KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY

A ROBUST PROCEDURE FOR THE FIT OF ONEWAY ANALYSIS OF VARIANCE (ANOVA) MODELS UNDER UNCORRELATED ERRORS

BY

EBENEZER AFRIFA-YAMOAH

A THESIS SUBMITTED TO THE DEPARTMENT OF MATHEMATICS, KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY IN PARTIAL FULFILMENT OF THE REQUIREMENT FOR THE DEGREE OF M.PHIL MATHEMATICAL STATISTICS

November, 2016

Declaration

I hereby declare that this submission is my own work towards the award of the M.Phil degree and that, to the best of my knowledge, it contains no material previously published by another person nor material which had been accepted for the award of any other degree of the university, except where due acknowledgement had been made in the text.

Ebenezer Afrifa-Yamoah	....	....
Student	Signature	Date

Certified by:

Dr. Gabriel Asare Okyere	....	....
Supervisor	Signature	Date

Certified by:

Dr. R. K. Avuglah	....	....
Head of Department	Signature	Date

Dedication

To my lovely parents,

Howard and *Agnes*.

Acknowledgement

I give the Almighty God glory and honour. Thank you, My Maker. I am greatly indebted to my supervisor, Dr. G.A. Okyere, for his technical support, understanding and encouragement. To the family: Howard, Agnes, Juliet, Victor, Emmanuel and Abigail, I appreciate your sacrifices, encouragement and prayers. A special thanks to Prof. Bashiru I.I. Saeed, Dean, Faculty of Applied Sciences, Kumasi Polytechnic, Silverius Kwasi Braku, André Flakke and Killian Asampana for their support.

Abstract

The possible dominance of basic assumption about underlying models on the analysis of data is of much concern to some statisticians (Anscombe (1967);Hogg (1974); Büning (1996)). The advocacy of distribution-free (nonparametric) tests for differences in location problems between samples has been emphasized over the past seven decades (Hao and Houser, 2011). This study develops a robust fitting procedure for one-way ANOVA models. Further investigation on Asymptotic Relative Efficiency (ARE) of this procedure and parametric F-test under class of distributions was carried out. In line with these objectives, 10,000 simulations were carried out for a one-way ANOVA model with three levels for size 5, 10, 15, and 20. Intralevel correlation coefficient $\rho = 0$ was considered in these simulations. The findings revealed that the parametric F-test for Oneway ANOVA model performed better than the non-parametric Adaptive test proposed for symmetric and moderate tailed distributions and then in symmetric and light tailed distributions with ARE between 2% and 55%. However, the Adaptive test outperformed the F-test in symmetric and asymmetric with varying tail weights distributions with ARE between 5% and 70%. Although, the F-test displayed superiority in efficiency in symmetric medium and light tailed distributions, the Adaptive test was more efficient in more broader class of continuous distribution.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Introduction

In our world of industrialization, competition for supremacy is the order of the day. Conjecturing has become a common playground and to establish the fanaticism or otherwise of these claims is a subject of global interest. Customers want to make a choice of the best products; companies are interested in, how effective are our products compared to others?; pharmaceuticals companies are interested in, which combinations of chemicals achieve the best results?; medical officer wants the results of clinical trials to certify a life and death procedures; the engineer wants to measure the efficiency of equipments; the agriculturist is interested in which fertilizer will give the maximum yield etc. Central in resolving these dilemmas is the statistician. Various techniques are employed in his quest to address these matters of global interest; notable among these techniques is the Analysis of Variance (ANOVA, hereafter) models.

ANOVA, is perhaps the most powerful statistical tool (Botton and Bon, 2009). It is a general method of analyzing data from designed experiments, whose objective is to test appropriate hypotheses about treatment means and to estimate them. The model for the data is represented as follows;

$$y_{ij} = \mu + \tau_i + \epsilon_{ij} \tag{1.1}$$

where $i = 1, 2, \cdots, a$, is the number of treatments considered for the experiment and $j = 1, 2, \cdots, n$, is the number of replications. The model (1.1) has μ as a parameter common to all treatments called the overall mean, τ_i is the parameter unique to the ith treatment called the ith treatment effect and ϵ_{ij} is the random error component that incorporates all other sources of variability in the experiment including measurement, variability arising from uncontrolled factors, differences between the experimental units to which the treatments are applied, and the general background in the process (such as variability over time, effects of the environment variables, and so forth). In testing hypothesis with this model, the errors are assumed to be normally and independently distributed with mean zero and variance σ^2, with constant variance for all levels of factor. This implies that the observations,

$$y_{ij} \sim N(\mu + \tau_i, \sigma^2)$$

and that the observations are mutually independent. The model describes two different situations with respect to the treatment effects. The choice of the treatments for the experiment could be specifically chosen or randomly selected from a population. If treatments are specifically chosen the resultant model is a fixed effect model otherwise random effect model (or components of variance model) (Montgomery, 2001). We are interested in testing the equality of the a treatment means, that is, $E(y_{ij}) = \mu + \tau_i = \mu_i, i = 1, 2, \cdots, a$. The appropriate hypothesis are;

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_a$$

$$H_1 : \mu_i \neq \mu_j$$

for at least one pair (i, j).

The statistical significance of the experiment is determined by the ratio of two variances which are independent of constant bias and scaling errors as well as the units used in expressing observation. The name ANOVA is derived from a partitioning of total variability into its component parts. ANOVA estimates three(3) sample variances: a total variance based on all the observation deviations from the grand mean, that is, sum of square total, SS_{Total}, an error variance based on all the observation deviations from their corresponding treatment means, that is, sum of square error, SS_{Error} and a treatment variance based on the deviations of treatment means from the grand mean, the result being multiplied by the number of observations in each treatment, that is, sum of square treatment, $SS_{Treatment}$. Therefore, we have $SS_{Total} = SS_{Error} + SS_{Treatment}$. Mathematically,

$$\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y_{..}}) = \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y_{i.}}) + n\sum_{i=1}^{a}(\bar{y_{i.}} - \bar{y_{..}}) \quad (1.2)$$

