DOKUZ EYLÜL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

GOODNESS-OF-FIT TESTS IN RANKED SET SAMPLING

by
Yusuf Can SEVİL

December, 2017
İZMİR

GOODNESS-OF-FIT TESTS IN 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 Master of
Science in Statistics, Statistics Program

by
Yusuf Can SEVİL

December, 2017
İZMİR

M.Sc. THESIS EXAMINATION RESULT FORM

We have read the thesis entitled "**GOODNESS-OF-FIT TESTS IN RANKED SET SAMPLING**" completed by **YUSUF CAN SEVİL** under supervision of **ASSOC. PROF. DR. TUĞBA YILDIZ** and we certify that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

.....

Assoc. Prof. Dr. Tuğba YILDIZ

Supervisor

.....

Assist. Prof. Dr. Özge Elmastaş Gültekin

Jury Member

.....

Assoc. Prof. Dr. Neslihan Demirel

Jury Member

Prof. Dr. Kadriye ERTEKİN
Director
Graduate School of Natural and Applied Sciences

ACKNOWLEDGEMENTS

Foremost, I would like to express my sincere gratitude to my advisor Assoc. Prof. Dr. Tuğba YILDIZ for the continuous support of my thesis, for her patience, motivation, enthusiasm and immense knowledge. It has been a great honor for me to be a student of her.

Besides my advisor, I would like to thank to all members in Department of Statistics.

Also, I would like to offer my special thanks to my schoolmates, Ayça ÖLMEZ, Aylin GÖÇOĞLU, Sami AKDENİZ and Tolga YAMUT for their everlasting support and motivation.

Finally, I want to thank to Perihan SEVİL, my mother, Hüseyin SEVİL, my father and Ceyhun SEVİL, my brother, for their endless support and patience.

Yusuf Can SEVİL

GOODNESS-OF-FIT TESTS IN RANKED SET SAMPLING

ABSTRACT

In literature, many authors have studied goodness-of-fit (GOF) tests based on ranked set sampling (RSS). In these studies, many different distribution function estimators have been suggested. In this thesis, empirical distribution function (EDF) estimators based on sampling designs which are level-0, level-1 and level-2 in RSS are proposed and GOF tests based on EDF are studied. Also, efficiencies of these EDF estimators are investigated with respect to EDF estimator of simple random sampling (SRS) under perfect and imperfect ranking for finite population. Moreover, powers of different EDF based GOF test statistics for the sampling designs are examined under perfect ranking for finite population. Besides the sampling designs, partially rank-ordered set (PROS) is used in RSS procedure. By using different simulation algorithms, powers, critical values for different GOF tests and efficiencies of the EDF estimators are obtained. Based on these efficiency and power values, in general, it is observed that RSS has higher performance than SRS. These results are presented in tables and illustrated in figures.

Keywords: Ranked set sampling, partially rank-ordered set, sampling designs, goodness-of-fit tests, empirical distribution functions

SIRALI KÜME ÖRNEKLEMESİNDE UYUM İYİLİĞİ TESTLERİ

ÖZ

Literatürde, birçok yazar sıralı küme örneklemesine (SKÖ) dayalı uyum iyiliği testleri çalışılmıştır. Bu çalışmalarda, birçok farklı dağılım fonksiyonu kestiricisi önerilmiştir. Bu tezde, SKÖ'de seviye-0, seviye-1 ve seviye-2 örneklem tasarımlarına dayalı ampirik dağılım fonksiyonu kestiricileri önerildi ve ampirik dağılım fonksiyonuna dayalı uyum iyiliği testleri çalışmıştır. Ayrıca, sonlu kitle için kusursuz ve kusurlu sıralamada bu ampirik dağılım fonksiyonu kestiricilerinin verimlilikleri basit rastgele örneklemenin (BRÖ) ampirik dağılım fonksiyonu kestiricisine göre incelenmiştir. Buna ek olarak, örneklem tasarımları için ampirik dağılım fonksiyonuna dayalı farklı uyum iyiliği test istatistiklerinin güçleri kusursuz sıralama altında sonlu kitle için incelenmiştir. Örnekleme tasarımlarının yanı sıra, SKÖ sürecinde kısmi sıralı küme örneklemesi kullanılmıştır. Farklı simülasyon algoritmaları kullanılarak, uyum iyiliği testleri için güç değerleri, kritik değerler ve ampirik dağılım fonksiyonu kestiricilerinin etkinlik değerleri elde edilmiştir. Bu sonuçlara dayanarak, genel olarak SKÖ'nün BRÖ'den daha iyi performans gösterdiği gözlemlenmiştir. Bu sonuçlar tablo halinde sunulmuş ve şekillerde gösterilmiştir.

Anahtar kelimeler: Sıralı küme örneklemesi, kısmi sıralı küme örneklemesi, örneklem tasarımları, uyum iyiliği testleri, ampirik dağılım fonksiyonları

CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER ONE
INTRODUCTION

In scientific researches, basic statistical principles play vital roles and one of these principles is that ensure experimental data for making valid judgments on the question(s) of interest under investigation. To obtain the experimental data, sampling methods are used in important research across all of the sciences-agricultural, biological, ecological, engineering, medical, physical, and social. The most fundamental of these sampling methods is simple random sampling (SRS). Via SRS, we select only a single random sample of size n, $X_1, \cdots, X_n$, from a fixed population of interest. To make valid statistical inference, we hope that the sample is to be representative of the population characteristic, say mean, median, etc., of interest. However, in practice there is no guarantee that the single random sample is truly representative of the entire population. In this case, sample size is usually increased by researcher. However, if sample size is increased, it may not be appropriate in terms of cost or time. To deal with the problem, McIntyre (1952) introduced ranked set sampling (RSS) as an advantageous alternative to SRS. By using RSS, SRS process is repeated over and over. Also, sample size is reduced and it is caused that time and cost is optimized. On the other hand, population is partitioned into several artificially strata under perfect ranking. Thus, a good sample is provided from population via RSS.

McIntyre (1952) benefited from RSS for seeking to estimate the yield of pasture in Australia, effectively. Because, making precise yield measurements requires harvesting the crops and so it is expensive. McIntyre (1952) described RSS as follows. First, a set of size k is drawn by using SRS from population and the sample observations are ranked by visual inspection. Then, the first smallest observation is identified and taken for full measurement. The other observations are discarded. Next, another set of size k is drawn by using SRS. The second smallest observation is measured and the other observations are discarded. This process is repeated until the kth smallest observation is measured in the kth set. With the kth smallest

observation, a cycle is completed. Then, the cycle repeats l times and a ranked set sample of size $n = kl$ is obtained. McIntyre (1952) showed that mean of the measured sample observations is an unbiased estimator of the population mean regardless of any error in ranking process. After McIntyre (1952), Halls & Dell (1966) published second study on RSS. The authors evaluated the performance of RSS for estimating the weights of browse and of herbage in a pine. They determine efficiency of RSS in the face of SRS. Also, they investigated the effect of ranking errors in practice. Takahasi & Wakimoto (1968) established the first theoretical result about RSS. It is shown that mean of ranked set sample is unbiased estimator and the variance of the estimator is always smaller than the variance of the mean of a simple random sample under perfect ranking. Dell & Clutter (1972) evaluated the effect of ranking errors on RSS. To examine the numerical effect of ranking error, Dell & Clutter (1972) used a model, $Y = X + \epsilon$, where X and ϵ are independent and $\epsilon \sim N(0, \sigma_\epsilon^2)$. In this model, X is the interested variable and is ranked according to Y. Using of concomitant variables is considered by Stokes (1977). In this study, it is assumed that each sampling unit has a bivariate response (X, Y) where X and Y are interested and concomitant variables, respectively. Stokes (1980a) gave an estimator of the population variance for the RSS. Stokes used only balanced allocations, but allowed ranking errors. Stokes (1980b) proposed the estimation of correlation coefficient of a bivariate normal population in the RSS. Neerchal et al. (1998) suggested an RSS estimator of population proportion and they showed that the proportion of the ranked set sample is unbiased and has smaller variance than the variance of the proportion of the simple random sample. See these studies, Kaur et al. (1995), Chen et al. (2003) and Al-Omari & Bouza (2014), in order to obtain some other results.

Having knowledge about the population distribution is needed to apply accurate tests in statistics. To check distributional assumptions, goodness-of-fit (GOF) tests have been used in scientific researches. GOF tests specify the distance between the theoretical distribution function and the empirical distribution function (EDF). Using GOF tests based on the EDF is introduced by Stephens (1974) under the SRS as a

practical guide. Stephens (1974) discussed Kolmogorov-Smirnov and Cramĕr-von Mises tests in this work. For extensive reviews about the GOF tests based on SRS, see D'Agostino (1986).

In literature, the estimation of cumulative distribution function (CDF) with various settings of the RSS has been studied by many authors. Stokes & Sager (1988) suggested an unbiased estimator for the population distribution function based on the EDF of RSS. Under perfect ranking, they considered the performance of Kolmogorov-Smirnov statistic by using the EDF. It is seen that the RSS can result in a substantial decrease in the width of the simultaneous confidence band for the CDF in this study. Shahabuddin et al. (2009) developed modification of Kolmogorov-Smirnov GOF test and investigated the power of several GOF tests by using the normal distribution with different parameters. Mahdizadeh & Arghami (2010) investigated entropy estimation in terms of bias and root mean square error (RMSE) in the RSS. They gave entropy-based GOF test for the inverse Gaussian distribution using RSS. Al-Subh et al. (2009) considered a new way to develop the power of the chi-square test for GOF was proposed based on selective ordered ranked set sample (SORSS). Frey & Wang (2014) suggested alternative GOF tests that are sensitive both to imperfect rankings and to departures from parametric family by using the RSS. Al-Omari & Zamanzade (2016) proposed the GOF tests for Rayleigh distribution using the RSS. Alizadeh Noughabi (2017) studied seven GOF tests for normality and presented the powers of tests under many alternative distributions in the RSS. Sevil & Yildiz (2017) examined the power of Kolmogorov-Smirnov test for standard normal and inverse Gaussian distribution. In the RSS process, they benefited from auxiliary informations, level-2 sampling design and partially rank ordered set (PROS). This study is referred in the thesis, thoroughly. More on GOF tests in the RSS, see Chen et al. (2003) and Al-Omari & Bouza (2014).

On the studies of GOF tests in RSS, it has been proved that RSS is more efficient than its SRS counterpart. In these studies, different RSS versions have been used and different estimators of population distribution have been suggested. In this thesis, we studied GOF tests based on EDF. Unlike these previous studies, we used multiple

auxiliary variables and PROS in ranking process. Also, different sampling designs are used to collect RSS data. Moreover, we purposed EDF estimator for each sampling designs and investigate powers of GOF tests based on these EDF. In Chapter two, RSS procedure is described thoroughly. Then, using multiple auxiliary informations and PROS in ranking process are introduced. Also, the performance of Kolmogorov-Smirnov GOF test based on RSS is compared with Kolmogorov-Smirnov based on SRS for standard normal distribution and inverse Gaussian distribution in this chapter. In chapter three, level-0, level-1 and level-2 sampling designs and the EDF estimators for each of them are described. Efficiencies of these estimators are examined for perfect and imperfect ranking. Chapter four presents the powers of GOF tests for the three sampling designs by using symmetric and asymmetric distributions. Finally, chapter five includes general conclusion and final remarks.

CHAPTER TWO
THE PERFORMANCE OF KOLMOGOROV-SMIRNOV GOODNESS-OF-FIT TEST UNDER RSS AND SRS

In the RSS, the ranking process plays vital role to construct a ranked set sample that represents to the population. Many different mechanisms are used to rank the observations such as expert opinion, visual inspections or auxiliary variable. In many studies on the GOF tests, the interested variable is ranked by using expert opinion or visual inspections. In cases which visual inspections are time consuming or costly, use of auxiliary variable is suggested by Stokes (1977) as an alternative these two ranking process. Husby et al. (2005) benefited from the multiple auxiliary variables to rank corn yields in a data set. In this study, they selected pairs of variables according to different correlations. Ozturk (2014) used the PROS to combine the auxiliary variables in the data set.

In this chapter, we study the EDF estimator based on the level-2 sampling design in RSS and investigate performance of Kolmogorov-Smirnov GOF test under RSS and SRS. In ranking procedure, we used multiple auxiliary informations and PROS. Also, the ranked set sample is constructed by using level-2 sampling design that is introduced with other sampling designs thoroughly in chapter three. Finally, the power of Kolmogorov-Smirnov GOF test based on RSS is compared with the power of Kolmogorov-Smirnov GOF test based on SRS for standard normal distribution and inverse Gaussian distribution.

2.1 Ranked Set Sampling

This popular data collection method has been used in many scientific research areas such as forestry Halls & Dell (1966), medicine Chen et al. (2005) and agriculture Husby et al. (2005). It is shown that the RSS has some advantages against the SRS in these studies. Since the number of sample observations (n) is reduced via the RSS, cost and time are optimized as well. Also, the RSS divide the population into homogeneous

groups of judgment strata under perfect ranking.

In RSS process, $X_{t_h,i}$ is notation for a unit in the ith set, $t_h \in \{1, 2, \cdots, N\}$; $i = 1, \cdots, k$ and $h = 1, \cdots, k$. First, a set of size k, $S_1 = \left\{X_{t_1,1} \quad , \quad X_{t_2,1} \quad , \quad \cdots \quad , \quad X_{t_k,1}\right\}$ is selected without replacement from the population of size N. Then, the units are ranked from smallest to largest by using one of the mechanisms, but not actual measurements. The smallest unit is identified to measure and denoted by $X_{[1]1}$. After the first set, second set of size k, $S_2 = \left\{X_{t_1,2} \quad , \quad X_{t_2,2} \quad , \quad \cdots \quad , \quad X_{t_k,2}\right\}$ is selected and the second smallest unit is measured, $X_{[2]2}$. The process of selecting set is continued until the kth smallest unit, $X_{[k]k}$, is measured in the kth set, $S_k = \left\{X_{t_1,k} \quad , \quad X_{t_2,k} \quad , \quad \cdots \quad , \quad X_{t_k,k}\right\}$. A cycle is completed with this unit. If the process is repeated l independent time, the data, $X_{[1]j}, X_{[2]j}, \cdots, X_{[k]j}$ is collected, for $j = 1, \cdots, l$ and can be shown as the following matrix,

$$\begin{pmatrix} X_{[1]1} & X_{[1]2} & \cdots & X_{[1]l} \\ X_{[2]1} & X_{[2]2} & \cdots & X_{[2]l} \\ \vdots & \vdots & & \vdots \\ X_{[k]1} & X_{[k]2} & \cdots & X_{[k]l} \end{pmatrix}.$$

In this matrix, $X_{[i]j}$ is denoted the ith smallest unit in the jth cycle, $i = 1, \cdots, k$ and $j = 1, \cdots, l$.

