DOKUZ EYLÜL UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

ANALYSIS OF ROBUST APPROACHES IN THE TWO INDEPENDENT SAMPLES CASE

by

Gözde NAVRUZ

March, 2019

İZMİR

ANALYSIS OF ROBUST APPROACHES IN THE TWO INDEPENDENT SAMPLES CASE

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

by
Gözde NAVRUZ

March, 2019
İZMİR

Ph.D. THESIS EXAMINATION RESULT FORM

We have read the thesis entitled "**ANALYSIS OF ROBUST APPROACHES IN THE TWO INDEPENDENT SAMPLES CASE**" completed by **GÖZDE NAVRUZ** under supervision of **ASSOC. PROF. DR. A. FIRAT ÖZDEMİR** and we certify that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Doctor of Philosophy.

Assoc. Prof. Dr. A. Fırat ÖZDEMİR

Supervisor

Prof. Dr. Burcu HÜDAVERDİ

Thesis Committee Member

Assoc. Prof. Dr. Çetin DİŞİBÜYÜK

Thesis Committee Member

Assoc. Prof. Dr. Sevcan DEMİR ATALAY

Examining Committee Member

Assoc. Prof. Dr. Hatice Sevim SAYDAM

Examining Committee Member

Prof. Dr. Kadriye ERTEKİN

Director

Graduate School of Natural and Applied Sciences

ACKNOWLEDGEMENTS

First and foremost, I would like to express my special thanks and deep appreciation to my supervisor Assoc. Prof. Dr. A. Fırat ÖZDEMİR for his continuous encouragement and immense knowledge. His guidance and suggestions played a very important role not only throughout this thesis, but also all my academic research. This thesis would not have been possible without his support and it is a great chance for me to be his student since my M.Sc. research.

I would like to acknowledge my dissertation committee members Prof. Dr. Burcu HÜDAVERDİ and Assoc. Prof. Dr. Çetin DİŞİBÜYÜK for their constructive comments and suggestions.

I would also like to thank TÜBİTAK (The Scientific and Technical Research Council of Turkey) for its generous financial support via "2211-A National Scholarship Programme for PhD Students" during my Ph.D. research.

I am also grateful to Uğur BİNZAT for his precious support and help whenever I needed during my research.

Words cannot express my thanks to my family for their endless love and encouragement throughout all my life. I wish to express my gratitude to my father Coşkun NAVRUZ for his continuing support without which this work would not be complete. I also would like to thank my sister Ece NAVRUZ for her patience and invaluable love during this thesis. Last, but not least, I would like to express my heartfelt thanks to my mother Rakibe NAVRUZ who has always believed in me and encouraged me to accomplish things that I intend to. This thesis is dedicated to her.

Gözde NAVRUZ

ANALYSIS OF ROBUST APPROACHES IN THE TWO INDEPENDENT SAMPLES CASE

ABSTRACT

Comparing two independent groups has been the subject of many classic studies in applied statistics. Although the most common approach for comparing two independent groups is on the basis of just one reference point, it may provide misleading results and cause missing the important differences among subpopulations of the groups. Determination of the differences in tails of the groups, namely quantiles, can also be of particular concern.

This thesis aims to compare two independent groups from a different point of view by utilizing quantiles. Therefore, firstly attention is focused on quantile estimation theory and the newly proposed NO quantile estimator is introduced. NO quantile estimator is computationally practical, has desirable asymptotic properties and is more efficient than some quantile estimators that are commonly used. Following this, NO quantile estimator is used in conjunction with a percentile bootstrap to propose a method that compares two independent groups through multiple quantiles simultaneously.

By conducting a series of simulation studies and using some real datasets, the performances of NO quantile estimator and the proposed method for comparing two independent groups are investigated. The proposed method could control the actual type I error rates in almost all cases that are investigated. This method is more preferable than the conventional approaches that considers a single measure of location since it makes a more sensitive comparison by using more than one reference points.

Keywords: NO quantile estimator, two independent groups, quantile estimators, asymptotic distribution, percentile bootstrap

İKİ BAĞIMSIZ ÖRNEKLEM İÇİN DAYANIKLI YAKLAŞIMLARIN ANALİZİ

ÖZ

İki bağımsız grubu karşılaştırmak, uygulamalı istatistikte birçok klasik çalışmanın konusu olmuştur. İki bağımsız grubu karşılaştırmak için en yaygın yaklaşım yalnızca bir referans noktasına dayanır. Ancak bu yaklaşım yanıltıcı sonuçlar verebilir ve grupların alt popülasyonları arasındaki önemli farkların gözden kaçırılmasına neden olabilir. Grupların kuyruklarındaki, yani kantillerdeki farklılıkların belirlenmesi de özel ilgi konusu olabilir.

Bu tez, iki bağımsız grubu kantillerden yararlanarak farklı bir bakış açısıyla karşılaştırmayı amaçlamaktadır. Bu nedenle öncelikle dikkat kantil tahmin teorisine odaklanmıştır ve yeni önerilen NO kantil tahmin edicisi tanıtılmıştır. NO kantil tahmin edicisi, pratik bir şekilde hesaplanabilir, istenen asimptotik özelliklere sahiptir ve yaygın olarak kullanılan bazı kantil tahmin edicilerinden daha etkindir. Bunu takiben NO kantil tahmin edicisi, iki bağımsız grubu eşzamanlı olarak birden fazla kantil aracılığıyla karşılaştıran bir yöntem önermek için bir yüzdelik bootstrap ile birlikte kullanılmıştır.

Bir dizi simülasyon çalışması yapılarak ve bazı gerçek veri setleri kullanılarak, NO kantil tahmin edicisinin ve iki bağımsız grubu karşılaştırmak için önerilen yöntemin performansları incelenmiştir. Önerilen yöntem, incelenen durumların neredeyse tamamında gerçekleşen 1. tip hata oranlarını kontrol edebilmiştir. Bu yöntem, birden fazla referans noktası kullanarak daha hassas bir karşılaştırma yaptığından, tek bir konum ölçüsünü göz önünde bulunduran geleneksel yaklaşımların yerine tercih edilebilir.

Anahtar kelimeler: NO kantil tahmin edicisi, iki bağımsız grup, kantil tahmin edicisi, asimptotik dağılım, yüzdelik bootstrap

CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER ONE
INTRODUCTION

One of the main objectives in applied statistics is to determine whether two independent groups differ, and if so, to understand how they do. It is necessary here to clarify exactly what is meant by "independent groups". If any of the observations in the first group is not related to anyone in the second group, two groups are said to be independent.

Detecting differences between two independent groups is frequently of interest in the fields such as psychology, medicine, environment, biology, marketing etc. Is there any significant difference between the cholesterol levels of the participants who take an experimental drug and a placebo? Is there any significant difference between the weight gains of rats that live in an ozone environment and an ozone-free environment? Is there any significant difference between the reading ability of children who watch thirty hours or more of television per week and watch ten hours or less? Is there any significant difference between the birth weights of newborns whose mothers smoke and do not smoke? Number of such research questions would be increased and if a researcher encounters any of them, firstly has to decide how to compare two independent groups (Wilcox, 2016).

A natural way of comparing two independent groups considers a single reference point, where it is the mean in most of the cases. Student's t test dates back to early 20'th century and it is the first method that a researcher tries to apply when working on two independent groups comparison. Although it has some advantages in particular cases, obtaining consistent results depends on whether the underlying assumptions are satisfied. Controlling the Type I error probability and providing a reasonable power are highly affected by violation of assumptions such as normality and homoscedasticity. However, unfortunately these assumptions are rarely met in real data. In fact, the populations that the samples are drawn from are never normally distributed, rather they can be asymmetric and/or heavy tailed. Together with, even a

single outlier can inflate the sample variance and correspondingly the results may be misleading anymore.

Momentarily, suppose the aim is to compare two independent groups by one reference point in a similar manner with the classical method. Due to the restrictive assumptions behind Student's t test, there is a need for more robust approaches. Welch's heteroscedastic methods for means deals with the violation of homoscedasticity assumption (Welch, 1938), whereas Yuen's method for trimmed means also allows non-normality (Yuen, 1974). Also, it is possible to consider median or an M measure of location as the reference point instead of mean or trimmed mean. To improve the control over Type I error probabilities, it is possible to use any of these reference points in conjunction with a bootstrap-t or a percentile bootstrap approach.

However, this thesis has a broader purpose which is to develop a different perspective for comparing two independent groups. Because, giving a chance to different perspectives will surely lead to get a deeper and more accurate insight. From a different point of view, a researcher may be interested in understanding how subpopulations of the groups compare. Among all subpopulations, most especially the lower and upper tails of the groups are of interest in many working areas. In applied statistics, the lower and upper tails of the groups correspond to lower and upper quantiles of the groups. Hence, the research question may have a more formal setting such as "Do the lower and upper quantiles of the two independent groups differ?".

To understand better, consider an experimental method which is being compared with a control group. In such a case, the experimental method may be efficient for the participants with low scores. That is, participants with low scores in the experimental group have higher results than participants with low scores in the control group. Conversely, the experimental method may be ineffective or detrimental for the participants with high scores. Namely, participants with high scores in the experimental group have lower results than participants with high scores in the control group. In brief, lower and upper tails of populations can respond to the experimental method differently, which means the utility of the experimental method may be related

to which quantiles are being compared. Furthermore, if an experimental method is expensive or invasive, knowing how different subpopulations compare might affect the strategy that a researcher will adopt when dealing with a particular problem.

Since it is thought that comparing two independent groups through the tails, namely the quantiles is more informative and consistent; firstly the attention is focused on the quantile estimation. A quantile estimator which is actually a mathematical formula, estimates the population quantile by using the observations in the sample. An important class of existing quantile estimators are investigated by considering both advantageous and disadvantageous features of them. Dielman, Lowry & Pfaffenberger (1994); Harrell & Davis (1982); Kaigh & Lachenbruch (1982); Parrish (1990); Sheather & Marron (1990) can be listed as examples for the most cited studies in the literature. Furthermore, Harrell & Davis (1982) is found to be nearly the best performing quantile estimator according to many papers. Yet still it has some inadequacies that can cause problems not only when estimating a population quantile but also when using it along with in some hypothesis testing procedure.

With the intention of making contribution to quantile estimation theory, a new quantile estimator is introduced in this thesis. The new estimator is desired to have the edge over available quantile estimators even with small sample sizes. This detail might be pretty important in certain studies, for instance a researcher has to work with very few patients when working on a medicine that can cause adverse effects.

Thereafter, following the idea of Wilcox (1995) and Wilcox, Erceg - Hurn, Clark & Carlson (2014) two independent groups are compared through quantiles by using the new quantile estimator along with a percentile bootstrap method. In addition to the newly proposed quantile estimator, fundamental finding of this thesis is to construct a method for comparing two independent groups through more than one quantiles simultaneously. This approach is based on data depth and bootstrap and it is appropriate to use it without any necessity of some underlying assumptions as in the classical approaches. On the other side, the new method can be used under situations that classic approaches already doing well.

This thesis is composed of five chapters. In the second chapter, the methods that compare two independent groups by one reference point are examined. The existing quantile estimators in literature are given and also a newly proposed quantile estimator NO is introduced in Chapter 3. In the fourth chapter, attention is focused on comparing two independent groups by using more than one reference point and for this purpose a new approach is defined. Finally, the last chapter includes the simulation studies that investigate the performances of the new quantile estimator and the new approach for comparing two independent groups. Additionally, some concluding remarks and recommendations are also given in the last chapter.

CHAPTER TWO
COMPARING GROUPS BY USING ONE REFERENCE POINT

The classic and most common way of comparing two independent groups is based on just one reference point such as arithmetic mean as in Student's t test. Although the mean is appropriate as a reference point for reasonably symmetric distributions, attention might be focused on a more robust measure of location when skewed distributions are of interest (Wilcox et al., 2014). Welch's heteroscedastic method for means, Yuen's method for trimmed means, bootstrap based methods that consider a single measure of location are some of the most known examples for robust alternatives of Student's t test (Welch, 1938; Yuen, 1974).

2.1 Student's T Test

The Student's t test is the traditional and most commonly known method for comparing two independent groups. To be more precise, it determines whether the two independent groups differ through a single reference point such as mean. Let X_{ij} be a random sample of size n_j, where $i = 1,...,n_j$ and $j = 1,2$. Besides, μ_j and σ_j^2 are the population mean and variance corresponding to j'th group, respectively. The goal is to test

$$H_0 : \mu_1 = \mu_2 \tag{2.1}$$

namely the hypothesis that the equality of two population means. In relation to that, another goal is to construct a confidence interval for $\mu_1 - \mu_2$, the difference between population means. Note that, Student's t test has some restrictive assumptions where the probability of Type I error rate could be controlled if the underlying assumptions are satisfied. The stated assumptions are listed below:

- The observations are randomly sampled for both groups.

- The observations are sampled from a normally distributed population for both groups.
- $\sigma_1^2 = \sigma_2^2$, that is the two groups have equal variances. This assumption is called homoscedasticity, where the reverse $\sigma_1^2 \neq \sigma_2^2$ corresponds to the case of heteroscedasticity.

Note that, Student's t test has some advantages and drawbacks in particular situations. If two independent groups follow an identical distribution, Student's t test can control the probability of Type I error even the distributions are non-normal. Because, even though the distributions are skewed, their difference is symmetric about zero. In addition to this when the distributions are identical, both their mean and variance will be the same (Wilcox, 2011). On the other side, consider two independent groups have the same sample size, they are sampled from a normal distribution but their variances are not equal. In such a situation, Student's t test performs reasonably well provided the sample sizes are large enough (Ramsey, 1980). However sampling from a normal distribution with unequal sample sizes, sampling from a non-normal distribution with equal or unequal sample sizes can cause problems in terms of controlling the probability of Type I error rate (Wilcox, 2017).

2.2 Dealing With Heteroscedasticity: Welch Test

It is already stated that, Student's t test cannot perform well when the underlying assumptions are violated. Welch (1938) proposed a method for testing the equality of population means that allows heteroscedasticity, but still needs normality assumption. The null hypothesis is the same as Equation (2.1) and the estimates of the squared standard errors of sample means $\bar{X}_1$ and $\bar{X}_2$ are given in Equation (2.2) for j=1, 2.

$$q_j = \frac{s_j^2}{n_j} \tag{2.2}$$

Here, s_j^2 and n_j represent the sample variances and sample sizes for the j'th group respectively. Welch (1938) estimates the distribution of corresponding test statistics W as a Student's t distribution with degrees of freedom υ.

$$W = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \tag{2.3}$$

Namely, υ is estimated with $\hat{\upsilon}$ in Equation (2.4), where it is evaluated by using sample variances and sample sizes of both groups.

$$\hat{\upsilon} = \frac{\left(q_1 + q_2\right)^2}{\frac{q_1^2}{n_1 - 1} + \frac{q_2^2}{n_2 - 1}} \tag{2.4}$$

If $|W| \geq t$, the null hypothesis of equal means is rejected where t is the $1-\alpha/2$ quantile of the Student's t distribution with $\hat{\upsilon}$ degrees of freedom. Following this, a $1-\alpha$ confidence interval for the difference between population means $\mu_1 - \mu_2$ is obtained as

$$\left(\bar{X}_1 - \bar{X}_2\right) \pm t\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}\,. \tag{2.5}$$

It should not be forgotten that, Welch method cannot control the probability of Type I error rate when the groups are differing in skewness. Moreover, when there are outliers among observations Welch's test produce misleading results. That is, when sample mean is considered as the reference point for detecting differences between two independent groups, neither Student's t test nor Welch test is appropriate in the presence of outliers.

2.3 Yuen's Method

Another method for comparing two independent groups by using one reference point is derived by Yuen (1974). Yuen's method tests the equality of population trimmed means.

$$H_0 : \mu_{t1} = \mu_{t2} \tag{2.6}$$

It can also be of interest to compute a $1-\alpha$ confidence interval for $\mu_{t1} - \mu_{t2}$, the difference between two population trimmed means. In advance of explaining the method, the definitions of sample trimmed mean, sample Winsorized mean and sample Winsorized variance are given.

Let $X_1, ..., X_n$ be a random sample and $X_{(1)}, ..., X_{(n)}$ be the corresponding order statistics of the sample, where the observations are written in ascending order as $X_{(1)} \leq ... \leq X_{(n)}$. The amount of trimming γ is determined for $0 \leq \gamma < 0.5$. Note the trimming proportion γ is chosen as 20%, generally. Let $g = [\gamma n]$, where g is the integer part of γn. After obtaining the value g, the g largest and g smallest observations are removed from the sample and the sample trimmed mean $\bar{X}_t$ is calculated on the basis of n-2g remaining observations.

$$\bar{X}_t = \frac{X_{(g+1)} + ... + X_{(n-g)}}{n - 2g} \tag{2.7}$$

Additionally, to get the squared standard error of $\bar{X}_{tj}$, there is a need for defining the sample Winsorized variance. For this purpose, the Winsorization of a random sample is demonstrated in Equation (2.8). Instead of removing the smallest and the largest g observations from the sample, Winsorization changes the smallest g observations with $X_{(g+1)}$ whereas changes the largest g observations with $X_{(n-g)}$.

$$W_i = \begin{cases} X_{(g+1)}, & \text{if } X_{(i)} \le X_{(g+1)} \\ X_{(i)}, & \text{if } X_{(g+1)} < X_{(i)} < X_{(n-g)} \\ X_{(n-g)}, & \text{if } X_{(i)} \ge X_{(n-g)} \end{cases} \tag{2.8}$$

The Winsorized sample mean and Winsorized sample variance are given in Equation (2.9) and (2.10), respectively.

$$\bar{X}_w = \frac{1}{n}\sum_{i=1}^{n} W_i \tag{2.9}$$

$$s_w^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(W_i - \bar{X}_w\right)^2 \tag{2.10}$$

The explanation of Yuen's method starts with determining $g_j = [\gamma n_j]$ for j=1, 2. Let $h_j = n_j - 2g_j$ be the number of observations that remain after trimming. Also, let $\bar{X}_{tj}$ and s_{wj}^2 denote the sample trimmed mean and the Winsorized sample variance, respectively. By means of s_{wj}^2 and h_j, the estimated squared standard error of the sample trimmed mean is obtained as in Equation (2.11).

$$d_j = \frac{\left(n_j - 1\right)s_{wj}^2}{h_j\left(1 - h_j\right)} \tag{2.11}$$

Yuen's test statistics T_y has an approximate Student's t distribution with estimated degrees of freedom $\hat{\upsilon}_y$.

$$T_y = \frac{\bar{X}_{t1} - \bar{X}_{t2}}{\sqrt{d_1 + d_2}} \tag{2.12}$$

$$\hat{\upsilon}_y = \frac{\left(d_1 + d_2\right)^2}{\frac{d_1^2}{h_1 - 1} + \frac{d_2^2}{h_2 - 1}} \tag{2.13}$$

The null hypothesis in Equation (2.6) is rejected if $|T_y| \geq t$, where t is the $1-\alpha/2$ quantile of the Student's t distribution with $\hat{\upsilon}_y$ degrees of freedom. A corresponding $1-\alpha$ confidence interval for $\mu_{t1} - \mu_{t2}$ is obtained as

$$\left(\bar{X}_{t1} - \bar{X}_{t2}\right) \pm t\sqrt{d_1^2 + d_2^2}\,. \tag{2.14}$$

It is also noted that, Yuen test allows heteroscedasticity and when the trimming proportion $\gamma = 0$, this method will be equivalent to Welch's heteroscedastic method for means. However, when compared to Welch (1938) method, Yuen's method with 20% trimming has some superiorities in particular situations. Yuen's method not only provides better control over the probability of Type I error rate but also improves power (Wilcox, 2011).