There are $an = N$ total observations, thus SS_{Total} has $N-1$ degrees of freedom. There are a levels of the factors (and a treatment means), so $SS_Treatment$ has $a-1$ degrees of freedom. Finally, within any treatment there are n replicates providing $n-1$ degrees of freedom with which to estimate the experimental error. Because there are a treatments, we have $a(n-1) = an - a = N - a$ degrees of freedom for error. In performing statistical analysis, we investigate a formal test of the hypothesis of no difference in treatment means. It is worthy to note that, because of the assumption of normality of the errors, ϵ_{ij} and its implied effect on the observations, y_{ij} (because y_{ij} is a linear combination of ϵ_{ij}), the SS_{Total} is a sum of squares is a normally distributed random variable; consequently SS_{Total}/σ^2 is distributed as chi-square with $N-1$ degrees of freedom; SS_{Error}/σ^2 is chi-square with $N-a$ degrees of freedom and that $SS_{Treatment}/\sigma^2$ is also chi-square with $a-1$ degrees of freedom if the null hypothesis $H_0 : \mu_i = 0$ is true. The $SS_{Treatment}/\sigma^2$ and SS_{Error}/σ^2 are independently distributed chi-square random variables are implied by Cochran's theorem, as the degrees of freedom for $SS_{Treatment}$ and SS_{Error} add to $N-1$, the total number of freedom. Therefore, if the null hypothesis of no difference in treatment means is true, the ratio,

$$F_0 = \frac{(SS_{Treatment}/\sigma^2)/(a-1)}{(SS_{Error}/\sigma^2)/(N-a)} = \frac{MS_{Treatment}}{MS_{Error}} \quad (1.3)$$

is distributed as F with $a-1$ and $N-a$ degrees of freedom.

1.2 Background of study

According to O'Gorman (2004), simulated studies have confirmed that most traditional tests like the F test and their confidence intervals are robust, at least as far as validity is concerned. That is, the traditional tests maintain their actual significance level close to their nominal significance level, and their confidence intervals maintain their coverage probabilities close to their nominal coverage probabilities and thus are relatively unaffected by violations of assumptions . Nevertheless, the traditional tests have a serious defect, if the distribution of the errors is non-normal, thus not robust for efficiency (O'Gorman, 2004). There is no doubt about the important role of the normal distribution in model formulation in the realms of statistics. The standard normal theory is good, provided the normal distribution is reasonably close to the real model for the problem at hand (Hogg, 1974). However, if the normal distribution is extreme with reference to the data at hand, the model formulated would be a poor one, a typical example was the poor performance of the least square estimator, $\bar{X}$, in the Princeton study of robust estimates of location in which 68 estimates were compared (Andrews et al., 1972), because the normal distribution was an extreme one in the broad class of models which were studied. The efficiency of the parametric version of hypothesis testing mostly depend on the assumption of the underlying distribution of the data, for instance, the assumption of normality will require the use of optimal test for one-, two- and k-sample location or scale problem such as t-test, F-test and Chi-square tests. Notably, there seems to be over-reliance on the normal distribution and its implied assumptions by the practising statistician in model formulation especially in ANOVA applications, as several works on the data at hand are swept under the carpet, because of the assumption of normality, which is often violated in practice (Büning, 1996).

Models are generally believed to be simplified representation of reality, they are abstract in nature and may be deterministic or probabilistic. Statistical models are made up with probabilistic component. Thus, much effort should be devoted into reducing to the bearest minimum the level of uncertainty associated with the model. Although, there does not exist a perfect model, yet the onus falls on us to formulate near perfect models. It is a well known fact that, in practice most models will seldomly fit exactly real situations. Thus, for the sake of applications, it seems ridiculous to try to get the last ounce of mathematical efficiency out of some assumed situation. According to Hogg (1974) a more realistic approach would be to seek statistical procedures good for a broad class of underlying models, but which are not necessarily best for any one of them. Such procedures are robust (that is, exhibit strength).

1.3 Statement of the Problem

Some statisticians, according to Hogg (1974), are now concerned that basic assumptions about underlying models might very well dominate the analysis of the data in many cases. The advocacy for distribution-free (nonparametric) tests for differences in locations problems between samples has been emphasized over the past seven decades (Hao and Houser, 2011). Anscombe (1967) stated emphatically that the disposition of the present-day statistician theorists to suppose that all error distributions are exactly normal can be ascribed to their ontological perception that normality is too good not to be true. Huber (1972) remarked that the dogma that measurement errors should be distributed according to the normal law is still widespread among the users of least squares and suggests that a more rational action would be to check whether they were compatible with a normal distribution and, if not, to develop a different theory of estimation. The reliability of models should be questioned, because no model in the realms of statistical analysis is sacred (Hogg, 1974).

1.4 Objective of the study

The thesis aims to develop a robust adaptive procedure for hypothesis testing of Oneway ANOVA models with uncorrelated errors and compare the asymptotic relative efficiency of the traditional parametric approach (F-test) to the adaptive (non-parametric) method.

1.5 Methodology

In this thesis, the data generated are assumed to have been derived from unknown continuous distributions. The data will be used to reveal the underlying distribution by assessing the associated values for skewness and tail weight proposed by (Hogg et al., 1975). This initial classifications will determine the specific scores functions on which inferences will be based. Simple adaptive procedures established by (Hogg et al., 1975) will be used in the estimation of the scores and will be classified according to the family of winsorised wilcoxons. These procedures will classify the data as; skewed left and light-tailed (LL), left skewed and moderate-tailed (LM), left skewed and heavy-tailed (LH), symmetric and light-tailed (SL), symmetric and moderate-tailed (SM), symmetric and heavy-tailed (SH), right skewed and light-tailed (RL), right skewed and moderate-tailed (RM) and right skewed and Heavy-tailed (RH). Four procedures are considered in this research. Three adaptive procedures and one parametric procedure. The adaptive procedures are the Pure-Hogg where adaption is done on the samples. We also considered adaption on residuals from the ordinary least square (OLS) and Wilcoxon fit. The F-test is the parametric procedure considered. Simulation studies will be performed to prove the dominance or otherwise of the adaptive procedures over the parametric approaches over a wide range of continuous distributions.

1.6 Justification of work

The asymptotic properties of statistical estimates and tests solely rely on the Central Limit Theorem (CLT), however, in practice sample sizes are not often large. One notable area of application of ANOVA models is in Medical Statistics. Clinical trials are conducted in the medical fields to compare the efficacy of methods and drugs. The high sensitivity of such process has often resulted in low sample sizes usage. This is just one of the important areas of application, thus, one could argue that, could the use of the F-test, which employs the assumption of normality of the data, be the optimal test to be conducted in such situations?

