, Y_{(n:n)}$.

(6) To obtain the CP of $(Y_{(i:n)}, Y_{(j:n)})$, $i < j$, we obtain

$$E(X^*_{(j:n)}) - E(X^*_{(i:n)}).$$

(7) To obtain the $E(R_{ij})$ of $(Y_{(i:n)}, Y_{(j:n)})$, $i < j$, we calculate

$$E(Y_{(j:n)}) - E(Y_{(i:n)}) = E\left(F^{-1}(X^*_{(j:n)})\right) - E\left(F^{-1}(X^*_{(i:n)})\right).$$

(8) After that, the $RCP_{ij} = \frac{\text{The coverage probability}}{\text{Expected range}} = \frac{CP\ of\ (Y_{(i:n)}, Y_{(j:n)})}{E(R_{ij})}$.

Table (3.4) and Table (3.5) contain the values of the CP, $E(R_{ij})$ and RCP using RSS for the distributions: U(0,1), N(0,1), Exp(1) and Cauchy(0,1), when $n = 21$ with set size $m = 3,7$, respectively.

Based on these tables we can conclude the following (effect of set size):

(1) The CP for $(X_{(1)}, X_{(n)})$, which does not depend on the distribution, increases from 0.91 to 0.92. The same can be said about other divisions of the "5-number summary" $(Q_1, Q_2), (Q_2, Q_3), (Q_3, X_{(n)})$.

(2) The $E(R)$, based on normal distribution for each division, also increases. For exponential distribution, we note that the $E(R)$ increases for the divisions $(X_{(1)}, X_{(n)})$, $(Q_3, X_{(n)})$, but for the other divisions tends to decrease. Also, the $E(R)$ for Cauchy distribution decreases for divisions (Q_1, Q_2) and (Q_2, Q_3).

(3) The RCP for the normal distribution, decreases for divisions $(X_{(1)}, X_{(n)})$, $(X_{(1)}, Q_1)$, $(Q_3, X_{(n)})$, and the RCP for the divisions (Q_1, Q_2) and (Q_2, Q_3) are almost the same.

For the exponential distribution, we note that, RCP for the divisions $(X_{(1)}, X_{(n)}), (Q_3, X_{(n)})$ tends to decrease, and for divisions $(X_{(1)}, Q_1), (Q_1, Q_2), (Q_2, Q_3)$, RCP tends to increase.

For the Cauchy distribution, the RCP increases for the divisions (Q_1, Q_2) and (Q_2, Q_3).

(4) For N(0,1), it can be seen that for $(Q_1, Q_2), (Q_2, Q_3)$, the $E(R)$ differs from the range of the population by about 0.03 (bias) for $m = 3, 7$.

For Exp(1), for the divisions $(X_{(1)}, Q_1)$ and (Q_1, Q_2), the $E(R)$ differs from the range of the population by about 0.02 (bias) for $m = 3, 7$, and for (Q_2, Q_3) it differ is about 0.3 for $m = 3, 7$.

For Cauchy(0,1), we note that, for (Q_1, Q_2), the $E(R)$ differs from the range of the population by about 0.02 (bias), for $m = 3, 7$. For (Q_2, Q_3), the difference is 0.01 when $m = 3$, and 0.02 when $m = 7$.

For U(0,1), the $E(R)$ for $(X_{(1)}, Q_1), (Q_3, X_{(n)})$ differs from the range of the population by about 0.03, 0.02 (bias), for $m = 3$ and 7, respectively and for $(Q_1, Q_2), (Q_2, Q_3)$, $E(R)$ differs from the range of the population by about 0.02 (bias) for $m = 3, 7$.

Tables (3.6 – 3.9) contain the values of the CP, $E(R_{ij})$ and RCP for the 4 distributions, when $n = 45$, with set size $m = 3,5,9$ and 15, respectively. Based on these tables, we can conclude the following :

(1) The CP for each division tends to increase.

(2) The $E(R)$ based on normal distribution for each division, $(X_{(1)}, X_{(n)})$, $(X_{(1)}, Q_1)$, $(Q_3, X_{(n)})$ increases, and are almost the same for the divisions (Q_1, Q_2),(Q_2, Q_3).

For the exponential distribution, we note that $E(R)$ increases for the divisions $(X_{(1)}, X_{(n)})$,$(Q_3, X_{(n)})$, are almost the same for $(X_{(1)}, Q_1)$, and decreases for (Q_1, Q_2),(Q_2, Q_3). Also, it can be seen that, $E(R)$ for Cauchy distribution tends to decrease for divisions (Q_1, Q_2),(Q_2, Q_3).

(3) The RCP for the normal distribution decreases for divisions $(X_{(1)}, X_{(n)})$, $(X_{(1)}, Q_1)$, $(Q_3, X_{(n)})$, and tends to increase for divisions (Q_1, Q_2),(Q_2, Q_3). For exponential distribution, the RCP tends to decrease for divisions $(X_{(1)}, X_{(n)})$, $(Q_3, X_{(n)})$ but for other divisions, tends to increase.

For Cauchy distribution, RCP tends to increase for the divisions (Q_1, Q_2),(Q_2, Q_3).

(4) For N(0,1), it can be seen that, for (Q_1, Q_2), (Q_2, Q_3), $E(R)$ differs from the range of the population by about 0.01 (bias).

For Exp(1), for $(X_{(1)}, Q_1)$, (Q_1, Q_2), $E(R)$ differs from the range of the population by about 0.01 (bias). But for (Q_2, Q_3), the difference is about 0.01 for $m = 3,5$, and about 0.02 for $m = 9,15$.

For Cauchy(0,1), we note that for $(Q_1, Q_2), (Q_2, Q_3)$, $E(R)$ differs from the range of the population by about 0.01.

For U(0,1), for $(X_{(1)}, Q_1), (Q_1, Q_2), (Q_2, Q_3), (Q_3, X_{(n)})$, $E(R)$ differs from the range of the population by about 0.01 (bias), for all set size (m).

Based on Tables (3.4 – 3.9) we can conclude the following :

(1) The CP for $(X_{(1)}, X_{(n)})$, which does not depend on the distribution, increases from 0.92 to 0.96. and CP for (Q_2, Q_3), increases from 0.23 to 0.24. The same can be said about other divisions of the "5-number summary" $\left((X_{(1)}, Q_1), (Q_1, Q_2), (Q_3, X_{(n)})\right)$.

(2) The $E(R)$ for normal distribution for each division, tends to increase, and $E(R)$ for exponential distribution, for each division, tends to increase, except for $(Q_3, X_{(n)})$, it decreases.

For Cauchy distribution, for each division (Q_1, Q_2),(Q_2, Q_3), and (Q_1, Q_3), $E(R)$ tends to decrease.

(3) The RCP for the normal distribution, increases for each of (Q_1, Q_2),(Q_2, Q_3), but tends to decrease for divisions $(X_{(1)}X_{(n)})$, $(Q_3, X_{(n)})$, and $(X_{(1)}, Q_1)$. For exponential distribution, RCP, tend to increase for $(X_{(1)}, Q_1)$, (Q_1, Q_2),(Q_2, Q_3), and decreases for divisions $(X_{(1)}X_{(n)})$, and $(Q_3, X_{(n)})$.

For Cauchy distribution, for each division (Q_1, Q_2),(Q_2, Q_3), RCP tends to increase.