2.4 Comparing Medians

Since the median is insensitive to outliers, a possible strategy is to compare two independent groups through medians. Here a method based on the McKean Schrader estimate of the standard errors is given which performs reasonably well in terms of controlling the probability of Type I error rate provided that there are no tied values.

A random sample of size n is putted in ascending order as in $X_{(1)} \leq \ldots \leq X_{(n)}$. The McKean-Schrader estimate of the squared standard error of sample median M is given as

$$\left(\frac{X_{(n-k+1)} - X_{(k)}}{2z_{0.995}}\right)^2 \tag{2.15}$$

where $z_{0.995}$ is the 0.995 quantile of a standard normal distribution and the value of k is calculated as in Equation (2.16)

$$k = \frac{n+1}{2} - z_{0.995}\sqrt{\frac{n}{4}} \tag{2.16}$$

Let M_1 and M_2 be the sample medians of two independent groups where the corresponding McKean Schrader estimates of squared standard errors are represented as S_1^2 and S_2^2. Then, the null hypothesis of equal population medians is rejected if,

$$\frac{|M_1 - M_2|}{\sqrt{S_1^2 + S_1^2}} \geq c \tag{2.17}$$

where c is the $1-\alpha/2$ quantile of a standard normal distribution. Furthermore, an approximate $1-\alpha$ confidence interval for the difference between population medians is obtained as

$$(M_1 - M_2) \pm c\sqrt{S_1^2 + S_1^2}\,. \tag{2.18}$$

Note that, if there are tied values in any of the groups the probability of Type I error rate cannot be controlled. Because, in the presence of tied values the estimates of the standard errors of sample medians are inaccurate and the sampling distribution does not converge a normal distribution as the sample size increases (Wilcox, 2017).

2.5 A Bootstrap-t Method for Comparing Trimmed Means

Although Yuen method with 20% trimming performs better in terms of controlling the probability of Type I error in comparison with no trimming, problems might occur even when 20% trimming is considered. For instance, when the sample sizes are unequal and a one – sided test is conducted, the actual probability of Type I error may exceed the nominal level. In such situation, a bootstrap-t method can give more appropriate results.

Basically a bootstrap-t method tries to estimate the lower and the upper critical values of the test statistics T_y by generating bootstrap samples. More clearly, let $X_{1j}^*,...,X_{n_j j}^*$ be a bootstrap sample corresponding to j'th group for j=1, 2. Firstly, the empirical distributions of two independent groups are shifted by setting $C_{ij}^* = X_{ij}^* - \bar{X}_{tj}$ so that they have identical trimmed means, where $i = 1,...,n_j$. In other words, C_{ij}^* corresponds to a sample that comes from a population with trimmed mean equals zero. Clearly, the null hypothesis that the equality of two population trimmed means will be true for C_{ij}^* values. The remaining steps of bootstrap-t procedure are listed as follows:

- Yuen method is applied to C_{ij}^* values, that is the test statistics T_y which are obtained on the basis of C_{ij}^* are represented as T_y^*.
- The previous step is repeated B times yielding T_{yb}^* for b=1, 2, …, B.
- Put the T_{yb}^* values in ascending order, namely $T_{y(1)}^* \le ... \le T_{y(B)}^*$.
- Let $\ell = \alpha B/2$ where ℓ is rounded to nearest integer and $u = B - \ell$.
- The estimated lower and upper critical values of T_y are obtained as $T_{y(\ell+1)}^*$ and $T_{y(u)}^*$.

If $T_y < T_{y(\ell+1)}^*$ or $T_y > T_{y(u)}^*$, then $H_0 : \mu_{t1} = \mu_{t2}$ is rejected. Furthermore, a corresponding $1-\alpha$ confidence interval is obtained as

$$\left(\bar{X}_{t1} - \bar{X}_{t2} - T_{y(u)}^* \sqrt{d_1 + d_2}, \bar{X}_{t1} - \bar{X}_{t2} - T_{y(\ell+1)}^* \sqrt{d_1 + d_2}\right) \qquad (2.19)$$

where d_j is the estimate of the squared standard error of $\bar{X}_{tj}$ given in Equation (2.11).

Note that when testing at $\alpha = 0.05$ level, B=599 is sufficient in terms of controlling the probability of Type I error rates, whereas B=999 might be better in terms of power. There are some cases that using 10% or 15% trimming is satisfactory for bootstrap-t, but with small and unequal sample sizes the method cannot perform well with 10%

trimming. Following this, sometimes using 20% trimming is advantageous in terms of controlling the probability of Type I error rates (Wilcox, 2017).

2.6 Inferences Based on a Percentile Bootstrap Method

When comparing two independent groups, a percentile bootstrap method can be applied to any measure of location. But yet, the percentile bootstrap method is not recommended to compare means (Wilcox, 2011). In opposition to bootstrap-t method, percentile bootstrap does not need to estimate the standard error of the estimator of interest. The lower and the upper limits of the confidence interval are obtained by means of the bootstrap distribution of the related parameter. The aim is to test

$$H_0 : \theta_1 = \theta_2 \tag{2.20}$$

where θ_j is any parameter of interest for the j'th group with j=1,2. The percentile bootstrap approach for testing the null hypothesis in Equation (2.20) is as follows:

- Let $\hat{\theta}_j^*$ be the bootstrap estimate of θ_j, and define $D^* = \hat{\theta}_1^* - \hat{\theta}_2^*$.
- Repeat the first step B times yielding $D_1^*, ..., D_B^*$, for b=1, 2, …, B.
- Put the D^* values in ascending order yielding $D_{(1)}^* \leq ... \leq D_{(B)}^*$.
- Let $\ell = \alpha B / 2$ where ℓ is rounded to nearest integer and $u = B - \ell$.
- The approximate $1-\alpha$ confidence interval for $\theta_1 - \theta_2$ is obtained as

$$\left(D_{(\ell+1)}^*, D_{(u)}^* \right) \tag{2.21}$$

Additionally, define $p^* = P\left(\hat{\theta}_1^* > \hat{\theta}_2^* \right)$ for the bootstrap estimates $\hat{\theta}_1^*$ and $\hat{\theta}_2^*$. To estimate p^*, let A be the number of times that $D^* > 0$. Then an estimate of p^* is obtained as

$$\hat{p}^* = \frac{A}{B} \tag{2.22}$$

where a generalized p-value is estimated by using p^* as in Equation (2.23):

$$\text{a generalized p-value} = 2\min\left(p^*, 1-p^*\right) \tag{2.23}$$

Clearly, if $\text{p-value} < \alpha$, H_0 in Equation (2.20) is rejected.

Thus far, the percentile method is summarized for a general parameter of interest θ. When the goal is to compare two independent groups by using one reference point with a percentile bootstrap based method, the parameter θ can be of any M-measures of location, trimmed mean and median. If the M-measures of locations are compared, the percentile bootstrap is the best performing approach. For comparing trimmed means with at least 20% trimming, percentile bootstrap is more preferable than bootstrap-t.

On the other hand, for comparing medians by a percentile bootstrap approach in the presence of tied values, a modification is needed. Namely, let M_1^* and M_2^* be the bootstrap sample medians for two groups. Let A be the number of cases that $M_1^* > M_2^*$, C be the number of cases that $M_1^* = M_2^*$ and B be the number of bootstrap samples. Define p^* as

$$p^* = P\left(M_1^* > M_2^*\right) + 0.5P\left(M_1^* = M_2^*\right) \tag{2.24}$$

and it is estimated with

$$\hat{p}^* = \frac{A}{B} + 0.5\frac{C}{B}. \tag{2.25}$$

The corresponding p-value is $2\min\left(\hat{p}^*, 1-\hat{p}^*\right)$ (Wilcox, 2017).

CHAPTER THREE
ESTIMATION OF POPULATION QUANTILE

Quantiles have a wide range of working areas in both theoretical and applied statistics and accordingly using an appropriate quantile estimator is an important issue. Let X be a continuous random variable having the cumulative distribution function F(x). The q'th quantile of a population is represented as $Q(q)$ and defined in terms of the functional inverse of the cumulative distribution function obtained at q, namely $Q(q) = F^{-1}(q) = \inf\{x : F(x) \geq q\}$ for predetermined $0<q<1$. This statement can be also given as $P(X \leq Q(q)) = q$, which means the 100q percent of the observations are less than or equal to the population quantile Q(q). For instance, if X has a standard normal distribution, its 0.95'th quantile is 1.645 where it is represented as $Q(0.95) = P(X \leq 1.645) = 0.95$.

A quantile estimator is a formula that estimates the quantile of a distribution by using a sample. There are three main approaches for obtaining a quantile estimator: using a single order statistic, taking the weighted average of two order statistics and taking the weighted average of all order statistics. Although second approach has a little advantage relative to the first one, these two approaches are not desired because using one or two order statistics cause a larger variance and consequently lower efficiency. Using a weighted average of all order statistics is a better idea which improves efficiency.

3.1 Traditional Quantile Estimators

The quantile estimators that are based on a single order statistics and that use a weighted average of two order statistics are considered as traditional or conventional quantile estimators. In a study which compares ten nonparametric quantile estimators, Parrish (1990) expressed the following traditional quantile estimators as well. Note that, these definitions don't depend on the underlying population distribution in which the observations are sampled from.

For any random sample of size n, let $X_{(1)} \le X_{(2)} \le ... \le X_{(n)}$ denote the order statistics and q is the quantile value to be estimated where $0<q<1$. In the below definitions, the j, g, h, k and i values are used in order to simplify the notation. Here, nq=j+g with j=[nq], that is j corresponds to integral part of nq whereas g corresponds to the fractional part. In a similar manner, k=[(n+1)q] and h=(n+1)q-k are evaluated.

QE_1, QE_2, QE_3 and QE_4 are the quantile estimators that are obtained by using a single order statistic. The empirical CDF can be defined as $F_n(X) = i/n$, where $X_{(i)} \le X \le X_{(i+1)}$. Following this, QE_1 and QE_2 are based on the infimum and supremum of the interval related to empirical CDF, respectively.

$$QE_1 = X_{(j)} \tag{3.1}$$

$$QE_2 = X_{(j+1)} \tag{3.2}$$

QE_3 is also associated with empirical CDF, while QE_4 is based on the observation numbered closest to nq.

$$QE_3 = \begin{cases} X_{(j)} & ,\text{if } g=0 \\ X_{(j+1)}, & \text{if } g>0 \end{cases} \tag{3.3}$$

$$QE_4 = \begin{cases} X_{(j)} & ,\text{if } g<0.5 \\ X_{(j+1)}, & \text{if } g \ge 0.5 \end{cases} \tag{3.4}$$

On the other side, QE_5, QE_6, QE_7 and QE_8 are the weighted averages of two consecutive order statistics. Note that, in Equation (3.8), i=[nq+0.5].

$$QE_5 = (1-g)X_{(j)} + gX_{(j+1)} \tag{3.5}$$

$$QE_6 = (1-h)X_{(k)} + hX_{(k+1)} \tag{3.6}$$

$$QE_7 = \begin{cases} (X_{(j)} + X_{(j+1)})/2 & ,\text{if } g=0 \\ X_{(j+1)} & ,\text{if } g>0 \end{cases} \tag{3.7}$$

$$QE_8 = (0.5 + i - nq)X_{(i)} + (0.5 - i + nq)X_{(i+1)} \quad (3.8)$$

In addition to traditional quantile estimators considered by Parrish (1990), Kaigh & Lachenbruch (1982) use $QE_9 = X_{(k)}$.

Hyndman & Fan (1996) compared the sample quantile definitions in common statistical computer packages by writing them in an identical notation, analyzing their motivation and investigating some properties of them. Their study is restricted with the quantile estimators based on one or two order statistics. Namely, the quantile estimators have the form

$$QE = (1-\gamma)X_{(j)} + \gamma X_{(j+1)} \quad (3.9)$$

where γ is a function of j=[nq+m] and g=nq+m-j. They draw attention to the difficulty of comparing definitions since there are many equivalent ways of defining quantiles in various packages. They proposed the median-unbiased estimator which corresponds to Definition 8, to avoid confusion between statistical packages. This estimator uses $m = \frac{q+1}{3}$ and $\gamma = g$ in order to obtain a definition as in the Equation (3.9).

Note that, the default quantile estimator of R programming language is given in Definition 7 of Hyndman & Fan (1996). By substituting m=1-q and $\gamma = g$ in Equation (3.9), the corresponding quantile estimator is obtained and can be denoted as R_q.

3.2 Using a Weighted Average of All Order Statistics

When the purpose is to get a quantile estimator for estimating the population quantile, the reasonable approach is using a weighted average of all order statistics. Here, the motivation is reducing the variance and accordingly improving efficiency.

There are a great number of ways for obtaining weights for order statistics in the literature that leads to get different quantile estimators. Of course, the existing quantile estimators have some advantageous and disadvantageous properties in particular situations. The studies both defining an approach for obtaining a quantile estimator and comparing the performance of the existing quantile estimators are investigated in this part of the thesis.

In 1982, Harrell & Davis derived a distribution-free quantile estimator which uses all order statistics to estimate a population quantile. They also calculate a jackknife variance estimator and stated that their estimator is more efficient than the traditional one that is based on one or two order statistics. For instance, they considered QE_6 in Equation (3.6) which uses a weighted average of two consecutive order statistics and Harrell Davis quantile estimator is generally 1.1 times more efficient than QE_6. However, using Harrell Davis quantile estimator for small sample sizes and extreme quantiles is not recommended (Harrell & Davis, 1982).

To be more precise, let Y be a random variable that follows a beta distribution with parameters a=(n+1)q and b=(n+1)(1-q). The probability density function of Y is

$$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} y^{a-1}(1-y)^{b-1} \quad , 0 \le y \le 1 \tag{3.10}$$

where Γ is the gamma function. Given that,

$$W_i = P(\frac{i-1}{n} \le Y \le \frac{i}{n}) \tag{3.11}$$

the Harrell Davis estimate of q'th quantile is obtained as

$$HD_q = \sum_{i=1}^{n} W_i X_{(i)} \tag{3.12}$$

(Harrell & Davis, 1982).

Among the quantiles, estimation of the population median is of frequent interest in many working areas. The classical median estimator is defined as

$$m = \begin{cases} X_{((n+1)/2)} & ,\text{if n is odd} \\ \dfrac{X_{(n/2)} + X_{(n/2+1)}}{2} & ,\text{if n is even} \end{cases} \tag{3.13}$$

for a random sample having length n (Yoshizawa, Sen & Davis, 1985). There are various alternatives for the classical median estimator which are weighted averages of order statistics. For example, by substituting q=0.5 in the formula of HD_q, Harrell Davis median estimator $\hat{\theta}_{HD}$ is obtained. Yoshizawa et al. (1985) represented the asymptotic equivalence of Harrell Davis median estimator and the classical median estimator. By means of asymptotic equivalence of these estimators, it is possible to apply asymptotic properties of classical median estimator to the Harrell Davis median estimator. More clearly, $\hat{\theta}_{HD}$ has an asymptotically normal distribution. Furthermore, $\hat{\theta}_{HD}$ is almost 1.1 times as efficient as the classical median estimator (Harrell & Davis, 1982). According to Yoshizawa et al. (1985), $\hat{\theta}_{HD}$ is a better choice for small and medium sample sizes, however the classical median estimator can be used in large samples due to its easy calculation and asymptotic equivalence of the stated estimators.

Stewart (1989), expanded the Harrell Davis quantile estimator to the case of sampling from a finite population by considering both sampling with and without replacement. Consider a finite population of size N and a random sample of size n. A simulation study with various N, n, q values is conducted to examine the relative efficiency of Harrell Davis estimator over the conventional quantile estimator QE_6. For the case of sampling with replacement, Harrell Davis estimator performs well for the median, but when extreme quantiles are of interest the results are unsatisfactory. On the other hand, when sampling without replacement is the case, it is seen that an increasing number of the order statistics are getting useless as n goes to N. Hence, the order statistics are needed to be trimmed where the trimming proportion depends on the values of N, n and q. Let $k_{min} = \max(1,[Nq]-N+n)$, $k_{max} = \min(n,[Nq])$ and

$n^* = 1 + k_{max} - k_{min}$. After trimming, the remaining subset of the order statistics is obtained as $X^* = X_{(k_{min})} \leq ... \leq X_{(k_{max})}$. The modified Harrell Davis estimator HD_q^* is achieved based on the sample X^*. By this means the bias can be reduced without inflating the variance, and again HD_q^* is more efficient than QE_6.

While in the same year with Harrell & Davis , Kaigh & Lachenbruch (1982) derived an alternative to conventional quantile estimator which is found by averaging a proper subsample quantile over all possible $\binom{n}{m}$ subsamples of a fixed size m, where $1 \leq m \leq n$. The proposed estimator in Equation (3.14) uses all order statistics and the weights involve an arbitrary sample size whose optimal value depends on the sample size and the quantile to be estimated (Kaigh & Lanchenbruch, 1982).

$$KL_q = \sum_{s=u}^{u+n-m} \binom{s-1}{u-1}\binom{n-s}{m-u}\binom{n}{m}^{-1} X_{(s)} \qquad (3.14)$$

where $u = [(m+1)q]$. Moreover, results of their study show KL_q has a smaller mean squared error when compared to conventional quantile estimator QE_9, under various distributions with sufficient sample sizes.