Researchers do not readily know the underlying distribution of the data at hand by mere observation to apply the most powerful rank tests. This thesis employs adaptive procedures which first reveals differences in the underlying distributions of the data, which its parametric counterparts do not, and further gives the appropriate rank test to be used in making inferences about the data. Hogg et al. (1975) and Büning (1996) have confirmed adaptive procedures displaying significant power over the parametric t-test with non-normal error distributions with large or moderate sample sizes and considerably about the same power as the most appropriate parametric test for even normal error distributions in their works. Further work in this direction will hopefully bring to light more useful tools that will help the statistician to be more efficient in data analysis.

1.7 Organisation of the research

This thesis is composed of five chapters, references and appendix.

Chapter 1 introduces the research problem, research objectives and the significance of the research.

Chapter 2 reviews literature on adaptive procedures in the analysis of one-sample, k-sample and other forms of data.

Chapter 3 covers the methodology of the thesis.

Chapter 4 covers data generation, simulations and discussion.

Chapter 5 presents the summary, and conclusions of the study.

Chapter 2

Literature Review

2.1 Introduction

This chapter presents a review of literature on adaptive procedures for statistical methods of hypothesis testing.

2.2 Adaptive Statistical Methods

Adaptive procedures use the data to ascertain which statistical method or technique is most conducive and efficient. Generally, it is conducted in two phases. In the first phase, a selection statistic is computed from the estimate of skewness and tail-weight, that is, the shape of the error distribution of the data. In the second phase, the selector statistic is used to determine the appropriate statistical procedure for the analysis. This procedure has been proven to have several advantages. Notably, it can increase the power of the test if the error distribution is skewed and makes narrow confidence internals, are robust for both validity and efficiency and automatically downweight outliers, which has the effect of making the results less sensitive to observations that do not agree with the model (O'Gorman, 2004).

2.2.1 Single Location Adaptive Procedure

Bandyopadhyay and Dutta (2007) proposed two adaptive tests for a single location problem without making assumptions about the symmetry of the continuous distribution of the data. Whereas one is based on a measure of symmetry, used as a standard of deciding between the Wilcoxon signed rank test(W^+) and the signed test(S^+) (the deterministic approach), the other (the probabilistic approach) is a combination of the signed test and the Wilcoxon signed rank test based on evidence of asymmetry provided by the p-value from the triples test defined as

$$\hat{\eta} = \frac{1}{(\binom{n}{3})} = \sum_{i<j<k} h(X_i, X_j, X_k) \tag{2.1}$$

and

$$h(x_1, x_2, x_3) = \frac{1}{3}[\text{sign}(x_1 + x_2 - 2x_3) + \text{sign}(x_1 + x_3 - 2x_2) + \text{sign}(x_2 + x_3 - 2x_1)]$$

where sign(x) $= 1, 0, -1$ according as $x >, =, < 0$ based on equation 2.1 proposed by (**?**).

For the probabilistic approach, Bandyopadhyay and Dutta (2007) used p to denote the p-value associated with the observed value of $\hat{\eta}$ in equation 2.1, using the p-value to denote the amount of symmetry of the distribution present in the data. For any value of p, a Bernoulli trial with probability of success p is performed. If a success was realized, the Wilcoxon signed rank test was used otherwise, the sign test was used. The adaptive test rule was: Reject H_0 with probability p if $W^+ > w^+$ and with probability $(1-p)$ if $S^+ > s^+$ are the upper α-critical values of W^+ and S^+

However with the deterministic approach, a sample measure of symmetry on which a preliminary test was based was used. The proposed measure of symmetry was given as;

$$Q = \frac{X_{(n)} - 2\tilde{X} + X_{(1)}}{X_{(n)} - X_{(1)}}$$

where $-1 \leq Q \leq 1$, $\tilde{X}$ is the median of the data and $X_{(i)}$ is the order statistics of the data. The median was equidistant from both extremes if the distribution of the data was symmetric, closer to the minimum value for a positively skewed distribution and closer to a maximum value for a negatively skewed distribution. The test statistic was then proposed as;

$$T = S^{+}I(|Q| > c) + W^{+}(|Q| \leq c)$$

where $I(y)$ is an indicator function assuming value 1 or 0 depending on whether y is true or false. For all values of c considered, $c = 0.075$ was regarded the best in terms of robustness of the test (Bandyopadhyay and Dutta, 2007).

Consequently, from simulation studies, when the two adaptive methods were compared, Bandyopadhyay and Dutta (2007) concluded that the probabilistic approach was in general found to be very robust and had high power over the deterministic approach, thus concluding that when nothing was known about the skewness of the distribution, the probabilistic approach should be used.

2.2.2 Two-Sample Rank Test Statistics

Hao and Houser (2011) presented a seven decade advances of adaptive procedures for non-parametric test and extensively narrated the progress made in this area. This section employs their material as useful reference for this review. For a given $f(.)$, as the probability function of the cumulative distribution function function $F(.)$. Let R_j represent the rank of observation Y_j $(j = 1, 2, \ldots, n_2)$ in the order statistics of the combined sample $N = n_1 + n_2$ observations with $1 \leq R_j \leq N$. Hájek and Sidak, (1967) as found in Hao and Houser (2011) showed that, in general the asymptotically most powerful rank test statistic S depends on the inverse c.d.f F^{-1};

$$S = \sum_{j=1}^{n_2} a(R_j) \tag{2.2}$$

and

$$a(R_j) = -\frac{f'[F^{-1}(u)]}{f[F^{-1}(u)]} \tag{2.3}$$

where $u = \frac{R_j}{N+1}$ is the $Y_j's$ rank normalised in the combined sample particularly $u \in (0,1)$ and $a(R_j)$ defined as the scores or the ranks, since it maps the observation Y_j to the rank of Y_j in the combined sample. As $N \to \infty$, $F^{-1}(u)$ shows the corresponding observation using its rank R_j and the inverse c.d.f of the data, that is $a(u)$ provides the information in the ranks.