Under perfect ranking, it is expected that the observations in the matrix build k strata in the population and number of measured observations equal l in each strata. In this case, the $X_{[i]j}$ has probability density function (PDF) $f_{[i]}(x)$ and the CDF $F_{[i]}(x)$ for a fixed $j \in \{1, 2, \cdots, l\}$. The following (2.1), (2.3) and (2.4) are noted by Chen et al. (2003).

$$f_{[i]}(x) = \frac{k!}{(i-1)!(k-i)!}[F(x)]^{i-1}[1 - F(x)]^{k-i} f(x), \tag{2.1}$$

$$F_{[i]}(x) = P[X_{[i]} \leq x] = \sum_{r=i}^{k} \binom{k}{r} [F(x)]^r [1 - F(x)]^{k-r}, \tag{2.2}$$

$$f(x) = \frac{1}{k}\sum_{i=1}^{k} f_{[i]}(x) \tag{2.3}$$

and

$$F(x) = \frac{1}{k}\sum_{i=1}^{k} F_{[i]}(x) \tag{2.4}$$

for all x. Let's illustrate case of the perfect ranking by an example for a single cycle. We assume that a ranked set sample of size five is collected from a population (X) having standard normal distribution, $X_{[1]}$, $X_{[2]}$, $X_{[3]}$, $X_{[4]}$ and $X_{[5]}$. Densities of these observations, $f_{[i]}(x)$, $i = 1, \cdots, 5$, are shown by the five individual marginal density curves that are mutually independent in Figure 2.1, Wolfe (2012). As seen in the Figure

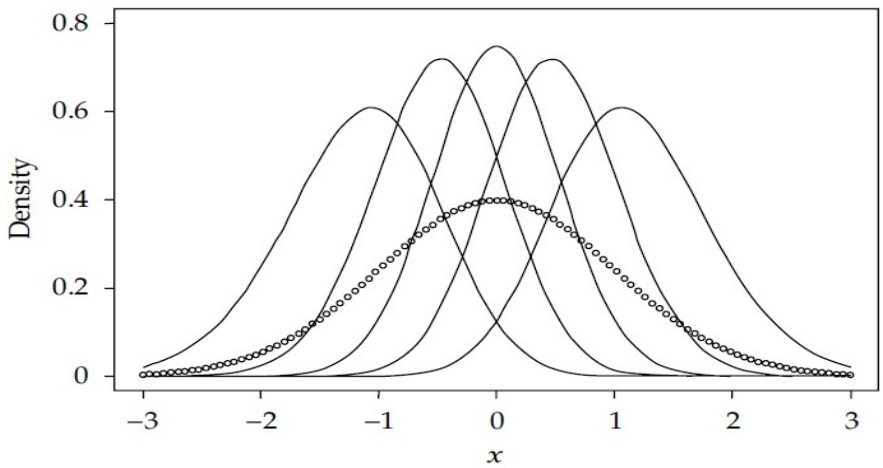

Figure 2.1 The population with standard normal density (dotted curve) and the groups of judgment strata for the five order statistics $X_{[1]}, X_{[2]}, X_{[3]}, X_{[4]}$ and $X_{[5]}$ (solid curves)

2.1, the population is partitioned into five judgment strata. For details about distribution of order statistics, see Arnold et al. (2008).

Under imperfect ranking, the PDF of the ith order statistic is no longer $f_{[i]}(x)$, but the CDF $F_{[i]}(x)$ is expressed in the following form by Chen et al. (2003),

$$F_{[i]}(x) = \sum_{r=1}^{k} p_{ir} F_{(r)}(x), \tag{2.5}$$

where p_{ir} denotes the probability that the unit actually has rank r in the set is selected as ith judgment order statistic.

2.1.1 Auxiliary Variables

As a motivation, we first consider the RSS with a single auxiliary variable. X and Y are denoted, the interested variable and auxiliary variable, respectively. The auxiliary variable is highly correlated, either positively or negatively, to the variable of interest. Also, the interested variable is hard to measure and difficult to rank as well, but an auxiliary variable can be easily measured. So, the auxiliary variable can be used to rank the sampling units. RSS with the single auxiliary variable goes as follows. Each time, a set of size k is selected from the population and the set includes two variables which are X and Y. Then, rank values are assigned to all of the Xs according to order of related Ys. A single auxiliary variable is investigated in many studies and some of them are Dell & Clutter (1972), Stokes (1977), Kaur et al. (1996) and Al-Saleh & Al-Ananbeh (2007).

In practice, there can be two or more auxiliary variables. The each auxiliary variable has relatively high correlation with the interested variable. So, any of the auxiliary variables can be used as a ranking criterion. Chen (2002) and Chen & Shen (2003) developed the ranking mechanisms based on the multiple auxiliary variables. Husby et al. (2005) studied the multiple auxiliary variables under different correlations. Ozturk (2014) suggested combining multiple auxiliary variables instead of single auxiliary variable to obtain ranking information.

Now, let's explain the ranking process with an example. We assume a finite population of size N, $X_1, \cdots, X_N$. The population includes M auxiliary variables, $Y_1, \cdots, Y_M$. They have high correlation with the interested variable X, $\rho_m = corr(X, Y_m)$, $m = 1, \cdots, M$. A set of size k is selected from the population and the values of $Y_{m,i}$, $i = 1, \cdots, k$, corresponding to each $X_{t_h,i}$, $h = 1, \cdots, k$ and $t_h \in \{1, 2, \cdots, N\}$, are measured. Then, an operator that is suggested by Ozturk (2014) is applied to each auxiliary variables and this operator assign rank values ($O_{s,m,i}$) to all of the $X_{t_h,i}$, $s = 1, \cdots, k$,

$$O_{m,i} = O(Y_{t_1,m,i}, \cdots, Y_{t_k,m,i}) = \{O_{1,m,i}, \cdots, O_{k,m,i}\} \tag{2.6}$$

Here, $O_{k,m,i}$ is assigned as a rank value to the $X_{t_m,i}$ by using the operator $O_{m,i}$ and $Y_{m,i} = (Y_{t_1,m,i}, \cdots, Y_{t_k,m,i})$, are values of the auxiliary variables in the ith set, $m = 1, \cdots, M$ and $i = 1, \cdots, k$. This example is given in Table 2.1.

Table 2.1 For the first set, $i = 1$, auxiliary variables $(Y_{1,1}, \cdots, Y_{M,1})$ and their rank values

X	$Y_{1,1}$	$Y_{2,1}$	$\cdots$	$Y_{M,1}$	$O_{1,1}$	$O_{2,1}$	$\cdots$	$O_{M,1}$
$X_{t_1,1}$	$Y_{t_1,1,1}$	$Y_{t_1,2,1}$	$\cdots$	$Y_{t_1,M,1}$	$O_{1,1,1}$	$O_{1,2,1}$	$\cdots$	$O_{1,M,1}$
$X_{t_2,1}$	$Y_{t_2,1,1}$	$Y_{t_2,2,1}$	$\cdots$	$Y_{t_2,M,1}$	$O_{2,1,1}$	$O_{2,2,1}$	$\cdots$	$O_{2,M,1}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$		$\vdots$
$X_{t_k,1}$	$Y_{t_k,1,1}$	$Y_{t_k,2,1}$	$\cdots$	$Y_{t_k,M,1}$	$O_{k,1,1}$	$O_{k,2,1}$	$\cdots$	$O_{k,M,1}$

2.1.2 Partially Rank Ordered Sets

In RSS, rankers aim to rank the all units in the sets accurately even with low confidence. However, in practice, the units in the set are ranked inaccurately if the rankers have low confidence. Also, if there are two or more tied units in selected set, this case makes it difficult to rank the units in the set. This situation reduces the efficiency of RSS. PROS is suggested by Ozturk (2011) against the situation. Nonparametric inference is developed for one and two sample problems in PROS by Ozturk (2012a, 2012b). Ozturk (2014) used PROS in a data including multiple auxiliary variables.

PROS provides some flexibilities to rankers in RSS process. In RSS, increased the set size usually causes ranking error, but PROS allows that the set size is increased due to ranking the units in each set partially. That means, a full ranking of all units is not required in each set. As increasing set size, ranking error is under control the PROS. Also, the multiple auxiliary variables or rankers can be combined by using PROS in RSS.

Ozturk (2014) used PROS to develop estimators of population mean and total for United States Department of Agriculture (USDA) 1992 Ohio corn data. This data includes five variables such as corn yields ($bushels$, X), farm size ($acreage$, Y_1),

group size (Y_2), acre planted (Y_3) and acre harvested (Y_4). The group size (Y_2) is an integer valued random variable and takes values 1, 2, 3.

We now define the PROS process with an example. Suppose that a set of size five is selected, $S_1 = \{X_{76,1}, X_{147,1}, X_{87,1}, X_{119,1}, X_{48,1}\}$ from the Ohio corn data. Informations about the set are in the Table 2.2.

Table 2.2 Auxiliary measurements of randomly selected five farms (the numbers have been changed to make the farms not identifiable)

t_h	Y_1	Y_2	Y_3	Y_4	X	O_1	O_2	O_3	O_4
76	55	1	16	16	x_{76}	$1-2$	$1-4$	2	2
147	1280	3	389	389	x_{147}	5	5	5	5
87	55	1	9	9	x_{87}	$1-2$	$1-4$	1	1
119	135	1	77	77	x_{119}	3	$1-4$	4	4
48	855	1	55	55	x_{48}	4	$1-4$	3	3

The units in the set S_1 are ranked by using the operator $O_{m,i}$, $m = 1, \cdots, 4$ and $i = 1$ and the ranks, $O_{s,m,i}$, $s = 1, \cdots, 5$ are assigned to the units by each of the $Y_{m,i}$. In the auxiliary variables Y_1 and Y_2, there are tied values. Thus, it is seen that dashed integers $a - b$ ($a < b$) including all ranks between a and b are assigned to the set units ($X_{t_h,i}$) for corresponding the tied values in the Table 2.2. That means, the set units are ranked partially by using the auxiliary variables Y_1 and Y_2. For example, the units $X_{t_{76},1}, X_{t_{87},1}, X_{t_{119},1}, X_{t_{48},1}$ has same rank values according to Y_2. In the auxiliary variables Y_3 and Y_4, there is no tied unit so different ranks are assigned to the set units. Also, these auxiliary variables can be combined as a ranking information ($\bar{W}_i$, $i = 1, \cdots, k$) via PROS. Ozturk (2014) suggested two equations to obtain a ranking information.

$$\bar{W}_i = \sum_{m=1}^{M} \alpha_m W_{m,i} \tag{2.7}$$

where $\sum_{m=1}^{M} \alpha_m = 1$. These weights α_m show the quality of auxiliary variables and are calculated by Equation 2.8 if the correlation coefficients ρ_m, $m = 1, \cdots, M$, are known.

$$\alpha_m = \frac{|\rho_m|}{\sum_{m=1}^{M} |\rho_m|} \tag{2.8}$$

If the ranking process is performed by using multi-ranker design, α_m can be specified based on their experiences. In our example, these weights α_m are given in vector α, $\alpha = \{0.252, 0.208, 0.263, 0.276\}$. $W_{m,i}$ is k by k weight matrix that is computed for each judgment rank vector $O_{m,i}$. In this matrix, the rows and columns represent the units and the assigned judgment ranks, respectively. This matrix includes one and zero if mth auxiliary variable has no tied units like $W_{3,1}$ and $W_{4,1}$,

$$W_{3,1} = W_{4,1} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}.$$

If the mth auxiliary variable has t tied values for hth unit, weights $1/t$ are assigned to the unit corresponds to the tied values in hth row and the other weights in this row are zero like $W_{1,1}$ and $W_{2,1}$,

$$W_{1,1} = \begin{bmatrix} 1/2 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1/2 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix},$$

$$W_{2,1} = \begin{bmatrix} 1/4 & 1/4 & 1/4 & 1/4 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1/4 & 1/4 & 1/4 & 1/4 & 0 \\ 1/4 & 1/4 & 1/4 & 1/4 & 0 \\ 1/4 & 1/4 & 1/4 & 1/4 & 0 \end{bmatrix}.$$

By the Equation 2.7, the auxiliary variables are combined as an ranking information

matrix $\bar{W}_1$ to rank the units in the set S_1,

$$\bar{W}_1 = \begin{bmatrix} 0.178 & 0.717 & 0.052 & 0.052 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 & 1.000 \\ 0.717 & 0.178 & 0.052 & 0.052 & 0.000 \\ 0.052 & 0.052 & 0.304 & 0.591 & 0.000 \\ 0.052 & 0.052 & 0.591 & 0.304 & 0.000 \end{bmatrix}.$$

For instance, the ranks 1, 2, 3, 4 and 5 are assigned to the unit X_{76} with weights 17.8%, 71.7%, 5.2%, 5.2% and 0%. It can be said that the unit X_{76} is the second judgment ordered unit according to the largest weight. In this set, X_{87} is selected for full measurement as the first smallest unit with the weight 71.7%.

2.2 Kolmogorov-Smirnov Test Statistic

In RSS, the most basic EDF estimator is suggested by Stokes & Sager (1988). In this study, they showed that the EDF estimator ($F^*(t)$) of RSS is unbiased and has smaller variance than the EDF estimator of SRS. Kolmogorov-Smirnov test statistic based on the EDF of RSS, D^*, is given by Stokes & Sager (1988).

$$D^* = \sup_t \left[F^*(t) - F_0(t)\right] \tag{2.9}$$

where $F_0(t)$ is a specific distribution function. By using the D^*, a confidence band that is narrower than the corresponding band based D calculated from a simple random sample. Many Kolmogorov-Smirnov test statistics based different EDF estimators are investigated in the literature. Some of them are Al-Subh et al. (2009, 2012), Frey & Wang (2014) and Al-Omari & Zamanzade (2016).

If a simple random sample of size n, $X_1, \cdots, X_n$ is selected from a specific population having CDF $F(x)$, and $H_0 : F(x) = F_0(x)$ is tested against $H_1 : F(x) \neq F_0(x)$ by GOF tests based on SRS EDF ($\hat{F}(x)$).

$$\hat{F}(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \leq x) \tag{2.10}$$

By using $\hat{F}(x)$, Kolmogorov-Smirnov test statistic is expressed in the following form.

$$D_K \equiv \sup_x \left|\hat{F}(x) - F_0(x)\right| \tag{2.11}$$

Under RSS, a sample of size kl, $X_{[i]j}$, $i = 1, \cdots, k$ and $j = 1, \cdots, l$, is selected by using level-2 sampling design from a population having CDF $F(x)$. The sampling observations and the EDF that belongs to the level-2 sampling design are denoted by $X^{(2)}_{[i]j}$ and $\hat{F}^*_{L-2}(x)$, respectively. Then, the EDF is given by Equation (2.12).

$$\hat{F}^*_{L-2}(x) = \frac{1}{lk}\sum_{j=1}^{l}\sum_{i=1}^{k} I(X^{(2)}_{[i]j} < x) \tag{2.12}$$

and it is shown that $\hat{F}^*_{L-2}(x)$ is more efficient than $\hat{F}(x)$ in the following chapter. The Kolmogorov-Smirnov test statistic is given in Equation (2.13).

$$D^*_K \equiv \sup_x \left|\hat{F}^*_{L-2}(x) - F_0(x)\right| \tag{2.13}$$

To show the behavior of the test statistics (D_K and D^*_K), the efficiencies of the test statistics are calculated by using the following equations.

$$EF(D_K, D^*_K) = \frac{power\ of\ D^*_K}{power\ of\ D_K} \tag{2.14}$$

If $EF(D_K, D^*_K) > 1$, it can be said that D^*_K is more powerful than D_K.