(4) For N(0,1), it can be seen that, for (Q_1, Q_2), (Q_2, Q_3), $E(R)$ differs from the range in the population by about 0.03 (bias), for $n = 21$, and about 0.01 for $n = 45$.

Also, for Exp(1), it can be seen that, for $(X_{(1)}, Q_1)$, (Q_1, Q_2), $E(R)$ differs from the range in the population by about 0.02 (bias) for $n = 21$, and about 0.01 for $n = 45$, but for (Q_2, Q_3), the difference is about 0.03 for $n = 21$, and $n = 45$, the difference is 0.01 ($m = 3,5$), and 0.02 ($m = 9,15$).

For Cauchy(0,1), we note that, for (Q_1, Q_2), (Q_2, Q_3), $E(R)$ differs from the range in the population by about 0.02 (bias), for $n = 21$ ($m = 7$), and about 0.01 for $n = 45$.

With respect U(0,1), nearly, $E(R)$ differs from the range in the population by about 0.02 (bias), for $n = 21(m = 7)$, and about 0.01 for $n = 45$, for all divisions $(X_{(1)}, Q_1)$, (Q_1, Q_2), (Q_2, Q_3), $(Q_3, X_{(n)})$, with different set size (m).

Table (3.4): The coverage probability, expected range and relative coverage probability for each division of "5-number summary" w. r. t. RSS when $n = 21$.

Set Size (m)	Divisions	CP	$E(R)$ and (RCP)			
			U(0,1)	**N(0,1)**	**Exp(1)**	**Cauchy(0,1)**
3	$(X_{(1)}, X_{(n)})$	0.91441	0.91441 (1)	3.8142 (0.2397)	3.6435 (0.2509)	Does not exist (-)
	$(X_{(1)}, Q_1)$	0.22406	0.22406 (1)	1.2739 (0.1758)	0.27262 (0.8218)	Does not exist (-)
	(Q_1, Q_2)	0.23301	0.23301 (1)	0.63779 (0.3653)	0.39109 (0.5957)	1.0234 (0.2276)
	(Q_2, Q_3)	0.23330	0.23330 (1)	0.63938 (0.3648)	0.65817 (0.3544)	1.0184 (0.2290)
	$(Q_3, X_{(n)})$	0.22433	0.22433 (1)	1.2665 (0.1771)	2.3142 (0.0969)	Does not exist (-)

Table (3.5): The coverage probability, expected range and relative coverage probability for each division of "5-number summary" w. r. t. RSS when $n = 21$.

Set Size (m)	**Divisions**	CP	$E(R)$ and (RCP)			
			U(0,1)	**N(0,1)**	**Exp(1)**	**Cauchy(0,1)**
7	$(X_{(1)}, X_{(n)})$	0.92188	0.92188 (1)	3.8760 (0.2378)	3.6933 (0.2496)	Does not exist (-)
	$(X_{(1)}, Q_1)$	0.22576	0.22576 (1)	1.3045 (0.1730)	0.27133 (0.8320)	Does not exist (-)
	(Q_1, Q_2)	0.23460	0.23460 (1)	0.64245 (0.3651)	0.39076 (0.6003)	0.98407 (0.2383)
	(Q_2, Q_3)	0.23431	0.23431 (1)	0.64210 (0.3649)	0.65498 (0.3577)	0.98611 (0.2376)
	$(Q_3, X_{(n)})$	0.22654	0.22654 (1)	1.2957 (0.1748)	3.3944 (0.0667)	Does not exist (-)

Table (3.6): The coverage probability, expected range and relative coverage probability for each division of "5-number summary" w. r. t. RSS when $n = 45$.

Set Size (m)	Divisions	CP	$E(R)$ and (RCP)			
			U(0,1)	N(0,1)	Exp(1)	Cauchy(0,1)
3	$(X_{(1)}, X_{(n)})$	0.95822	0.95822 (1)	4.4308 (0.2162)	4.4006 (0.2177)	Does not exist (-)
	$(X_{(1)}, Q_1)$	0.23690	0.23690 (1)	1.5574 (0.1521)	0.28020 (0.8454)	Does not exist (-)
	(Q_1, Q_2)	0.24212	0.24212 (1)	0.65778 (0.3680)	0.39951 (0.6060)	1.0077 (0.2402)
	(Q_2, Q_3)	0.24188	0.24188 (1)	0.65869 (0.3672)	0.67654 (0.3575)	1.0073 (0.2401)
	$(Q_3, X_{(n)})$	0.23662	0.23662 (1)	1.5608 (0.1516)	3.0338 (0.0779)	Does not exist (-)

Table (3.7): The coverage probability, expected range and relative coverage probability for each division of "5-number summary" w. r. t. RSS when $n = 45$.

Set Size (m)	**Divisions**	CP	$E(R)$ and (RCP)			
			U(0,1)	**N(0,1)**	**Exp(1)**	**Cauchy(0,1)**
5	$(X_{(1)}, X_{(n)})$	0.95944	0.95944 (1)	4.4510 (0.2155)	4.4170 (0.2172)	Does not exist (-)
	$(X_{(1)}, Q_1)$	0.23708	0.23708 (1)	1.5702 (0.1509)	0.27881 (0.8503)	Does not exist (-)
	(Q_1, Q_2)	0.24257	0.24257 (1)	0.65855 (0.3683)	0.39894 (0.6080)	0.99950 (0.2426)
	(Q_2, Q_3)	0.24274	0.24274 (1)	0.66016 (0.3676)	0.67527 (0.3594)	0.99882 (0.2430)
	$(Q_3, X_{(n)})$	0.23667	0.23667 (1)	1.5647 (0.1512)	3.0539 (0.0774)	Does not exist (-)

Table (3.8): The coverage probability, expected range and relative coverage probability for each division of "5-number summary" w. r. t. RSS when $n = 45$.

Set Size (m)	Divisions	CP	$E(R)$ and (RCP)			
			U(0,1)	N(0,1)	Exp(1)	Cauchy(0,1)
9	$(X_{(1)}, X_{(n)})$	0.96132	0.96132 (1)	4.4769 (0.2147)	4.4293 (0.2170)	Does not exist (-)
	$(X_{(1)}, Q_1)$	0.23738	0.23738 (1)	1.5799 (0.1502)	0.27850 (0.8523)	Does not exist (-)
	(Q_1, Q_2)	0.24285	0.24285 (1)	0.65924 (0.3683)	0.39839 (0.6095)	0.99207 (0.2447)
	(Q_2, Q_3)	0.24309	0.24309	0.65795 (0.3694)	0.67400 (0.3606)	0.99157 (0.2451)
	$(Q_3, X_{(n)})$	0.23767	0.23767 (1)	1.5854 (0.1499)	3.0747 (0.0772)	Does not exist (-)

Table (3.9): The coverage probability, expected range and relative coverage probability for each division of "5-number summary" w. r. t. RSS when $n = 45$.