In later development, Hájek et. al.(1999) established that for any particular distribution of interest equations (2.2) and (2.3) provides the most powerful rank test. Thus as n_1, $n_2 \to \infty$, $\frac{S-E(S)}{\sqrt{Var(S)}} \sim N(0,1)$. Hence they provided three examples:

1. The normal score test denoted S_{nor} was considered to be the most powerful rank test when the distribution of the data was normal with the test defined by;

$$S_{nor} = \sum_{j=1}^{n_2} \Phi^{-1}(\frac{R_j}{N+1}) \tag{2.4}$$

   where Φ is the c.d.f of the standard normal distribution with mean and variance of the normal score test defined as $E(S_{nor}) = 0$ and $Var(S_{nor}) = \frac{n_2 n_1}{N(N-1)} \sum_{j=1}^{N} \left[\Phi^{-1}(\frac{j}{N+1})\right]^2$.

2. The Mann-Whitney-Wilcoxon (MWW, hereafter) test was selected as the most powerful rank test once the data was known to have been drawn from a logistic distribution, with the test statistic given by;

$$S_{log} = \frac{2n_2}{N+1} \sum_{j=1}^{n_2} R_j - n_2 \tag{2.5}$$

   with the linear transformation of equation (2.5) given by $S_{WMW} = \sum_{j=1}^{n_2} R_j$. The mean and variance are respectively given as $E[S_{WMW}] = \frac{1}{2} n_2(N+1)$ and $Var[S_{WMW}] = \frac{1}{12} n_2 n_1 (N+1)$.

3. The median test which was adjourned the most powerful test when the data was from a Laplace (double exponential) distribution, with the test statistic

defined as;

$$S_{lap} = \sum_{j=1}^{n_2} \text{sign}(R_j - \frac{N+1}{2}) \tag{2.6}$$

where

$$\text{sign(x)} = \begin{cases} 1, & \text{if } x > 0 \\ 0, & \text{if } x = 0 \\ -1, & \text{if } x < 0 \end{cases}$$

Equation (2.6) is practically the same as the test that counts the number of $Y_j^\prime s$ above the median of the combined sample and increases by $\frac{1}{2}$ when the median falls in the sample of $Y_j^\prime s$.

Therefore,

$$\begin{aligned} S_{median} &= \sum_{j=1}^{n_2} \frac{1}{2}\left[\text{sign}(R_j - \frac{N+1}{2}) + 1\right] \\ &= \frac{1}{2} S_{lap} + n_2 \end{aligned}$$

and the mean and variance of the median test given as $E[S_{median}] = \frac{n_2}{2}$ and $Var[S_{median}] = \frac{n_2 n_1}{4(N-1)}$ if N is even, and $Var[S_{median}] = \frac{n_2 n_1}{4N}$ if N is odd.

To ascertain the comparative strength of efficiency of these statistics to their parametric counterparts, the Asymptotic Relative Efficiencies (ARE) (that is, for two consistent test statistics, A and B under H_0, ARE is the reciprocal of the ratio of sample sizes needed to derive similar power against the same alternative hypothesis H_1, taking the limit as the sample size $N \to \infty$ as $H_1 \to H_o$). For instance, Edwin J.G Pitman in 1949 computed the asymptotic relative efficiency $(A.R.E)$ of the MWW test relative to the t-test as;.

$$A.R.E_{w,t} = 12\sigma^2 \left[\int f^2(x)\mathrm{d}x\right]^2$$

where σ is the standard deviation of the underlying distribution of $f(x)$.

2.3 Gastwirth's Modification of the Two Sample Rank Test

A much simpler approach to increasing the efficiency of the MWW test with respect to relatively easy-to-discover features of the underlying distribution of the data at hand is discussed. His suggestion ensured that the flaws and difficulties that would have evolved in the evaluation of the score function, $-\frac{f'[F^{-1}(u)]}{f[F^{-1}(u)]}$ was dealt with. The test was modified by only including the top p and the bottom r fraction of the combined sample $(0 < p, r < 1)$ where the optimal values of p and r depend on the underlying distribution.

Let $Z_1, Z_2, \ldots, Z_N$ be the combined sample and $Z_1 < Z_2 < \ldots < Z_N$ as the order statistics of the combined sample, where $N = n_1 + n_2$. T_p defined as the total scores of the top t fraction and B_r defined as the total scores of the bottom b fraction. The test statistic was given as $T_p - B_r$, which begins with the median of the $Z_j's$ and scores with increasing positive integers for bigger ranks, and symmetric negative integers for smaller ranks. If N is odd, let $k = \frac{N+1}{2}$, and T_p and B_r respectively defined as;

$$B_r = \sum_{j=1}^{k}(k-j)\delta_j; T_p = \sum_{j=k}^{N}(j-k)\delta_j \tag{2.7}$$

However, if N is even, let $k = \frac{N}{2}$, then B_r and T_p are also respectively defined as;

$$B_r = \sum_{j=1}^{k}(k-j+\frac{1}{2})\delta_j; T_p = \sum_{j=k+1}^{N}(j-k-\frac{1}{2})\delta_j, \tag{2.8}$$

where

$$\delta_j = \begin{cases} 1 \text{ if } Z_j \text{ belongs to } Y_j's \\ 0 \text{ otherwise .} \end{cases}$$

Thus, the test statistic $T_p - B_r$ with $p = r = 50\%$ is similar to the WMW test for the change in location parameter. After deciding the values of p and r to be used, the data that fall outside the top p and bottom r were given zero weights. Hence, if N is odd then,

$$B_r = \sum_{j=1}^{R}(R - j + 1)\delta_j; T_p = \sum_{j=N-P+1}^{N}(i - (N - P))\delta_j. \tag{2.9}$$

If N is even, then

$$B_r = \sum_{j=1}^{R}(R - j + \frac{1}{2})\delta_j; T_p = \sum_{j=N-P+1}^{N}(j - (N - P) - \frac{1}{2})\delta_j, \tag{2.10}$$

where $R = [Nr] + 1$ and $P = [Np] + 1$ with $[x]$ defined as the nearest integer of x.