Steinberg & Davis (1985) stated both Harrell Davis and Kaigh Lanchenbruch estimators are at least as efficient as a conventional quantile estimator based on one or two order statistics. The attention of their work is concentrated on the confidence interval estimation for quantiles. Harrell Davis and Kaigh Lanchenbruch estimators can be used to construct confidence intervals for quantiles in large samples, since both are linear combinations of order statistics and follow an asymptotically normal distribution. They compared the confidence intervals based on different approaches in terms of ability to get desired confidence level. According to Steinberg & Davis (1985), the confidence interval that is obtained via Harrell Davis quantile estimator cannot save the stated confidence level in small samples.

As a matter of fact, the quantile estimators that take a weighted average of all the order statistics are distinguished by their weight functions. Kaigh & Cheng (1991), stated that the weight functions can be obtained either by subsampling or by a kernel approach, where Harrell Davis and Kaigh Lachenbruch quantile estimators are important examples of subsampling considerations. Kaigh & Cheng (1991) followed the subsampling approach and they derived a new statistic which is closely associated with Harrell Davis and Kaigh Lachenbruch estimators. By determining $1 \le k \le n$, their new estimator is obtained as

$$Q_{r:k;n} = \sum_{j=1}^{n} [\binom{r+j-2}{r-1}\binom{n-j+k-r}{k-r}\binom{n+k-1}{k}]X_{(j)} \qquad (3.15)$$

where $r = [(k+1)q]$. $Q_{r:k;n}$ was found to be successful in efficiency comparisons (Kaigh & Cheng, 1991).

Parrish (1990) compared the performances of ten nonparametric quantile estimators in terms of bias and mean squared error. The comparison involves the quantile estimators that are based on one or two order statistics given in Equation (3.1) – (3.8), Harrell Davis estimator and Kaigh Lachenbruch estimator. Their performance characteristics are investigated under the restriction of sampling from a normal distribution. In general, Harrell Davis estimator is the best one in terms of mean squared error whereas QE_8 in Equation (3.8) is the minimum biased one.

Furthermore; Dielman et al. (1994) widened Parrish's work for non-normal distributions. They used symmetric light tailed, symmetric heavy tailed and skewed distributions where uniform, Laplace and lognormal distributions are examples for each of them respectively. The performance criteria involve mean squared error, mean absolute deviation and bias. Harrell Davis estimator performs best in many cases, but in extreme quantiles there are some exceptions.

One of the most current quantile estimators were derived by Sfakianakis & Verginis (2008). They used weighted averages of all order statistics and aimed to get efficient

estimators with small samples. Although Harrell Davis estimator was nearly the best estimator according to literature, Sfakianakis and Verginis estimators outperform Harrell Davis and conventional quantile estimators in some cases studied. Their quantile estimators are denoted as $SV1_q, SV2_q$ and $SV3_q$ where B(i; n, q) are the binomial probabilities for i=0, 1, …,n.

$$\begin{aligned} SV1_q = {} & \frac{2B(0;n,q)+B(1;n,q)}{2}X_{(1)} + \frac{B(0;n,q)}{2}X_{(2)} - \frac{B(0;n,q)}{2}X_{(3)} \\ & + \sum_{i=2}^{n-1}\frac{B(i;n,q)+B(i-1;n,q)}{2}X_{(i)} - \frac{B(n;n,q)}{2}X_{(n-2)} \\ & + \frac{B(n;n,q)}{2}X_{(n-1)} + \frac{2B(n;n,q)+B(n-1;n,q)}{2}X_{(n)} \end{aligned} \tag{3.16}$$

$$SV2_q = \sum_{i=0}^{n-1} B(i;n,q)X_{(i)} + (2X_{(n)} + X_{(n-1)})B(n;n,q) \tag{3.17}$$

$$SV3_q = \sum_{i=1}^{n} B(i;n,q)X_{(i)} + (2X_{(1)} + X_{(2)})B(0;n,q) \tag{3.18}$$

3.2.1 Kernel Quantile Estimators

As stated previously, the quantile estimators distinguished by their weight functions where the weight functions can be formed either by subsampling consideration or by a kernel approach. The papers on kernel approach are briefly explained, such as Azzalini (1981), Cune & Cune (1991), Falk (1985), Kenani & Yu (2012), Parzen (1979), Sheather & Marron (1990), Yang (1985) and Zelterman (1990). Before literature review, some information about kernel approach is given in general terms.

When the aim is estimating f(x), the probability density function, a reasonable approach is to make use of kernel estimation theory. Kernel estimation theory points out a weight function known as kernel function K(x). To detail, K is some probability density function with the properties $\int_{-\infty}^{\infty} K(x)dx = 1$, symmetric about zero and $K(x) \geq 0$. A kernel density estimator is defined as

$$\hat{f}_n(x;h) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{x - X_i}{h}) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - X_i) \tag{3.19}$$

where $K_h(x) = h^{-1}K(x/h)$ and the bandwidth $h \to 0 \text{ as } n \to \infty$ (Efromovich, 2008; Wilcox, 2017).

Some commonly used kernel functions are listed below, where $\mathbf{1}_{\{...\}}$ is the indicator function.

- Uniform kernel: $K(u) = \frac{1}{2}\mathbf{1}_{\{|u|\le 1\}}$
- Triangular kernel: $K(u) = (1 - |u|)\mathbf{1}_{\{|u|\le 1\}}$
- Epanechnikov kernel: $K(u) = \frac{3}{4}(1 - u^2)\mathbf{1}_{\{|u|\le 1\}}$
- Gaussian kernel: $K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}\mathbf{1}_{\{|u|\le 1\}}$

According to Parzen (1979), both exploratory data analysis and conventional parametric statistical inference should aim to estimate the quantile function $Q(q) = F^{-1}(q) = \inf\{x : F(x) \ge q\}$ where F(x) is the distribution function and 0<q<1. The main purpose of his study is to present a density estimation approach that also allows estimation of Q. The nonparametric approach estimates Q by using a kernel estimator with an appropriate kernel function K and bandwidth h. It is also noted that the kernel approach may be practically difficult since it is important to choose optimal h.

Let $X_1, X_2, ..., X_n$ be independent and identically distributed random variables. It is possible to estimate a distribution function by means of integrating a kernel estimator of the density function f(x). Following this, Azzalini (1981) estimated quantiles by inverting the estimate of the distribution function. The classical kernel estimator of the density function is obtained as

$$\hat{f}(x) = \frac{1}{nb}\sum_{i=1}^{n} w\left(\frac{x - X_i}{b}\right) \tag{3.20}$$

where w is some bounded density function with w(t)=w(-t) for all real t, and b>0. The corresponding estimate of the distribution function F(x) is obtained as

$$\hat{F}(x) = \frac{1}{n}\sum_{i=1}^{n} W\left(\frac{x - X_i}{b}\right) \tag{3.21}$$

where $W(t) = \int_{-\infty}^{t} w(u)du$. Clearly, inverting $\hat{F}(x)$ enables estimating the quantiles.

In the later years, Yang (1985) proposed a nonparametric kernel quantile estimator which follows the same asymptotic distribution with the traditional quantile estimator QE1. The proposed quantile estimator $S_n(q)$ is given in Equation (3.22) where, K is some probability distribution function and $h(n) \to 0$ as $n \to \infty$.

$$S_n(q) = \frac{1}{n}\sum_{i=1}^{n} X_{(i)} h(n)^{-1} K\left[\frac{i/n - q}{h(n)}\right] \tag{3.22}$$

Actually, $S_n(q)$ weights the $X_{(i)}$ for that i/n is close to q more heavily than $X_{(i)}$ for that i/n is far from q. $S_n(q)$ is motivated from the kernel estimator $T_n(q)$, where it is in the form of a weighted average of all order statistics. Yang (1985), showed that $S_n(q)$ and $T_n(q)$ are asymptotically equivalent in mean squares.

$$\begin{aligned} T_n(q) &= \int_0^1 F_n^{-1}(t) h^{-1}(n) K\left(\frac{t-h}{h(n)}\right) dt \\ &= \sum_{i=1}^{n} X_{(i)} h^{-1}(n) \int_{(i-1)/n}^{i/n} K\left(\frac{t-h}{h(n)}\right) dt \end{aligned} \tag{3.23}$$

Besides, $T_n(q)$ appeared in Falk (1984), Parzen (1979) and Reiss (1980). Reiss (1980), studied the asymptotic relative deficiency (Hodges & Lehmann, 1970) of the conventional quantile estimator relative to $T_n(q)$. On the other side, Reiss (1980) used $T_n(q)$ for testing hypothesis related with population quantiles.

Yang (1985) investigated the asymptotic normality and consistency of the estimators $S_n(q)$ and $T_n(q)$. Furthermore, via a simulation study the QE_1, $S_n(q)$, $T_n(q)$ and Kaigh Lanchenbruch estimator KL_q are compared. The weight functions for $S_n(q)$ and $T_n(q)$ are chosen as the triangular density function. Furthermore, an estimate for the optimal bandwidth h(n) is proposed that minimizes the bootstrap mean squared error estimate of $S_n(q)$.

Based on the fairly general assumptions about distribution function and kernel, Falk (1985) proved the asymptotic normality of kernel quantile estimators of the form in Equation (3.23).

Sheather & Marron (1990) emphasize choosing the weight function when obtaining a quantile estimator as weighted average of all order statistics, since using a weighted average of order statistics leads to improve efficiency. As is known, kernel quantile estimators are an important class of those type estimators.

Let K is a density function symmetric about zero such that, $K_h(\cdot) = h^{-1}K(\cdot/h)$ where the bandwidth $h \to 0$ as $n \to \infty$. One example of kernel quantile estimators is obtained by

$$KQ_q = \sum_{i=1}^{n}\left[\int_{i-1/n}^{i/n} K_h(t-q)dt\right]X_{(i)} \qquad (3.24)$$

as in Falk (1984), Parzen (1979), Reiss (1980) and Yang (1985). Furthermore, there are different approximations to KQ_q where they are denoted as $KQ1_q$, $KQ2_q$, $KQ3_q$ and $KQ4_q$.

$$KQ1_q = \sum_{i=1}^{n} [n^{-1} K_h(\frac{i}{n} - q)] X_{(i)} \tag{3.25}$$

$$KQ2_q = \sum_{i=1}^{n} [n^{-1} K_h((i - \frac{1}{2})/n - q)] X_{(i)} \tag{3.26}$$

$$KQ3_q = \sum_{i=1}^{n} [n^{-1} K_h(\frac{i}{n+1} - q)] X_{(i)} \tag{3.27}$$

However, the weights of the estimators $KQ1_q$, $KQ2_q$ and $KQ3_q$ do not sum to 1. Hence to standardize their weights, the weights are divided by their sum. Following this approach, $KQ2_q$ becomes $KQ4_q$.

$$KQ4_q = \sum_{i=1}^{n} [n^{-1} K_h(\frac{i-1/2}{n} - q)] X_{(i)} \Big/ \sum_{j=1}^{n} K_h(\frac{j-1/2}{n} - q)] \tag{3.28}$$

Sheather & Marron (1990) stated the asymptotic equivalence between the kernel quantile estimators KQ_q, $KQ1_q$, $KQ2_q$, $KQ3_q$ and $KQ4_q$. They proposed a data based bandwidth selection method on the basis of minimizing the asymptotic mean squared error of KQ_q and data based bandwidths for the kernel quantile estimators are compared. Besides, the performance of the kernel quantile estimators with Harrell Davis estimator and Kaigh Lachenbruch estimator is compared by conducting a simulation study.

Yet another finding of Sheather & Marron (1990) is on the relationship between the kernel quantile estimators with Harrell Davis and Kaigh Lachenbruch estimators. In large samples, HD_q is equivalent to KQ_q with the bandwidth in Equation (3.29) whereas KL_q is equivalent to $KQ1_q$ with the bandwidth in Equation (3.30) when standard normal density is chosen as the kernel function.

$$h = [q(1-q)/(n+1)]^{1/2} \tag{3.29}$$

$$h = [q(1-q)/k]^{1/2} \tag{3.30}$$

where $i/n = q + O(k^{-1/2})$ and $k = o(n)$.

Zelterman (1990) proposed another smooth kernel estimator $R_n(q)$ for nonparametric estimation of a population.

$$R_n(q) = \sum_{i=1}^{n} r_{in} X_{(i)} \tag{3.31}$$

where the weights are given as

$$r_{in} = r_{in}(q) = [n\, g(n,q)]^{-1} K_N\left(\frac{i - nq - 1/2}{n\, g(n,q)}\right) \tag{3.32}$$

with the normal kernel function $K_N(u) = (2\pi)^{-1/2} \exp(-\frac{1}{2}u^2)$ for $-\infty < u < \infty$ and bandwidth function $g(n,q) = [q(1-q)/n]^{1/2}$. The weights in Equation (3.32) do not sum to 1, so standardized weights can be used instead. This situation has a little effect in the middle of the distribution, but provides improvement in the tails.

Zelterman (1990) additionally investigated the asymptotic equivalence of $R_n(q)$ with Harrell Davis and Kaigh Lanchenbruch estimators. Namely, when n is large enough the weights r_{in} are approximately the same as the weights of Harrell Davis and Kaigh Lanchenbruch estimators provided q is not too close to zero or one.

Another estimator is obtained by jackknifing the kernel quantile estimator (Cune & Cune, 1991). Suppose that $KQ_q(K, h_1)$ and $KQ_q(K, h_2)$ are two kernel quantile estimators which use the same kernel function K with different bandwidths h_1 and h_2 where KQ_q is the kernel quantile estimator given in Equation (3.24). Then, for some

constant $R \neq 1$, Cune & Cune (1991) proposed the generalized jackknife estimator as below.

$$G\left(Q_n(q;K,h_1),Q_n(q;K,h_2)\right)=\frac{Q_n(q;K,h_1)-RQ_n(q;K,h_2)}{1-R} \tag{3.33}$$

Besides, they considered the asymptotic relative deficiency of the kernel quantile estimator relative to the jackknifed quantile estimator. As a result of that study, using the jackknife estimator reduces the bias and leads an asymptotically less deficient estimator (Cune & Cune, 1991; Hodges & Lehmann, 1970).

Another study on kernel quantile estimators is more recent dated than others. As the estimation of population quantiles is of great interest, determining the bandwidth parameter in kernel quantile estimators is also very important. Kenani & Yu (2012), proposed a cross validation method which determines the optimal bandwidth parameter. A numerical study is performed in order to compare the performance of their bandwidth selection method with the method propose by Sheather & Marron (1990). Throughout the numerical study, Gaussian kernel is chosen as the kernel function. Comparison criteria are mean squared error and relative efficiency. According to the results of their study, the proposed bandwidth selection method provides accurate estimates, plus the new method is asymptotically unbiased as well. (Kenani & Yu, 2012).

3.2.2 Quantile Estimators Using Bernstein Polynomials

Bernstein polynomials defined explicitly by $B_i^n(t)=\binom{n}{i}t^i(1-t)^{n-i}$ where the binomial coefficients are given by Equation (3.34).

$$\binom{n}{i}=\begin{cases}\dfrac{n!}{i!(n-i)!} & \text{if } 0\le i\le n\\ 0 & \text{else}\end{cases} \tag{3.34}$$

Some important properties of these polynomials can be listed as follows:

- They satisfy the recursion, $B_i^n(t) = (1-t)B_i^{n-1}(t) + tB_{i-1}^{n-1}(t)$ with $B_0^0(t) = 1$ and $B_i^n(t) = 0$ for $i \notin \{0,...,n\}$.
- They form a partition of unity, $\sum_{i=1}^{n} B_i^n(t) = 1$.
- B_i^n are nonnegative over the interval [0,1] (Farin, 2002).

In addition to wide range of quantile estimators, it is also possible to obtain the weights by using Bernstein polynomials when the quantile estimators are weighted averages of all order statistics.

Brewer (1986) proposed the Bernstein polynomial quantile estimator as below.

$$\hat{Q}_n^B(q) = \sum_{j=1}^{n} \left[\binom{n-1}{j-1} q^{j-1}(1-q)^{n-j} \right] X_{(j)} \tag{3.35}$$

Furthermore, Perez & Palacin (1987) studied the mean squared error $\hat{Q}_n^B$ while Cheng (1995) investigated asymptotic properties of this estimator.

Moreover, let $Q_m(x)$ be the empirical quantile function and by applying the Bernstein-Durrmeyer smoothing operator to the quantile estimator, the Bernstein-Durrmeyer estimator of the population quantile is estimated as

$$D_N(Q_m(x)) = (N+1)\sum_{i=0}^{N} \tilde{a}_i B_i^N(x) \tag{3.36}$$

for $x \in [0,1]$ where $\tilde{a}_i$ corresponds to weights and simplified by

$$\hat{a}_i = \frac{1}{m}\sum_{j=1}^{n} X_{(j)} B_i^N \binom{j-1}{m-1} \tag{3.37}$$

and $B_i^N(x)$ are Bernstein polynomials of degree N. Pepelyshev, Rafajłowicz & Steland (2014) studied these type quantile estimators in terms of mean squared error, rates of convergence and asymptotic distribution.

3.3 The New Quantile Estimator: NO

3.3.1 Description of the New Quantile Estimator NO

Firstly, $S_0 = (L, X_{(1)})$, $S_1 = [X_{(1)}, X_{(2)})$, …, $S_{n-1} = [X_{(n-1)}, X_{(n)}), S_n = [X_{(n)}, U)$ are set where $X_{(1)},...,X_{(n)}$ are order statistics of the sample, U and L are upper and lower bounds for X respectively. Certainly, the q'th quantile of the population, Q_q, is in one of these intervals. Define the random variables

$$\delta_i = \begin{cases} 1, & X_i \le Q_q \\ 0, & X_i > Q_q \end{cases} \tag{3.38}$$

where $\delta_i \sim \text{Bernoulli}(q)$. Then, their sum $N = \delta_1 + \delta_2 + ... + \delta_n$ has a binomial distribution with probability of success q, supposing δ_i are independent. So, $P(Q_q \in S_i) = P(N = i) = B(i; n, q)$, i=0,1,…,n. Note that, the binomial probabilities $B(i;n,q)$ are explicitly given in Equation (3.39).

$$B(i;n,q) = \binom{n}{i} q^i (1-q)^{n-i} \tag{3.39}$$

Consider the point estimator of Q_q, namely $Q'_{q,i}$, conditioned on the event $Q_q \in S_i$ for i=0,1,…,n. An estimator of Q_q is found by evaluating the expected value $E(Q'_{q,i})$.

Here, four different $Q'_{q,i}$ are applied and clearly each of them yields different quantile estimators. The first three of them are denoted as $NO_q(1)$, $NO_q(2)$ and $NO_q(3)$ but it is important to say that the attention is focused on the fourth one NO_q.