2.3 Simulation Results

In this section, our aim is to make power comparison for Kolmogorov-Smirnov between RSS and SRS. In RSS, sample data is obtained by using level-2 sampling design. In ranking procedure, we benefited from multiple auxiliary variables and PROS.

Two simulation studies are described to compute critical values and powers for the Kolmogorov-Smirnov GOF test under RSS. Suppose that population size (N) is 300 and the population has four variables which are interested variable (X) and auxiliary variables (Y_m, $m = 1, 2, 3$). For the two simulation studies, we take correlation coefficients $\rho_m = \{1.00, 0.75, 0.50\}$, set sizes $k = \{2, 3, 4\}$ and cycle sizes $l = \{5, 10\}$. The critical values and powers are obtained by simulating 10,000 samples.

The critical values are presented for standard normal and inverse Gaussian distributions. These distributions can be easily generated by using R statistical software with packages MASS and SuppDists. Distributions of interested variables are illustrated in the Figure 2.2. Under these distributions, critical values are

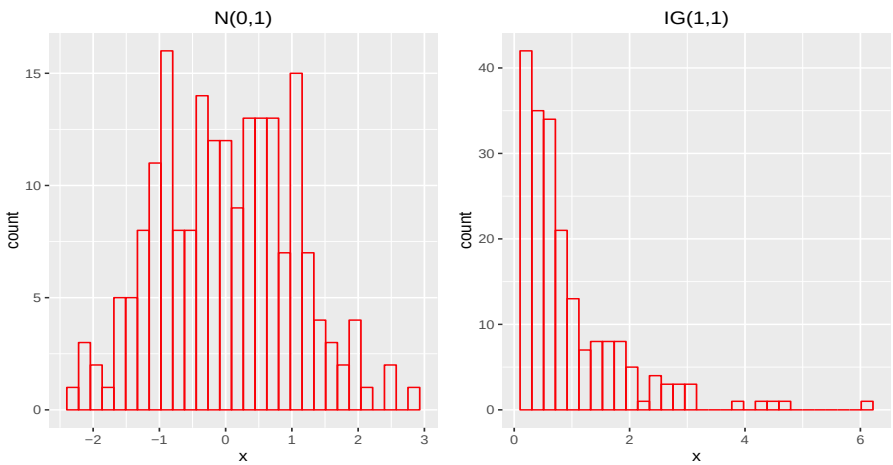

Figure 2.2 The population with standard normal ($N(0, 1)$) and inverse Gaussian distributions ($IG(1, 1)$)

computed for $\alpha = 0.05$ by using the following algorithm,

(1) Select a RSS sample $X_{[i]j}$ from $F_0(x)$, $i = 1, 2, \cdots, k$, $j = 1, 2, \cdots, l$.

(2) Calculate the D_K^* according to Equation 2.13.

(3) This steps (1)-(2) are repeated to get $D_{K,1}^*, \cdots, D_{K,10,000}^*$.

(4) The critical value C_α^* is the $100(1-\alpha)$ percentage point of D_K^*.

In Table 2.3, critical values that are used to simulate powers of D_K and D_K^* are given. For D_K, actual critical values are used. In this table, it is seen that the critical values for D_K^* are lower than the critical values for D_K except for $\rho = 0.50$ and this situation is increased the power of D_K^* against D_K. Even, Stokes & Sager (1988) was drawn attention to it. Also, critical values for D_K^* are obtained by setting cycle sizes $l = 1, \cdots, 10$ and set sizes $k = 2, 3, 4$. These values are presented in Table A.1 and A.2.

Table 2.3 The critical values for RSS (D_K^*) and SRS (D_K) under standard normal and inverse Gaussian

			$N(0,1)$		$IG(1,1)$	
ρ	l	k	RSS	SRS	RSS	SRS
1.00	5	2	0.362	0.410	0.361	0.409
		3	0.276	0.338	0.273	0.341
		4	0.224	0.294	0.222	0.295
	10	2	0.258	0.294	0.255	0.295
		3	0.192	0.240	0.192	0.244
		4	0.151	0.184	0.151	0.212
0.75	5	2	0.379	0.410	0.374	0.409
		3	0.292	0.338	0.297	0.341
		4	0.252	0.294	0.252	0.295
	10	2	0.267	0.294	0.266	0.295
		3	0.210	0.240	0.207	0.244
		4	0.167	0.184	0.171	0.212
0.50	5	2	0.410	0.410	0.407	0.409
		3	0.334	0.338	0.331	0.341
		4	0.290	0.294	0.282	0.295
	10	2	0.297	0.294	0.296	0.295
		3	0.242	0.240	0.240	0.244
		4	0.208	0.184	0.204	0.212

The powers for D_K^* and D_K are obtained by using different alternative distributions. When the null distribution is standard normal, the alternatives are

different parameters in the normal distribution such as $N(0.25, 1)$, $N(0.50, 1)$, $N(0.75, 1)$, $N(1, 1)$, $N(1.25, 1)$, $N(0, 2)$, $N(0, 4)$, $N(0, 8)$, $N(0, 16)$ and $N(0, 32)$. Figure 2.3 illustrates these alternative distributions. When the null distribution is

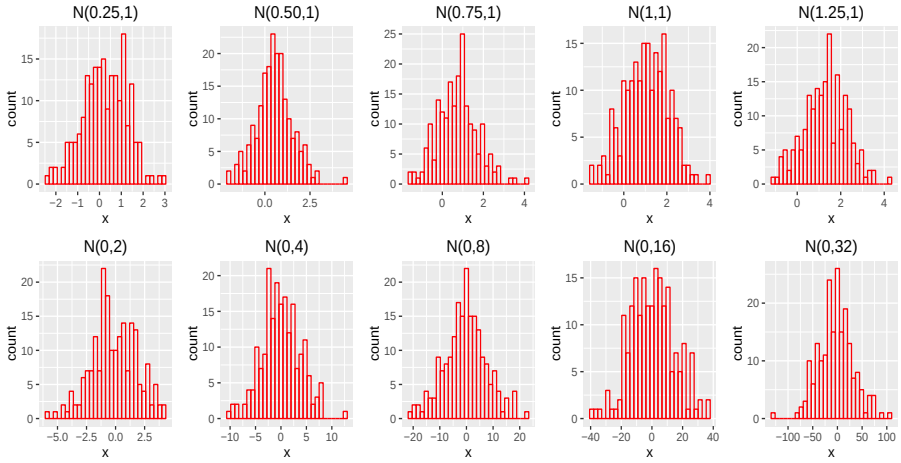

Figure 2.3 Alternative distributions for standard normal

inverse Gaussian, we take the alternative distributions such as exponential(1), chi-square(1), lognormal(0,2), Weibull(2,1) (with shape parameter 2 and scale parameter 1) and beta(5,2). Figure 2.4 illustrates these alternative distributions. By

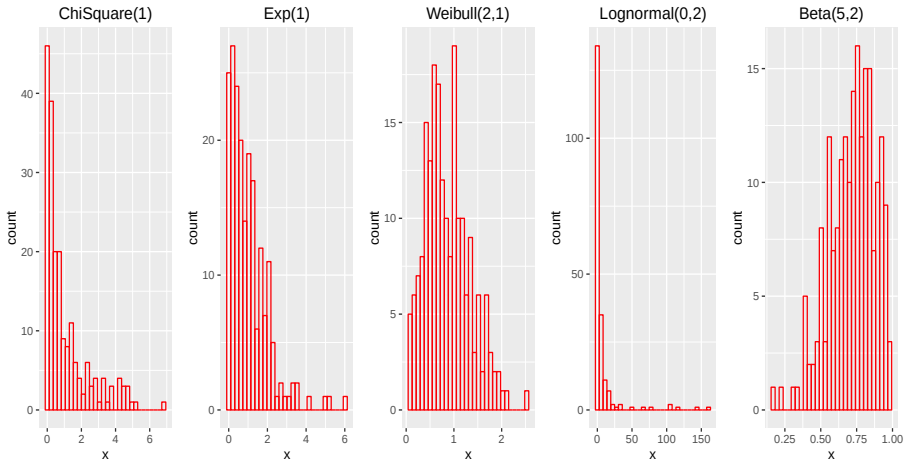

Figure 2.4 Alternative distributions for inverse Gaussian

using these alternative distributions, the following algorithm is performed in order to calculate the powers of D_K^*.

(1) Select a sample $X_{[i]j}$ from a distribution under H_1, $i = 1, \cdots, m$, $j = 1, \cdots, n$.

(2) Calculate D_K^* according to Equation 2.13.

(3) This steps (1)-(2) are repeated to get $D_{K,1}^*, \cdots, D_{K,10,000}^*$.

(4) $Power\ of\ D_K^* \approx \frac{1}{10,000} \sum_{h=1}^{10,000} I(D_{K,h} > C_\alpha^*)$

In first five columns of Table 2.4, the results are given for the fixed variance. As the correlations coefficients are decreased, the efficiencies are approaching to 1 for fixed set and cycle sizes. Even, some efficiencies are lower than 1 for $\rho = 0.5$. As the set and cycle are increased, the efficiencies increase for $N(0.25, 1)$ and $N(0.50, 1)$, but the efficiencies close to 1 for $N(1, 1)$ and $N(1.25, 1)$. It can be said that the set and cycle sizes affect the efficiencies as the mean parameter is approaching to 0.

Table 2.4 The efficiencies when the null hypothesis is standard normal distribution

ρ	$Cycle(l)$	$Set(k)$	$N(0.25,1)$	$N(0.50,1)$	$N(0.75,1)$	$N(1,1)$	$N(1.25,1)$	$N(0,2)$	$N(0,4)$	$N(0,8)$	$N(0,16)$	$N(0,32)$
1.00	5	2	1.171	1.318	1.308	1.193	1.077	1.315	1.416	1.479	1.478	1.338
		3	1.787	1.673	1.319	1.085	1.016	1.584	1.749	1.493	1.273	1.119
		4	1.957	1.743	1.191	1.023	1.002	2.152	1.936	1.394	1.098	1.017
	10	2	1.441	1.385	1.144	1.023	1.003	1.577	1.384	1.229	1.071	1.014
		3	1.855	1.371	1.048	1.001	1.000	2.115	1.547	1.137	1.010	1.000
		4	2.597	1.245	1.011	1.000	1.000	2.972	1.417	1.026	1.000	1.000
0.75	5	2	1.139	1.204	1.184	1.159	1.027	1.140	1.326	1.360	1.253	1.178
		3	1.238	1.464	1.347	1.131	1.010	1.498	1.398	1.245	1.093	1.026
		4	1.268	1.520	1.296	1.074	1.000	1.514	1.424	1.124	1.018	1.002
	10	2	1.175	1.377	1.203	1.053	1.001	1.312	1.205	1.099	1.015	1.000
		3	1.423	1.414	1.117	1.009	1.000	1.522	1.243	1.018	1.000	1.000
		4	2.090	1.458	1.047	1.001	1.000	1.835	1.184	1.003	1.000	1.000
0.50	5	2	0.829	0.943	0.987	1.010	1.010	0.855	0.940	0.952	0.978	0.983
		3	0.848	0.988	1.045	1.024	1.005	0.902	0.924	0.986	1.003	0.996
		4	0.914	1.043	1.044	1.014	1.005	0.953	0.999	1.029	1.006	1.000
	10	2	0.899	0.982	1.008	1.005	0.999	0.858	0.929	0.960	0.994	0.998
		3	0.893	1.010	1.017	1.001	1.002	0.865	0.973	0.997	0.999	1.000
		4	0.910	1.041	1.005	1.000	1.000	0.949	0.997	1.004	1.000	1.000

In the other five columns of Table 2.4, the efficiencies are presented for fixed mean. While the correlation coefficients are decreased, the efficiencies approach to 1 as well. Also, some efficiencies are lower than 1 for $\rho = 0.50$. As the set and cycle sizes increased, the efficiencies increase for $N(0, 2)$, but the efficiencies decrease for $N(0, 16)$ and $N(0, 32)$. While the alternative distributions expand, the efficiencies decrease. Generally, it can be said that D_K^* is more powerful than D_K for both fixed variance and mean. In the Table 2.5, the outcomes appear for asymmetric alternatives and the null distribution is inverse Gaussian. According to the results, D_K^* has higher performance than D_K when $\rho = 1.00$ and $\rho = 0.75$ except for $Beta(5, 2)$. Mostly,

Table 2.5 The efficiencies when the null hypothesis is inverse Gaussian

ρ	$Cycle(l)$	$Set(k)$	$ChiSquare(1)$	$Exp(1)$	$Weibull(2,1)$	$Lognormal(0,2)$	$Beta(5,2)$
1.00	5	2	1.583	1.266	1.502	1.494	3.476
		3	1.589	1.723	1.845	1.538	1.000
		4	1.543	2.235	2.168	1.469	1.000
	10	2	1.229	1.344	1.533	1.268	1.000
		3	1.226	2.285	2.102	1.184	1.000
		4	1.092	2.886	2.121	1.061	1.000
0.75	5	2	1.325	1.183	1.368	1.333	2.607
		3	1.338	1.666	1.310	1.337	1.000
		4	1.165	1.647	1.478	1.277	1.000
	10	2	1.128	1.384	1.472	1.195	1.000
		3	1.059	1.451	1.458	1.082	1.000
		4	1.015	1.825	2.151	1.022	1.000
0.50	5	2	0.967	0.818	0.839	0.967	0.804
		3	1.027	0.867	0.875	1.017	1.053
		4	1.050	1.020	0.890	1.013	1.000
	10	2	0.967	0.804	0.904	0.975	1.000
		3	1.015	0.894	0.863	0.993	1.000
		4	1.018	1.038	0.972	1.000	1.000

D_K^* and D_K have equal power for $Beta(5,2)$. Finally, EDFs are simulated for population, SRS and RSS. These functions are constructed under perfect ranking ($\rho = 1.00$) for RSS by setting $l = 5, 10$ and $k = 2, 3, 4$. In these figures, the RSS is better than the SRS even if the cycle size is 5 and set size is 2. While the set and cycle sizes are increased, the distances among the lines reduce.