The $A.R.E$ of the two sample rank tests, as $N \to \infty$ under certain regularity conditions, the standardised limiting efficacy (S.L.E) of $T_p - B_r$ is;

$$\frac{12\left[\int_{-\infty}^{l} f^2(x)\mathrm{d}x + \int_{u}^{\infty} f^2(x)\mathrm{d}x\right]^2}{4(p^3 + r^3) - 3(r^2 - p^2)^2} \tag{2.11}$$

and that S.L.E of the t-test is $\frac{1}{\sigma^2}$. Thus, the asymptotic relative efficiency of the Gastwirth's test statistic denoted G with respect to the t-test t is;

$$A.R.E_{G,t} = \frac{12\sigma^2\left[\int_{-\infty}^{l} f^2(x)\mathrm{d}x + \int_{u}^{\infty} f^2(x)\mathrm{d}x\right]^2}{4(p^3 + r^3) - 3(r^2 - p^2)^2} \tag{2.12}$$

where $p = 1 - F(u)$ and $r = F(l)$, and $F(\cdot)$ is the c.d.f of the combined sample's underlying distribution.

The $A.R.E$ of the rank test depends on the distribution function of equation (2.7) and the values of p and r. Below is a summary of the $A.R.E$ of the test statistic $T_p - B_r$ to the t-test for the double exponential, the standard normal and the uniform distributions.

Table 2.1: Formulae for the $A.R.E_{G,t}$ of $T_p - B_r$ to the t-test for some Distributions

Distribution	Density	$A.R.E$ formula
Double Exponential	$\frac{1}{2}\exp^{-\lvert x\rvert}$	3p
Standard normal	$\frac{1}{\sqrt{2\pi}}\exp^{-\frac{x^2}{2}}$	$3\Phi^2\frac{(\sqrt{2}u)}{2\pi p^2}$
Uniform	1 if $x \in [0,1]$	$\frac{1}{2p}$

Clearly indicating that for light-tailed models, the $A.R.E_{G,t}$ for the uniform distribution is higher for small values of p and for heavy-tailed distributions like the exponential distribution, the $A.R.E_{G,t}$ is higher when p is large based on the A.R.E formulae in Table 2.1.

2.3.1 Two-Sample Adaptive Procedure for Location

Hogg et al. (1975) proposed a two-sample adaptive test that did not require symmetric error distributions. This two-sample test, which will be called the HFR test, used selection statistics that were measures of asymmetry and tailweight to select one of several rank tests. If the selection statistics fell into one of the regions defined by the adaptive procedure, then a certain set of rank scores was selected, whereas if the selection statistic fell into a different region, then a different set of rank scores would be used in the test. The HFR test played an important role in the development of rank-based adaptive tests because it was easy to compute and was the first practical two-sample test that was robust for validity and efficiency.

Hogg et al. (1975) extensively advocated for an adaptive testing procedures in performing two-sample statistical analysis. The reason for their advocacy was that the adaptive procedures are robust and even maintains the level of significance than the parametric procedures used in analysing the two sample data. Hogg et al. (1975) proposed formulae for skewness and tail weight denoted Q_1 and Q_2 respectively as selector statistics. HFR's measure of skewness (Q_1) is the ratio of the distance between the mean of the upper 5% of the data and the mid-mean of the data to the distance between the mid-mean of the data and the mean of the lower 5% of the data. Mathematically,

$$Q_1 = \frac{\overline{U}_{5\%} - \overline{M}_{50\%}}{\overline{M}_{50\%} - \overline{L}_{5\%}}.$$

HFR's Measure of Tail weight (Q_2) is the ratio of the distance between the mean of the upper 5% and the mean of the lower 5% of the data to the distance between the averages of the upper and lower halves of the data. Mathematically,

$$Q_2 = \frac{\overline{U}_{5\%} - \overline{L}_{5\%}}{\overline{U}_{50\%} - \overline{L}_{50\%}}.$$

Q_1 and Q_2 were defined such that when $Q_1 \leq 2$ and $2 \leq Q_2 \leq 7$, indicate a distribution with heavier-tailed and so the WMW scores were used as a test statistic, when $Q_1 > 7$, there was an indication of the distribution of the data being very heavy-tailed, thus proposing the use of the median test as a test statistic, when $Q_1 < 2$ and $1 \leq Q_2 < 2$ then the distribution is light-tailed, and finally when $Q_1 > 2$ and $Q_2 \leq 7$, then the distribution is right-skewed and scores for distributions skewed right were used (Hogg et al., 1975) .

Hogg et al. (1975) known as HFR, hereafter, however worked with the assumption that the data could not be skewed to the left, but proposed methods to be used should there be evidence that the data was skewed left. That was precisely when $Q_1 < \frac{1}{2}$ and $Q_2 > 7$. Through Monte Carlo simulations, they concluded that the adaptive test performs powerfully over a broad class of distributions and is to be preferred over some popular non-adaptive tests including parametric ones (Hogg et al., 1975).

2.3.2 Büning's Adaptive Tests

Büning (1996) proposed an adaptive tests for the one-way layout. In this test, he first ordered the observations of the $k-$samples, that $N = n_1 + n_2 + \cdot + n_k$. Then, choose a test statistic based on a vector of selector statistics, $Q = (Q_S, Q_T)$, where Q_S is a measure of skewness and Q_T is a measure of tailweight were given as

$$Q_S = \frac{\hat{x}_{0.975} - \hat{x}_{0.5}}{\hat{x}_{0.5} - \hat{x}_{0.025}}$$

and

$$Q_T = \frac{\hat{x}_{0.975} - \hat{x}_{0.025}}{\hat{x}_{0.875} - \hat{x}_{0.125}}$$

with the k-quantile $\hat{x}_k$ given by;

$$\hat{x}_k = \begin{cases} X_{(1)} \text{ if } k \leq \frac{0.5}{N} \\ (1-\lambda)X_{(j)} + \lambda X_{(j+1)} \text{ if } \frac{0.5}{N} < k < 1 - \frac{0.5}{N} \\ X_{(N)} \text{ if } k > 1 - \frac{0.5}{N} \end{cases}$$

where $X_{(1)}, X_{(2)}, \cdots, X_{(N)}$ are the order statistics of the combined k-samples and $j = [nk + 0.5]$; $\lambda = nk + 0.5 - j$. Clearly, $Q_S < 1$, if F, the distribution function is skewed to the left, $Q_S = 1$, if F is symmetric and $Q_S > 1$, if F is skewed to the right.