Set Size (m)	Divisions	CP	$E(R)$ and (RCP)			
			U(0,1)	N(0,1)	Exp(1)	Cauchy(0,1)
15	$(X_{(1)}, X_{(n)})$	0.96348	0.96348 (1)	4.4987 (0.2141)	4.4725 (0.2154)	Does not exist (-)
	$(X_{(1)}, Q_1)$	0.23798	0.23798 (1)	1.5916 (0.1495)	0.27914 (0.8525)	Does not exist (-)
	(Q_1, Q_2)	0.24324	0.24324 (1)	0.65783 (0.3697)	0.39818 (0.6108)	0.98367 (0.2472)
	(Q_2, Q_3)	0.24345	0.24345 (1)	0.65780 (0.3700)	0.67452 (0.3609)	0.98590 (0.2469)
	$(Q_3, X_{(n)})$	0.23886	0.23886 (1)	1.5978 (0.1494)	3.1228 (0.0764)	Does not exist (-)

3.4 The Coverage Probability, Expected Range, and Relative

Coverage Probability for MERSS

In this procedure, we choose the i^{th} order statistic from the i^{th} set size i, $i = 1, 2,..., m$. This may be repeated r times to obtain a MERSS of size $n = rm$. The MERSS is $\{X^1_{(1:1)}, X^1_{(2:2)}, \dots, X^1_{(m:m)}\}$, where $X^1_{(i:i)}$ be the maximum of the i^{th} sample in the first cycle $i = 1,2,\dots,m$. For r cycles, let $X^k_{(i:i)}$ be the maximum of the i^{th} sample in k^{th} cycle.

The main objective in this section is to find the CP for any division of the "5-number summary" based on MERSS.

We calculate the CP for any division of 5-number summary, in MERSS by simulation as described in the following steps :

(1) Simulate a random sample of size r from $beta(1,1)$.

(2) Simulate a random sample of size r from $beta(2,1)$, etc.

(3) Continue in this way, till obtaining a random sample of size r from $beta(m,1)$.

(4) Denote the sample obtained by $\{X^k_{(1:1)}, X^k_{(2:2)}, \dots, X^k_{(m:m)}, k = 1,2, \dots, r\}$. This is a MERSS of size $n = rm$ from the $U(0,1)$. Denote the elements of this MERSS (after ordered) by $\{X^{**}_{(1:n)}, X^{**}_{(2:n)}, \dots, X^{**}_{(n:n)}\}$.

The corresponding ordered MERSS from F is

$$\{Y_{(1:n)}, Y_{(2:n)}, \dots, Y_{(n:n)}\} = \{F^{-1}(X^{**}_{(1:n)}), F^{-1}(X^{**}_{(2:n)}), \dots F^{-1}(X^{**}_{(n:n)})\}.$$

(5) To obtain the CP of $(Y_{(i:n)}, Y_{(j:n)})$, $i < j$, we obtain

$$E\left(X^{**}_{(j:n)}\right) - E\left(X^{**}_{(i:n)}\right).$$

(6) To obtain the $E(R)$ of $(Y_{(i:n)}, Y_{(j:n)})$, $i < j$, we calculate

$$E\left(Y_{(j:n)}\right) - E\left(Y_{(i:n)}\right) = E\left(F^{-1}\left(X^{**}_{(j:n)}\right)\right) - E\left(F^{-1}\left(X^{**}_{(i:n)}\right)\right).$$

(7) After that, the $RCP_{ij} = \frac{\text{The coverage probability}}{\text{Expected range}} = \frac{CP \text{ of } (Y_{(i:n)}, Y_{(j:n)})}{E(R_{ij})}$.

Table (3.10) and Table (3.11) contain the values of the CP, $E(R_{ij})$ and RCP for the distributions, U(0,1), N(0,1), Exp(1) and Cauchy(0,1), when $n = 21$, with set size $m = 3,7$, respectively.

Based on these two tables we can conclude the following :

(1) The CP for $(X_{(1)}, Q_1)$, which does not depend on the distribution, increases from 0.36 to 0.45. But, CP for other divisions $(X_{(1)}, X_{(n)})$,(Q_1, Q_2),(Q_2, Q_3), $\left(Q_3, X_{(n)}\right)$ decreases.

(2) The $E(R)$ based on normal distribution for the division $(X_{(1)}, Q_1)$ increases, but for the other divisions it decreases.

(3) For the exponential distribution, it is clear that, $E(R)$ for all divisions increases.

(4) Also, the $E(R)$ for Cauchy distribution, for (Q_1, Q_2), (Q_2, Q_3) increases.

(5) The RCP, for the normal distribution, increases for division $(X_{(1)}, Q_1)$, but for other divisions decreases. For exponential distribution, RCP decreases for all divisions of the "5-number summary". And, for Cauchy distribution for each division $(Q_1, Q_2), (Q_2, Q_3)$, the RCP decreases.

(6) For N(0,1), it can be seen that for (Q_1, Q_2) the $E(R)$ differs from the range of the population by about 0.09 (bias) for $m = 3$, and about 0.12 for $m = 7$. For (Q_2, Q_3), the difference is about 0.12 for $m = 3$, and about 0.17 for $m = 7$.

(7) For Exp(1), it can be seen that for $(X_{(1)}, Q_1)$ the $E(R)$ differs from the range of the population by about 0.24 (bias) for $m = 3$, and about 0.51 for $m = 7$. For (Q_1, Q_2),it is about 0.14 and 0.27 for $m = 3, 7$, respectively. Also, for (Q_2, Q_3), it is about 0.07 and 0.14 for $m = 3, 7$, respectively.

(8) For Cauchy(0,1), we note that for (Q_1, Q_2), the $E(R)$ differs from the range of the population by about 0.16 (bias), for $m = 3$ or $m = 7$, but for (Q_2, Q_3), it is about 0.54 and 1.75 for $m = 3, 7$, respectively.

(9) For U(0,1), for all divisions, the $E(R)$ differs for the range of the population, increases for (Q_1, Q_2), 0.03 when $m = 3$, and at about 0.07 for $m = 7$. Similar things can be said about other divisions.

Tables (3.12 – 3.15) contain the values of the CP, $E(R_{ij})$ and RCP, for the distributions, U(0,1), N(0,1), Exp(1) and Cauchy(0,1), when $n = 45$, with set size $m = 3,5,9$ and 15, respectively. Similar conclusions can be seen from these tables.

Table (3.10): The coverage probability, expected range and relative coverage probability for each division of "5-number summary" w. r. t. MERSS when $n = 21$.

Set Size (m)	Divisions	CP	$E(R)$ and (RCP) U(0,1)	N(0,1)	Exp(1)	Cauchy(0,1)
3	$(X_{(1)}, X_{(n)})$	0.86898	0.86898 (1)	3.5997 (0.2414)	4.2077 (0.2065)	Does not exist (-)
	$(X_{(1)}, Q_1)$	0.36066	0.36066 (1)	1.3346 (0.2702)	0.53113 (0.6790)	Does not exist (-)
	(Q_1, Q_2)	0.21891	0.21891 (1)	0.58361 (0.3750)	0.54615 (0.4008)	0.84461 (0.2591)
	(Q_2, Q_3)	0.16051	0.16051 (1)	0.55817 (0.2875)	0.75958 (0.2113)	1.5439 (0.1039)
	$(Q_3, X_{(n)})$	0.12937	0.12937 (1)	1.1202 (0.1154)	2.3728 (0.0545)	Does not exist (-)

Table (3.11): The coverage probability, expected range and relative coverage probability for each division of "5-number summary" w. r. t. MERSS when $n = 21$.