- If $Q'_{q,i} = 1/4\,X_{(i)} + 3/4\,X_{(i+1)}$,

$$NO_q(1) = \frac{2B(0;n,q)+B(1;n,q)}{4}X_{(1)} + \frac{5B(0;n,q)}{4}X_{(2)} - \frac{3B(0;n,q)}{4}X_{(3)} + \sum_{i=1}^{n-2}\frac{3B(i;n,q)+B(i+1;n,q)}{4}X_{(i+1)} - \frac{B(n;n,q)}{4}X_{(n-2)} - \frac{B(n;n,q)}{4}X_{(n-1)} + \frac{3B(n-1;n,q)+6B(n;n,q)}{4}X_{(n)} \tag{3.40}$$

is obtained.

- If $Q'_{q,i} = 3/4\,X_{(i)} + 1/4\,X_{(i+1)}$,

$$NO_q(2) = \frac{6B(0;n,q)+3B(1;n,q)}{4}X_{(1)} - \frac{B(0;n,q)}{4}X_{(2)} - \frac{B(0;n,q)}{4}X_{(3)} + \sum_{i=1}^{n-2}\frac{B(i;n,q)+3B(i+1;n,q)}{4}X_{(i+1)} - \frac{3B(n;n,q)}{4}X_{(n-2)} + \frac{5B(n;n,q)}{4}X_{(n-1)} + \frac{B(n-1;n,q)+2B(n;n,q)}{4}X_{(n)} \tag{3.41}$$

is obtained.

- If $Q'_{q,i} = (1-q)X_{(i)} + qX_{(i+1)}$,

$$NO_q(3) = [B(0;n,q)(2-2q) + B(1;n,q)(1-q)]X_{(1)} + B(0;n,q)(3q-1)X_{(2)} - B(0;n,q)qX_{(3)} + \sum_{i=1}^{n-2}[B(i;n,q)q + B(i+1;n,q)(1-q)]X_{(i+1)} - B(n;n,q)(1-q)X_{(n-2)} + B(n;n,q)(2-3q)X_{(n-1)} + [B(n-1;n,q)q + B(n;n,q)(2q)]X_{(n)} \quad (3.42)$$

is obtained.

$NO_q(1)$, $NO_q(2)$, $NO_q(3)$ and NO_q are analyzed in terms of relative efficiencies and the ability to control actual type I error rate when the quantile estimator is used in some hypothesis testing on comparing two independent groups. The relative efficiencies of these estimators are calculated over both Harrell Davis quantile estimator HD_q and the default quantile estimator of R programming language R_q. NO_q performs very well for every experimental cases considered, but the others cannot control the nominal significance level with small sample sizes especially when the extreme quantiles are of interest. In addition, NO_q is more efficient than HD_q and R_q in most of the cases while the others are not as satisfactory as NO_q. Due to these reasons, the main interest is concentrated on NO_q and its calculation is expressed step by step.

Let $Q'_{q,i} = qX_{(i)} + (1-q)X_{(i+1)}$ and suppose the following proxies in Equation (3.43) and Equation (3.44).

$$Q'_{q,0} - Q'_{q,1} = Q'_{q,1} - Q'_{q,2} \quad (3.43)$$

$$Q'_{q,n} - Q'_{q,n-1} = Q'_{q,n-1} - Q'_{q,n-2} \quad (3.44)$$

Then, $E(Q'_{q,i}) = E\left(qX_{(i)} + (1-q)X_{(i+1)}\right)$ is evaluated as follows:

$$E(qX_{(i)}+(1-q)X_{(i+1)})=\sum_{i=0}^{n}B(i;n,q)\left(qX_{(i)}+(1-q)X_{(i+1)}\right)= \underbrace{B(0;n,q)\left(qX_{(0)}+(1-q)X_{(1)}\right)}_{*}+ \underbrace{\sum_{i=1}^{n-1}B(i;n,q)\left(qX_{(i)}+(1-q)X_{(i+1)}\right)}_{***}+ \underbrace{B(n;n,q)\left(qX_{(n)}+(1-q)X_{(n+1)}\right)}_{**} \tag{3.45}$$

By using the proxy in Equation (3.43),

$$Q'_{q,0}=2Q'_{q,1}-Q'_{q,2} \tag{3.46}$$

$$qX_{(0)}+(1-q)X_{(1)}=2[qX_{(1)}+(1-q)X_{(2)}]-[qX_{(2)}+(1-q)X_{(3)}] \tag{3.47}$$

$$qX_{(0)}+(1-q)X_{(1)}=2qX_{(1)}+2(1-q)X_{(2)}-qX_{(2)}-(1-q)X_{(3)} \tag{3.48}$$

$$qX_{(0)}+(1-q)X_{(1)}=2qX_{(1)}+(2-3q)X_{(2)}-(1-q)X_{(3)} \tag{3.49}$$

the part denoted with (*) is obtained as:

$$B(0;n,q)[qX_{(0)}+(1-q)X_{(1)}]=B(0;n,q)[2qX_{(1)}+(2-3q)X_{(2)}-(1-q)X_{(3)}] \tag{3.50}$$

By using the proxy in Equation (3.44),

$$Q'_{q,n}=2Q'_{q,n-1}-Q'_{q,n-2} \tag{3.51}$$

$$qX_{(n)}+(1-q)X_{(n+1)}=2[qX_{(n-1)}+(1-q)X_{(n)}]-[qX_{(n-2)}+(1-q)X_{(n-1)}] \tag{3.52}$$

$$qX_{(n)}+(1-q)X_{(n+1)}=2qX_{(n-1)}+2(1-q)X_{(n)}-qX_{(n-2)}-(1-q)X_{(n-1)} \tag{3.53}$$

$$qX_{(n)}+(1-q)X_{(n+1)}=2(1-q)X_{(n)}+(3q-1)X_{(n-1)}-qX_{(n-2)} \tag{3.54}$$

the part denoted with (**) is obtained as:

$$B(n;n,q)[qX_{(n)}+(1-q)X_{(n+1)}]=B(n;n,q)[(2-2q)X_{(n)}+(3q-1)X_{(n-1)}-qX_{(n-2)}] \tag{3.55}$$

Finally, the part denoted with (***) is evaluated:

$$\sum_{i=1}^{n-1} B(i;n,q)\left(qX_{(i)}+(1-q)X_{(i+1)}\right) = B(1;n,q)\left(pX_{(1)}+(1-q)X_{(2)}\right)+$$
$$B(2;n,q)\left(qX_{(2)}+(1-q)X_{(3)}\right)+\ldots+$$
$$B(n-2;n,q)\left(qX_{(n-2)}+(1-q)X_{(n-1)}\right)+$$
$$B(n-1;n,q)\left(qX_{(n-1)}+(1-q)X_{(n)}\right) \quad (3.56)$$

$$= B(1;n,q)qX_{(1)} + \sum_{i=1}^{n-2}[B(i;n,q)(1-q)+B(i+1;n,q)q]X_{(i+1)} + B(n-1;n,q)(1-q)X_{(n)} \quad (3.57)$$

Combining the Equations (3.50), (3.55) and (3.57) yields the statement in Equation (3.58):

$$B(0;n,q)2qX_{(1)} + B(0;n,q)(2-3q)X_{(2)} - B(0;n,q)(1-q)X_{(3)} + B(1;n,q)qX_{(1)} +$$
$$\sum_{i=1}^{n-2}[B(i;n,q)(1-q)+B(i+1;n,q)q]X_{(i+1)} + B(n-1;n,q)(1-q)X_{(n)}$$
$$+B(n;n,q)(2-2q)X_{(n)} + B(n;n,q)(3q-1)X_{(n-1)} - B(n;n,q)qX_{(n-2)} \quad (3.58)$$

and the new quantile estimator NO_q is obtained as below, where n represents the sample size, q is the considered quantile where $0<q<1$ and B (i; n, q) are the binomial probabilities for i=0, 1, …,n.

$$NO_q = [B(0;n,q)2q + B(1;n,q)q]X_{(1)} + B(0;n,q)(2-3q)X_{(2)} - B(0;n,q)(1-q)X_{(3)} +$$
$$\sum_{i=1}^{n-2}[B(i;n,q)(1-q)+B(i+1;n,q)q]X_{(i+1)} - B(n;n,q)qX_{(n-2)} + B(n;n,q)(3q-1)X_{(n-1)} +$$
$$[B(n-1;n,q)(1-q)+B(n;n,q)(2-2q)]X_{(n)} \quad (3.59)$$

3.3.2 Asymptotic Properties

When the objective is to investigate the asymptotic properties of the new estimator, firstly the asymptotic distribution of the order statistics $X_{(i)}$ is studied where i=[nq]+1. Note that, [nq] is the integer part of nq for $0<q<1$ and n denotes the sample size.

Let F(x) be an absolutely continuous distribution function with the probability density function f(x) which is positive and continuous at the point $F^{-1}(q)$ for 0<q<1. Then,

$$\sqrt{n} f(F^{-1}(q)) \frac{X_{(i)} - F^{-1}(q)}{\sqrt{q(1-q)}} \to N(0,1) \tag{3.60}$$

as $n \to \infty$ where i=[nq]+1.

Firstly, the statement in Equation (3.60) is proved when F(x) is a Uniform (0,1) cumulative distribution function. Then the inverse probability integral transform $X_{(i)} \overset{d}{=} F^{-1}(U_{(i)})$ is used to prove it for a general F, where $\overset{d}{=}$ represents the equality in distribution. To show that,

$$\sqrt{n}(U_{(i)} - q) \xrightarrow{d} N(0, q(1-q)) \tag{3.61}$$

as $n \to \infty$ and clearly $i \to \infty$.

Note that, $U_{(i)}$ has a beta distribution with parameters i and n-i+1, which leads the expression $U_{(i)} = \frac{A_n}{A_n + B_n}$ with $A_n = \sum_{r=1}^{i} Z_r$ and $B_n = \sum_{r=i+1}^{n+1} Z_r$ where Z_r's are independent and identically distributed random variables having the distribution Exponential (1).

By Central Limit Theorem $((A_n - i)/\sqrt{n}) \xrightarrow{d} A$ where A is N(0,q) and $(B_n - (n-i+1)/\sqrt{n}) \xrightarrow{d} B$ where B is N(0,1-q). The limit random variables A and B are independent, because A_n and B_n are independent for all n. Hence,

$$C_n = \frac{1}{\sqrt{n}}\{(1-q)(A_n - i) - q(B_n - (n-i+1))\} \xrightarrow{d} N(0, q(1-q)). \qquad (3.62)$$

Thereafter,

$$\sqrt{n}(U_{(i)} - q) = \frac{C_n - \{(i - nq - 1)/\sqrt{n}\}}{\{(A_n + B_n)/\sqrt{n}\}} \qquad (3.63)$$

where the numerator converges in distribution to N(0,q(1-q)) whereas the denominator converges to 1 in probability. So by Slutsky's Theorem, Equation (3.61) is verified.

Thus far, the asymptotic distribution of $X_{(i)}$ is obtained when F is a cumulative distribution function having a uniform probability distribution. For an arbitrary F, Taylor series expansion for $X_{(i)}$ is used and

$$X_{(i)} \overset{d}{=} F^{-1}(q) + (U_{(i)} - q)\{f(F^{-1}(D_n))\}^{-1} \qquad (3.64)$$

is written where D_n is between $U_{(i)}$ and q. This can be organized as

$$\sqrt{n}(X_{(i)} - F^{-1}(q)) \overset{d}{=} \sqrt{n}(U_{(i)} - q)\{f(F^{-1}(D_n))\}^{-1} \qquad (3.65)$$

If f is continuous at $F^{-1}(q)$, as $n \to \infty$, $f(F^{-1}(D_n)) \xrightarrow{p} f(F^{-1}(q))$. Then using Equation (3.61) and Slutsky's Theorem in Equation (3.65), proof of the asymptotic normality of $X_{(i)}$ in Equation (3.60) is accomplished (Arnold, Balakrishnan & Nagaraja, 1992).

Now, the asymptotic results for functions of order statistics are investigated. Let

$$S_n = \frac{1}{n}\sum_{i=1}^{n} J\left(\frac{i}{n+1}\right)X_{(i)} \tag{3.66}$$

where J is a suitable weight function.

If $E(X_i^2) < \infty$ and J(u) is a bounded and continuous almost everywhere F^{-1}, then $\sigma^2(J,F) > 0$ implies

$$\frac{S_n - E(S_n)}{\sigma(S_n)} \xrightarrow{d} N(0,1) \tag{3.67}$$

as $n \to \infty$ where

$$\sigma^2(J,F) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} J(F(x))J(F(y))[F(\min(x,y)) - F(x)F(y)]dxdy \tag{3.68}$$

To prove this statement, recall that $\sqrt{n}(S_n - E(S_n))$ and $\sqrt{n}(T_n - E(T_n))$ are asymptotically equivalent in mean square and also, $\sqrt{n}(T_n - E(T_n))$ and $\sqrt{n}(\hat{T}_n - E(\hat{T}_n))$ are asymptotically equivalent in mean square.

So, it is enough to prove that $\sqrt{n}(\hat{T}_n - E(\hat{T}_n))$ is asymptotically normally distributed where

$$T_n = \frac{1}{n}\sum_{i=2}^{n-1} J\left(\frac{i}{n+1}\right)X_{(i)} \tag{3.69}$$

$$\hat{T}_n - E(\hat{T}_n) = \frac{1}{n}\sum_{k=1}^{n} Z_{kn} \tag{3.70}$$

with

$$Z_{kn} = \sum_{i=2}^{n-1} J\left(\frac{i}{n+1}\right) \int_{-\infty}^{\infty} (F(y) - I_{(x_k \le y)}) P_i(y) dy \tag{3.71}$$

$$= \int_{-\infty}^{\infty} (F(y) - I_{(x_k \le y)}) Q_n(y) dy \tag{3.72}$$

where

$$Q_n(y) = \sum_{i=2}^{n-1} J\left(\frac{i}{n+1}\right) P_i(y) \tag{3.73}$$

and

$$P_i(y) = \binom{n-1}{i-1} F(y)^{i-1} (1 - F(y))^{n-i} \tag{3.74}$$

for $1 < i < n$. Let

$$Z_k = \int_{-\infty}^{\infty} (F(y) - I_{(x_k \le y)}) J(F(y)) dy \,. \tag{3.75}$$

Then it is obtained that $\frac{1}{\sqrt{n}} \sum_{k=1}^{n} Z_{kn}$ and $\frac{1}{\sqrt{n}} \sum_{k=1}^{n} Z_k$ are asymptotically equivalent in mean square.

Z_k 's are independent and identically distributed with finite variance $\sigma^2(J, F)$ and hence the Central Limit Theorem shows that $\frac{1}{\sqrt{n}} \sum_{k=1}^{n} Z_k$ is asymptotically normally distributed, so $\sqrt{n}(\hat{T}_n - E(\hat{T}_n))$ is also (Stigler, 1974).

Clearly, the new estimator is also a linear function of the order statistics, with suitable weights which consist of binomial probabilities. Accordingly, following the

explanation by Stigler (1974), the new estimator is asymptotically normally distributed.

CHAPTER FOUR
COMPARING GROUPS BY USING MORE THAN ONE REFERENCE POINT

Of course, the best known and the most commonly applied method for comparing two independent groups is based on one measure of location intended to reflect the population. As reviewed in Chapter 2, for reasonably symmetric distributions the mean is appropriate, whereas a robust measure of location like median or trimmed mean can be used when the distributions are skewed (Wilcox et al., 2014). Nevertheless, not only the conventional method Student's t test, but also its more robust alternatives as Welch test or Yuen test are designed to consider only one reference point (Welch, 1938; Yuen, 1974).

On the other side, when comparing two independent groups, it is often of interest to determine whether the differences occur in the tails of distributions or not. In other saying, to understand how the high scoring observations in first group differs compared to the high scoring observations in the second group might be the purpose of a researcher. Similarly, comparing the low scoring observations can be the research interest as well. Note that, the high and low scoring observations correspond to the upper and lower quantiles, respectively (Wilcox et al., 2014). Comparing the tails of distributions, namely the quantiles, is commonly used in the fields such as psychology, biology, medicine, economics, etc.

For instance, think about a study that aims to assess the effectiveness of intervention in terms of reducing depressive symptoms in older adults. In such a study, lower and upper tails of populations may respond to the experimental method differently. That is, the intervention can be more effective for the participants having higher levels of depression that corresponds to upper quantiles of the groups. Conversely, may be less effective or harmful for the participants having lower levels of depression that corresponds to lower quantiles of the groups. Since there may be such situations, comparing the control and experimental groups through quantiles is more informative (Wilcox et al., 2014).

Yet another example considers a study on teaching reading which compares a standard method with an experimental method. Suppose that, these groups have equal means and clearly methods for comparing means should not reject. It is possible that, students who do poorly under the standard method benefit from the experimental method, whereas the experimental method is detrimental to students who already do well using the standard technique (Wilcox, 1995). Hence, comparing groups by using more than one reference points which correspond different quantiles of the populations allows a deeper insight.

In a study that is conducted by Erceg-Hurn & Steed (2011), level of negative emotional reactions of the smokers exposed to anti-smoking warnings are considered. The individuals that form two independent groups see text or picture warnings. The emotional responses are scored between the values 0 and 16 leading a great number of tied values. Comparing two independent groups based on mean, trimmed mean or median found significant differences, however the researcher is interested in comparison through quantiles in order to understand how the subgroups compare. Due to the tied values, there could be some problems with available methods. Choosing the appropriate method when dealing with tied values is an important issue.

For an instant, let the purpose be to compare two independent groups through medians. In the presence of tied values, the estimates of the standard error of sample median is not accurate and the estimates of the population median do not necessarily converge a normal distribution as the sample size gets larger. In general, a percentile bootstrap based method performs reasonably well when there are tied values. So, attention might be focused on applying the same percentile bootstrap approach when other quantiles of interest.

Another motivation for comparing quantiles is that, two populations may differ in a way that is not discovered by considering one reference point. When comparing two independent groups by using Student's t test, the consequence of the test may state that there is no significant difference between populations. Also, Student's t test has the assumptions of homoscedasticity and normality of two population distributions, while

these assumptions are met uncommonly. Correspondingly, the results can be misleading when the assumptions are not hold. Moreover, the robust alternatives of Student's t test such as Welch's heteroscedastic method for means and Yuen's method for trimmed means can result with a p-value yielding the groups are not differing. In brief, considering a single reference point may miss important differences among subpopulations of groups.