The adaptive test proposed and used based on the distribution from which the data was deemed to have been drawn denoted was A and defined as;

$$A = \begin{cases} G \text{ if } 0 \leq \hat{Q}_1 \leq 2; \quad 1 \leq \hat{Q}_2 \leq 1.5 \\ KW \text{ if } 0 \leq \hat{Q}_1 \leq 2; \quad 1.5 \leq \hat{Q}_2 \leq 2 \\ LT \text{ if } \hat{Q}_1 \geq 0; \quad \hat{Q}_2 > 2 \\ HFR \text{ if } \hat{Q}_1 \geq 2; \quad 1 \leq \hat{Q}_2 \leq 2, \end{cases}$$

where G, KW, LT and HFR are the Gastwirth, Kruskal Wallis, Light-tailed and the HFR scores respectively used to make inferences. A Monte Carlo simulation study by Büning (1996) revealed no significant difference between results based on Büning's measure of skewness and tailweight and that of HFR, and performed significantly better than their parametric counters in asymmetric distributions.

2.3.3 Other Adaptive Tests

This section discusses adaptive procedures based on subsets of regression coefficients, one-way layout and paired data. Hill et al. (1988) agreed that the formulae proposed by (Hogg et al., 1975) as measures of skewness and tail weight through Monte Carlo studies have proved to be good indicators when the shift parameter Δ is close to zero but then may indicate wrong statistics when the shift parameter is large. Thus for statistical hypothesis testing the measures proposed by Hogg et al. (1975) does not pose serious implications as it is only a test of location, however from the point of view of estimation, it poses serious problems as wrong or unsuitable test statistic may lead to wide confidence intervals. Subsequently two adaptive schemes were proposed: one that was similar to that of Hogg et al. (1975), but with the measures of skewness and tail weight denoted $\overline{Q}_1$ and $\overline{Q}_2$ and defined respectively as

$$\overline{Q}_1 = \frac{n_1 Q_{1,1} + n_2 Q_{1,2}}{n_1 + n_2}$$

and

$$\overline{Q}_2 = \frac{n_1 Q_{2,1} + n_2 Q_{2,2}}{n_1 + n_2}$$

where $Q_{1,1}$ and $Q_{1,2}$ are obtained from HFR measure of skewness for sample 1 and 2 respectively, and $Q_{2,1}$ and $Q_{2,2}$ are obtained from HFR measure of tailweight for sample 1 and 2 respectively. This was apparently to cater for the challenges of large shift parameter. The second one addressed the situation where there were ties in the order statistics of observations in a two sample problem. The same average score was assigned to each of the observations in this tie. Consequently, Hill et al. (1988) the use of both procedures on lung cancer data demonstrates the dominance of the adaptive procedures over the parametric and rank based procedures when the size of each sample was at least 20.

O' Gorman (1997) proposed an adaptive test for the one-way layout. The procedure proposed, used the order statistics of the combined sample to derive "selector statistics" this time not through the measures of skewness and tail weight but through the use of percentiles which served as a basis to choose suitable sets of rank scores for the one-way test statistic. This procedure served as a generalization for the already existing two-sample adaptive procedure. The power and significance level of the adaptive procedure proposed were evaluated and compared with procedures such as the F-test, Kruskal-Wallis test and the normal scores test through Monte Carlo simulations. The results revealed that all the tests maintained their level of significance for datasets with at least 24 observations, but the adaptive tests proved to be more powerful for distributions that were skewed when the total number of observations were at least 24. This procedure also proved to be even more powerful than the F-test in particular for some symmetric distributions when the data set was at least 60 (O' Gorman, 1997).

Freidlin et al. (2003) introduced the concept of a two-stage adaptive procedure using the p-value of the Shapiro-Wilk test as the "selector statistics" in the first stage and then using the p-value to select the second stage test, denoted AD_3. The power of AD_3 was compared with the approach of the United States Environmental Protection Agency (EPA) and a natural adaptive test AD_4. The EPA was faced with data from heavy tails and thus used the sign test and t-test for paired differences and recommended that both tests yielded non-significant results when applied to the differences in order for the null hypothesis to be accepted and the AD_4 procedure however tested for normality and the use of a distribution-free method when normality is rejected. One such procedure is to apply the Shapiro-Wilk (SW) test for the differences, if normality is accepted, the t-test is used, otherwise the Wilcoxon test is used (Freidlin et al., 2003).

The adaptive procedure AD_3 however chose to use the Wilcoxon test in the second stage if at the first stage the data revealed moderate tails and then used an appropriate nonparametric test that has high power for the data where the distribution indicates heavy tails at the first stage (Freidlin et al., 2003). This procedure as is common with most nonparametric test proposed the use of the linear signed rank test given by

$$T = \frac{1}{n}\sum_{j=1}^{n} a(R_j^+)\delta(j)$$

with $a(j) = I(\frac{j}{n+1})$, where R_j^+ is the absolute rank of the paired difference, $I(u)$ is a normalised score function defined on $(0,1)$ and satisfying $\int_0^1 I(u)du = 0$ and $\int_0^1 I^2(u)du = 1$ with $\delta(Y_j)$ an indicator function of value 1 if y is positive and 0 otherwise (Freidlin et al., 2003). AD_3 used the linear rank tests with three score functions denoted $I_1(u)$, $I_2(u)$ and $I_3(u)$ for selection and testing. As stated, the choice of the second stage test depends on the p-value of the Shapiro-Wilk test denoted P_{SW}. Thus, the Wilcoxon test was used with score function $I_1(u)$ when $P_{SW} > 0.01$, the t-distribution scores with two degrees of freedom and score function $I_2(u)$ used when $0.0001 < P_{SW} \leq 0.01$ and the Cauchy scores with score function $I_3(u)$ used when $P_{SW} \leq 0.0001$ (Freidlin et al., 2003).

Consequently in comparing the power robustness of the three approaches via simulations; AD_3's, AD_4's and the EPA's, results showed that AD_3 had power robustness than the other methods over a set of distributions with sample sizes at most 50, but less power than the others using a contaminated normal distribution with a sample size of 300 paired differences which is unrealistic in practice. In addition, in almost all other cases, the new adaptive scheme AD_3 had power robustness inside some small percentage points of the optimal test for each distribution, and for heavier-tailed distributions, the power of AD_3 exceeds the AD_4's and the EPA test by about 15% (Freidlin et al., 2003).