Set Size (m)	Divisions	CP	$E(R)$ and (RCP)			
			U(0,1)	N(0,1)	Exp(1)	Cauchy(0,1)
7	$(X_{(1)}, X_{(n)})$	0.80540	0.80540 (1)	3.4859 (0.2310)	4.7652 (0.1690)	Does not exist (-)
	$(X_{(1)}, Q_1)$	0.44972	0.44972 (1)	1.4069 (0.3196)	0.82057 (0.5480)	Does not exist (-)
	(Q_1, Q_2)	0.17771	0.17771 (1)	0.55157 (0.3221)	0.68263 (0.2603)	1.1611 (0.1530)
	(Q_2, Q_3)	0.10360	0.10360 (1)	0.50368 (0.2056)	0.83426 (0.1241)	2.7517 (0.0376)
	$(Q_3, X_{(n)})$	0.07242	0.07242 (1)	1.0264 (0.0705)	2.4365 (0.0297)	Does not exist (-)

Table (3.12): The coverage probability, expected range and relative coverage probability for each division of "5-number summary" w. r. t. MERSS when $n = 45$.

Set Size (m)	**Divisions**	CP	$E(R)$ and (RCP)			
			U(0,1)	**N(0,1)**	**Exp(1)**	**Cauchy(0,1)**
3	$(X_{(1)}, X_{(n)})$	0.93249	0.93249 (1)	4.2409 (0.2198)	5.0153 (0.1859)	Does not exist (-)
	$(X_{(1)}, Q_1)$	0.40359	0.40359 (1)	1.6707 (0.2415)	0.56503 (0.7142)	Does not exist (-)
	(Q_1, Q_2)	0.22861	0.22861 (1)	0.59999 (0.3810)	0.56097 (0.4075)	0.84088 (0.2718)
	(Q_2, Q_3)	0.16573	0.16573 (1)	0.57631 (0.2875)	0.78883 (0.2100)	1.5119 (0.1096)
	$(Q_3, X_{(n)})$	0.13331	0.13331 (1)	1.3933 (0.0956)	3.1115 (0.0428)	Does not exist (-)

Table (3.13): The coverage probability, expected range and relative coverage probability for each division of "5-number summary" w. r. t. MERSS when $n = 45$.

Set Size (m)	Divisions	CP	$E(R)$ and (RCP)			
			U(0,1)	**N(0,1)**	**Exp(1)**	**Cauchy(0,1)**
5	$(X_{(1)}, X_{(n)})$	0.90689	0.90689 (1)	4.1429 (0.2189)	5.3847 (0.1684)	Does not exist (-)
	$(X_{(1)}, Q_1)$	0.48946	0.48946 (1)	1.7357 (0.2819)	0.77526 (0.6313)	Does not exist (-)
	(Q_1, Q_2)	0.19824	0.19824 (1)	0.57064 (0.3473)	0.63992 (0.3097)	0.97153 (0.2040)
	(Q_2, Q_3)	0.12540	0.12540 (1)	0.53465 (0.2345)	0.83490 (0.1501)	2.1150 (0.0592)
	$(Q_3, X_{(n)})$	0.09224	0.09224 (1)	1.3100 (0.0704)	3.1365 (0.0294)	Does not exist (-)

Table (3.14): The coverage probability, expected range and relative coverage probability for each division of "5-number summary" w. r. t. MERSS when $n = 45$.

Set Size (m)	Divisions	CP	$E(R)$ and (RCP)			
			U(0,1)	N(0,1)	Exp(1)	Cauchy(0,1)
9	$(X_{(1)}, X_{(n)})$	0.85739	0.85739 (1)	4.0546 (0.2114)	5.8291 (0.1470)	Does not exist (-)
	$(X_{(1)}, Q_1)$	0.56680	0.56680 (1)	1.7978 (0.3152)	1.0778 (0.5258)	Does not exist (-)
	(Q_1, Q_2)	0.15082	0.15082 (1)	0.52546 (0.2870)	0.71933 (0.2096)	1.3637 (0.1105)
	(Q_2, Q_3)	0.08331	0.08331 (1)	0.48830 (0.1706)	0.87208 (0.0955)	3.4304 (0.0242)
	$(Q_3, X_{(n)})$	0.05725	0.05725 (1)	1.2255 (0.0467)	3.1581 (0.0181)	Does not exist (-)

Table (3.15): The coverage probability, expected range and relative coverage probability for each division of "5-number summary" w. r. t. MERSS when $n = 45$.

Set Size (m)	Divisions	CP	$E(R)$ and (RCP)			
			U(0,1)	**N(0,1)**	**Exp(1)**	**Cauchy(0,1)**
15	$(X_{(1)}, X_{(n)})$	0.80150	0.80150 (1)	3.9429 (0.2032)	6.2321 (0.1286)	Does not exist (-)
	$(X_{(1)}, Q_1)$	0.60011	0.60011 (1)	1.8367 (0.3267)	1.3757 (0.4362)	Does not exist (-)
	(Q_1, Q_2)	0.10914	0.10914 (1)	0.48597 (0.2245)	0.77178 (0.1414)	2.0211 (0.0540)
	(Q_2, Q_3)	0.05503	0.05503 (1)	0.45275 (0.1215)	0.89959 (0.0611)	5.3511 (0.0102)
	$(Q_3, X_{(n)})$	0.03647	0.03647 (1)	1.1577 (0.0315	3.1940 (0.0114)	Does not exist (-)

3.5 Comparison of the CP, $E(R)$, RCP, Among the Three Sampling Techniques

3.5.1 Coverage Probability

Based on the tables of coverage probability, for the three sampling techniques, SRS, RSS, MERSS, we can conclude the following :

(1) For the lower quarter, $(X_{(1)}, Q_1)$, the CP w. r. t. SRS, RSS, MERSS, is 0.22727, 0.22576, 0.44972, respectively, for $n = 21$ ($m = 7$), and for $n = 45$ ($m = 15$), is 0.23913, 0.23798, 0.60011, respectively.

i.e. The closest CP to 0.25 is that obtained based on SRS followed by RSS.

(2) For the second quarter, (Q_1, Q_2), the CP w. r. t. SRS, RSS, MERSS is 0.22727, 0.23460, 0.17771, respectively, for $n = 21$ ($m = 7$), and for $n = 45$ ($m = 15$), is 0.23913, 0.24324, 0.10914, respectively.

i.e. The closest CP to 0.25 is that obtained based on RSS, followed by SRS, and last is MERSS.

(3) For the third quarter, (Q_2, Q_3), the CP w. r. t. SRS, RSS, MERSS is 0.22727, 0.23431, 0.10360, respectively, for $n = 21$ ($m = 7$), and for $n = 45$ ($m = 15$), is 0.23913, 0.24345, 0.05503, respectively.

i.e. The closest CP to 0.25 is that obtained based on RSS, followed by SRS, and last is MERSS.

(4) For the fourth (upper) quarter, $(Q_3, X_{(n)})$, the CP w. r. t. SRS, RSS, MERSS is 0.22727, 0.22654, 0.0.07242, respectively, for $n = 21$ ($m = 7$), and for $n = 45$ ($m = 15$), is 0.23913, 0.23886, 0.03647, respectively.

i.e. The closest CP to 0.25 is that obtained based on RSS and SRS are almost the same.

(5) The CP of $(X_{(1)}, X_{(n)})$, w. r. t. SRS, RSS, MERSS is 0.90909, 0.92188, 0.80540, respectively, for $n = 21$ ($m = 7$), and for $n = 45$ ($m = 15$), is 0.95652, 0.96348, 0.80150, respectively.

i.e. The closest CP to 1 is that obtained based on RSS, followed by SRS, and last is MERSS.