Wilcox (1995) proposed to compare two independent groups via multiple quantiles rather than comparing in terms of a single measure of location. Thereafter, Wilcox et al. (2014) compared two independent groups by using lower and upper quantiles with a method that uses a percentile bootstrap in conjunction with Harrell Davis quantile estimator. As a result of the corresponding study, the actual type I error rates were controlled for both continuous and discrete distributions even in the presence of tied values. However, there is an issue about the minimum sample size when the extreme quantiles are of interest. For example, when the aim is to compare two independent groups via 0.25'th or 0.75'th quantiles the minimum sample size that is sufficient is 20. Additionally, when comparing 0.9'th quantiles as the minimum sample size n=20 is needed, whereas n=50 is enough when the 0.95'th quantiles are being compared.

Briefly, comparing two independent groups by using one reference point, not only provides inappropriate results but also leads missing the important differences between two groups especially in the tails of them. Therefore, comparing two independent groups through more than one reference point, namely quantiles, is a better idea. For this purpose, two different methods are introduced. The method in Section 4.1 follows the same manner with Wilcox et al. (2014) and compares two independent groups through the q'th quantile where $0<q<1$. On the other hand, the newly proposed method in Section 4.2 compares two independent groups by considering q=0.1, q=0.5 and q=0.9 at the same time.

4.1 Comparing Two Independent Groups Through Quantiles Individually

An appropriate quantile estimator is used in conjunction with a percentile bootstrap method in order to get a reasonable control over the Type I error rate probabilities. The purpose is to test

$$H_0 : \theta_{q1} = \theta_{q2} \tag{4.1}$$

where θ_{q1} and θ_{q2} are the q'th quantiles of the first and second populations respectively. Let X_{ij} be a random sample from the j'th group where j=1,2 and $i = 1, ..., n_j$. The percentile bootstrap approach is as follows:

- Generate a bootstrap sample from each group by resampling with replacement yielding X^*_{ij}.
- θ^*_{qj} is the estimate of q'th quantile estimate for group j based on this bootstrap sample that is obtained by means of a quantile estimator.
- Define $d^* = \theta^*_1 - \theta^*_2$ and repeat this process B times to obtain d^*_b, b=1,…,B. In this study B is chosen as 2000.
- Put the d^*_b values in ascending order, that is $d^*_{(1)} \leq ... \leq d^*_{(B)}$.
- Let $\ell = \alpha B/2$ is rounded to nearest integer and $u = B - \ell$. An approximate confidence interval for $\theta_1 - \theta_2$ is achieved as $(d^*_{(\ell+1)}, d^*_{(u)})$.

Additionally, a generalized p-value can be evaluated. Let A be the number of times that $d^* < 0$ and let C be the number of times that $d^* = 0$. Then, $p^* = \frac{A + 0.5C}{B}$ and a generalized p-value is $2\min(p^*, 1 - p^*)$.

Obviously, if p-value $< \alpha$, H_0 is rejected where α is the nominal significance level (Wilcox et al. 2014; Wilcox, 2017).

4.2 The Proposed Method: Comparing Two Independent Groups Through Multiple Quantiles Simultaneously

NO quantile estimator is used in conjunction with a percentile bootstrap method to allow a deeper insight into comparing two independent groups. To describe the proposed method in formal terms, the goal is to test

$$H_0 : \theta_{q1} = \theta_{q2} \tag{4.2}$$

where θ_{qj} is the q'th quantile corresponding to the j'th group (j = 1, 2). The comparison is done by considering three quantiles for each groups: q=0.1 as a lower quantile, q=0.5 as the middle and q=0.9 as an upper quantile.

The explanation of the method begins with estimating the quantiles for the first and second groups. Then, the differences between the estimated quantile values are calculated by $NO_{1q} - NO_{2q}$, where q=0.1, 0.5, 0.9. These three estimates are used as the original differences.

Let X_{ij} be a random sample from the j'th group where j=1, 2 and $i = 1, ..., n_j$. Generate a bootstrap sample from each group by resampling with replacement yielding X_{ij}^*. Compute $NO_{1q} - NO_{2q}$, where q=0.1, 0.5, 0.9 for each bootstrap sample. Here, the number of bootstrap samples are chosen as B=2000 which leads a three by two thousand matrix. The first row consists of differences between the 0.1'th quantiles, the second consists of differences between the 0.5'th quantiles and the third row consists of differences between the 0.9'th quantiles. Also, each bootstrap sample corresponds a column. To determine whether these two groups differ, the general strategy is to determine how deeply **0** = (0, 0, 0) is nested within the bootstrap values where **0** is a vector having length 3 since three different quantiles are being considered.

For this purpose, Mahalanobis distance is computed for each point among all of the B=2000 bootstrap points. Roughly, Mahalanobis distance is a standardized distance

from the center of the data cloud that depends in part on the covariances among the bootstrapped quantile difference values. As a general formula, the Mahalanobis distance of an observation $x = (x_1, x_2, ..., x_n)^T$ from a set of observations with $\mu = (\mu_1, \mu_2, ..., \mu_n)^T$ and covariance matrix S is defined as:

$$D(x) = \sqrt{(x - \mu)^T S^{-1} (x - \mu)} \tag{4.3}$$

The three by two thousand matrix that consists the differences of quantile values from bootstrap samples is transposed and the vector that contains means of its rows is subtracted. Then the original difference values of quantiles are added. Evaluating the variance of the final matrix yields the variance covariance matrix for the computation of Mahalanobis distance.

The distance of **0** = (0, 0, 0) from the center of the cloud is computed where the center is original difference values of quantiles. Also, the distances of the bootstrap difference values of quantiles from the center are computed. The number of cases that the distance of **0** = (0, 0, 0) from the center of the cloud is smaller than the distances of the bootstrap difference values of quantiles from the center is counted and divided by the number of bootstrap samples. This final ratio gives the significance level and clearly if this significance level is smaller than the nominal significance level $\alpha = 0.05$, the null hypothesis is rejected (Liu & Singh, 1997; Wilcox, 2011).

CHAPTER FIVE
SIMULATION STUDY AND CONCLUSION

5.1 Design of the Simulation Study

5.1.1 Asymptotic properties

A series of simulation studies are conducted to examine the asymptotic properties of NO quantile estimator practically. Asymptotic consistency, asymptotic unbiasedness, asymptotic normality, asymptotic mean squared error and asymptotic variance are the properties that are to be examined through these simulation studies.

First aim was to check whether the estimated quantile value converge to the population quantile and whether the difference between the estimated quantile value and population quantile converges to zero as the sample size increases. By measuring these two, the asymptotic consistency and asymptotic unbiasedness will be questioned numerically.

In addition to theoretical findings concerning asymptotic normality, simulation results are also used in order to test the normality of estimated quantile values. Shapiro-Wilk and Anderson-Darling normality tests are conducted where the null hypothesis states the estimated quantile values follow a normal distribution:

$$H_0 : \text{The estimated quantile values are normally distributed.} \qquad (5.1)$$

At the nominal significance level $\alpha = 0.05$, p–values that are obtained from the mentioned normality tests are saved. Clearly, when p-value>0.05 there is no evidence to reject the null hypothesis.

Two other criteria about NO quantile estimator are asymptotic variance and asymptotic mean squared error. It is expected that, the variance and the mean squared error of the estimator decrease as the sample size n increases.

Four different g-and-h distributions with different parameters are considered. Namely, standard normal (g=h=0), asymmetric and heavy-tailed (g=h=0.5), asymmetric and light-tailed (g=0.5, h=0), symmetric and heavy-tailed (g=0, h=0.5). g-and-h distribution allows to understand how a distribution differs from normal distribution by parameters g and h which are concerned with skewness and kurtosis, respectively. Let Z be a random variable that has a standard normal distribution. By using the transformations in Equation (5.2) when $g \neq 0$ and in Equation (5.3) when g=0, the data is generated from g-and-h distribution (Hoaglin, 1985).

$$X = \frac{(\exp(gZ) - 1)\exp(hZ^2/2)}{g} \tag{5.2}$$

$$X = Z\exp(hZ^2/2) \tag{5.3}$$

The skewness and kurtosis values of these distributions are given in the Table 5.1 where they are estimated with $\hat{\kappa}_1$ and $\hat{\kappa}_2$, respectively. These estimated skewness and kurtosis values are based on 100000 observations. When g=0, it is known that skewness equals to zero, so in such cases it is not estimated (Wilcox, 2017). For g and h values in Table 5.1, the shapes of the corresponding distributions are given in a series of figures through Figure 5.1 to Figure 5.4.

Note that, the nominal significance level is set at α=0.05. Both lower, middle and upper quantiles are considered: q=0.05, q=0.1, q=0.5, q=0.9 and q=0.95. Also, sample sizes varied from 10 to 10000 in order to show the effect of increasing the sample size. All simulations except the one that controls the asymptotic normality are based on 10000 replications and done by using R programming language version 3. 5. 1. p - values from normality tests are obtained on the basis of 1000 replications.

Table 5.1 g-and-h distributions

	$\hat{\kappa}_1$	$\hat{\kappa}_2$	Type
g=0 and h=0	0	0	Symmetric and light tailed
g=0 and h=0.5	0	10760.45	Symmetric and heavy tailed
g=0.5 and h=0	1.767607	6.323596	Asymmetric and light tailed
g=0.5 and h=0.5	58.37469	5863.357	Asymmetric and heavy tailed

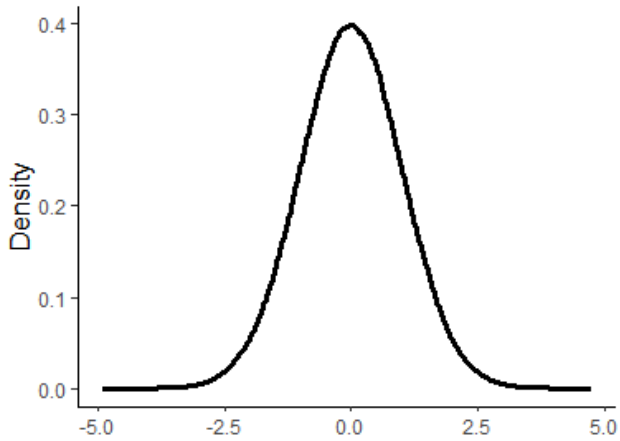

Figure 5.1 Shape of the g-and-h distribution, g=0 and h=0

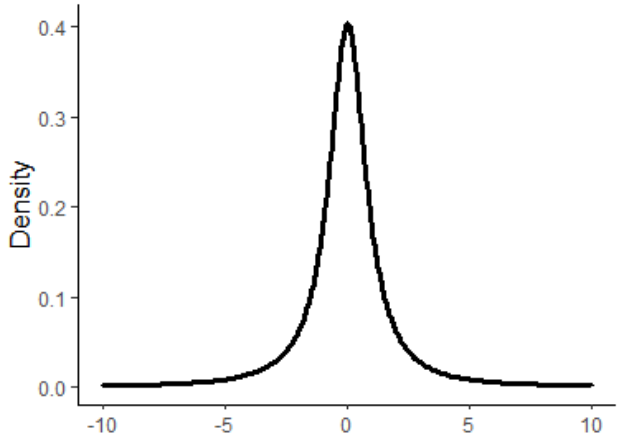

Figure 5.2 Shape of the g-and-h distribution, g=0 and h=0.5

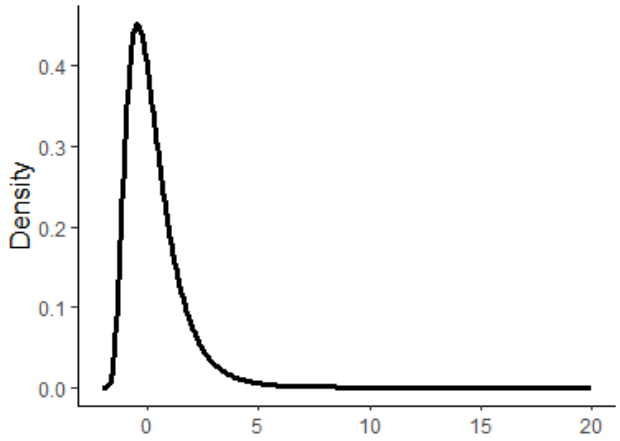

Figure 5.3 Shape of the g-and-h distribution, g=0.5 and h=0

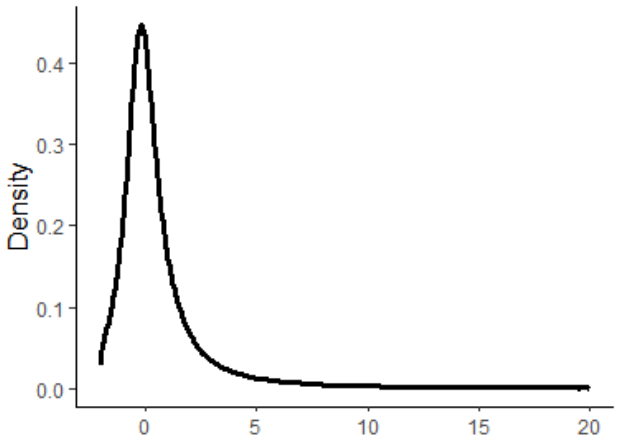

Figure 5.4 Shape of the g-and-h distribution, g=0.5 and h=0.5

5.1.2 Relative efficiencies

In the second part of the simulation study, the relative efficiency of the NO quantile estimator over both Harrell Davis estimator in Equation (3.12) and the default quantile estimator of R programming language are calculated for different experimental settings. To remind, the default quantile estimator of R programing language is a weighted average of two consecutive order statistics which is obtained by substituting m=1-q and $\gamma = g$ in Equation (3.9). Note that, nq=j+g with j=[nq], that is j corresponds to integral part of nq whereas g corresponds to the fractional part.

Relative efficiencies are calculated by the ratios of mean squared errors as in the Equations (5.4) and (5.5).

$$\text{relative efficiency over R} = \frac{\text{MSE(R)}}{\text{MSE(NO)}} \tag{5.4}$$

$$\text{relative efficiency over HD} = \frac{\text{MSE(HD)}}{\text{MSE(NO)}} \tag{5.5}$$

Sample sizes are chosen as n=20 and n=40, where q=0.1, q=0.5 and q=0.9 are considered quantiles with the distributions in Section 5.1.1. All simulations are based on 10000 replications and done by using R programming language version 3. 5. 1.

5.1.3 Comparing Two Independent Groups Through Quantiles

When comparing two independent groups through quantiles with a percentile bootstrap based method, the performance of NO quantile estimator is investigated. In the first instance, with the aim of comparing two independent groups through quantiles individually with the method given in Section 4.1, the performances of NO quantile estimator, Harrell Davis quantile estimator and the default quantile estimator of R programming language are compared. In addition to this, the performance of NO quantile estimator with the proposed method (Section 4.2) that simultaneously compares two independent groups through multiple quantiles is investigated.

The criterion was to control the actual type I error rates at the α=0.05 nominal significance level. To interpret the results, Bradley's liberal criterion of robustness is considered (Bradley, 1978). According to this criterion, actual type I error rates should fall within the interval $(0.5\alpha, 1.5\alpha)$ and consequently at α=0.05 an actual type I error rate should fall within the interval (0.025, 0.075).

Note that, sample sizes are taken as n=10, n=20 and n=40 with the aim of investigating the effect of both small and large sample cases. Two independent groups are compared through the quantiles q=0.05, q=0.1, q=0.5, q=0.9 and q=0.95 separately, so that the differences occur not only in the middle but also in the tails can be detected.

Data are generated from g-and-h distributions, however in addition to given parameters in Section 5.1.1, additional parameters are also considered to offer variety in some cases. In addition to g-and-h distributions, a Log normal distribution (mean = 0, standard deviation = 1) and a Gamma distribution (rate = 1, shape = 1) are also used since they are more common skewed distributions. The estimated kurtosis and skewness values of these distributions are given in Table 5.2 and their shapes are shown in Figures 5.5 through 5.9.

Table 5.2 g-and-h, Log normal, Gamma distributions

	$\hat{\kappa}_1$	$\hat{\kappa}_2$	Type
g=0 and h=0.2	0	28.0846	Symmetric and heavy tailed
g=0.2 and h=0	0.614015	0.685143	Asymmetric and light tailed
g=0.2 and h=0.2	3.260014	73.58213	Asymmetric and heavy tailed
Log normal (mean=0, standard deviation=1)	6.108049	91.41896	Asymmetric and heavy tailed
Gamma (rate=1, shape=1)	2.045652	6.34953	Asymmetric and heavy tailed