A new adaptive procedure for analysing paired data was proposed by Miao and Gastwirth (2009). As with most adaptive procedures, the new adaptive procedure used a function of the ordered absolute values of the differences as a measure of tail heaviness ("selector statistics") of the underlying distribution denoted sM and given by;

$$sM = \frac{\tilde{s}}{\widetilde{M}}$$

where $\tilde{s}$ is the sample standard deviation and $\widetilde{M}$ is the median absolute deviation. The test statistic subsequently adopted in this new adaptive procedure was the signed rank test given by;

$$S = \sum_{j=1}^{n} a(R_j^+)\delta(X_j) \text{ with } a(j) = I(\frac{j}{n+1})$$

where R_j^+ is the rank of $|X_j|$ among $|X_1|, |X_2|, \ldots, |X_n|$ and

$$\delta(X_j) = \begin{cases} 1 \text{ if } X_j > 0 \\ 0 \text{ otherwise} \end{cases}$$

and I defined as the score function suitable for the data (Miao and Gastwirth, 2009).

Four score functions were further used for purposes of testing. These were the normal scores, wilcoxon scores, t_2 scores and the Cauchy scores, some of which are extreme members of the family of t distributions (Miao and Gastwirth, 2009). The choice of the scores however depended on the distribution of the data. Miao and Gastwirth (2009) found that when $sM \geq 2.7$, the appropriate scores test was the Cauchy, when $1.2 \leq sM < 2.7$, the t_2 scores were suitable, when $1.02 \leq sM < 1.2$, then the Wilcoxon scores were appropriate and when $sM < 1.02$, the normal scores were the best scores to be used. Results from simulation studies proved that the new adaptive procedure maintained its nominal level of significance for all continuous distributions even for sample sizes as small as 10 and has almost the same power as the best signed rank test for a broad class of distribution functions (Miao and Gastwirth, 2009).

O' Gorman (2002) developed an adaptive test for a subset of regression coefficients and compared the power of this adaptive test with that of an F-test for a subset of regression coefficients which is often used in order to compare two nested linear models. This comparison was made when the distribution of residual terms are not normally distributed. The adaptive test here used the studentized deleted residuals for calculating the weight for each observation and permutation procedure used so that the appropriate test maintained its level of significance near the nominal value. Results from simulation studies of two independent samples n_1 and n_2 indicate that, at a level of significance of 0.05, only the F-test and the proposed adaptive test maintained their size near the nominal value (O' Gorman, 2002).

An adaptive test that used the adaptive weighting scheme but not the permutation approach had a size of between 3.0 per cent and 7.1 per cent, and the proposed adaptive test (the adaptive scheme that used both the weighted scheme and the permutation test) had its size between 4.5 per cent and 5.4 per cent for a sample size of $n_1 + n_2 = 60$. Results for power also proved that the proposed test in general had greater power than the F-test for distributions that are not normal (O' Gorman, 2002).

O' Gorman (2006) suggested an adaptive multivariate test for a subset of regression coefficients in a linear model that was compared to the non-adaptive tests such as the likelihood ratio test. This procedure was based on the use of studentized deleted residuals to evaluate an appropriate weight for each observation as was the case in O' Gorman (2002). The weight were subsequently used to compute

Wilk's lambda for the weighted model. The adaptive test was then conducted through the permutation of independent variables representing the parameters that are assumed to be equal zero under the null hypothesis. The permuted variables are then weighted to find a permutation test used to calculate the p-value that served as a basis for the rejection or acceptance of the null hypothesis, i.e if p-value $< \alpha$- the level of significance, the null hypothesis is rejected. Results from simulation studies showed that the multivariate adaptive test maintains its level of significance for the three multivariate error distributions used, including normal, log-normal and $\sinh^{-1}$-normal error distributions, and the power of the adaptive test was almost about the same as the non-adaptive test where the models used normal errors whereas it exhibited substantially more power than the non-adaptive test when the error terms were non-normal (O' Gorman, 2006).

Chapter 3

Methodology

3.1 Introduction

The present study involves a simulation study and details of the relevant theories and techniques are presented in the subsequent sections in this chapter.

3.2 Hypothesis Testing

A hypothesis is generally believed to be an "educational guess", which acceptance or otherwise is based on empirical evidence. Hypothesis is generated via a number of means, but is usually the result of a process of inductive reasoning where observations lead to the formation of a theory. Scientists then use a large battery of deductive methods to arrive at a hypothesis that is testable, falsifiable and realistic. Suppose that a researcher wants to compare the time of relief from using different pain relief agents produced by k companies. Let $x_{11}, x_{12}, \cdots, x_{1n_1}$ represent n_1 observations from usage of product from company 1, $x_{21}, x_{22}, \cdots, x_{1n_2}$ represent that from company 2, and $x_{k1}, x_{k2}, \cdots, x_{kn_k}$ representing the observations noted to have used company k product. A model describing the data from such experiment, with the assumptions that the samples are drawn at random from k independent normal populations is given by;

$$x_{ij} = \mu_i + \varepsilon_{ij} \begin{cases} i = 1, 2, \cdots, k \\ \\ j = 1, 2, \cdots, n_i \end{cases} \tag{3.1}$$

where x_{ij} is the j^{th} observation from company i, μ_i is the time of relief from using the product of the i^{th} company, and ε_{ij} is the error margin associated with the j^{th}

observation. The assumption made regarding the $\varepsilon'_{ij}s$ is that they are normally and independently distributed with mean 0 and variance σ_i^2, $i = 1, 2, \cdots, k$. Since $\mu_1, \mu_2, \cdots, \mu_k$ are constants it implies that $x_{ij} \sim N(\mu_i, \sigma_i^2)$, $i = 1, 2, \cdots, k$ (Montgomery, 2001).

3.2.1 Test of k-sample Means

Example-Balanced One-way ANOVA 1 *This example was extracted from (?) for purposes of illustration. The data in Table 3.1 come from an agricultural experiment. We wish to test for different mean yields for the different fertilisers. Test at $\alpha = 0.05$ level of significance if the mean yield of the fertilizers are significantly different.*

Table 3.1: Type of fertilizer and amount of yield data.