Overall, we can say that RSS is the best with respect to CP while the MERSS has a very poor performance.

3.5.2 Expected Range

Based on the tables of expected range, for the three sampling techniques, SRS, RSS, MERSS, we can conclude the following :

(1) For **N(0,1)** :

- For the second quarter, (Q_1, Q_2), $n = 21 (m = 7)$, the $E(R_{6,11})$ w. r. t. SRS, RSS, MERSS, is 0.62944, 0.64245, 0.55157, respectively, and for $n = 45$ ($m = 15$), $E(R_{12,23})$ w. r. t. SRS, RSS, MERSS, is 0.65715, 0.65783, 0.48597, respectively. Note that, the actual range in the population for this quarter is 0.6745 which is closest to $E(R)$ obtained based on RSS, followed by SRS, for $n = 21(m = 7)$, but for $n = 45$ $(m = 15)$, RSS and SRS are almost the same.

- For the third quarter, (Q_2, Q_3), $n = 21(m = 7)$, $E(R_{11,16})$ w. r. t. SRS, RSS, MERSS, is 0.62452, 0.64210, 0.50368, respectively. The closest to 0.6745 is that obtained based on RSS, followed by SRS.

Similar comments can be concluded for $E(R_{23,34})$ when $n = 45$ $(m = 15)$.

(2) For **Exp(1)** :

- For the lower quarter, $(X_{(1)}, Q_1)$, $n = 21$ $(m = 7)$, the $E(R_{1,6})$ w. r. t. SRS, RSS, MERSS, is 0.27861, 0.27133, 0.82057, respectively, and for $n = 45$ $(m = 15)$, w. r. t. SRS, RSS, MERSS, the $E(R_{1,12})$ is 0.28318, 0.27914, 1.3757, respectively.

The closest to the actual range, 0.2877 is that obtained based on SRS, followed by RSS.

- For the second quarter, (Q_1, Q_2), $n = 21(m = 7)$, the $E(R_{6,11})$ w. r. t. SRS, RSS, MERSS, and $E(R_{12,23})$ for $n = 45$ $(m = 15)$. The closest to the actual range, 0.4054, is that obtained based on RSS and SRS (are almost the same).

- For the third quarter, (Q_2, Q_3), $n = 21$ $(m = 7)$, the $E(R_{11,16})$ w. r. t. SRS, RSS, MERSS, is 0.66427, 0.65498, 0.83426, respectively, and $E(R_{23,34})$ for $n = 45$ $(m = 15)$. w. r. t. SRS, RSS, MERSS, is 0.66958, 0.67452, 0.89959, respectively.

The actual range in the population for this quarter is 0.6932, which is closest to $E(R)$, obtained based on SRS, followed by RSS, for $n = 21$ ($m = 7$), but for $n = 45$ ($m = 15$), SRS and RSS are almost the same.

(3) For **Cauchy(0,1)** :

- For the second quarter, (Q_1, Q_2), $n = 21$ ($m = 7$), the $E(R_{6,11})$ w. r. t. SRS, RSS, MERSS, is 1.02774, 0.98407, 1.1611, respectively, and $E(R_{12,23})$ for $n = 45$ ($m = 15$) w. r. t. SRS, RSS, MERSS, is 1.01198, 0.98367, 2.0211, respectively.

We note that, the actual range in the population for this quarter is 1, closest to $E(R)$ obtained based on RSS, followed by SRS, for $n = 21$($m = 7$), but for $n = 45$ ($m = 15$), the closest is that obtained based on SRS.

- For the third quarter, (Q_2, Q_3), $n = 21$($m = 7$), the $E(R_{11,16})$ w. r. t. SRS, RSS, MERSS, is 1.0307, 0.98611, 2.7517, respectively, and $E(R_{23,34})$ for $n = 45$ ($m = 15$). w. r. t. SRS, RSS, MERSS, is 1.01302, 0.98590, 5.3511, respectively.

Note that, the actual range in the population for this quarter is 1, closest to $E(R)$, obtained based on RSS, followed by SRS, for $n = 21$($m = 7$). Almost no difference between RSS and SRS for $n = 45$.

(4) For **U(0,1)** :

The actual range of each of the 4 divisions in the population is 0.25.

- For the lower quarter, $(X_{(1)}, Q_1)$. $E(R)$ obtained based on SRS is the closest, followed by RSS, for $n = 21$ ($m = 7$), but for $n = 45$ ($m = 15$), RSS and SRS are almost the same.

Similar comments can be stated for the other divisions.

3.5.3 The Relative Coverage Probability

The RCP, of each of the 4 quarters in the population of the 4 distributions are given in the following table :

Quarter	U(0,1)	N(0,1)	Exp(1)	Cauchy(0,1)
Lower quarter	1	(-)	0.8689	(-)
Second quarter	1	0.3706	0.6166	0.25
Third quarter	1	0.3706	0.3606	0.25
Fourth quarter	1	(-)	(-)	(-)

These RCP in each population can be compared to the RCP of the corresponding divisions of the "5-number summary".

Based on the tables of RCP, for the three sampling techniques, SRS, RSS, MERSS, we can conclude the following:

(1) For the lower quarter, $(X_{(1)}, Q_1)$, we note that:

- For Exp(1), $n = 21(m = 7)$, $RCP_{1,6}$ w. r. t. SRS, RSS, MERSS, is 0.8157, 0.8320, 0.5480, respectively.

i.e. The closed RCP to 0.8689 is that obtained based on RSS, followed by SRS.

Similar comment can be concluded about $RCP_{1,12}$ for $n = 45$.

(2) For the second quarter, (Q_1, Q_2), it can seen that:

- For N(0,1), $n = 21(m = 7)$, the $RCP_{6,11}$ w. r. t. SRS, RSS, MERSS, is 0.3610, 0.3651, 0.3221, respectively.

- For Exp(1), $n = 21(m = 7)$, the $RCP_{6,11}$ w. r. t. SRS, RSS, MERSS, is 0.5798, 0.6003, 0.2603, respectively.

- For Cauchy(0,1), $n = 21(m = 7)$, the $RCP_{6,11}$ w. r. t. SRS, RSS, MERSS, is 0.2211, 0.2383, 0.1530, respectively.

i.e. For N(0,1), Exp(1), and Cauchy(0,1), the closed RCP to 0.3706, 0.6166, 0.25, respectively, is that obtained based on RSS, followed by SRS. Similar comment, can be said about $RCP_{12,23}$, for $n = 45$.

(3) For the third quarter, (Q_2, Q_3), it can be seen that:

- For N(0,1), $n = 21(m = 7)$, the $RCP_{11,16}$ w. r. t. SRS, RSS, MERSS, is 0.3639, 0.3649, 0.2056, respectively.

i.e. The closed RCP, to 0.3706 is that obtained based on RSS, and SRS.

- For Exp(1), $n = 21(m = 7)$, the $RCP_{11,16}$ w. r. t. SRS, RSS, MERSS, is 0.3421, 0.3577, 0.1241, respectively.

- For Cauchy(0,1), $n = 21(m = 7)$, the $RCP_{11,16}$ w. r. t. SRS, RSS, MERSS, is 0.2205, 0.2376, 0.0376, respectively.

i.e. For Exp(1), and Cauchy(0,1), the closed RCP, to 0.3606, 0.25, respectively, is that obtained based on RSS, followed by SRS.