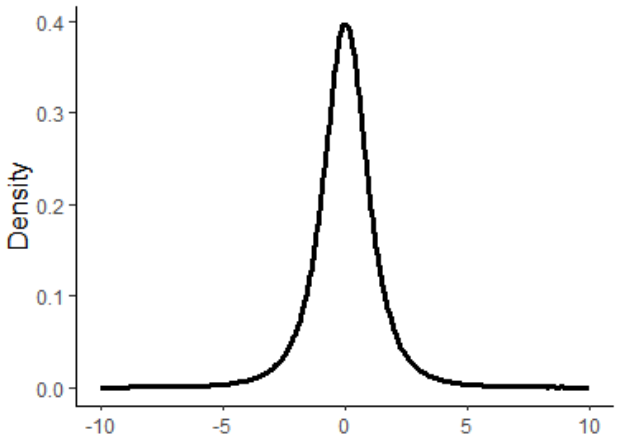

Figure 5.5 g-and-h distribution, g=0 and h=0.2

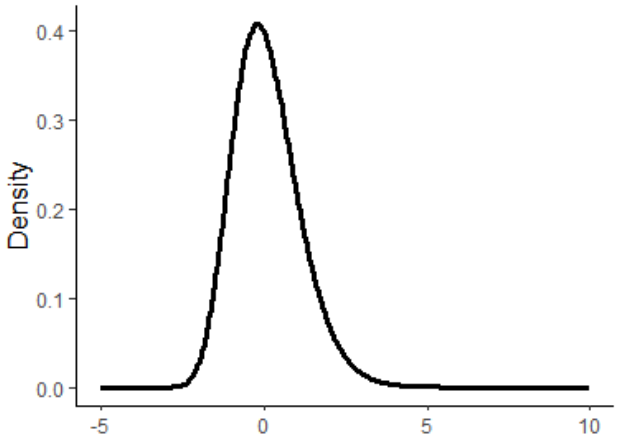

Figure 5.6 g-and-h distribution, g=0.2 and h=0

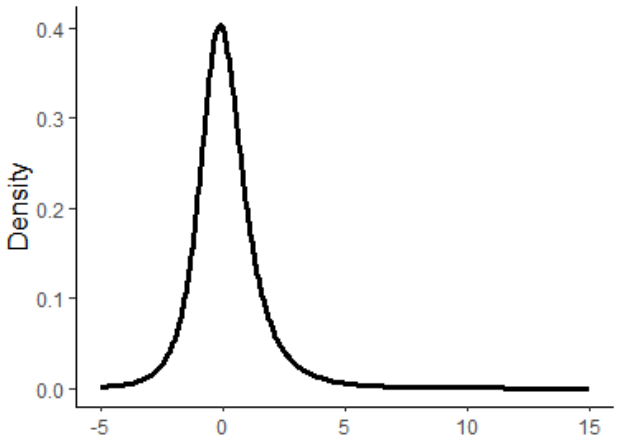

Figure 5.7 g-and-h distribution, g=0.2 and h=0.2

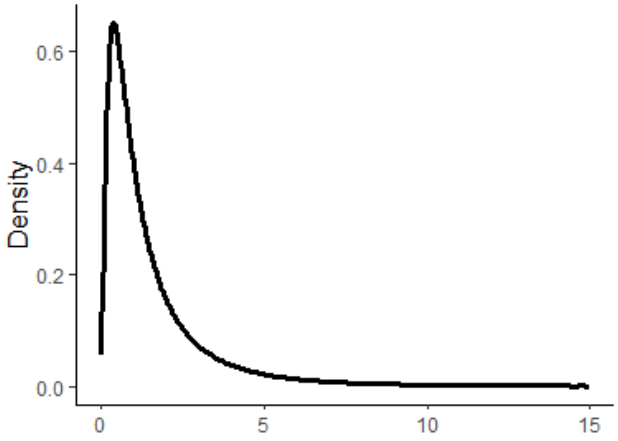

Figure 5.8 Log normal distribution, mean=0 and sd=1

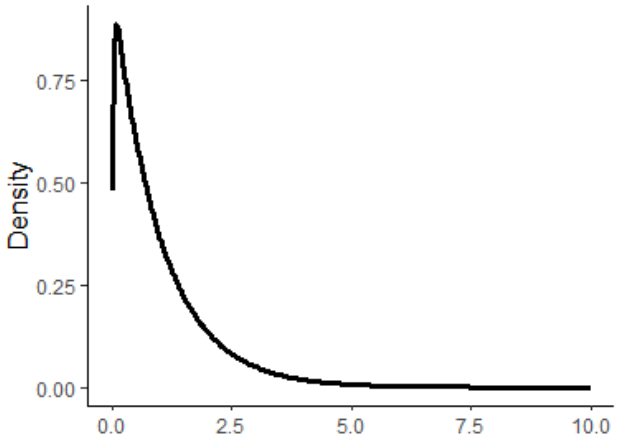

Figure 5.9 Gamma distribution, rate=1 and shape=1

There may occur some problems in hypothesis testing when there are tied values among observations. In addition to continuous distributions, the beta–binomial distribution is used for the discrete case since it enables to observe the effect of tied values. A variable with beta – binomial distribution is distributed as binomial with parameter p and the probability of success p has a beta distribution with parameters r and s. It has the following probability function where B is the complete beta function:

$$P(X = x) = \frac{B(m - x + r, x + s)}{(m+1)B(m - x + 1, x + 1)B(r, s)} \tag{5.6}$$

In this study m=10 is used which means the possible values for X are the integers 0, 1, …, 10. Also, the values for r and s are taken as r=1, 2, 3, 9 and s=9. Note that, with r=s=9 the distribution is bell shaped and symmetric with mean 5.

Number of bootstrap samples are chosen as B=2000 and all simulations are done by 10000 replications.

5.2 Results of the Simulation Study

5.2.1 Asymptotic Properties Results

The estimated quantile values are given in a series of tables through Table 5.3 to Table 5.6, where the tables are organized according to population distributions. For example, in Table 5.3 the population quantiles are estimated for the standard normal distribution. In a similar manner with Table 5.3; the population quantiles are estimated for symmetric and heavy tailed distribution in Table 5.4, for asymmetric and light tailed distribution in Table 5.5 and for asymmetric and heavy tailed distribution in Table 5.6. In each table, population quantiles for q=0.05, q=0.1, q=0.5, q=0.9 and q=0.95 are estimated with different sample sizes whereas the true quantile values are given in the last column of tables. For all population distributions that are considered, as the sample size goes to infinity, the estimated quantile values converge to the real value which demonstrates the asymptotic consistency of NO quantile estimator.

Table 5.3 Estimated quantile values g=0 and h=0

	n=10	**n=100**	**n=1000**	**n=10000**	**population quantile**
q=0.05	-1.211244	-1.633582	-1.643446	-1.64448	-1.644854
q=0.1	-1.061011	-1.274546	-1.28098	-1.281652	-1.281552
q=0.5	0.001266	0.000811	-0.000710	-0.000057	0
q=0.9	1.062493	1.275200	1.280973	1.281580	1.281552
q=0.95	1.209876	1.630211	1.643179	1.644827	1.644854

Table 5.4 Estimated quantile values g=0 and h=0.5

	n=10	**n=100**	**n=1000**	**n=10000**	**population quantile**
q=0.05	-2.174337	-3.577317	-3.264517	-3.237544	-3.235371
q=0.1	-1.978180	-2.029658	-1.941722	-1.932571	-1.931531
q=0.5	0.003219	-0.000131	-0.000093	-0.000003	0.000305
q=0.9	1.885704	2.045903	1.941484	1.933334	1.933806
q=0.95	2.120770	3.573750	3.268832	3.237695	3.237635

Table 5.5 Estimated quantile values g=0.5 and h=0

	n=10	**n=100**	**n=1000**	**n=10000**	**population quantile**
q=0.05	-0.885871	-1.104736	-1.119427	-1.121025	-1.121447
q=0.1	-0.781917	-0.934168	-0.945010	-0.946109	-0.946658
q=0.5	0.090718	0.008516	0.000529	0.000033	0.000097
q=0.9	1.502852	1.809993	1.796072	1.795663	1.795802
q=0.95	1.742760	2.575039	2.554862	2.552083	2.551496

Table 5.6 Estimated quantile values g=0.5 and h=0.5

	n=10	**n=100**	**n=1000**	**n=10000**	**population quantile**
q=0.05	-1.512341	-2.369799	-2.216067	-2.206738	-2.204211
q=0.1	-1.348625	-1.478920	-1.432176	-1.427140	-1.426551
q=0.5	0.151886	0.007774	0.000345	0.000098	0.000238
q=0.9	3.047103	2.923225	2.726090	2.708180	2.706450
q=0.95	3.405389	5.838127	5.082168	5.023742	5.019086

In Table 5.7, Table 5.8, Table 5.9 and Table 5.10 the mean squared error (MSE), variance and bias values of the estimated quantiles are given. The organization of the tables is similar with the asymptotic consistency considerations in previous part. For each tables, the mean squared error and variance values decreased as the sample size goes to infinity. In addition to this, the bias values converge to zero as well and it allows to establish asymptotic unbiasedness of NO quantile estimator numerically.

Finally, the p-values from two different normality tests are given in Table 5.11, Table 5.12, Table 5.13 and Table 5.14. Once more tables are organized on the basis of population distributions that the observations are sampled from. The p-values from Shapiro-Wilk and Anderson-Darling normality tests are given in two columns. In each table, the p-values that are greater than the nominal significance α=0.05 level are highlighted. According to the results, the estimates begin to follow a normal distribution as the sample size goes to infinity, so that asymptotic normality of NO quantile estimator is verified practically.

Table 5.7 MSE, variance and bias values g=0 and h=0

		n=10	**n=100**	**n=1000**	**n=10000**
MSE	q=0.05	0.48649	0.03570	0.00411	0.00044
	q=0.1	0.26576	0.02444	0.00272	0.00028
	q=0.5	0.11494	0.01424	0.00151	0.00015
	q=0.9	0.26082	0.02432	0.00272	0.00029
	q=0.95	0.48967	0.03513	0.00423	0.00044
Variance	q=0.05	0.29847	0.03557	0.00411	0.00044
	q=0.1	0.21712	0.02439	0.00272	0.00028
	q=0.5	0.11494	0.01424	0.00151	0.00015
	q=0.9	0.21283	0.02428	0.00272	0.00029
	q=0.95	0.30046	0.03492	0.00422	0.00044
Bias	q=0.05	0.43361	0.01127	0.00141	0.00037
	q=0.1	0.22054	0.00701	0.00057	-0.00010
	q=0.5	0.00127	0.00081	-0.00071	-0.00006
	q=0.9	-0.21906	-0.00635	-0.00058	0.00003
	q=0.95	-0.43498	-0.01464	-0.00167	-0.00003

Table 5.8 MSE, variance and bias values g=0 and h=0.5

		n=10	**n=100**	**n=1000**	**n=10000**
MSE	q=0.05	6.53017	1.41661	0.09376	0.00935
	q=0.1	4.38935	0.24014	0.02087	0.00212
	q=0.5	0.24366	0.01497	0.00154	0.00015
	q=0.9	3.00491	0.24911	0.02141	0.00215
	q=0.95	4.98637	1.40911	0.09370	0.00929
Variance	q=0.05	5.40438	1.29969	0.09291	0.00934
	q=0.1	4.38718	0.23051	0.02077	0.00212
	q=0.5	0.24366	0.01497	0.00154	0.00015
	q=0.9	3.00260	0.23654	0.02135	0.00215
	q=0.95	3.73899	1.29622	0.09273	0.00929
Bias	q=0.05	1.06103	-0.34195	-0.02915	-0.00217
	q=0.1	-0.04665	-0.09813	-0.01020	-0.00104
	q=0.5	0.00291	-0.00043	-0.00040	-0.00031
	q=0.9	-0.04810	0.11210	0.00768	-0.00047
	q=0.95	-1.11687	0.33612	0.03120	0.00060

Table 5.9 MSE, variance and bias values g=0.5 and h=0

		n=10	**n=100**	**n=1000**	**n=10000**
MSE	q=0.05	0.15892	0.00707	0.00082	0.00086
	q=0.1	0.10314	0.00707	0.00077	0.00079
	q=0.5	0.13392	0.01447	0.00148	0.00016
	q=0.9	0.77465	0.09138	0.00998	0.00102
	q=0.95	1.65525	0.19695	0.02161	0.00222
Variance	q=0.05	0.10342	0.00679	0.00081	0.00085
	q=0.1	0.07600	0.00692	0.00077	0.00079
	q=0.5	0.12570	0.01440	0.00148	0.00016
	q=0.9	0.68883	0.09117	0.00998	0.00102
	q=0.95	1.00120	0.19639	0.02160	0.00222
Bias	q=0.05	0.23558	0.01671	0.00202	0.00042
	q=0.1	0.16474	0.01249	0.00165	0.00055
	q=0.5	0.09062	0.00842	0.00043	-0.00064
	q=0.9	-0.29295	0.01419	0.00027	-0.00014
	q=0.95	-0.80874	0.02354	0.00337	0.00059

Table 5.10 MSE, variance and bias values g=0.5 and h=0.5

		n=10	n=100	n=1000	n=10000
MSE	q=0.05	2.00253	0.40213	0.03081	0.00319
	q=0.1	1.10596	0.08665	0.00819	0.00084
	q=0.5	0.29699	0.01519	0.00154	0.00015
	q=0.9	18.74674	0.74941	0.05918	0.00596
	q=0.95	24.49199	6.09278	0.32991	0.03235
Variance	q=0.05	1.52385	0.37471	0.03067	0.00318
	q=0.1	1.09989	0.08665	0.00816	0.00084
	q=0.5	0.27399	0.01513	0.00154	0.00015
	q=0.9	18.6307	0.70242	0.05879	0.00595
	q=0.95	21.88797	5.42195	0.32593	0.03232
Bias	q=0.05	0.69187	-0.16559	-0.01186	-0.00253
	q=0.1	0.07793	-0.05237	-0.00562	-0.00059
	q=0.5	0.15165	0.00754	0.00011	-0.00014
	q=0.9	0.34065	0.21677	0.01964	0.00173
	q=0.95	-1.61370	0.81904	0.06308	0.00466

Table 5.11 p-values from normality tests g=0 and h=0

		Shapiro-Wilk	Anderson-Darling
q=0.05	n=10	1.731e-09	4.944e-07
	n=100	0.0922	0.4015
	n=1000	0.5935	0.7173
	n=10000	0.5025	0.7858
q=0.1	n=10	0.8677	0.6029
	n=100	0.7037	0.5232
	n=1000	0.3237	0.3428
	n=10000	0.6782	0.5846
q=0. 5	n=10	0.4627	0.5076
	n=100	0.2851	0.1207
	n=1000	0.1328	0.1475
	n=10000	0.9745	0.8867
q=0.9	n=10	0.0020	0.0085
	n=100	0.1052	0.2111
	n=1000	0.8392	0.7883
	n=10000	0.6734	0.4494
q=0.95	n=10	1.266e-06	3.519e-05
	n=100	0.0333	0.0415
	n=1000	0.4509	0.1089
	n=10000	0.3288	0.2617

Table 5.12 p-values from normality tests g=0 and h=0.5

		Shapiro-Wilk	**Anderson-Darling**
q=0.05	n=10	< 2.2e-16	< 2.2e-16
	n=100	< 2.2e-16	< 2.2e-16
	n=1000	1.384e-07	5.319e-05
	n=10000	0.7301	0.7327
q=0.1	n=10	< 2.2e-16	< 2.2e-16
	n=100	7.289e-14	5.335e-15
	n=1000	0.0006	0.0019
	n=10000	0.8504	0.6501
q=0. 5	n=10	7.714e-14	3.056e-09
	n=100	0.5643	0.1521
	n=1000	0.6240	0.6051
	n=10000	0.6481	0.2963
q=0.9	n=10	< 2.2e-16	< 2.2e-16
	n=100	1.371e-11	4.959e-12
	n=1000	2.712e-06	9.85e-07
	n=10000	0.08922	0.8657
q=0.95	n=10	< 2.2e-16	< 2.2e-16
	n=100	< 2.2e-16	< 2.2e-16
	n=1000	0.0213	0.0433
	n=10000	0.0700	0.4143

Table 5.13 p-values from normality tests g=0.5 and h=0

		Shapiro-Wilk	**Anderson-Darling**
q=0.05	n=10	0.0033	0.0149
	n=100	0.7101	0.4085
	n=1000	0.5490	0.4045
	n=10000	0.7973	0.8303
q=0.1	n=10	5.391e-11	6.438e-07
	n=100	0.2045	0.3044
	n=1000	0.4429	0.2150
	n=10000	0.2100	0.2380
q=0. 5	n=10	7.114e-12	1.257e-06
	n=100	0.0133	0.1143
	n=1000	0.9272	0.8761
	n=10000	0.9946	0.9307
q=0.9	n=10	< 2.2e-16	< 2.2e-16
	n=100	0.0001	0.0171
	n=1000	0.0492	0.0359
	n=10000	0.5100	0.7947
q=0.95	n=10	< 2.2e-16	< 2.2e-16
	n=100	2.885e-08	1.628e-06
	n=1000	0.0092	0.0700
	n=10000	0.8803	0.9478

Table 5.14 p-values from normality tests g=0.5 and h=0.5

		Shapiro-Wilk	**Anderson-Darling**
q=0.05	n=10	< 2.2e-16	< 2.2e-16
	n=100	< 2.2e-16	< 2.2e-16
	n=1000	5.519e-05	2.318e-06
	n=10000	0.7570	0.9061
q=0.1	n=10	< 2.2e-16	< 2.2e-16
	n=100	6.421e-09	4.099e-09
	n=1000	0.1138	0.1411
	n=10000	0.6533	0.8504
q=0. 5	n=10	< 2.2e-16	< 2.2e-16
	n=100	0.0037	0.0449
	n=1000	0.1663	0.0599
	n=10000	0.7781	0.6127
q=0.9	n=10	< 2.2e-16	< 2.2e-16
	n=100	2.949e-13	2.909e-15
	n=1000	0.4621	0.5934
	n=10000	0.5540	0.1803
q=0.95	n=10	< 2.2e-16	< 2.2e-16
	n=100	< 2.2e-16	< 2.2e-16
	n=1000	1.165e-06	2.374e-05
	n=10000	0.1400	0.7832

5.2.2 Relative Efficiency Results

The relative efficiency results are given in Table 5.15, Table 5.16, Table 5.17 and Table 5.18 where each table represents results for different population distributions. In each table, the relative efficiencies of NO quantile estimator over both Harrell Davis quantile estimator and the default quantile estimator of R programming language were measured. Note that, starting from this part Harrell Davis quantile estimator and the default quantile estimator of R programming language are symbolized as HD and R, respectively.

The relative efficiency ratios that are greater than 1 are highlighted that means NO quantile estimator is more efficient than the other one. In Table 5.15, when the population distribution is standard normal, NO quantile estimator is more efficient than the others not only for median, but also for extreme quantiles.

Table 5.15 Relative efficiencies, g=0 and h=0

		g=0, h=0	
		over HD	over R
n=20	q=0.1	1.24822	1.21651
	q=0.5	1.13870	1.17301
	q=0.9	1.06228	1.19235
n=40	q=0.1	1.28075	1.25219
	q=0.5	1.13071	1.14274
	q=0.9	1.11119	1.23146

Table 5.16 Relative efficiencies, g=0 and h=0.5

		g=0, h=0.5	
		over HD	over R
n=20	q=0.1	0.81206	1.02860
	q=0.5	1.00466	1.01084
	q=0.9	1.49783	1.05383
n=40	q=0.1	0.92285	1.05482
	q=0.5	1.00308	1.00416
	q=0.9	1.21385	1.07138

When the population distribution is symmetric and heavy tailed, the results are given in Table 5.16. NO quantile estimator is more efficient than R quantile estimator for all cases, however there are two exceptions for HD quantile estimator. Nevertheless, the relative efficiency results for q=0.1 is not so far from 1.

Table 5.17 shows the relative efficiency results for asymmetric and light tailed distribution. Unfortunately, NO quantile estimator is more efficient than the others in three experimental settings.

Lastly, the relative efficiencies for asymmetric and heavy tailed distribution are given in Table 5.18. For this distribution, NO quantile estimator is more efficient than HD and R for all cases, except for three cases. Note that, these exceptions are the values that are very close to 1 such as 0.99902. In other saying, for these cases NO quantile estimator is almost as efficient as the other estimators.

Table 5.17 Relative efficiencies, g=0.5 and h=0

		g=0.5, h=0	
		over HD	over R
n=20	q=0.1	0.51581	0.46756
	q=0.5	0.99870	1.03346
	q=0.9	0.14283	0.46256
n=40	q=0.1	0.65111	0.56500
	q=0.5	1.06729	1.08051
	q=0.9	0.43523	0.58080

Table 5.18 Relative efficiencies, g=0.5 and h=0.5

		g=0.5, h=0.5	
		over HD	over R
n=20	q=0.1	4.29891	4.97690
	q=0.5	0.98569	0.99296
	q=0.9	13.01179	5.30639
n=40	q=0.1	1.47630	1.51427
	q=0.5	0.99902	1.00073
	q=0.9	1.58808	1.49951

In general, NO quantile estimator is more efficient than Harrell Davis quantile estimator in 15 times out of 24 cases, whereas more efficient than the default quantile estimator of R programming language in 19 times out of 24 cases. In addition to this, there are four results that are smaller than, but very close to 1 like 0.99902. Shortly, NO quantile estimator is more efficient than other quantile estimators for most of the cases, and for the remaining cases the efficiency of NO quantile estimator can compete well with the other estimators.