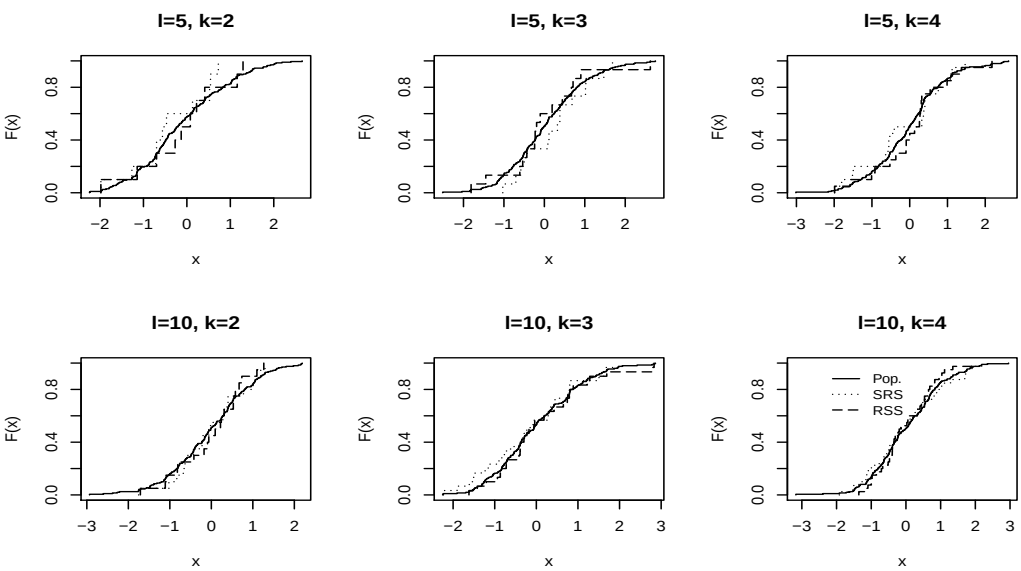

Figure 2.5 EDFs for standard normal

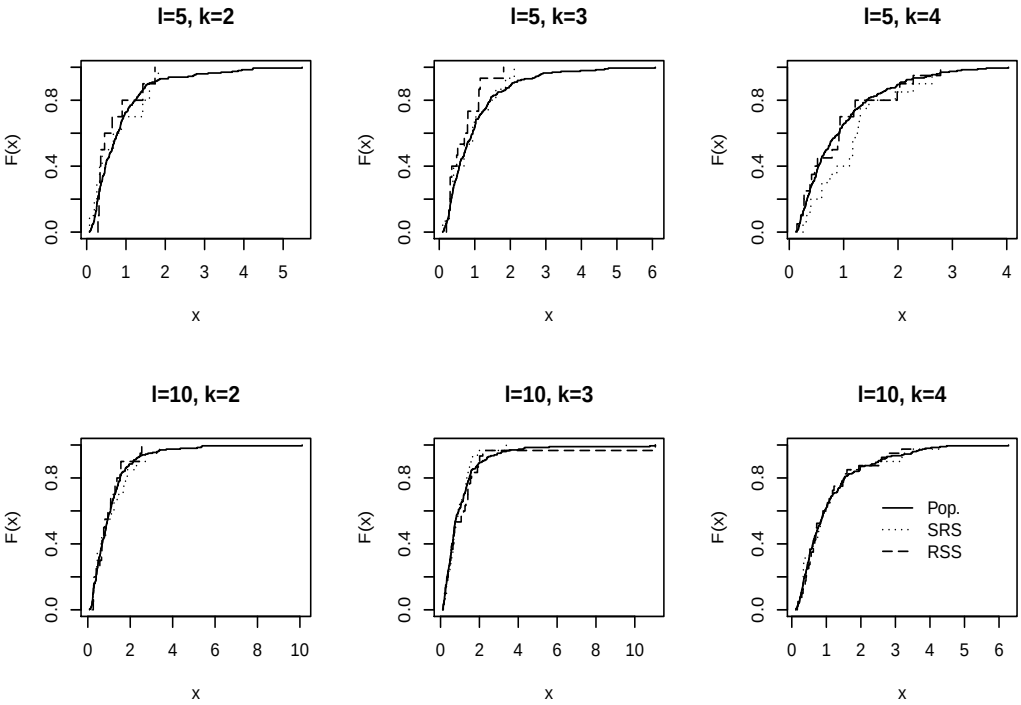

Figure 2.6 EDFs for inverse Gaussian

CHAPTER THREE
EMPIRICAL DISTRIBUTION FUNCTION ESTIMATORS FOR SAMPLING DESIGNS IN RANKED SET SAMPLING

A single sampling procedure is generalized the infinite population, but it is not generalized the finite population. Instead of using single sampling procedure in RSS, Deshpande et al. (2006) described three different sampling designs which are level-0, level-1 and level-2 in RSS. They simulated nonparametric confidence intervals for population median in finite populations. Also, Ozturk (2014) developed mean estimators for the sampling designs in PROS. In this chapter, we investigate EDF estimators for each sampling designs in RSS. Also, it is shown that these estimators are more efficient than SRS EDF estimator.

3.1 Sampling Designs

In this section, we define algorithms that are introduced by Deshpande et al. (2006) for each the sampling designs in RSS. These designs depend on with or without replacement sampling protocol as in SRS.

- Level-0 sampling design

  (1) Select k units without replacement from the population, $S_i = \left\{X^{(0)}_{t_1,i}, X^{(0)}_{t_2,i}, \cdots, X^{(0)}_{t_k,i}\right\}$.

  (2) Rank all units in the S_i and measured the ith order statistic.

  (3) Units in S_i are replaced back into the population.

  (4) (1)-(3) are repeated for $i = 1, \cdots, k$.

  (5) (1)-(4) are repeated for $i = 1, \cdots, l$.

Level-1 has two types that are level-1 ascending and level-1 descending. In this thesis, level-1 ascending is used.

- Level-1 ascending (A1) sampling design

  (1) Select k units without replacement from the population, $S_i = \left\{X_{t_1,i}^{(A1)}, X_{t_2,i}^{(A1)}, \cdots, X_{t_k,i}^{(A1)}\right\}$.

  (2) Rank all units ascending order in the S_i and measured the ith order statistic.

  (3) Other $(k-1)$ units are replaced back into the population.

  (4) (1)-(3) are repeated for $i = 1, \cdots, k$.

  (5) (1)-(4) are repeated for $i = 1, \cdots, l$.

- Level-1 descending (D1) sampling design

  (1) Select k units without replacement from the population, $S_i = \left\{X_{t_1,i}^{(D1)}, X_{t_2,i}^{(D1)}, \cdots, X_{t_k,i}^{(D1)}\right\}$.

  (2) Rank all units descending order in the S_i and measured the $(k+1-i)th$ order statistic.

  (3) Other $(k-1)$ units are replaced back into the population.

  (4) (1)-(3) are repeated for $i = 1, \cdots, k$.

  (5) (1)-(4) are repeated for $i = 1, \cdots, l$.

- Level-2 sampling design

  (1) Select k units without replacement from the population, $S_i = \left\{X_{t_1,i}^{(2)}, X_{t_2,i}^{(2)}, \cdots, X_{t_k,i}^{(2)}\right\}$.

  (2) Rank all units ascending order in the S_i and measured the ith order statistic.

  (3) None of the set units are replaced back into the population.

  (4) (1)-(3) are repeated for $i = 1, \cdots, k$.

  (5) (1)-(4) are repeated for $i = 1, \cdots, l$.

In the level-0 sampling design, same unit can be measured more than one in RSS sample. Via Level-0, all units in RSS sample are independent and selected with equal

probability. In the level-1 sampling design, it is allowed that same unit is used in ranking process. However, different units are measured for RSS sample. By using level-2 sampling design, different units are appeared in both ranking process and RSS sample. Thus, this design contains more information about population than other sampling design. Also, strong correlation occurs among the units in RSS sample via the level-2 design.

3.2 Empirical Distribution Function Estimators

EDF is basically a CDF. However, EDF models empirical data while CDF is a hypothetical model of a distribution. That means, EDF is used for making inference about entire distribution function. In many studies, different EDF estimators are suggested for different sampling methods. The most basic EDF for SRS which is given in Equation (2.10) is introduced by Stephens (1974). Then, EDF estimators are considered by Stokes & Sager (1988) for RSS, by Samawi & Al-Sageer (2001) for extreme and median RSS methods, by Abu-Dayyeh et al. (2002) for double RSS, by Kim et al. (2005) for extreme median RSS, by Al-Subh et al. (2009) for SORSS and by Nazari et al. (2016) for PROS. In this section, we describe EDF estimators based on the sampling designs which are level-0, level-1 and level-2.

If the RSS sample, $X^{(0)}_{[1]j}, \cdots, X^{(0)}_{[k]j}$ and $j = 1, \cdots, l$, is selected by using Level-0 from $F(x)$, then the EDF estimator ($\hat{F}^*_{L-0}(x)$) is

$$\hat{F}^*_{L-0}(x) = \frac{1}{lk} \sum_{i=1}^{l} \sum_{j=1}^{k} I(X^{(0)}_{[i]j} \leq x). \tag{3.1}$$

where $I(.)$ is indicator function. If the RSS sample, $X^{(1)}_{[1]j}, \cdots, X^{(1)}_{[k]j}$, is obtained by using Level-1, then the EDF estimator ($\hat{F}^*_{L-1}(x)$) is

$$\hat{F}^*_{L-1}(x) = \frac{1}{lk} \sum_{i=1}^{l} \sum_{j=1}^{k} I(X^{(1)}_{[i]j} \leq x). \tag{3.2}$$

If the RSS sample, $X^{(2)}_{[1]j}, \cdots, X^{(2)}_{[k]j}$, is obtained by using Level-2, then the EDF estimator ($\hat{F}^*_{L-2}(x)$) is

$$\hat{F}^*_{L-2}(x) = \frac{1}{lk} \sum_{i=1}^{l} \sum_{j=1}^{k} I(X^{(2)}_{[i]j} \leq x). \tag{3.3}$$

In the following section, it is shown that these EDFs are more efficient than SRS EDF by numerical results.

3.3 Simulation Results

In this section, numerical results are provided to compare performances of EDFs based on SRS and the sampling designs in RSS. Also, behaviors of sampling designs are investigated. To evaluate RSS EDF estimators respect to SRS EDF estimator, their MSEs are used.

In order to generate population of size $N = 120$, bivariate standard normal distribution is used. In this population, there are interested variable (X) and auxiliary variable (Y). Between these variables, a fixed correlation coefficient (ρ) is existed.

Auxiliary variable is used in the ranking procedure, so the quality of the ranking depends on the correlation coefficient between X and Y. If the correlation is high, the interested variable can be ranked accurately. On the other hand, the ranking is performed randomly if the correlation is low. In Table 3.1, we present the results under perfect ranking ($\rho = 1.00$) and imperfect ranking ($\rho = 0.25$) for the sampling designs.

In simulation study, RSS EDFs are constructed for $k = 2, 3, 4$ and 5 and $l = 2$ and 3. Also, we take sample size as $n = 4, 6, 8, 9, 10, 12$ and 15. MSEs of EDFs are computed based on $10,000$ samples in R software. For SRS, mean squared error (MSE) of EDF estimator is given by

$$MSE(\hat{F}(x)) = \frac{1}{10,000} \sum_{i=1}^{10,000} (\hat{F}_i(x) - F(x))^2. \tag{3.4}$$

Following equation is also MSE for the sampling designs,

$$MSE(\hat{F}^*_{L-t}(x)) = \frac{1}{10,000} \sum_{i=1}^{10,000} (\hat{F}^*_{L-t,i}(x) - F(x))^2. \tag{3.5}$$

where $t = 0, 1, 2$. To compare these MSEs, relative efficiency (RE) is calculated by using following equation,

$$RE_t = \frac{MSE(\hat{F}(x))}{MSE(\hat{F}^*_{L-t}(x))}. \tag{3.6}$$

These relative efficiencies are computed for some probabilities (**P**), **P** $= \{0.05, 0.10, 0.15, 0.20, 0.30, 0.40, 0.50\}$. According to these probabilities, $MSE(\hat{F}^{-1}(p))$ and $MSE(\hat{F}^{-1}_{L-t}(p))$ are obtained for SRS and RSS, respectively. Table 3.1 shows that EDFs based on the sampling designs are more efficient than EDF based on SRS for perfect ranking ($\rho = 1.00$). However, efficiencies are approximately equal to 1 when ranking is imperfect ($\rho = 0.25$). While the set size is increased, the efficiencies increase for perfect ranking. However, the efficiencies do not monotone increase or decrease while the cycle size is increased.

Finally, efficiencies based on perfect ranking ($\rho = 1.00$) are illustrated for each sampling designs in Figure 3.1 and Figure 3.2. According to Figure 3.1, EDF based the on level-2 sampling design is more efficient than EDFs based on the level-0 and level-1 except for $k = 2$. In Figure 3.2, efficiencies of EDF based on the level-2 sampling design is the highest among three different sampling designs. In the next chapter, powers of different GOF tests are compared using these EDFs for the sampling design.

Table 3.1 Efficiencies for the sampling designs

			$\rho = 1.00$							$\rho = 0.25$						
Designs	l	k	0.05	0.10	0.15	0.20	0.30	0.40	0.50	0.05	0.10	0.15	0.20	0.30	0.40	0.50
$Level-0$	2	2	0.971	1.042	1.100	1.158	1.215	1.306	1.279	0.986	0.953	0.986	0.987	0.988	0.968	0.968
		3	1.057	1.116	1.209	1.286	1.431	1.479	1.541	0.927	0.968	0.927	0.960	0.966	0.946	0.934
		4	1.107	1.211	1.290	1.451	1.581	1.728	1.762	0.992	0.914	0.992	0.997	0.942	0.938	0.968
		5	1.074	1.240	1.370	1.560	1.780	1.825	1.905	0.986	0.915	0.986	0.944	0.933	0.939	0.925
	3	2	0.968	1.015	1.065	1.114	1.200	1.222	1.249	0.978	0.994	0.978	0.959	0.971	0.952	0.948
		3	1.032	1.117	1.181	1.276	1.382	1.437	1.514	0.954	0.938	0.954	0.926	0.968	0.958	0.952
		4	1.012	1.180	1.316	1.415	1.556	1.569	1.672	0.920	0.913	0.920	0.944	0.915	0.910	0.925
		5	0.997	1.209	1.355	1.437	1.637	1.787	1.737	0.925	0.911	0.925	0.900	0.880	0.914	0.914
$Level-1$	2	2	1.028	1.099	1.164	1.185	1.241	1.341	1.299	1.017	0.985	1.017	0.995	0.997	0.970	0.987
		3	1.085	1.165	1.266	1.351	1.518	1.572	1.642	0.989	1.012	0.989	1.000	0.999	0.991	0.978
		4	1.149	1.280	1.394	1.580	1.687	1.865	1.885	1.061	0.966	1.061	1.065	1.002	0.969	1.012
		5	1.172	1.350	1.497	1.711	1.944	2.040	2.167	1.032	0.987	1.032	1.020	1.005	1.008	1.001
	3	2	1.031	1.055	1.125	1.184	1.282	1.315	1.345	1.017	0.994	1.017	1.007	1.015	1.006	0.993
		3	1.110	1.176	1.251	1.386	1.518	1.576	1.602	1.016	0.996	1.016	1.006	1.032	0.993	0.987
		4	1.100	1.300	1.458	1.604	1.755	1.790	1.850	1.026	1.010	1.026	1.016	0.973	0.983	1.021
		5	1.164	1.362	1.569	1.639	1.916	2.005	1.988	1.039	1.024	1.039	1.020	0.996	1.010	1.000
$Level-2$	2	2	1.026	1.051	1.127	1.154	1.251	1.329	1.313	1.055	0.977	1.055	1.009	1.010	0.974	0.994
		3	1.062	1.182	1.258	1.361	1.520	1.577	1.646	1.013	1.029	1.013	1.025	1.015	0.989	0.989
		4	1.103	1.287	1.408	1.585	1.767	1.952	2.031	1.034	0.973	1.034	1.064	0.989	0.974	0.989
		5	1.195	1.393	1.537	1.785	2.098	2.179	2.297	1.044	0.967	1.044	1.006	1.002	1.013	1.009
	3	2	1.050	1.096	1.153	1.205	1.320	1.341	1.352	0.973	1.002	0.973	0.996	1.015	0.983	0.970
		3	1.109	1.188	1.285	1.408	1.535	1.625	1.652	0.998	0.996	0.998	1.036	1.040	1.010	1.004
		4	1.120	1.308	1.504	1.608	1.854	1.891	2.058	1.010	0.985	1.010	1.040	1.000	0.989	1.003
		5	1.170	1.451	1.730	1.817	2.139	2.329	2.344	1.061	1.000	1.061	1.024	1.015	1.017	1.019