The same is true for $RCP_{11,16}$, when $n = 45$.

For U(0,1), the closed RCP to 1 is that obtained based on RSS.

Overall, we can say that RSS is the best with respect to RCP.

3.6 Application to Real Data (Trees Data)

The data consists of the diameters and heights of 1083 trees, obtained from Prodan (1968). The numerical and graphical descriptions of the data are given in the following rectangle and figures. The correlation between the two variables is 0.721. For more details about the data, see **Al-Saleh and Al-Hadrami (2003).**

Description the data

(1)

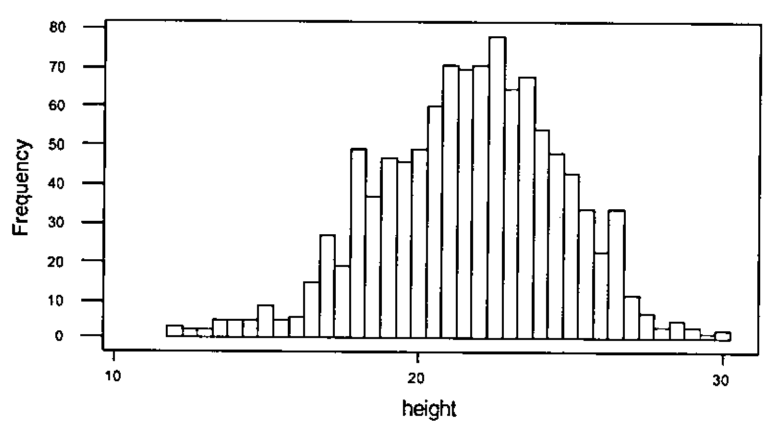
Descriptive Statistics

(c1: Diameter, c2: Height)

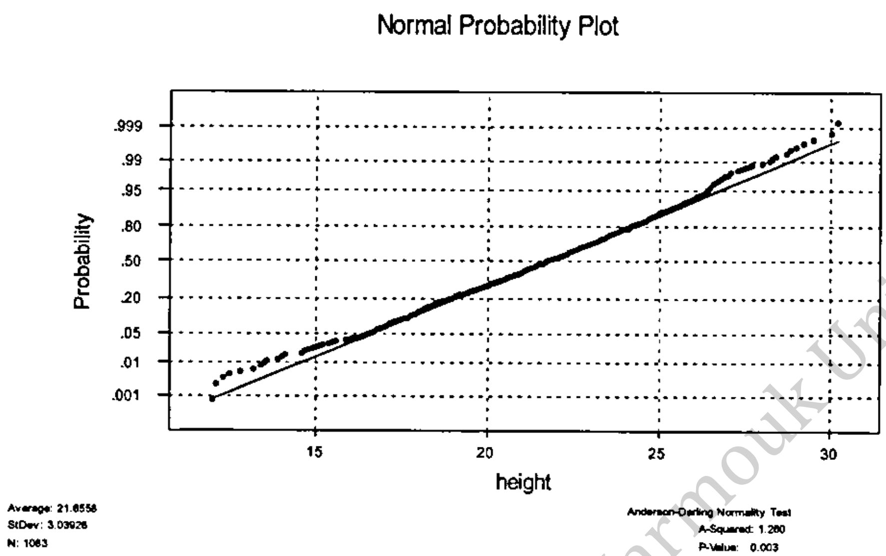
Variable	N	Mean	Median	Tr Mean	StDev	SE Mean
C1	1083	23.070	22.450	22.729	6.268	0.190
C2	1083	21.656	21.800	21.715	3.039	0.092

Variable	Min	Max	Q1	Q3
C1	11.650	57.250	18.350	26.700
C2	12.000	30.200	19.600	23.700

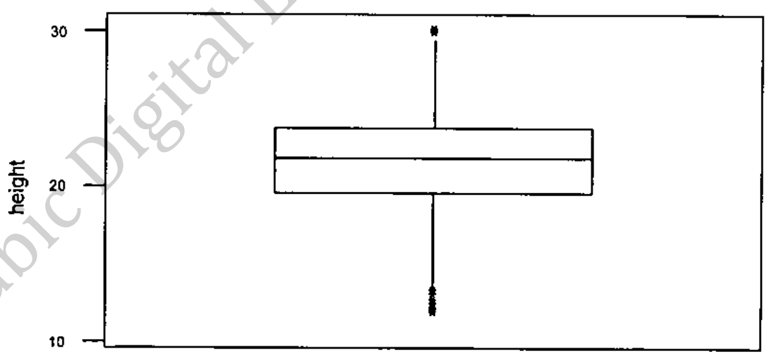
(2) The Histogram of height



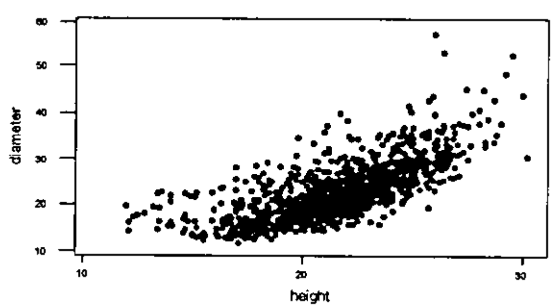
(3)



(4) Box plot



(5) The plot



Tables (3.16 – 3.21) contain the values of CP, when n =21, 45, for the three sampling techniques SRS, RSS, MERSS, for the divisions of "5-number summary" :

♣ Table (3.16) and Table (3.17) give the values of CP based on SRS, when n =21, and n = 45. We note that, CP for $(X_{(1)}, X_{(n)})$, increases from 0.91 to 0.95. Similar comments can be said about other divisions of the 5-number summary.

♣ Table (3.18) and Table (3.19) contain the values of CP based on RSS, when n =21(m = 3, 7), n = 45(m = 3, 5, 9, 15); it can be seen that the CP for each division increases with the number of cycles.

♣ Table (3.20) and Table (3.21) contain the values of CP based on MERSS, when n =21(m = 3, 7), n = 45(m = 3, 5, 9, 15); it can be seen that CP for each division, decreases when set size (m) increase within each sample n =21, and n = 45 .

Based on the Tables (3.16 – 3.21) of CP, for the three sampling techniques, SRS, RSS, MERSS, we can conclude the following:

(1) For the lower quarter, $(X_{(1)}, Q_1)$, the CP w. r. t. SRS, RSS, MERSS, is 0.23161, 0.22955, 0.46163, respectively , for n = 21(m = 7), and CP w. r. t. SRS, RSS, MERSS, for n = 45 (m = 15), is 0.24298, 0.24165, 0.60119, respectively. The closed CP to 0.25 is that obtained based on SRS, followed by RSS.

(2) For the second quarter, (Q_1, Q_2), the CP based on SRS, RSS, MERSS, is 0.22891, 0.23672, 0.16270, respectively, for $n = 21 (m = 7)$, and the CP w. r. t. SRS, RSS, MERSS, for $n = 45$ $(m = 15)$, is 0.24131, 0.24641, 0.10729, respectively. The closed CP to 0.25 is that obtained based on RSS, followed by SRS.

(3) For the third quarter, (Q_2, Q_3),the CP w. r. t. SRS, RSS, MERSS, is 0.22626, 0.23368, 0.09591, respectively, for $n = 21 (m = 7)$, and CP w. r. t. SRS, RSS, MERSS, for $n = 45$ $(m = 15)$, is 0.23823, 0.24200, 0.05458, respectively. The closed CP to 0.25 is that obtained based on RSS, followed by SRS.