5.2.3 Actual Type I Error Rates Results

In this section, the actual type I error rates that are obtained from different simulation studies are given. Two independent groups are compared through quantiles with two different approaches in which the quantiles are utilized individually or simultaneously. Note that, in each table the results that fall within the Bradley's interval highlighted (Bradley, 1978).

5.2.3.1 Comparison Through Quantiles Individually

When comparing two independent groups through quantiles individually, the method given in Section 4.1 is applied with the quantile estimators NO, HD and R. The performance of these quantile estimators are compared in terms of controlling the actual type I error rate within the stated bounds (0.025, 0.075) at the $\alpha = 0.05$ nominal significance level (Bradley, 1978).

The actual type I error rates for the continuous distributions are given in Table 5.19, Table 5.20, Table 5.21 and Table 5.22. The tables are organized on the basis of population distribution and for all of these four tables, two independent groups come from a population that have the same distribution. Note that, two independent groups are having the same sample size.

In Table 5.19, the actual type I error rates are given for standard normal distribution. When comparing medians of two independent groups, all estimators could control the actual type I error rates no matter the sample size is. On the other side, when comparing the extreme quantiles with small sample sizes, only NO quantile estimator could save the nominal significance level. Increasing the sample size could improve the results of R, however HD could not control the stated significance even with large samples for q=0.05 and q=0.95.

Table 5.19 Actual Type I error rates, g=0 and h=0

		q=0.05	**q=0.1**	**q=0. 5**	**q=0.9**	**q=0.95**
n=10	**R**	0.1142	0.0620	0.0346	0.0661	0.1125
	HD	0.1465	0.1249	0.0505	0.1222	0.1386
	NO	0.0418	0.0578	0.0484	0.0643	0.0372
n=20	**R**	0.0633	0.0346	0.0337	0.0344	0.0591
	HD	0.1157	0.0851	0.0467	0.0787	0.1140
	NO	0.0576	0.0616	0.0450	0.0591	0.0548
n=40	**R**	0.0315	0.0296	0.0351	0.0307	0.0365
	HD	0.0840	0.0515	0.0426	0.0492	0.0894
	NO	0.0596	0.0502	0.0466	0.0490	0.0553

The actual type I error rates for symmetric and heavy tailed distribution, asymmetric and light tailed distribution, asymmetric and heavy tailed distribution are given in Table 5.20, Table 5.21 and Table 5.22 respectively. For these distribution settings, the results are very similar with Table 5.19. Namely, increasing the sample size leads to control actual type I error rates with R estimator, whereas HD has some problems even with n=40. At the same time, NO quantile estimator could control its results within the interval (0.025, 0.075) even when the extreme quantiles are compared with small sample sizes.

Table 5.20 Actual Type I error rates, g=0 and h=0.5

		q=0.05	**q=0.1**	**q=0. 5**	**q=0.9**	**q=0.95**
n=10	**R**	0.1238	0.0702	0.0341	0.0746	0.1245
	HD	0.1488	0.1314	0.0432	0.1323	0.1469
	NO	0.0498	0.0702	0.0489	0.0729	0.0505
n=20	**R**	0.0695	0.0366	0.0338	0.0364	0.0733
	HD	0.1243	0.0983	0.0417	0.1001	0.1266
	NO	0.0648	0.0700	0.0423	0.0667	0.0630
n=40	**R**	0.0341	0.0299	0.0358	0.0319	0.0334
	HD	0.0993	0.0517	0.0466	0.0546	0.0996
	NO	0.0590	0.0563	0.0444	0.0550	0.0606

Table 5.21 Actual Type I error rates, g=0.5 and h=0

		q=0.05	**q=0.1**	**q=0. 5**	**q=0.9**	**q=0.95**
n=10	**R**	0.1047	0.0603	0.0337	0.0674	0.1172
	HD	0.1419	0.1136	0.0455	0.1331	0.1410
	NO	0.0398	0.0540	0.0512	0.0622	0.0445
n=20	**R**	0.0600	0.0363	0.0340	0.0387	0.0619
	HD	0.1086	0.0752	0.0415	0.0928	0.1167
	NO	0.0554	0.0591	0.0441	0.0618	0.0576
n=40	**R**	0.0349	0.0312	0.0376	0.0312	0.0332
	HD	0.0757	0.0507	0.0451	0.0497	0.0890
	NO	0.0578	0.0513	0.0434	0.0466	0.0555

Furthermore, in the presence of tied values, two independent groups are compared through quantiles individually by applying the method in Section 4.1 with NO quantile estimator. The actual type I error rates are given for the beta binomial distribution with parameters r=3, s=9, m=10 and r=9, s=9, m=10 in Table 5.23. The other settings of beta binomial distributions are omitted, since they are same as Table 5.23 except for 4

results out of 60 different experimental settings. That is, the ability of controlling the actual type I error rate is not affected by the existence of tied values.

Table 5.22 Actual Type I error rates, g=0.5 and h=0.5

		q=0.05	**q=0.1**	**q=0. 5**	**q=0.9**	**q=0.95**
n=10	**R**	0.1218	0.0711	0.0342	0.0729	0.1244
	HD	0.1427	0.1300	0.0428	0.1299	0.1436
	NO	0.0453	0.0723	0.0471	0.0697	0.0526
n=20	**R**	0.0661	0.0380	0.0340	0.0386	0.0678
	HD	0.1215	0.0952	0.0424	0.0996	0.1186
	NO	0.0618	0.0676	0.0425	0.0703	0.0718
n=40	**R**	0.0354	0.0299	0.0341	0.0313	0.0337
	HD	0.1015	0.0561	0.0429	0.0555	0.0934
	NO	0.0625	0.0524	0.0430	0.0529	0.0618

Table 5.23 Actual Type I error rates, Beta Binomial distributions

		n=10	**n=20**	**n=40**
r=3 s=9 m=10	**q=0.05**	0.0395	0.0467	0.0596
	q=0.1	0.0653	0.0674	0.0479
	q=0.5	0.0480	0.0466	0.0479
	q=0.9	0.0503	0.0598	0.0597
	q=0.95	0.0289	0.0356	0.0626
r=9 s=9 m=10	**q=0.05**	0.0338	0.0418	0.0648
	q=0.1	0.0565	0.0655	0.0516
	q=0.5	0.0498	0.0439	0.0384
	q=0.9	0.0548	0.0600	0.0497
	q=0.95	0.0368	0.0394	0.0618

As a different point of view, when comparing two independent groups through quantiles by using the method in Section 4.1 with NO quantile estimator, sampling from different distributions are considered. Corresponding actual type I error rates are given in Table 5.24 and Table 5.25 for n=10 and n=20, respectively. When the shapes of two population distributions are not very different from each other, the results could be saved within the bounds 0.025 and 0.075. When the groups are compared through medians, the actual type I error rates were close to the nominal level, even with small sample sizes. However, when comparing the tails there were some uncontrolled results. For example, when one the first group comes from a symmetric distribution whereas the second comes from an asymmetric one, the actual type I error rate could

exceed 0.2. So, when the populations are not identical the researcher should be careful especially with the lower and upper quantiles.

Table 5.24 Actual type I error rates, continuous distributions, n=10

Group 1	Group 2	q=0.1	q=0.5	q=0.9
g=0 & h=0.2	**g=0.2 & h=0.2**	0.0672	0.0544	0.0673
g=0 & h=0.2	**g=0.2 & h=0**	0.0854	0.0492	0.0641
g=0.2 & h=0	**g=0.2 & h=0.2**	0.0699	0.0477	0.0746
g=0 & h=0	**g=0.2 & h=0**	0.0740	0.0492	0.0698
g=0 & h=0	**g=0 & h=0.2**	0.0710	0.0481	0.0780
g=0 & h=0	**g=0.2 & h=0.2**	0.0626	0.0496	0.0891
g=0 & h=0	**g=0.5 & h=0.5**	0.0708	0.0532	0.2026
g=0.5 & h=0	**g=0 & h=0.5**	0.1808	0.0522	0.0724

Table 5.25 Actual type I error rates, continuous distributions, n=20

Group 1	Group 2	q=0.1	q=0.5	q=0.9
g=0 & h=0.2	**g=0.2 & h=0.2**	0.0610	0.0434	0.0689
g=0 & h=0.2	**g=0.2 & h=0**	0.0788	0.0489	0.0613
g=0.2 & h=0	**g=0.2 & h=0.2**	0.0713	0.0426	0.0694
g=0 & h=0	**g=0.2 & h=0**	0.0615	0.0413	0.0642
g=0 & h=0	**g=0 & h=0.2**	0.0679	0.0439	0.0643
g=0 & h=0	**g=0.2 & h=0.2**	0.0631	0.0390	0.0795
g=0 & h=0	**g=0.5 & h=0.5**	0.0716	0.0480	0.1254
g=0.5 & h=0	**g=0 & h=0.5**	0.1282	0.0448	0.0720

5.2.3.2 Comparison Through Multiple Quantiles

In this section, two independent groups are compared through multiple quantiles simultaneously with the proposed method in Section 4.2. The method compares two independent groups via the quantiles q=0.1, q=0.5 and q=0.9 by means of NO quantile estimator.

Table 5.26 represents the actual type I error rates for continuous distributions, whereas results for discrete distributions are given in Table 5.27.

Table 5.26 Actual Type I error rates, continuous distributions

	n=10	n=20	n=40
g=0 & h=0	0.0583	0.0662	0.0526
g=0 & h=0.5	0.0349	0.0319	0.0131
g=0.5 & h=0	0.0463	0.0571	0.0464
g=0.5 & h=0.5	0.0260	0.0290	0.0120
LN(0,1)	0.0303	0.0355	0.0267
Gamma(1,1)	0.0383	0.0442	0.0389

Table 5.27 Actual Type I error rates, discrete distributions

	n=10	n=20	n=40
Beta binomial r=1 s=9 m=10	0.0301	0.0431	0.0621
Beta binomial r=2 s=9 m=10	0.0491	0.0558	0.0373
Beta binomial r=3 s=9 m=10	0.0519	0.0692	0.0743
Beta binomial r=9 s=9 m=10	0.0578	0.0694	0.0581

In Table 5.26, 18 different experimental cases are obtained by combining different population distributions with different sample sizes. When comparing two independent groups through quantiles via NO estimator based on a percentile bootstrap method, the actual type I error rates fall within the stated interval in almost all cases. Using a heavy tailed distribution (h=0.5) with a large sample size (n=40) gives only exception.

In Table 5.27, the proposed method can control its actual type I error rates for all of the 12 different discrete distribution settings. That is, the ability of controlling the type I error rate is not effected by the existence of tied values.

When two independent groups follow different population distributions, the actual type I error rates are investigated for sample sizes n=10, n=20 and n=40 in Table 5.28. When the populations have different skewness values, the results could not be controlled within the limits 0.025 and 0.075 with small sample sizes. Hence, when comparing two independent groups with the proposed method given in Section 4.2 that uses NO quantile estimator in conjunction with a percentile bootstrap method, if two populations differ the minimum sample size should be n=40 in order to control actual type I error rates.

Table 5.28 Actual Type I error rates, different continuous distributions

Group 1	Group 2	n=10	n=20	n=40
g=0 & h=0.2	g=0.2 & h=0.2	0.0507	0.0549	0.0364
g=0 & h=0.2	g=0.2 & h=0	0.0674	0.0727	0.0489
g=0.2 & h=0	g=0.2 & h=0.2	0.0596	0.0687	0.0441
g=0 & h=0	g=0.2 & h=0	0.0599	0.0737	0.0538
g=0 & h=0	g=0 & h=0.2	0.0612	0.0715	0.0492
g=0 & h=0	g=0.2 & h=0.2	0.0692	0.0782	0.0502
g=0 & h=0	g=0.5 & h=0	0.0804	0.0924	0.0635
g=0 & h=0	g=0 & h=0.5	0.1133	0.1026	0.0500
g=0 & h=0	g=0.5 & h=0.5	0.1595	0.1134	0.0510
g=0.5 & h=0	g=0 & h=0.5	0.1188	0.0971	0.0459
g=0.5 & h=0	g=0.5 & h=0.5	0.1035	0.0873	0.0386
g=0 & h=0.5	g=0.5 & h=0.5	0.0604	0.0451	0.0176
g=0 & h=0	LN(mean=0, sd=1)	0.1742	0.1206	0.0695
g=0 & h=0.5	LN(mean=0, sd=1)	0.1373	0.0978	0.0430
g=0.5 & h=0	LN(mean=0, sd=1)	0.0742	0.0677	0.0439
g=0.5 & h=0.5	LN(mean=0, sd=1)	0.0687	0.0643	0.0318
g=0.2 & h=0	LN(mean=0, sd=1)	0.1299	0.1022	0.0605
g=0 & h=0.2	LN(mean=0, sd=1)	0.1457	0.1118	0.0565
g=0.2 & h=0.2	LN(mean=0, sd=1)	0.1040	0.0848	0.0505
g=0 & h=0	GAMMA(rate=1, shape=1)	0.1180	0.1026	0.0720
g=0 & h=0.5	GAMMA(rate=1, shape=1)	0.1687	0.1101	0.0498
g=0.5 & h=0	GAMMA(rate=1, shape=1)	0.0534	0.0603	0.0490
g=0.5 & h=0.5	GAMMA(rate=1, shape=1)	0.1437	0.1062	0.0463
g=0.2 & h=0	GAMMA(rate=1, shape=1)	0.0803	0.0827	0.0624
g=0 & h=0.2	GAMMA(rate=1, shape=1)	0.1296	0.1030	0.0574
g=0.2 & h=0.2	GAMMA(rate=1, shape=1)	0.0989	0.0888	0.0517

On the other side, the effect of heteroscedasticity on actual type I error rates is investigated in Table 5.29. In total 21 different experimental settings are considered that are combining different population distributions with different sample sizes. In each of these settings, two independent groups come from a population having the same distribution whereas the second group has the variance which is 2.25 times of the variance of first one. The actual type I error rates could be controlled under heteroscedasticity except in four cases out of 21.

Table 5.29 Actual type I error rates, continuous distributions, heteroscedastic case

	g=0 h=0	**g=0 h=0.2**	**g=0.2 h=0**	**g=0.2 h=0.2**	**g=0 h=0.5**	**g=0.5 h=0**	**g=0.5 h=0.5**
n=10	0.0737	0.0672	0.0710	0.0616	0.0560	0.0671	0.0615
n=20	0.0781	0.0700	0.0781	0.0649	0.0419	0.0698	0.0382
n=40	0.0619	0.0395	0.0645	0.0354	0.0173	0.0511	0.0139

Some additional simulations were conducted by using the same theoretical distributions with unequal sample sizes. 21 experimental cases were investigated by using g-and-h distributions where $n_1 = 10$ and $n_2 = 20$, $n_1 = 10$ and $n_2 = 40$, $n_1 = 20$ and $n_2 = 40$ were chosen as sample sizes. When the sample sizes are too small or too far from each other, the actual type I error rates could not be controlled, hence using the proposed method in balanced designs is suggested.

5.3 Real Data Examples

For comparing two independent groups through quantiles by using NO quantile estimator in conjunction with a percentile bootstrap, the method given in Section 4.1 and the proposed method in Section 4.2 are applied to two real datasets.

The first dataset is obtained from a study by Johnson, Tateishi & Hoan (2013) in which they developed a new method for detecting diseased pine and oak trees in high - resolution multispectral satellite imaginary (Johnson et al., 2013). The data consists of image segments that are generated by segmenting the pansharpened image. For each image segment, mean spectral values for the green, red and near infrared bands were calculated. However, in this thesis two variables were considered which are related to green and near infrared bands and they are called as Mean Green and Mean NIR, respectively. Some descriptive statistics are given in Table 5.30 and their histograms are given in Figure 5.10 (Johnson et al., 2013).

For both Mean NIR and Mean Green data sets, sample sizes are chosen as n=10, n=20 and n=40. In addition to median (q=0.5), q=0.1 is compared as a lower quantile and q=0.9 is compared as an upper quantile. The actual type I error rates are obtained

by 1000 replications for both methods and corresponding results are given in Table 5.31.

Table 5.30 Descriptive Statistics of Mean Green and Mean NIR

	N	Mean	Median	Standard deviation	Skewness	Kurtosis
Mean Green	4339	233.907	221.455	60.758	5.270	40.976
Mean NIR	4339	534.105	528.500	154.495	0.103	-0.406

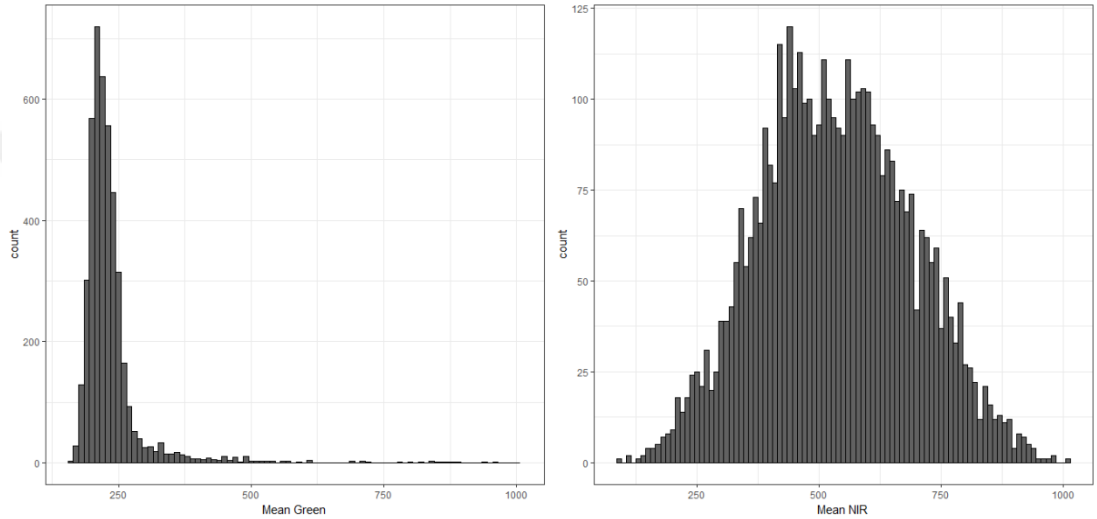

Figure 5.10 Histograms of Mean Green and Mean NIR

Table 5.31 Actual Type I error rates, Mean Green and Mean NIR datasets

		Proposed method	**q=0.1**	**q=0.5**	**q=0.9**
Mean Green	**n=10**	0.030	0.052	0.049	0.069
	n=20	0.041	0.062	0.031	0.064
	n=40	0.030	0.043	0.043	0.045
Mean NIR	**n=10**	0.040	0.058	0.049	0.074
	n=20	0.065	0.064	0.041	0.062
	n=40	0.067	0.056	0.055	0.049

In Table 5.31, the actual type I error rates that are obtained by the proposed method in Section 4.2 are given in the first column, whereas the next three columns consist of the results that are obtained by comparing two independent groups through quantiles individually. NO quantile estimator could save the actual type I error rates within the Bradley's interval in all cases with two methods (Bradley, 1978). Namely, the actual

type I error rates were controlled for not only the Mean NIR data, but also for the Mean Green data. In other words, NO quantile estimator performs well in real data applications even if the data is non-normal.