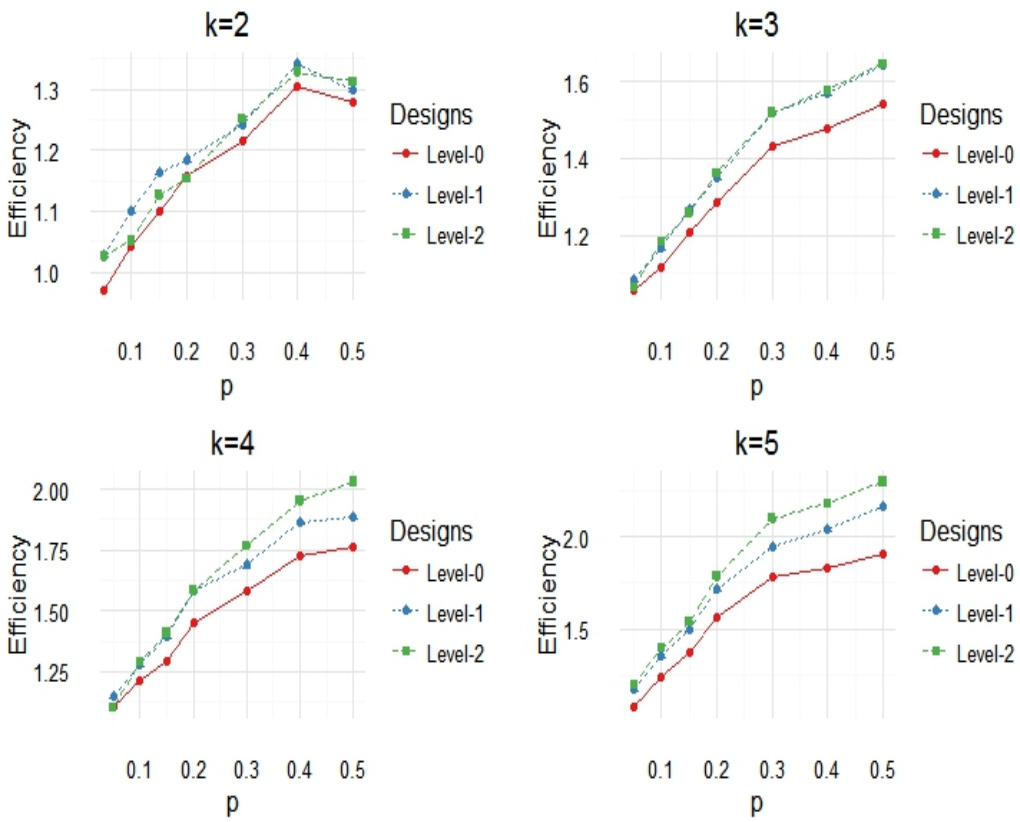

Figure 3.1 When $l = 2$, efficiencies of EDFs based on the sampling designs

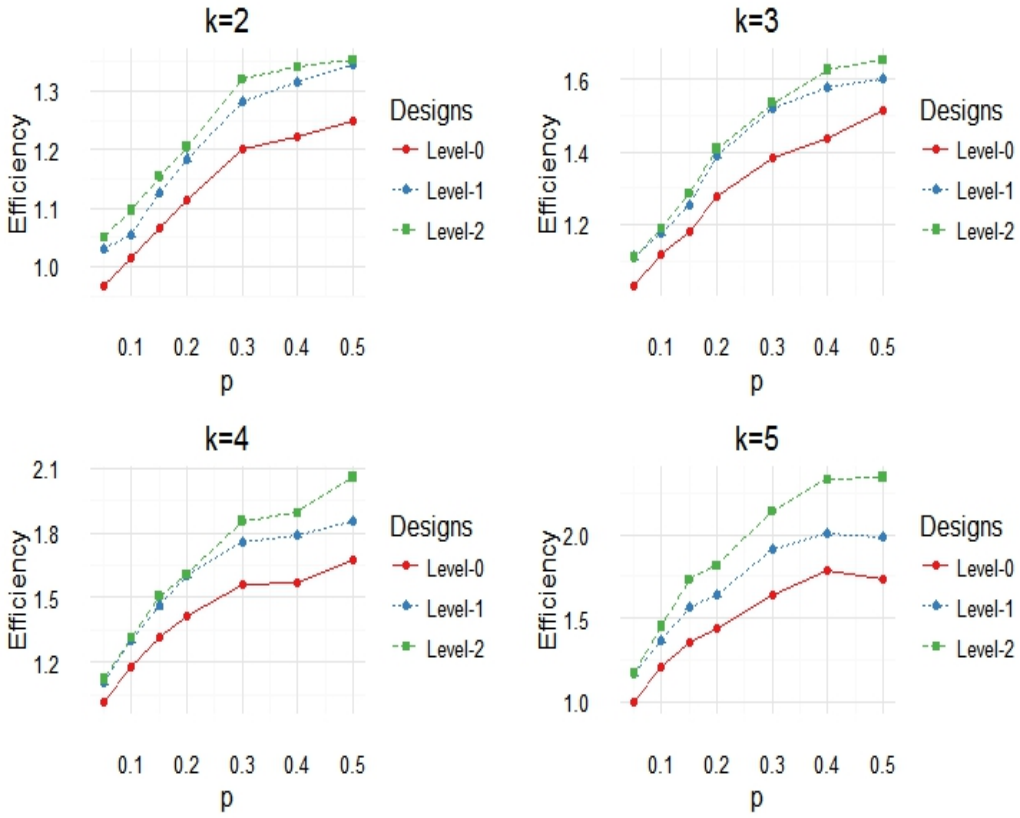

Figure 3.2 When $l = 3$, efficiencies of EDFs based on the sampling designs

CHAPTER FOUR
POWER COMPARISONS OF SOME GOODNESS-OF-FIT TESTS FOR SAMPLING DESIGNS IN RANKED SET SAMPLING

In statistics, parametric tests are used under some assumptions and one of them is normality. Normality assumption is an important GOF problem. To test this assumption, many GOF techniques are suggested, see D'Agostino (1986).

In this chapter, we consider five GOF tests based on the EDF for RSS. Powers of these tests based on RSS are simulated and compared with the test based on SRS. Also, the sampling designs are used to construct RSS samples. In ranking process, the units are ranked perfectly. To generate finite populations, Tukey's g-h distribution is used.

RSS, the sampling designs and the EDF estimators are discussed in previous chapter. Therefore, we only emphasize GOF tests and Tukey's g-h distribution in this chapter. Finally, results are given in tables and illustrated as figures.

4.1 Goodness-of-fit Tests

GOF tests examine how well a sample of observations agrees with a specific distribution as its population. That means these tests are used for making inference about the population distribution. Mostly, it is tested whether sampling observations are obtained from a population having normal distribution or not. In this situation, null hypothesis H_0 is simple hypothesis if we know parameters. On the other hand, H_0 is composite hypothesis when the parameters don't be known. In this case, the parameters are estimated by using sampling observations. Also, alternative hypothesis H_1 is mostly composite hypothesis since we have little or no information about distribution of the data. Actually, researchers hope that H_0 is true.

In this study, we investigate the powers of Kolmogorov-Smirnov, Kuiper, Lilliefors, Cramĕr-von Misses and Anderson-Darling GOF tests under SRS and RSS. These are

GOF tests based on EDF and divided two different classes. Kolmogorov-Smirnov, Kuiper and Lilliefors test statistics are in the supremum class. Anderson-Darling and Cramer-von Mises tests belong to the quadratic class.

Firstly, we introduce these tests for SRS. It is assumed that a random sample of size n, $X_1, \cdots, X_n$ from a population and CDF of this population is $F(x)$. We test the null hypothesis $H_0 : F(x) = F_0(x)$ against $H_1 : F(x) \neq F_0(x)$. Kolmogorov-Smirnov is suggested by Kolmogorov (1933) and Smirnov (1939). Kolmogorov (1933) gave the following test statistic D,

$$D_K = \sup_x \left| \hat{F}(x) - F_0(x) \right| = max(D^+, D^-) \tag{4.1}$$

where $\hat{F}(x)$ is EDF and

$$D_K^+ \equiv \sup_x \left\{ \hat{F}(x) - F_0(x) \right\}, \tag{4.2}$$

$$D_K^- \equiv \sup_x \left\{ F_0(x) - \hat{F}(x) \right\}. \tag{4.3}$$

Smirnov (1939) proposed the following equations,

$$D^+ = \max(\frac{i}{n} - Z_i), \tag{4.4}$$

$$D^- = \max(Z_i - \frac{i-1}{n}). \tag{4.5}$$

where $Z_i = \phi(\frac{X_{(i)} - \bar{X}}{S_X})$ is CDF of standard normal distribution. Another test statistic V is introduced by Kuiper (1962). This test statistic is used both Kolmogorov-Smirnov test statistics, D^+ and D^-. That means this test statistic is also modified version of Kolmogorov-Smirnov,

$$V = D^{+} + D^{-}. \tag{4.6}$$

When population parameters are unknown, Kolmogorov-Smirnov and Kuiper tests cannot be applied. To solve this problem, Lilliefors (1967) gave a test statistic D,

$$D = \max \left| \hat{F}(x) - F_0(x) \right| \tag{4.7}$$

where parameters of $F_0(x)$ are $\mu = \bar{X}$ sample mean and $\sigma^2 = S^2$ sample variance. Anderson-Darling test statistic A^2 is proposed by Anderson-Darling (1954).

$$A^2 = n \int_{-\infty}^{\infty} \{\hat{F}(x) - F_0(x)\}^2 \psi(x) dF_0(x) \tag{4.8}$$

where $\psi(x)$ is weight function and $\psi(x) = \frac{1}{F_0(x)(1-F_0(x))}$. When $\psi(x) = 1$, Cramer-von Mises test statistic is obtained. This test statistic is suggested by von Mises (1931).

$$W^2 = n \int_{-\infty}^{\infty} \{\hat{F}(x) - F_0(x)\}^2 dF_0(x) \tag{4.9}$$

Suppose that a RSS sample of size kl, $X_{[i]j}^{(t)}$, $i = 1, \cdots, k$ and $j = 1, \cdots, l$ is selected by using Level-t sampling design where $t = 0, 1, 2$. Then, these five GOF tests are as follows.

- Kolmogorov-Smirnov test statistics:

$$D_K^{+} \equiv \sup_x \left\{ \hat{F}_{L-t}^{*}(x) - F_0(x) \right\} \tag{4.10}$$

$$D_K^{-} \equiv \sup_x \left\{ F_0(x) - \hat{F}_{L-t}^{*}(x) \right\} \tag{4.11}$$

$$D_K^{*} \equiv \sup_x \left| \hat{F}_{L-t}^{*}(x) - F_0(x) \right| \tag{4.12}$$

- Kuiper's test statistic:

$$V = D_K^+ + D_K^- \tag{4.13}$$

- Lilliefors Test statistic:

$$D = \max \left| \hat{F}^*_{L-t}(x) - F_0(x) \right| \tag{4.14}$$

where $F_0(x)$ is standard normal CDF with $\mu = \bar{x}$ sample mean and $\sigma^2 = s^2$ sample variance.

- Cramĕr-von Mises test statistic:

$$W^2 = n \int_{-\infty}^{\infty} \{\hat{F}^*_{L-t}(x) - F_0(x)\}^2 dF_0(x) \tag{4.15}$$

- Anderson-Darling Test statistic:

$$A^2 = n \int_{-\infty}^{\infty} \{\hat{F}^*_{L-t}(x) - F_0(x)\}^2 \psi(x) dF_0(x) \tag{4.16}$$

where $\hat{F}^*_{L-t}(x)$ is EDF for level-t sampling design, $t = 0, 1, 2$.

4.2 Tukey's g-h Distribution

A new class of distribution function is introduced by Tukey (1977) and it is named as Tukey's g-h distribution function. The Tukey's g-h distribution arises as a non-linear transformation of a standard normal random variable Z. That means, symmetric and asymmetric distributions can be produced and PDF and CDF can be expressed in parametric form by using this function. A g-h random variable is denoted by $Y_{a,b,g,h}$ where a and b are location and scale parameters and g and h are skewness and the kurtosis, respectively.

$$Y_{a,b,g,h} = a + bZ\left(\frac{e^{gZ} - 1}{gZ}\right)e^{hZ^2/2}, \tag{4.17}$$

where $a \in \mathbb{R}$, $b \in \mathbb{R}^+$, $g \in \mathbb{R}$ and $h \in \mathbb{R}$. When $h = 0$ the Tukey's g-h distribution reduces to

$$Y_{a,b,g} = a + b\left(\exp\left(gZ\right) - 1\right)/g, \tag{4.18}$$

which is Tukey's g distribution. Similarly, when $g \to 0$, the Tukey's g-h distribution is given by

$$Y_{a,b,h} = a + bZ\exp\left(hZ^2/2\right) \tag{4.19}$$

known as the h distribution. When h is negative, a left skewed distribution is obtained. On the other hand, a right skewed distribution is obtained when h is positive; see Jorge & Boris (1984) and Dutta & Babbel (2002) for more details. For specific values of g and h, the distribution approximates a selected set of well known distributions as given in Table 4.1.

Table 4.1 Values of g and h for some distributions

Distribution	Parameters	a	b	g	h
Cauchy	$\mu, \gamma > 0$	μ	σ	0	1
Exponential	$\lambda > 0$	$\frac{1}{\lambda}\ln 2$	$\frac{g}{\lambda}$	0.773	-0.09445
Laplace	$\alpha, \beta > 0$	α	β	0	0
Logistic	$\alpha, \beta > 0$	α	β	0	1.7771×10^{-3}
Lognormal	$\mu, \sigma^2, C > 0$	C^{μ}	gC^{μ}	$\sigma \ln C$	0
Normal	μ, σ^2	μ	σ	0	0
t_{10}	$v = 10$	0	1	0	5.7624×10^{-2}

In the case of $a = 0$ and $b = 1$, $\tau_{g,h}$ is denoted as univariate $g - h$ distribution

function. By using this function, a data set that is distributed for $a = 0$ can be generated.

$$Y_{g,h} = \tau_{g,h}(Z) = \frac{(e^{g^z} - 1)}{g} e^{hZ^2/2} \tag{4.20}$$

The multivariate form of the function is introduced by Field & Genton (2006). In this thesis, we used bivariate form of the function (**Y**) to construct the alternative hypothesis.

$$\mathbf{Y} = (\tau_{g_1,h_1}(Z_1), \tau_{g_2,h_2}(Z_2)) = \tau_{\mathbf{g},\mathbf{h}}(\mathbf{Z}) \tag{4.21}$$

where $\tau_{g_1,h_1} = X$ (interested variable) and $\tau_{g_2,h_2} = C$ (auxiliary variable). The **Z** is distributed bivariate standard normal. Finally, the **Y** is defined by the following equation.

$$\mathbf{Y} = \sigma^{1/2} \tau_{\mathbf{g},\mathbf{h}}(\mathbf{Z}) + \mu \tag{4.22}$$

where $\sigma = \begin{bmatrix} \sigma_{(Z_1)} = 1 & \sigma_{(Z_1,Z_2)} \\ \sigma_{(Z_1,Z_2)} & \sigma_{(Z_2)} = 1 \end{bmatrix}$ and $\mu = \begin{bmatrix} 0, 0 \end{bmatrix}$.