(4) For the fourth (upper) quarter, $(Q_3, X_{(n)})$, the CP w. r. t. SRS, RSS, MERSS, is 0.21866, 0.21792, 0.06434, respectively, for $n = 21 (m = 7)$, and the CP w. r. t. SRS, RSS, MERSS, for $n = 45$ $(m = 15)$, is 0.23257, 0.23113, 0.03274, respectively. The closed CP to 0.25 is that obtained based on RSS, and SRS (are almost the same).

(5) The CP for $(X_{(1)}, X_{(n)})$, w. r. t. SRS, RSS, MERSS, is 0.90574, 0.91752, 0.78596, respectively, for $n = 21 (m = 7)$, and CP w. r. t. SRS, RSS, MERSS, for $n = 45$ $(m = 15)$, is 0.95483, 0.96131, 0.79469, respectively.

Overall, we can say that RSS is the best with respect to CP.

Table (3.16): The coverage probability w. r. t. SRS when $n = 21$ for trees data

$(X_{(1)}, X_{(n)})$	$(X_{(1)}, Q_1)$	(Q_1, Q_2)	(Q_2, Q_3)	$(Q_3, X_{(n)})$
0.90574	0.23161	0.22891	0.22626	0.21866

Table (3.17): The coverage probability w. r. t. SRS when $n = 45$ for trees data

$(X_{(1)}, X_{(n)})$	$(X_{(1)}, Q_1)$	(Q_1, Q_2)	(Q_2, Q_3)	$(Q_3, X_{(n)})$
0.95483	0.24298	0.24131	0.23823	0.23257

♣ The coverage probability when $n = 21$. The "5-number summary" are

$$(X_{(1)}, Q_1, Q_2, Q_3, X_{(n)}) = (Y_{(1:21)}, Y_{(6:21)}, Y_{(11:21)}, Y_{(16:21)}, Y_{(21:21)}),$$

respectively.

Table (3.18): The coverage probability w. r. t. RSS when $n = 21$ for trees data

Set Size (m)	$(X_{(1)}, X_{(n)})$	$(X_{(1)}, Q_1)$	(Q_1, Q_2)	(Q_2, Q_3)	$(Q_3, X_{(n)})$
3	0.91109	0.22675	0.23546	0.23176	0.21624
7	0.91752	0.22955	0.23672	0.23368	0.21792

♣ The coverage probability when $n = 45$. The "5-number summary" are

$$\left(X_{(1)}, Q_1, Q_2, Q_3, X_{(n)}\right) = \left(Y_{(1:45)}, Y_{(12:45)}, Y_{(23:45)}, Y_{(34:45)}, Y_{(45:45)}\right),$$

respectively.

Table (3.19): The coverage probability w. r. t. RSS when $n = 45$ for trees data

Set size (m)	$(X_{(1)}, X_{(n)})$	$(X_{(1)}, Q_1)$	(Q_1, Q_2)	(Q_2, Q_3)	$(Q_3, X_{(n)})$
3	0.95565	0.24099	0.24411	0.24072	0.22974
5	0.95657	0.24027	0.24501	0.24194	0.22939
9	0.95828	0.24105	0.24606	0.24191	0.22986
15	0.96131	0.24165	0.24641	0.24200	0.23113

♣ The coverage probability when $n = 21$. The "5-number summary" are

$$\left(X_{(1)}, Q_1, Q_2, Q_3, X_{(n)}\right) = \left(Y_{(1:21)}, Y_{(6:21)}, Y_{(11:21)}, Y_{(16:21)}, Y_{(21:21)}\right),$$

respectively.

Table (3.20): The coverage probability w. r. t. MERSS when $n = 21$ for trees data

Set Size (m)	$(X_{(1)}, X_{(n)})$	$(X_{(1)}, Q_1)$	(Q_1, Q_2)	(Q_2, Q_3)	$(Q_3, X_{(n)})$
3	0.86450	0.36371	0.21720	0.15854	0.12579
7	0.78596	0.46163	0.16270	0.09591	0.06434

♣ The coverage probability when $n = 45$. The "5-number summary" are

$$\left(X_{(1)}, Q_1, Q_2, Q_3, X_{(n)}\right) = \left(Y_{(1:45)}, Y_{(12:45)}, Y_{(23:45)}, Y_{(34:45)}, Y_{(45:45)}\right),$$

respectively.

Table (3.21): The coverage probability w. r. t. MERSS when $n = 45$ for trees data

Set size (m)	$(X_{(1)}, X_{(n)})$	$(X_{(1)}, Q_1)$	(Q_1, Q_2)	(Q_2, Q_3)	$(Q_3, X_{(n)})$
3	0.92907	0.40765	0.22892	0.16387	0.12853
5	0.90176	0.49358	0.19652	0.12326	0.08809
9	0.85358	0.56802	0.14802	0.08209	0.05318
15	0.79469	0.60119	0.10729	0.05458	0.03274

3.7 Concluding Remarks

In this chapter, we investigate some properties of the "5-number summary". We consider the 4 divisions of the "5-number summary" $[X_{(1)}, Q_1]$, $(Q_1, Q_2]$, $(Q_2, Q_3]$ and $(Q_3, X_{(n)}]$. Three properties are investigated and compared with respect to the three sampling techniques SRS, RSS and MERSS. The main results can be summarized as follows:

(1) The coverage probability is compared for each division w. r. t. the three sampling techniques. Overall, RSS gives the closest CP to that of the actual coverage. We noted that CP is independent of the underlying distribution.

(2) The expected range of each division is compared w. r. t. the three sampling techniques for the distributions, U(0,1), N(0,1), Exp(1) and Cauchy(0,1). Note that $E(R)$ is not distribution free. No general conclusion can be obtained in this case.

(3) Taking into account the coverage probability and the range, we use the relative coverage probability to compare each division w. r. t. the three sampling techniques for U(0,1), N(0,1), Exp(1) and Cauchy(0,1). Overall, RSS gives the closest RCP to the actual one. Tree data are used for more illustration.

Chapter Four

Conclusions and Suggestion for Further Research

4.1 General Concluding Remarks

In this thesis, our main concern was to investigate some properties (joint densities, probability of different ordering, overlapping coefficient, etc.) of the elements in each sample for the three sampling techniques SRS, RSS and MERSS. Also, we investigated some properties of the "5-number summary" divisions obtained based on these sampling techniques.

The main results of the thesis can be summarized as follows :

(1) The joint density (likelihood) for the elements of RSS tends to contain more information than likelihood of the elements of MERSS and SRS. Through the probability of different possible ordering of the samples' elements, we note the power of RSS in obtaining a spread out elements.

(2) The overlapping coefficient in the case of SRS is 1. But it is much smaller in the case of RSS and MERSS.

(3) The coverage probability is compared for each division w. r. t. the three sampling techniques. Overall, RSS gives the closest CP to that of the actual coverage. It is noted that CP is independent of the underlying distribution.

(4) The expected range of each division is compared w. r. t. the three sampling techniques for the distributions, U(0,1), N(0,1), Exp(1) and Cauchy(0,1). Note that $E(R)$ is not distribution free. No general conclusion could be obtained in this case.