As a second example, another real data is considered to rearticulate the emphasis of comparing two independent groups through quantiles instead of a single measure of location. The corresponding data set is obtained from a work related with soil water content (% water by volume) for independent random samples of soil from two experimental fields growing bell peppers. Some descriptive statistics of soil water contents for field 1 and field 2 are given in Table 5.32. Their histograms are also given in Figure 5.11 (Two Independent Data Sets, 2018).

Table 5.32 Descriptive statistics of field 1 and field 2

	N	Mean	Median	Standard deviation	Skewness	Kurtosis
Field 1	72	11.41528	11.05	2.081417	0.6365722	-0.3124996
Field 2	80	10.65125	10.7	3.027469	1.538486	5.940143

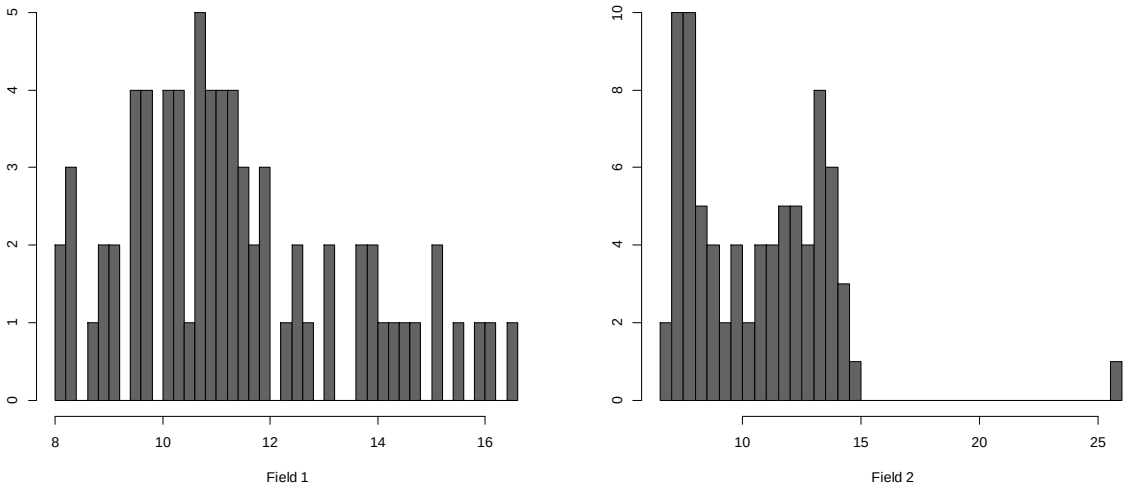

Figure 5.11 Histograms of field 1 and field 2

With the intent of comparing two independent groups, firstly two samples t test is conducted and p-value is found as 0.07493 which demonstrates there is no significant difference between two population means at the significance level α=0.05. Conversely, when these groups are compared via quantiles using NO estimator using the proposed method in Section 4.2, p-value is obtained as 0.0015. Accordingly, $H_0 : \theta_{q1} = \theta_{q2}$ where θ_{qj} is the q'th quantile corresponding to the j'th group (j = 1, 2) is rejected and the proposed method states that these groups differ in a manner that is not discovered by a single measure of location. After stating that the groups are differing in somehow, a researcher may be interested in to detect the difference occur in which quantile. This interest is addressed to comparing these two groups through quantiles individually with the method given in Section 4.1 by using NO quantile estimator. According to related results, NO quantile estimator detects significant differences between the lower tails of the groups, namely for q=0.1with the p-value equals 0.

5.4 Some Numerical Illustrations

In this section of the thesis, two numerical examples are used to emphasize the utility of comparing two independent groups through quantiles instead of considering one reference point. Actually, the first example aims to illustrate the effect of outliers on Student's t test and Welch's test, whereas the second one is concerned with the situation that tied values occur.

The first example is about two independent samples which are called as A and B. Suppose there are two independent groups A and B where Group A = 4, 5, 6, 7, 8, 9, 10, 11, 12, 13 and Group B = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10. $\bar{X}_1 = 8.5$ and $\bar{X}_2 = 5.5$ are the sample means of the groups respectively. When A and B are compared by Student's t test, p-value is found as 0.03985 at nominal level $\alpha = 0.05$ which provides rejecting the hypothesis of equal means. Moreover, with Welch's heteroscedastic test results in the same p-value. On the other side, these samples are compared through quantiles by using the approach given in Section 4.1, namely NO quantile estimator is used in conjunction with a percentile bootstrap method. NO quantile estimator found

significant difference for q=0.1 and q=0.9 with the corresponding p-values 0.05 and 0.044.

If A is updated as A*= 4, 5, 6, 7, 8, 9, 10, 11, 12, 23, in which the last observation increased ten units, again a significant difference would expected from the Student's t test. However, it resulted in a p-value=0.05597 which indicates there is no significant difference. In a similar manner, Welch's heteroscedastic method for means neither could find significant difference with p-value equals 0.06014. However, there were significant differences between the tails of A* and B. NO quantile estimator rejected the null hypothesis that equality of two population quantiles for q=0.1 and q=0.9 with p-values 0.05 and 0.041, respectively.

Finally, A is updated to A**= 4, 5, 6, 7, 8, 9, 10, 11, 12, 33. Once again a significant difference would expected from the Student's t test, but the p-value was obtained as 0.09088. Besides, Welch's heteroscedastic method for means yielded the p-value equals 0.1007. If quantiles are considered, clearly there were significant differences between the lower and upper quantiles as expected. p-values of 0.05 and 0.039 were obtained by NO quantile estimator for q=0.1 and q=0.9, respectively.

In substance, by increasing the last observation in first group the difference between sample means gets large. However, this situation also inflates the sample variance and that's why a single outlier can mask differences between two independent groups when the reference point is sample mean. That is, when sample mean is considered as the reference point for detecting differences between two independent groups both Student's t test and Welch test could give misleading results.

As mentioned before when comparing two independent groups, presence of tied values can cause problems. To demonstrate, the following example which consists of two independent samples including tied values is investigated. Let X= 1, 1, 2, 2, 1, 1, 1, 9, 1, 1, 1, 1, 1, 1, 2, 1, 1, 1, 1, 1, 1, 2, 1 and Y= 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1 with $n_1 = 23$ and $n_2 = 19$ (Two Independent Data Sets, 2017). With the aim of comparing two independent groups, Student's t test was conducted and the p-

value was found as 0.1831. Similarly, Welch test gave the p-value=0.1495. However, when comparison through quantiles was considered, NO quantile estimator found significant differences between the upper tails (q=0.9) and the middle (q=0.5). Hence, when there are tied values among the observations, comparing two independent groups through quantiles instead of one reference point is a better idea which avoids missing important differences.

5.5 Conclusion

In this thesis, comparing two independent groups is investigated from a different point of view. In the first instance, the methods for comparing two independent groups based on one reference point are studied where the Student's t test is the most commonly known approach. After explaining its advantageous and disadvantageous properties, the need for a more robust approach discussed and its existing robust alternatives in the literature such as Welch's test and Yuen test are briefly explained (Welch, 1938; Yuen, 1974).

However, this thesis aims drawing attention to importance of comparing two independent groups through more than one reference point, instead of considering only a single measure of location. Comparing two independent groups by using one reference point, not only provides inappropriate results but also leads missing the important differences between two groups especially in the tails of them. Hence, considering more than one reference point, namely quantiles, provides a deeper insight into understanding how the groups compare.

In accordance with the main purpose of comparing two independent groups via quantiles, some important works on quantile estimation are investigated. Thereby, the idea behind obtaining a quantile estimator is investigated and possible drawbacks of existing quantile estimators are determined. Attention is focused on the need for a new quantile estimator and following this, NO quantile estimator is proposed.

After explaining some theoretical results on asymptotic normality of the proposed estimator, asymptotic properties such as normality, consistency, unbiasedness, variance and mean squared error are investigated practically by conducting a simulation study. The results verified that NO quantile estimator has very desirable asymptotic properties. On the other hand, relative efficiencies of NO quantile estimator over Harrell Davis quantile estimator and the default quantile estimator of R programming language are examined. In most of the simulation settings, NO quantile estimator is more efficient than these two estimators.

NO quantile estimator has satisfactory asymptotic properties and it is more efficient than some of the existing quantile estimators. Hence, it is possible to use it along with a hypothesis testing procedure. Firstly, it is used in conjunction with an available percentile bootstrap method on the purpose of comparing two independent groups through quantiles individually. NO outperforms Harrell Davis and R quantile estimators in terms of controlling the actual type I error rates.

On the other hand, the essential objective of this thesis is to propose a new method for comparing two independent groups that considers more than one reference point simultaneously. The newly proposed method compares two independent groups by considering three quantiles which consists of a lower, a middle and an upper one.

The performance of the proposed method with NO quantile estimator is investigated for wide range of experimental settings. The significant findings of the various simulation studies can be listed as follows. The proposed method could control the actual type I error rates even with small sample sizes. It would not be effected by heteroscedasticity if the ratio of the variances is not too large. If two independent groups come from populations that are having different theoretical distributions, the actual type I error rates could be controlled provided the sample sizes are large enough. Existence of the tied values among the observations is an important issue in applied statistics, but the proposed method could save its results within the stated bounds even in the presence of tied values.

Additionally, the proposed method is applied not only for theoretical distributions, but also for real data sets. As it would be expected, the proposed method could control the actual type I error rates for both symmetric shaped and skewed datasets. Furthermore, to emphasize the importance of comparing groups through quantiles rather than comparing based on one reference point, some numerical illustrations are given. According to related results, the new method can discover the differences between two groups that are not found by the methods considering one reference point.

To conclude, the newly proposed NO quantile estimator provides more efficient estimates for the population quantiles especially for extreme ones. It is computationally practical as well as it has desirable asymptotic properties. In addition, it is possible to use NO quantile estimator in applied statistics. It is appropriate to use it for comparing two independent groups to make a more sensitive comparison and to control the stated nominal significance level. Furthermore, the proposed method that compares two independent groups through multiple quantiles simultaneously via NO quantile estimator allows a deeper insight into comparing two groups. To avoid missing important differences between the subpopulations of groups, comparing them by the proposed method rather than comparing using a single measure of location is a better idea.

As a final conclusion, this thesis contributes to existing quantile estimation theory by introducing NO quantile estimator. Furthermore, the proposed method for comparing two independent groups by means of NO quantile estimator contributes to understanding of two group comparison from a different point of view. Using NO quantile estimator for estimating a population quantile and using it along with comparing two independent groups might be an appropriate alternative for all researchers.

REFERENCES

Al-Kenani, A., & Yu, K. (2012). New bandwidth selection for kernel quantile estimators. *Journal of Probability and Statistics, 2012,* 1-18.

Arnold, B. C., Balakrishnan, N., & Nagaraja, H. N. (1992). *A first course in order statistics* (1st ed.). New Jersey: John Wiley & Sons Inc.

Azzalini, A. (1981). A note on the estimation of a distribution function and quantiles by a kernel method. *Biometrika*, *68* (1), 326-328.

Bradley, J. V. (1978). Robustness?. *British Journal of Mathematical and Statistical Psychology*, *31*, 144-152.

Brewer, K. R. W. (1986). *Likelihood based estimation of quantiles and density estimation.* Unpublished manuscript.

Cheng, C. (1995). The Bernstein polynomial estimator of a smooth quantile function. *Statistics & Probability Letters*, *24*, 321-330.

Dielman, T., Lowry, C., & Pfaffenberger, R. (1994). A comparison of quantile estimators. *Communications in Statistics-Simulation and Computation*, *23*, 355-371.

Efromovich, S. (1999). *Nonparametric curve estimation: methods, theory, and applications* (1st ed.). New York: Springer Verlag.

Erceg-Hurn, D. M., & Steed, L. G. (2011). Does exposure to cigarette health warnings elicit psychological reactance in smokers?. *Journal of Applied Social Psychology*, *41* (1), 219-237.

Falk, M. (1984). Relative deficiency of kernel type estimators of quantiles. *The Annals of Statistics*, *12* (1), 261-268.

Falk, M. (1985). Asymptotic normality of the kernel quantile estimator. *The Annals of Statistics*, *13* (1), 428-433.

Farin, G. E. (2002). *Curves and surfaces for CAGD: a practical guide*. (5th ed.). San Francisco: Morgan Kaufmann.

Harrell, F. E., & Davis, C. E. (1982). A new distribution-free quantile estimator. *Biometrika, 69*, 635–640.

Hoaglin, D. C. (1985). Summarizing shape numerically: The g-and-h distribution. In D. Hoaglin, F. Mosteller, & J. Tukey, (Ed.). *Exploring data tables trends and shapes* (1st ed.) (461-513). New Jersey: Wiley.

Hodges, J. L., & Lehmann, E. L. (1970). Deficiency. *The Annals of Mathematical Statistics*, *41*, 783-801.

Hyndman, R. J., & Fan, Y. (1996). Sample quantiles in statistical packages. *The American Statistician, 50*, 361–365.

Johnson, B. A., Tateishi, R., & Hoan, N. T. (2013). A hybrid pansharpening approach and multiscale object-based image analysis for mapping diseased pine and oak trees. *International Journal of Remote Sensing*, *34*, 6969-6982.

Kaigh, W. D., & Cheng, C. (1991). Subsampling quantile estimators and uniformity criteria. *Communications in Statistics - Theory and Methods, 20*, 539–560.

Kaigh, W. D., & Lachenbruch, P. A. (1982). A generalized quantile estimator. *Communications in Statistics - Theory and Methods, 11*, 2217-2238.

Liu, R. Y., & Singh, K. (1997). Notions of limiting P values based on data depth and bootstrap. *Journal of the American Statistical Association, 92* (437), 266-277.

Mc Cune, E. D., & Luna Mc Cune, S. (1991). Jackknifed kernel quantile estimators. *Communications in Statistics - Theory and Methods, 20*, 2719-2725.

Parrish, R. S. (1990). Comparison of quantile estimators in normal sampling. *Biometrics, 46* (1), 247-257.

Parzen, E. (1979). Nonparametric statistical data modeling. *Journal of the American Statistical Association, 74* (365), 105-121.

Perez, J. M., & Palacín, A. F. (1987). Estimating the quantile function by Bernstein polynomials. *Computational Statistics & Data Analysis, 5*, 391-397.

Pepelyshev, A., Rafajłowicz, E., & Steland, A. (2014). Estimation of the quantile function using Bernstein–Durrmeyer polynomials. *Journal of Nonparametric Statistics, 26* (1), 1-20.

Ramsey, P. H. (1980). Exact Type I error rates for robustness of Student's t test with unequal variances. *Journal of Educational Statistics, 5*, 337–349.

Reiss, R. D. (1980). One-sided tests for quantiles in certain nonparametric models. *Nonparametric Statistical Inference, 8* (1), 759-772.

Sfakianakis, M. E., & Verginis, D. G. (2008). A new family of nonparametric quantile estimators. *Communications in Statistics - Simulation and Computation, 37*, 337-345.

Sheather, S. J., & Marron, J. S. (1990). Kernel quantile estimators. *Journal of the American Statistical Association, 85*, 410-416.

Steinberg, S. M., & Davis, C. E. (1985). Distribution-free confidence intervals for quantiles in small samples. *Communications in Statistics - Theory and Methods, 14,* 979-990.

Stewart, P. W. (1989). Extension of the harrell-davis quantile estimator to finite populations. *Communications in Statistics - Theory and Methods, 18,* 853-875.

Stigler, S. M. (1974). Linear functions of order statistics with smooth weight functions. *The Annals of Statistics, 2*, 676-693.

Two Independent Data Sets (large and small sample) Birth Rate. (n.d). Retrieved February 14, 2017, from http://college.cengage.com/mathematics/brase/understanding_basic/4e/assets/datasets/tvis/tvis08.html

Two Independent Data Sets (large and small sample) Agriculture, Water Content of Soil. (n.d). Retrieved August 30, 2018, from http://college.cengage.com/mathematics/brase/understanding_basic/4e/assets/datsets/tvis/tvis05.html

Welch, B. L. (1938). The significance of the difference between two means when the population variances are unequal. *Biometrika, 29,* 350-362.

Wilcox, R. R. (1995). Comparing two independent groups via multiple quantiles. *The Statistician, 44* (1), 91-99.

Wilcox, R. R. (2011). *Modern statistics for the social and behavioral sciences: A practical introduction* (1st ed.). New York: CRC press.

Wilcox, R. R., Erceg-Hurn, D. M., Clark, F., & Carlson, M. (2014). Comparing two independent groups via the lower and upper quantiles. *Journal of Statistical Computation and Simulation, 84,* 1543-1551.

Wilcox, R. R. (2016). *Understanding and applying basic statistical methods using R* (1st ed.). New Jersey: John Wiley & Sons.

Wilcox, R. R. (2017). *Introduction to robust estimation and hypothesis testing* (4th ed.). United States: Academic press.

Yang, S. S. (1985). A smooth nonparametric estimator of a quantile function. *Journal of the American Statistical Association*, *80*, 1004-1011.

Yoshizawa, C. N., Sen, P. K., & Davis, C. E. (1985). Asymptotic equivalence of the Harrell-Davis median estimator and the sample median. *Communications in Statistics - Theory and Methods*, *14*, 2129-2136.

Yuen, K. K. (1974). The two-sample trimmed t for unequal population variances. *Biometrika*, *61* (1), 165-170.

Zelterman, D. (1990). Smooth nonparametric estimation of the quantile function. *Journal of Statistical Planning and Inference*, *26*, 339-352.