Here, $\sigma_{(Z_1,Z_2)}$ is covariance and computed by Equation (4.22).

$$\sigma_{(Z_1,Z_2)} = \rho_{(Z_1,Z_2)} \, \sigma_{(Z_1)} \, \sigma_{(Z_2)} = \rho_{(Z_1,Z_2)} \tag{4.23}$$

If the **Y** is described $f(x, c)$, then the marginal distributions of the **Y** are given by following equation.

$$\begin{aligned} f(x) &= \tau_{g_1,h_1}(Z_1) = (\frac{e^{g_1^{Z_1}} - 1}{g_1}) e^{h_1 Z_1^2/2}, \\ f(c) &= \tau_{g_2,h_2}(Z_2) = (\frac{e^{g_2^{Z_2}} - 1}{g_2}) e^{h_2 Z_2^2/2}. \end{aligned} \tag{4.24}$$

In the simulation study, it is assumed that kurtosis parameter (h= 0) and skewness parameter (g= $0, 1, -1$) to generate three different alternative distribution. Descriptive statistics of alternative distributions are in Table 4.2. Also, these

alternative distributions can be seen in Figure 4.1.

Table 4.2 The alternative distributions

g	h	$Skewness$	$Kurtosis$	μ	σ^2
0	0	0.02	3.03	0.06	3.64
1	0	3.09	15.14	1.28	14.59
-1	0	-3.99	26.78	-1.07	13.77

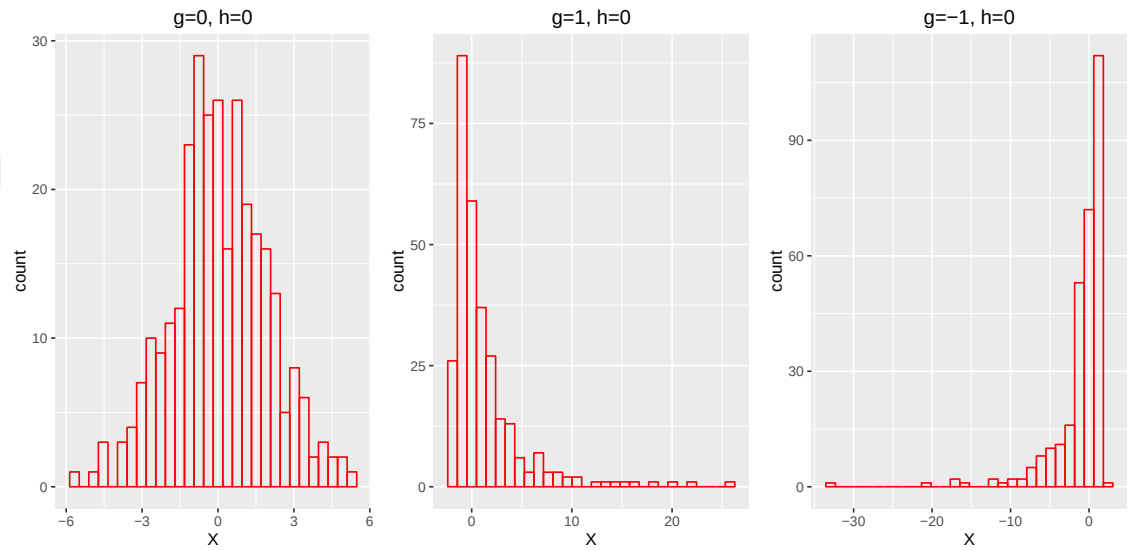

Figure 4.1 The alternative distributions

4.3 Simulation Results

In this chapter, we compare the performances of GOF tests which are Kolmogorov-Smirnov, Kuiper, Lilliefors, Cramĕr-von Mises and Anderson-Darling based on SRS and RSS in finite population. In RSS process, sample data is collected by using three different sampling designs which are Level-0, Level-1 and Level-2. Powers of GOF tests are simulated by using EDFs for the sampling designs and SRS.

In order to compute the critical values and powers of GOF tests, two different simulations are performed. Null hypothesis is assumed as bivariate standard normal

distribution and alternative hypothesis is obtained by using bivariate Tukey's g-h distribution. We consider the population size (N) 300, set sizes (k) 3, 5 and total sample sizes (n) 15, 30 and 45. For perfect ranking of interested variable (X), the correlation coefficient between interested variable and auxiliary variable (C) is assumed as 1, ($\rho = 1$). To generate critical values and powers of the GOF tests, simulations are repeated $5,000$ times.

For RSS, critical values are computed by using the following algorithm. T^* indicates a test statistic in Equations (4.10) - (4.16).

(1) Select a RSS sample $X_{[i]j}$ from $F_0(x)$, $i = 1, 2, \cdots, k$, $j = 1, 2, \cdots, l$.

(2) Calculate the T^* according to Equations (4.10) - (4.16).

(3) This steps (1)-(2) are repeated to get $T_1^*, \cdots, T_{5,000}^*$.

(4) The critical value C_α^* is the $100(1-\alpha)$ percentage point of T^*.

Table 4.3 Critical values for the GOF tests under RSS when $\alpha = 0.05$

$Designs$	l	k	D_K	V	D	W^2	A^2
	5	3	0.292	0.409	0.220	0.285	1.638
	10		0.211	0.301	0.160	0.292	1.705
$Level-0$	15		0.175	0.249	0.134	0.294	1.671
	3	5	0.269	0.397	0.216	0.222	1.305
	6		0.197	0.285	0.159	0.226	1.339
	9		0.161	0.241	0.134	0.224	1.350
	5	3	0.285	0.402	0.215	0.269	1.554
	10		0.200	0.286	0.155	0.258	1.463
$Level-1$	15		0.164	0.232	0.129	0.245	1.444
	3	5	0.263	0.383	0.210	0.210	1.244
	6		0.182	0.272	0.151	0.196	1.178
	9		0.151	0.225	0.126	0.195	1.164
	5	3	0.284	0.401	0.214	0.258	1.521
	10		0.197	0.286	0.153	0.249	1.436
$Level-2$	15		0.159	0.232	0.127	0.228	1.327
	3	5	0.259	0.381	0.210	0.199	1.169
	6		0.179	0.273	0.152	0.181	1.087
	9		0.142	0.222	0.126	0.166	1.015

Also, critical values (A.2 - A.6) are generated by setting the cycle sizes $l = 1, \cdots, 20$ and the set sizes $k = 2, \cdots, 5$ for each GOF tests. For these critical values, simulation

is repeated $2,000$ samples since a single replication takes a little over 40 seconds for $l = 15, \cdots, 20$. It can be benefited from these critical values in studies which sampling designs are used in RSS.

In order to calculate the powers of T^*, the following algorithm is used. This algorithm steps are also repeated $5,000$ times.

(1) Select a sample $X_{[i]j}$ from a distribution under H_1, $i = 1, \cdots, k$, $j = 1, \cdots, l$.

(2) Calculate T^* according to Equation (4.10)-(4.16).

(3) Repeat steps (1)-(2) to get $T_1^*, \cdots, T_{5,000}^*$.

(4) $Power\, of\, T^* \approx \frac{1}{5,000} \sum_{h=1}^{5,000} I(T_h^* > C_\alpha^*)$

Then, the results in Tables 4.4 - 4.6 are obtained for different set (k) and total sample sizes (n) under RSS and SRS. The power of GOF tests based on SRS are same for all sampling designs. According to powers of GOF tests, it can be said that GOF tests based on RSS have higher performance than the GOF tests based on SRS. Also, it is shown that the GOF tests based on $\hat{F}_{L-2}(t)$ have the best performance among the three sampling designs. As the set size (k) is increased, the power values increase. In general, Lilliefors test (D) has the best performance among the tests based supremum statistic. However, D has minimum power for symmetric distribution under both SRS and RSS. Among the five GOF tests, the best results are obtained for Anderson-Darling (A^2) test.

Finally, we illustrate EDFs based on SRS, level-0, level-1 and level-2 sampling designs for symmetric distribution. In these figures, it is shown that the EDFs of the sampling designs are almost identical for each set and total sample sizes.

Table 4.4 The power values for the Level-0 sampling design

$Distr.$	n	k	D_K		V		D		W^2		A^2	
			SRS	RSS	SRS	RSS	SRS	RSS	SRS	RSS	SRS	RSS
$g-h(0,0)$	15	3	0.268	0.375	0.553	0.606	0.064	0.059	0.288	0.456	0.785	0.881
		5		0.436		0.659		0.056		0.591		0.942
	30	3	0.475	0.647	0.860	0.878	0.058	0.074	0.562	0.766	0.966	0.986
		5		0.741		0.934		0.058		0.887		0.996
	45	3	0.746	0.838	0.980	0.973	0.069	0.071	0.829	0.924	0.996	0.998
		5		0.915		0.985		0.064		0.979		1.000
$g-h(-1,0)$	15	3	0.300	0.379	0.499	0.554	0.650	0.630	0.260	0.415	0.867	0.927
		5		0.468		0.606		0.678		0.533		0.960
	30	3	0.513	0.693	0.818	0.846	0.917	0.907	0.501	0.713	0.982	0.991
		5		0.778		0.904		0.914		0.848		0.998
	45	3	0.794	0.883	0.969	0.955	0.990	0.979	0.773	0.892	0.999	0.999
		5		0.935		0.973		0.986		0.963		1.000
$g-h(1,0)$	15	3	0.354	0.484	0.540	0.611	0.695	0.682	0.295	0.475	0.911	0.953
		5		0.564		0.661		0.717		0.616		0.978
	30	3	0.661	0.822	0.871	0.889	0.954	0.945	0.598	0.812	0.993	0.998
		5		0.891		0.937		0.955		0.920		0.999
	45	3	0.898	0.949	0.986	0.976	0.995	0.992	0.864	0.952	0.999	1.000
		5		0.985		0.990		0.993		0.988		1.000

Table 4.5 The power values for the Level-1 sampling design

$Distr.$	n	k	D_K		V		D		W^2		A^2	
			SRS	RSS	SRS	RSS	SRS	RSS	SRS	RSS	SRS	RSS
$g-h(0,0)$	15	3	0.268	0.388	0.553	0.635	0.064	0.058	0.288	0.486	0.785	0.901
		5		0.437		0.692		0.060		0.612		0.952
	30	3	0.475	0.707	0.860	0.920	0.058	0.065	0.562	0.832	0.966	0.991
		5		0.831		0.957		0.063		0.930		0.998
	45	3	0.746	0.900	0.980	0.989	0.069	0.060	0.829	0.969	0.996	0.999
		5		0.957		0.996		0.061		0.993		1.000
$g-h(-1,0)$	15	3	0.300	0.420	0.499	0.580	0.650	0.663	0.260	0.442	0.867	0.933
		5		0.506		0.647		0.686		0.557		0.959
	30	3	0.513	0.741	0.818	0.891	0.917	0.930	0.501	0.785	0.982	0.996
		5		0.866		0.933		0.944		0.904		0.998
	45	3	0.794	0.924	0.969	0.931	0.990	0.987	0.773	0.953	0.999	0.999
		5		0.974		0.993		0.994		0.984		1.000
$g-h(1,0)$	15	3	0.354	0.498	0.541	0.631	0.695	0.715	0.295	0.510	0.911	0.966
		5		0.627		0.693		0.762		0.648		0.986
	30	3	0.661	0.869	0.871	0.927	0.954	0.954	0.598	0.874	0.993	0.999
		5		0.942		0.958		0.969		0.951		1.000
	45	3	0.898	0.978	0.986	0.990	0.995	0.997	0.864	0.985	0.999	1.000
		5		0.997		0.997		0.998		0.997		1.000

Table 4.6 The power values for the Level-2 sampling design

$Distr.$	n	k	D_K		V		D		W^2		A^2	
			SRS	RSS	SRS	RSS	SRS	RSS	SRS	RSS	SRS	RSS
$g-h(0,0)$	15	3	0.268	0.383	0.553	0.631	0.064	0.064	0.288	0.504	0.785	0.909
		5		0.473		0.716		0.066		0.637		0.959
	30	3	0.475	0.736	0.860	0.912	0.058	0.070	0.562	0.843	0.966	0.994
		5		0.852		0.954		0.059		0.946		0.999
	45	3	0.746	0.916	0.980	0.987	0.069	0.057	0.829	0.976	0.996	0.999
		5		0.976		0.997		0.061		0.995		1.000
$g-h(-1,0)$	15	3	0.300	0.412	0.499	0.575	0.650	0.661	0.260	0.463	0.867	0.932
		5		0.512		0.661		0.705		0.592		0.972
	30	3	0.513	0.769	0.818	0.885	0.917	0.935	0.501	0.802	0.982	0.996
		5		0.882		0.937		0.942		0.922		0.999
	45	3	0.794	0.947	0.969	0.981	0.990	0.987	0.773	0.964	0.999	1.000
		5		0.987		0.994		0.991		0.994		1.000
$g-h(1,0)$	15	3	0.354	0.502	0.541	0.632	0.695	0.715	0.295	0.528	0.911	0.965
		5		0.640		0.711		0.747		0.670		0.987
	30	3	0.661	0.884	0.871	0.920	0.954	0.962	0.598	0.887	0.993	0.999
		5		0.960		0.961		0.971		0.971		1.000
	45	3	0.898	0.983	0.986	0.992	0.995	0.998	0.864	0.989	0.999	1.000
		5		0.999		0.997		0.997		0.999		1.000