(5) Taking into account the coverage probability and the range, we use the relative coverage probability to compare each division w. r. t. to the three sampling techniques for the distributions mentioned in (4). Overall, RSS gives the closest RCP to the actual ones.

(6) For more illustration, we use a real data set, (trees data).

4.2 Suggested Future Work

The following are some suggested future works:

(1) The overlapping for the 4 divisions of the "5-number summary" can be considered next. More effort is needed to obtain the density of these 5 order statistics.

(2) Construction of prediction and confidence interval based on the ordered RSS and MERSS.

(3) Investigation of the "5-number Summary" for other variation of RSS, such as DRSS, Multistage RSS, etc.

(4) Investigation of "5-number summary" when judgment ranking is inaccurate.

References

1. Al-Odat, M. and Al-Saleh, M.F. (2001). A variation of ranked set sampling. *J. Appl. Stat. Sci.* 10, 137-146.
2. Al-Omari, A.I. and Al-Saleh, M.F. (2010). Improvement in estimating the population mean using Two-stage balanced groups ranked set sampling. *Metron, LXVIII*, 85-196.
3. Al-Omari, A.I . and Al-Saleh, M.F. (2009). Quartile double rankled set sampling for estimating the population mean. *Economic Quality Control* 24, 243 -253.
4. Al-Saleh, M.F. and Ababneh, A. (2011). Test for perfect ranking in MERSS. Proceeding of ERCIM 2011, University of London, UK.
5. Al-Saleh, M.F. and Samawi, H. (2010). On estimating the odds using moving extreme ranked set sampling. *Statistical Methodology* 7, 133- 140.
6. Al-Saleh, M.F. and Diab, Y.A. (2009). Estimation of the parameters of Downtowns bivariate exponential distribution by using ranked set sampling scheme. *J. Stat. Plan. Inference* 139, 277-286.
7. Al-Saleh, M.F. (2007). On the similarity structure of order statistics. *Commun. Statist. Theor. Meth.* 36, 1433-39.

8. Al-Saleh, M.F. and Endeer, M. J. (2007). Modification of bivariate ranked set sampling using deep stratification. J. *Stat. Theory Appl.* 6, 388-407.

9. Al-Saleh, M.F. and Al-Hadrami, S. (2003). Estimation of the mean of the exponential distribution using moving extreme ranked set sampling. *Stat. Pap.* 44, 367-382.

10. Al-Saleh, M.F. and Abu-Hawwas, J. (2002). Characterization of ranked set sampling Bayes estimation with application to the normal distribution. *Soochow J. Math.* 28, 223-234.

11. Al-Saleh, M.F. and Al-Omari, A.I. (2002). Multistage ranked set sampling. J. *Statist. Plann. Inference* 102, 273-286.

12. Al-Saleh, M.F. and Zheng, G. (2002). Estimation of bivariate characteristics using ranked set sampling. *Austral. New Zealand J. Statist. 44*, 221-232.

13. Al-Saleh, M.F. and Al-Shrafat (2001). Estimation of average milk yield using RSS. *Environmetrics* 12, 395-399.

14. Al-Saleh, M.F. and Al-Kadri, M. (2000). Double ranked set sampling. *Stat. Prob. Lett.* 48, 205-212.

15. Al-Saleh, M.F. and Muttlak, H.A. (2000). Bayesian estimation using ranked set sampling. *Biom.* J. 42, 489-500.

16. Balakrishnan. and Li. (2008). Ordered ranked set samples and applications to inference. *J. Statist. Plann. Inference* 138, 3512-3524

17. Clemons, T.E., and Bradley, Jr. (2000). A nonparametric measure of the overlapping coefficient. *Comput. Statist. Data Anal.* 34. 51–61.

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

19. Hanandeh, A. (2011). Estimation of the parameters of Downton's bivariate exponential distribution using MERSS. Master Thesis. Department of Statistics, Yarmouk University, Jordan.

20. http://en.wikipedia.org/wiki/Five-number_summary. (Date: 28/5/2011)

21. Inman, H.F. and Bradley, E. L. (1989). The overlapping coefficient as a measure of agreement between probability distributions and point estimation of the overlap of two normal densities. *Commun. Statist. Theor. Meth.* 18, 3851–3874.

22. Kaur, A., Patil, G.P., Sinha, A.K. and Taillie, C. (1995). Ranked set sampling: an annotated bibliography. *Environmental and Ecological Statistics* 2, 25-54.

23. Lam, K., Sinha, B.K. and Wu, Z. (1994). Estimation of parameters in two-parameter exponential distribution using ranked set sampling. *Ann. Inst. Stat. Math.* 46, 723-736.

24. McIntyre, G.A. (1952). A method for unbiased selective sampling, using ranked sets. *Aust. J. Agr. Res.* 3, 385-390.

25. Mulekar, M.S. and Mishra, S.N. (2000). Confidence interval estimation of overlap: equal means case. *Comput. Statist. Data Anal.* 34, 121–137.

26. Muttlak, H.A. (1997). Median ranked set sampling with concomitant variables and a comparison with ranked set sampling and regression estimators. *Environmetrics* 9, 255–267.

27. Ozturk, O. and Deshpande. J.V. (2004). Ranked-set sample nonparametric quantile confidence intervals. J. *Statist. Plann. Inference* 136, 570-577.

28. Prodan, M. (1968). Forest Biometrics. Pergamon Press, London.

29. Reiser, B., and Faraggi, D. (1999). Confidence intervals for the overlapping coefficient: the normal equal variance case. *Statistician* 48, 413–418.

30. Richard L. Scheaffer, William Mendenhall and Lyman Ott. (1986). Elementary Survey Sampling, 3rd Ed, *Duxbury*.

31. Samawi, H.M. Ahmad, M.S. and Abu-Dayyeh, W. (1996). Estimating the population mean using extreme ranked sSet sample. *Biom.* J. 5, 577-586.

32. Samuh, M. and Al-Saleh, M.F. (2011). The effectiveness of multistage ranked set sampling in stratifying the population. *Comm. Stat. Theo. Methods* 40, 1063-1080.

33. Sinha, A. K. (2005). On some recent developments in ranked set sampling. *Bull. Inform. Cyberm.* 37, 137-160.

34. Stokes, S.L. and Sager, T.W. (1988). Characterization of a ranked set sample with application to estimating distribution function. *JASA.* 83, 374-381.

35. Stokes, S. L. (1977). Ranked set sampling with concomitant variables C*onnon. Statist. Theor. Meth.* 6, 1207-1211.

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

37. Takahasi, K. and Wakimoto, K. (1968). On unbiased estimates of the population mean based on the sample stratified by means of ordering. *Ann. Inst.* Statist. *Math.* 20, 421-428.

38. Weitzman, M.S. (1970). Measures of overlap of income distributions of white and Negro families in the United States.

Technical Paper No. 22, Washington, D.C., U.S.A: Department of Commerce, Bureau of Census.

39. Wolfe, D.A. (2004). Ranked set sampling: An approach to more efficient data collection. *Statist. Sci.* 19, 636-643.

40. Zheng and Al-Saleh, M.F. (2003). Improving the best linear unbiased estimator for the scale parameter of symmetric distributions using the absolute value of ranked set samples. *J. Appl. Stat. Sci.* 30, 253 - 265

41. Zhu, M. and Wang, Y. (2005). Quantile estimation from ranked set sampling data. *The Indian. J. Stat.* 67, 295-304.