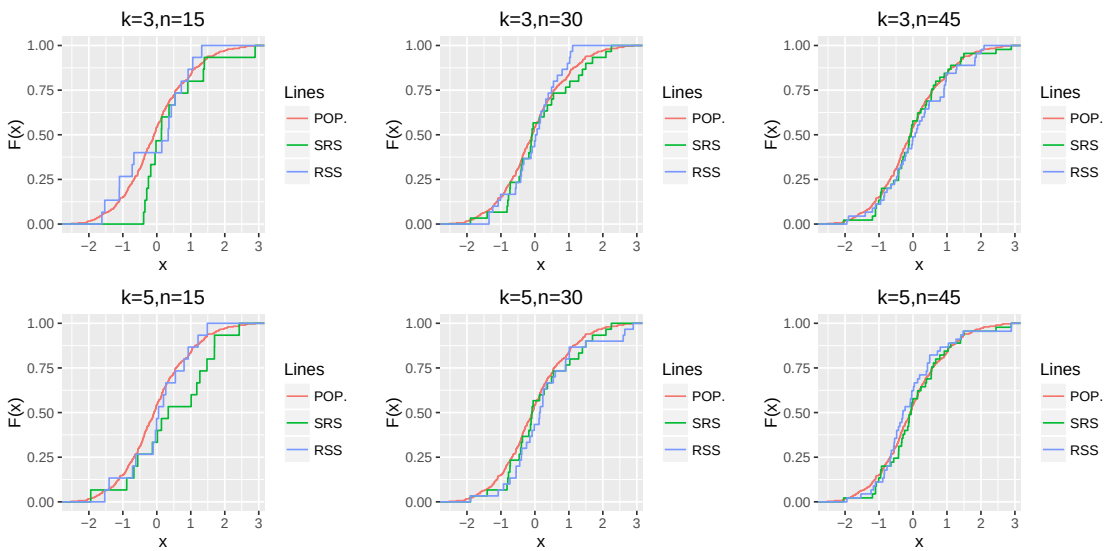

Figure 4.2 The EDFs under the level-0 sampling design

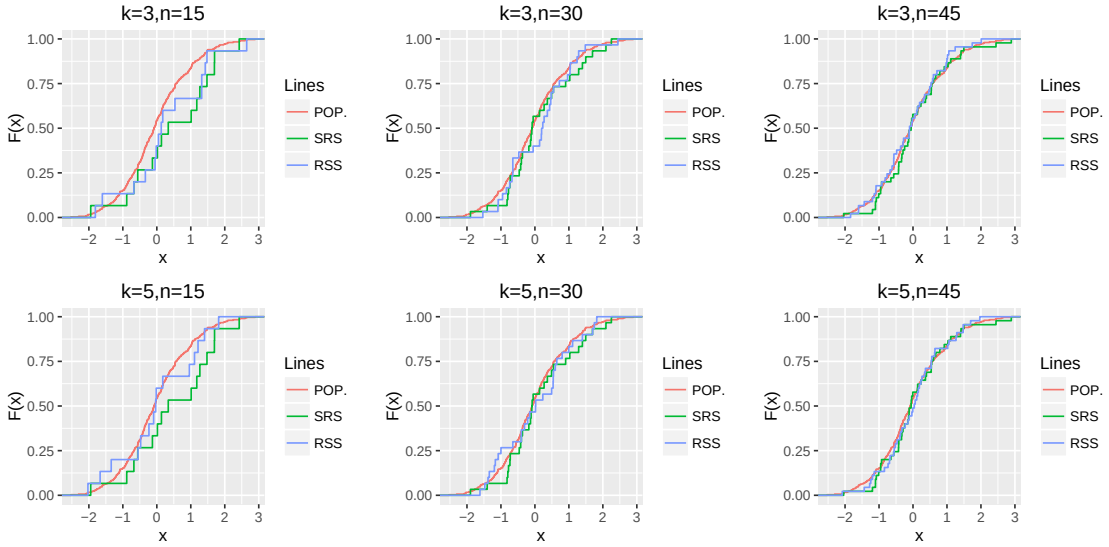

Figure 4.3 The EDFs under the level-1 sampling design

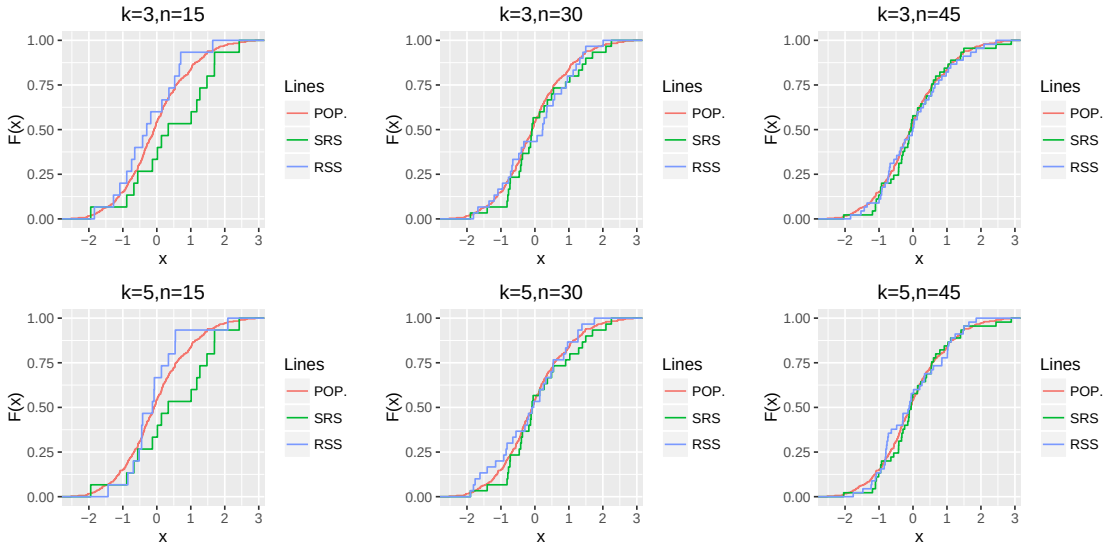

Figure 4.4 The EDFs under the level-2 sampling design

CHAPTER FIVE
CONCLUSION

In this thesis, EDFs based on the level-0, level-1 and level-2 are suggested and efficiencies of these EDFs are investigated in finite population. Also, we examined powers of GOF tests based on these three sampling designs.

In chapter two, power of Kolmogorov-Smirnov test statistic based on RSS is compared with power of the test statistic based on SRS. RSS data is collected by using level-2 sampling design. Moreover, we benefited from PROS in RSS and we showed that it is how useful procedure in ranking procedure. Multiple auxiliary variables are used in ranking process. Three different correlation coefficients are used to show the effects of ranking error. Efficiency of Kolmogorov-Smirnov GOF test based on RSS is computed by using different alternatives such as normal and asymmetric distributions. In the results, efficiency values depend on correlation coefficient between interested variable and auxiliary variable and these values decrease as the correlation coefficient are reduced. The best efficiency values are obtained when $l = 10$ and $k = 4$ for distributions $N(0.25, 1)$ and $N(0, 2)$. Also, the best efficiencies are obtained when $l = 5$ and $k = 2$ for distributions $N(1.25, 1)$ and $N(0, 32)$.

In chapter three, algorithms of the sampling designs are introduced and EDFs based on the level-0, level-1 and level-2 sampling designs are given. Two correlation coefficients are used for perfect and imperfect rankings. Under standard normal distribution, efficiency values are obtained by using ratio of MSEs of EDF estimators based on the sampling designs in RSS and based on SRS. According to the results, EDFs based on the sampling designs have higher performance than EDF based on SRS for perfect ranking. Also, the efficiencies increase as the set size is increased. EDF based on the level-2 sampling design has higher efficiencies than EDFs based on the level-0 and level-1 sampling designs except for $k = 2$ in Figure 3.1.

In chapter four, the powers of different GOF tests based on RSS and SRS are evaluated. RSS sample observations are collected by using level-0, level-1 and

level-2 sampling designs. Thus, powers of GOF tests are examined among the sampling designs. It is assumed that the ranking is perfect. Powers of GOF tests are obtained by using three alternative distributions. According to these powers, GOF tests based on RSS have higher power values than the GOF tests based on SRS. Mostly, GOF tests based on the level-2 sampling design have the highest power values. Among these five GOF tests, the best power is obtained for Anderson-Darling (A^2) test.

As a result, it is shown that the use of auxiliary variable is a good alternative to rank the interested variable under high correlation and PROS is a useful method to combine the multiple auxiliary variables. The EDFs based on the sampling designs in RSS have higher efficiencies than the EDF based on SRS under perfect ranking. Moreover, EDF based on the level-2 sampling design is the best estimator among the three sampling designs. GOF tests based on $\hat{F}^*_{L-0}(x)$, $\hat{F}^*_{L-1}(x)$ and $\hat{F}^*_{L-2}(x)$ have higher powers than GOF tests based on $\hat{F}(x)$.

REFERENCES

Abu-Dayyeh, W. A., Samawi, H. M., & Bani-Hani, L. A. (2002). On distribution function estimation using double ranked set samples with application. *Journal of Modern Applied Statistical Methods*, *1*(2), 443-451.

Al-Omari, A., & Zamanzade, E. (2016). Different goodness of fit tests for rayleigh distribution in ranked set sampling. *Pakistan Journal of Statistics and Operation Research*, *12*(1), 25-39.

Al-Omari, A. I., & Bouza, C. N. (2014). Review of ranked set sampling: modifications and applications. *Revista Investigación Operacional*, *3*, 215-240.

Al-Saleh, M. F., & Al-Ananbeh, A. M. (2007). Estimation of the means of the bivariate normal using moving extreme ranked set sampling with concomitant variable. *Statistical Papers*, *48*(2), 179-195.

Al-Subh, S., Alodat, M., Ibrahim, K., & Jemain, A. (2009). Edf goodness of fit tests of logistic distribution under selective order statistics. *Pakistan Journal of Statistics*, *25*(3), 265-274.

Al-Subh, S., Alodat, M., Ibrahim, K., & Jemain, A. A. (2012). Modified edf goodness of fit tests for logistic distribution under srs and rss. *Journal of Modern Applied Statistical Methods*, *11*(2), 385-395.

Alizadeh Noughabi, H. (2017). Efficiency of ranked set sampling in tests for normality. *Journal of Statistical Computation and Simulation*, *87*(5), 956-965.

Anderson, T. W., & Darling, D. A. (1954). A test of goodness of fit. *Journal of the American statistical association*, *49*(268), 765-769.

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

Chen, H., Stasny, E. A., & Wolfe, D. A. (2005). Ranked set sampling for efficient estimation of a population proportion. *Statistics in medicine*, *24*(21), 3319-3329.

Chen, Z. (2002). Adaptive ranked-set sampling with multiple concomitant variables: an effective way to observational economy. *Bernoulli*, *8*(3), 313-322.

Chen, Z., Bai, Z., & Sinha, B. (2003). *Ranked set sampling: theory and applications*. Springer Science & Business Media.

Chen, Z., & Shen, L. (2003). Two-layer ranked set sampling with concomitant variables. *Journal of Statistical Planning and Inference*, *115*(1), 45-57.

D'Agostino, R. B. (1986). *Goodness-of-fit-techniques*. CRC press.

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

Deshpande, J. V., Frey, J., & Ozturk, O. (2006). Nonparametric ranked-set sampling confidence intervals for quantiles of a finite population. *Environmental and Ecological Statistics*, *13*(1), 25-40.

Dutta, K. K., Babbel, D. F. et al. (2002). On measuring skewness and kurtosis in short rate distributions: The case of the us dollar london inter bank offer rates. *Center for Financial Institutions Working Papers 02*, *25*.

Field, C., & Genton, M. G. (2006). The multivariate g-and-h distribution. *Technometrics*, *48*(1), 104-111.

Frey, J., & Wang, L. (2014). Edf-based goodness-of-fit tests for ranked-set sampling. *Canadian Journal of Statistics*, *42*(3), 451-469.

Halls, L. K., & Dell, T. R. (1966). Trial of ranked-set sampling for forage yields. *Forest Science*, *12*(1), 22-26.

Husby, C. E., Stasny, E. A., & Wolfe, D. A. (2005). An application of ranked set sampling for mean and median estimation using usda crop production data. *Journal of agricultural, biological, and environmental statistics*, *10*(3), 354-373.

Jorge, M., & Boris, I. (1984). Some properties of the tukey g and h family of distributions. *Communications in Statistics-Theory and Methods*, *13*(3), 353-369.

Kaur, A., Patil, G., Shirk, S. J., & Taillie, C. (1996). Environmental sampling with a concomitant variable: a comparison between ranked set sampling and stratified simple random sampling. *Journal of Applied Statistics*, *23*(2-3), 231-256.

Kaur, A., Patil, G., Sinha, A., & Taillie, C. (1995). Ranked set sampling: an annotated bibliography. *Environmental and Ecological Statistics*, *2*(1), 25-54.

Kim, D., Kim, D., & Kim, H. (2005). On the estimation of the distribution function using extreme median ranked set sampling. *Journal of the Korean Data Analysis Society*, *7*(2), 429-439.

Kolmogorov, A. N. (1933). Sulla determinazione empirica di une legge di distribuzione. *Giornale dell'Intituta Italiano degli Attuari*, *4*, 83-91.

Kuiper, N. H. (1962). Test concerning random points on a circle. In *Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen, series A*. North-Holland.

Lilliefors, H. W. (1967). On the kolmogorov-smirnov test for normality with mean and variance unknown. *Journal of the American statistical Association*, *62*(318), 399-402.

Mahdizadeh, M., & Arghami, N. R. (2010). Efficiency of ranked set sampling in entropy estimation and goodness-of-fit testing for the inverse gaussian law. *Journal of Statistical Computation and Simulation*, *80*(7), 761-774.

McIntyre, G. A. (1952). A method for unbiased selective sampling, using ranked sets. *Australian Journal of Agricultural Research*, *3*(4), 385-390.

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

Neerchal, N. K., Sinha, B. K., & Lacayo, H. (1998). Ranked set sampling from a dichotomous population. *Journal of Applied Statistical Science*, *11*(1), 83-90.

Ozturk, O. (2011). Sampling from partially rank-ordered sets. *Environmental and Ecological statistics*, *18*(4), 757-779.

Ozturk, O. (2012a). Combining ranking information in judgment post stratified and ranked set sampling designs. *Environmental and ecological statistics*, *19*(1), 73-93.

Ozturk, O. (2012b). Quantile inference based on partially rank-ordered set samples. *Journal of Statistical Planning and Inference*, *142*(7), 2116-2127.

Ozturk, O. (2014). Estimation of population mean and total in a finite population setting using multiple auxiliary variables. *Journal of Agricultural, Biological, and Environmental Statistics*, *19*(2), 161-184.

Samawi, H. M., & Al-Sagheer, O. A. (2001). On the estimation of the distribution function using extreme and median ranked set sampling. *Biometrical Journal*, *43*(3), 357-373.

Sevil, Y. C., & Yildiz, T. O. (2017). Power comparison of the kolmogorov–smirnov test under ranked set sampling and simple random sampling. *Journal of Statistical Computation and Simulation*, *87*(11), 2175-2185.

Shahabuddin, F. A. A., Ibrahim, K., & Jemain, A. A. (2009). On the comparison of several goodness of fit tests under simple random sampling and ranked set sampling. *World Academy of Science, Engineering and Technology*, *54*, 77-80.

Smirnov, N. (1939). On the estimation of the discrepancy between empirical curves of distribution for two independent samples. *Bulletin de l'Univérsite de Moscou Sėrie International (MathėmatiQues)*, *2*, 3-16.

Stephens, M. A. (1974). Edf statistics for goodness of fit and some comparisons. *Journal of the American statistical Association*, *69*(347), 730-737.

Stokes, S. L. (1977). Ranked set sampling with concomitant variables. *Communications in Statistics Theory and Methods*, *6*(12), 1207-1211.

Stokes, S. L. (1980a). Inferences on the correlation coefficient in bivariate normal populations from ranked set samples. *Journal of the American Statistical Association*, *75*(372), 989-995.

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

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*(402), 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.

Tukey, J. W. (1977). Modern techniques in data analysis. In *Proceedings of the NSF-Sponsored Regional Research Conference*. Southern Massachusetts University.

von Mises, R. (1931). Wahrscheinlichkeitsrechnung. *Leipzig-Wien*.

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

APPENDICES

A.1: Critical values for Kolmogorov-Smirnov test

Table A.1 The critical values for RSS (D_K^*) when $\alpha = 0.05$ under standard normal distribution

	$m = 2$			$m = 3$			$m = 4$		
n	$\rho = 1.00$	$\rho = 0.75$	$\rho = 0.50$	$\rho = 1.00$	$\rho = 0.75$	$\rho = 0.50$	$\rho = 1.00$	$\rho = 0.75$	$\rho = 0.50$
1	0.751	0.816	0.798	0.570	0.621	0.651	0.484	0.562	0.565
2	0.548	0.582	0.581	0.432	0.471	0.483	0.351	0.408	0.426
3	0.460	0.485	0.492	0.355	0.386	0.403	0.291	0.332	0.349
4	0.405	0.429	0.433	0.307	0.338	0.354	0.252	0.284	0.304
5	0.362	0.379	0.410	0.276	0.292	0.334	0.224	0.252	0.290
6	0.330	0.342	0.369	0.253	0.266	0.296	0.203	0.225	0.258
7	0.311	0.314	0.334	0.230	0.245	0.276	0.186	0.208	0.243
8	0.290	0.296	0.324	0.215	0.229	0.261	0.173	0.195	0.225
9	0.272	0.282	0.301	0.202	0.217	0.243	0.161	0.183	0.212
10	0.258	0.267	0.297	0.192	0.210	0.242	0.151	0.167	0.208

A.2: Critical values tables for different sampling designs (Level-0, level-1, level-2) in Ranked Set Sampling

Table A.2 Critical Values of the Kolmogorov-Smirnov Test

	$Level-0$				$Level-1$				$Level-2$			
l	$k=2$	$k=3$	$k=4$	$k=5$	$k=2$	$k=3$	$k=4$	$k=5$	$k=2$	$k=3$	$k=4$	$k=5$
1	0.765	0.587	0.499	0.436	0.761	0.593	0.501	0.445	0.759	0.601	0.492	0.430
2	0.557	0.443	0.368	0.319	0.562	0.435	0.374	0.322	0.555	0.447	0.364	0.316
3	0.470	0.369	0.310	0.268	0.473	0.367	0.304	0.268	0.465	0.360	0.305	0.262
4	0.411	0.321	0.270	0.235	0.412	0.324	0.271	0.227	0.411	0.317	0.267	0.234
5	0.376	0.288	0.241	0.209	0.370	0.291	0.236	0.206	0.370	0.292	0.240	0.206
6	0.339	0.264	0.219	0.191	0.344	0.257	0.218	0.193	0.342	0.264	0.219	0.192
7	0.319	0.243	0.202	0.176	0.323	0.247	0.207	0.175	0.315	0.243	0.205	0.177
8	0.298	0.234	0.192	0.167	0.300	0.228	0.191	0.166	0.298	0.226	0.190	0.171
9	0.279	0.217	0.181	0.156	0.286	0.217	0.181	0.157	0.272	0.219	0.179	0.156
10	0.264	0.207	0.175	0.147	0.267	0.201	0.174	0.150	0.267	0.202	0.172	0.149
11	0.258	0.196	0.165	0.141	0.255	0.199	0.164	0.144	0.255	0.199	0.165	0.141
12	0.247	0.189	0.156	0.137	0.242	0.187	0.154	0.139	0.249	0.188	0.159	0.135
13	0.236	0.180	0.152	0.129	0.230	0.177	0.149	0.133	0.233	0.185	0.148	0.133
14	0.227	0.176	0.145	0.126	0.230	0.178	0.146	0.125	0.229	0.171	0.144	0.125
15	0.220	0.170	0.140	0.123	0.218	0.171	0.145	0.124	0.219	0.170	0.139	0.122
16	0.215	0.165	0.137	0.117	0.214	0.164	0.135	0.120	0.215	0.167	0.136	0.118
17	0.209	0.159	0.134	0.114	0.207	0.158	0.132	0.115	0.207	0.163	0.133	0.113
18	0.202	0.156	0.127	0.111	0.198	0.158	0.126	0.111	0.203	0.153	0.129	0.110
19	0.196	0.153	0.125	0.109	0.198	0.152	0.124	0.108	0.196	0.151	0.124	0.107
20	0.195	0.147	0.123	0.106	0.194	0.146	0.120	0.105	0.192	0.148	0.120	0.105

Table A.3 Critical Values of the Kuiper's Test

	$Level-0$				$Level-1$				$Level-2$			
l	$k=2$	$k=3$	$k=4$	$k=5$	$k=2$	$k=3$	$k=4$	$k=5$	$k=2$	$k=3$	$k=4$	$k=5$
1	0.966	0.824	0.702	0.632	0.967	0.817	0.707	0.631	0.968	0.825	0.696	0.629
2	0.737	0.614	0.523	0.463	0.740	0.605	0.524	0.463	0.744	0.617	0.517	0.463
3	0.626	0.508	0.439	0.386	0.629	0.502	0.440	0.385	0.622	0.500	0.434	0.392
4	0.556	0.450	0.383	0.340	0.550	0.448	0.383	0.336	0.557	0.456	0.381	0.341
5	0.499	0.405	0.345	0.308	0.501	0.407	0.343	0.303	0.510	0.405	0.344	0.304
6	0.462	0.369	0.317	0.280	0.464	0.373	0.314	0.278	0.462	0.375	0.316	0.279
7	0.430	0.339	0.295	0.258	0.426	0.343	0.300	0.260	0.437	0.347	0.293	0.259
8	0.401	0.324	0.276	0.246	0.405	0.321	0.274	0.241	0.404	0.318	0.271	0.247
9	0.380	0.305	0.261	0.230	0.384	0.305	0.266	0.229	0.378	0.305	0.257	0.230
10	0.364	0.290	0.249	0.218	0.364	0.287	0.253	0.221	0.358	0.291	0.244	0.218
11	0.350	0.277	0.237	0.209	0.345	0.279	0.239	0.208	0.350	0.281	0.238	0.206
12	0.334	0.265	0.227	0.199	0.326	0.264	0.228	0.200	0.330	0.266	0.229	0.199
13	0.320	0.256	0.217	0.191	0.315	0.251	0.217	0.195	0.317	0.260	0.217	0.195
14	0.309	0.244	0.209	0.185	0.313	0.250	0.212	0.184	0.305	0.245	0.211	0.186
15	0.298	0.241	0.204	0.181	0.293	0.246	0.204	0.179	0.296	0.240	0.206	0.182
16	0.291	0.235	0.198	0.174	0.289	0.232	0.198	0.175	0.290	0.228	0.197	0.173
17	0.282	0.228	0.194	0.167	0.282	0.223	0.190	0.169	0.285	0.224	0.192	0.168
18	0.273	0.218	0.186	0.163	0.270	0.224	0.184	0.162	0.269	0.220	0.185	0.162
19	0.268	0.214	0.182	0.160	0.267	0.215	0.181	0.160	0.266	0.213	0.179	0.157
20	0.262	0.207	0.179	0.156	0.261	0.203	0.177	0.156	0.259	0.210	0.178	0.154

Table A.4 Critical Values of the Lilliefors Test

	$Level-0$				$Level-1$				$Level-2$			
l	$k=2$	$k=3$	$k=4$	$k=5$	$k=2$	$k=3$	$k=4$	$k=5$	$k=2$	$k=3$	$k=4$	$k=5$
1	0.260	0.376	0.374	0.337	0.260	0.375	0.373	0.340	0.260	0.372	0.370	0.337
2	0.371	0.324	0.282	0.254	0.370	0.325	0.283	0.256	0.373	0.324	0.285	0.256
3	0.320	0.274	0.239	0.216	0.321	0.273	0.240	0.213	0.323	0.270	0.237	0.216
4	0.290	0.241	0.213	0.187	0.284	0.247	0.208	0.188	0.288	0.240	0.208	0.188
5	0.260	0.215	0.191	0.172	0.260	0.217	0.186	0.168	0.261	0.217	0.187	0.172
6	0.240	0.199	0.174	0.155	0.243	0.204	0.176	0.158	0.241	0.204	0.174	0.157
7	0.223	0.184	0.161	0.145	0.225	0.187	0.163	0.141	0.226	0.189	0.161	0.143
8	0.211	0.175	0.154	0.135	0.210	0.173	0.150	0.135	0.212	0.170	0.154	0.136
9	0.200	0.169	0.144	0.128	0.201	0.163	0.144	0.129	0.200	0.164	0.143	0.128
10	0.191	0.157	0.137	0.123	0.189	0.160	0.137	0.122	0.191	0.156	0.134	0.124
11	0.184	0.149	0.131	0.117	0.184	0.152	0.131	0.117	0.183	0.152	0.134	0.117
12	0.177	0.146	0.126	0.113	0.177	0.142	0.127	0.110	0.177	0.142	0.124	0.114
13	0.169	0.138	0.122	0.107	0.169	0.138	0.121	0.109	0.169	0.142	0.121	0.109
14	0.165	0.133	0.116	0.105	0.166	0.134	0.117	0.103	0.162	0.135	0.116	0.103
15	0.159	0.130	0.113	0.100	0.158	0.130	0.114	0.102	0.158	0.130	0.113	0.102
16	0.156	0.127	0.109	0.099	0.154	0.127	0.110	0.096	0.153	0.123	0.111	0.097
17	0.148	0.121	0.105	0.094	0.149	0.123	0.107	0.093	0.149	0.120	0.107	0.095
18	0.147	0.118	0.104	0.090	0.145	0.119	0.102	0.091	0.144	0.118	0.105	0.092
19	0.142	0.116	0.099	0.090	0.143	0.117	0.101	0.089	0.142	0.117	0.099	0.089
20	0.139	0.114	0.099	0.087	0.136	0.112	0.098	0.088	0.137	0.114	0.098	0.088

Table A.5 Critical Values of the Cramĕr-Von Mises Test

	$Level-0$				$Level-1$				$Level-2$			
l	$k=2$	$k=3$	$k=4$	$k=5$	$k=2$	$k=3$	$k=4$	$k=5$	$k=2$	$k=3$	$k=4$	$k=5$
1	0.328	0.260	0.238	0.212	0.331	0.260	0.237	0.227	0.321	0.275	0.233	0.210
2	0.323	0.284	0.242	0.215	0.335	0.279	0.250	0.220	0.323	0.288	0.238	0.222
3	0.338	0.277	0.245	0.221	0.345	0.276	0.242	0.222	0.337	0.269	0.234	0.221
4	0.334	0.287	0.249	0.222	0.343	0.294	0.240	0.206	0.326	0.285	0.244	0.223
5	0.339	0.281	0.240	0.224	0.342	0.290	0.257	0.210	0.343	0.288	0.241	0.218
6	0.331	0.276	0.289	0.220	0.350	0.273	0.237	0.224	0.345	0.280	0.245	0.221
7	0.346	0.274	0.242	0.215	0.355	0.288	0.256	0.215	0.329	0.277	0.247	0.215
8	0.334	0.288	0.249	0.221	0.337	0.268	0.242	0.223	0.330	0.271	0.237	0.231
9	0.326	0.279	0.249	0.222	0.338	0.296	0.241	0.225	0.321	0.289	0.241	0.221
10	0.336	0.283	0.247	0.216	0.347	0.273	0.262	0.220	0.340	0.261	0.242	0.212
11	0.345	0.281	0.244	0.220	0.344	0.286	0.250	0.229	0.332	0.280	0.245	0.212
12	0.348	0.288	0.244	0.221	0.337	0.280	0.227	0.223	0.356	0.282	0.255	0.221
13	0.340	0.275	0.244	0.212	0.322	0.268	0.246	0.233	0.326	0.297	0.244	0.225
14	0.347	0.282	0.245	0.220	0.346	0.295	0.246	0.223	0.363	0.264	0.237	0.216
15	0.345	0.278	0.241	0.222	0.321	0.278	0.253	0.227	0.336	0.285	0.241	0.220
16	0.344	0.291	0.252	0.212	0.338	0.279	0.242	0.228	0.346	0.300	0.241	0.218
17	0.350	0.283	0.251	0.213	0.349	0.273	0.246	0.215	0.355	0.294	0.243	0.215
18	0.345	0.283	0.245	0.220	0.330	0.288	0.238	0.223	0.349	0.276	0.251	0.214
19	0.345	0.282	0.244	0.215	0.338	0.294	0.249	0.217	0.336	0.274	0.234	0.217
20	0.362	0.274	0.250	0.219	0.358	0.279	0.238	0.206	0.345	0.283	0.244	0.211

Table A.6 Critical Values of the Anderson-Darling Test

	$Level-0$				$Level-1$				$Level-2$			
l	$k=2$	$k=3$	$k=4$	$k=5$	$k=2$	$k=3$	$k=4$	$k=5$	$k=2$	$k=3$	$k=4$	$k=5$
1	2.031	1.675	1.524	1.320	2.129	1.714	1.459	1.346	1.999	1.705	1.428	1.295
2	1.906	1.679	1.469	1.307	1.931	1.648	1.466	1.294	2.041	1.707	1.448	1.326
3	1.968	1.621	1.485	1.310	1.994	1.643	1.522	1.301	1.889	1.551	1.386	1.293
4	1.937	1.656	1.497	1.321	1.997	1.630	1.413	1.260	1.908	1.620	1.471	1.309
5	1.967	1.633	1.409	1.327	1.943	1.652	1.497	1.255	1.967	1.698	1.455	1.298
6	1.917	1.619	1.436	1.336	1.969	1.631	1.376	1.315	1.931	1.613	1.460	1.316
7	1.957	1.618	1.450	1.294	1.980	1.627	1.515	1.288	1.872	1.595	1.495	1.285
8	1.945	1.708	1.464	1.310	1.884	1.597	1.405	1.333	1.884	1.562	1.377	1.357
9	1.871	1.629	1.469	1.306	1.908	1.686	1.441	1.362	1.858	1.680	1.389	1.297
10	1.938	1.637	1.467	1.286	1.965	1.607	1.524	1.299	1.938	1.606	1.430	1.270
11	1.946	1.642	1.424	1.321	1.983	1.667	1.445	1.336	1.910	1.647	1.441	1.251
12	1.975	1.651	1.447	1.333	1.944	1.640	1.381	1.298	1.936	1.634	1.534	1.267
13	1.944	1.621	1.425	1.253	1.860	1.534	1.422	1.389	1.866	1.666	1.427	1.339
14	1.979	1.635	1.425	1.286	1.945	1.674	1.497	1.313	2.046	1.592	1.414	1.268
15	1.943	1.657	1.437	1.319	1.862	1.707	1.471	1.297	1.931	1.642	1.414	1.321
16	1.955	1.685	1.472	1.281	1.942	1.629	1.406	1.350	1.937	1.671	1.437	1.279
17	1.972	1.655	1.494	1.250	1.962	1.570	1.467	1.257	1.988	1.695	1.446	1.314
18	1.945	1.655	1.431	1.306	1.918	1.707	1.397	1.285	1.955	1.588	1.424	1.272
19	1.930	1.637	1.424	1.287	1.949	1.670	1.463	1.317	1.921	1.544	1.395	1.283
20	2.028	1.585	1.455	1.305	1.985	1.609	1.363	1.294	2.005	1.639	1.439	1.250