Fertilizer	Yield
A	14.5,12.0,9.0,6.5
B	13.5,10.0,9.0,8.5
C	11.5,11.0,9.0,8.5
D	13.0,13.0,13.5,7.5
E	15.0,12.0,8.0,7.0
F	12.5,13.5,14.0,8.0

Assumptions

1. The samples from the k-populations are independent and normally distributed.

2. The populations have equal variances.

We are testing;

$H_0 : \mu_A = \mu_B = \mu_C = \mu_D = \mu_E = \mu_F$ vrs $H_1 : \mu_i \neq \mu_j$, at least for some $i \neq j$

We have six treatments so 6-1 = 5 degrees of freedom (df) for treatments. The total number of degrees of freedom is the number of observations minus one, hence 23. This leaves 18 degrees of freedom for the within-treatments sum of squares. The total sum of squares can be calculated routinely as

$$\sum (x_{ij} - \bar{x})^2 = \sum {x_i}^2 - \bar{x}^2$$

, which is often most efficiently calculated as

$$\sum x_{ij}^2 - (\frac{1}{n})(\sum {x_{ij}}^2)$$

. This calculation gives

$$SS = 3119.251 - (\frac{1}{24})(266.5)^2 = 159.990$$

The easiest next step is to calculate SST, which means we can then obtain SSE by subtraction as above. The formula for SST is relatively simple and reads

$$\sum_i \frac{T_i}{n_i} - \frac{T^2}{n}$$

where T_i denotes the sum of the observations corresponding to the ith treatment and $T = \sum_{ij} x_{ij}$. Here this gives

$$SST = \frac{1}{4}(42^2 + 41^2 + 46.5^2 + 47^2 + 42^2 + 48^2) - \frac{1}{24}(266.5)^2 = 11.802$$

Table 3.2 below gives the summary estimates for the test statistic:

Table 3.2: One-way ANOVA table for fertilizer yield

Source	df	Sum of Squares	Mean Square	F
Between Fertilizers	5	11.802	2.360	0.287
Residual	18	148.188	8.233	
Total	23	159.990		

The following were obtained as follows;

$$MST = \frac{SST}{df} = \frac{11.802}{5} = 2.360$$

$$MSE = \frac{SSE}{df} = \frac{148.188}{18} = 8.233$$

and

$$F = \frac{MST}{MSE} = \frac{2.360}{8.233} = 0.287$$

This gives a non-significant p-value compared with $F_{3,16}(0.95) = 3.239$. R calculates the p-value to be 0.914. In conclusion, we have no evidence for difference between the various types of fertiliser.

This test is carried out on the assumption that the data is normally distributed. However, the practicality and applicability of this assumption is questionable, especially when sample sizes are small, as the data analyst or the researcher has no clear knowledge of the distribution from which the data was derived. The adaptive tests proposed and performed in this thesis studies the data properties before deciding the test to run.

3.3 Model specification for Two Sample Location Test

Let $x_1, x_2, \ldots, x_{n_1}$ be an independent and identically distributed (iid) random sample with cumulative distribution function $F(x)$ and density function $f(x)$. Also, let $y_1, y_2, \ldots, y_{n_2}$ be another random sample, independent and identically distributed from the cumulative distribution function $F(X - \delta)$ and density function $f(X - \delta)$, where $\delta = \mu_y - \mu_x$ represents a shift in location between the two distributions, μ_y and μ_x are the means of $F(X - \delta)$ and $F(X)$ respectively, and F is unknown. The objective is to test the hypothesis $H_0 : \delta = 0$ versus $H_1 : \delta \neq 0$. Hence let $n = n_1 + n_2$ represent the combined sample of both $X_i's$ and $Y_i's$. Let $Z' = (X_1, X_2, \ldots, X_{n_1}, Y_1, Y_2, \ldots, Y_{n_2})$ denote the vector of observations. Then the location model can be written as;

$$Z_i = v_i \delta + e_i, 1 \leq i \leq n \tag{3.2}$$

where $e_1, e_2, \ldots, e_n$ are iid with distribution function $F(x)$,

$$v_i = \begin{cases} 0 & \text{if } 1 \leq i \leq n_1 \\ 1 & \text{if } n_1 + 1 \leq i \leq n \end{cases} \tag{3.3}$$

and

$$Z_{(j)} = \begin{cases} X_i & \text{for } 1 \leq i \leq n_1 \\ Y_{i-n_1} & \text{for } n_1 + 1 \leq i \leq n_1 + n_2 = n, \end{cases} \tag{3.4}$$

where $Z_{(j)}$ is the order statistics for Z_j. Then under the null hypothesis H_0, for $Z_{(j)}$, the conditional distribution of $Z_{(j)}$ is discrete with probability $\frac{1}{n!}$ for all permutations of the vector Z_j (Okyere, 2011). This implies that the conditional distribution does not depend on $F(x)$, hence the order statistics are sufficient for F.

3.4 Order Statistics

The smallest, largest or middle observation can be a useful summary for a random variable. For example, the highest flood waters or lowest winter temperature recorded during last year might be useful for planning for future emergencies. These are examples of order statistics. The order statistics of a random sample $x_1, x_2, \ldots, x_n$ are the sample values places in ascending order. They are denoted by $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ and satisfy $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$

$X_{(1)} = \min_{1 \leq i \leq n} \mathrm{X}_i$

$X_{(2)}$= second smallest X_i

$\vdots$

$X_{(n)} = \max_{1 \leq i \leq n} \mathrm{X}_i$

If $X_{(1)}, \cdots, X_{(n)}$ are random samples from continuous population, the probability density function (pdf, hereafter) of the order statistics is derived by Theorem 3.3.1.

Theorem 3.4.1 (PDF of Order Statistics) *Let $X_{(1)}, \cdots, X_{(n)}$ denote the order statistics of a random sample $X_1, \cdots, X_n$ from continuous population with cumulative density function (cdf) $F_X(x)$ and pdf $f_X(x)$. Then the pdf of $X_{(j)}$ is;*

$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} f_X(x)[F_X(x)]^{(j-1)}[1 - F_X(x)]^{(n-j)}$$

see (Casella and Berger, 2002) for proof.

3.5 Properties of Order Statistics

3.5.1 Sufficiency

Let $T = t(X)$ where X is a random sample defined as $X_1, X_2, \ldots, X_n$ be a statistic taking values in a set T. Basically, T is a sufficient statistic for θ if T contains all the information about θ that is available in the entire data variable $\mathbf{X}$.

Let $X_1, X_2, \ldots, X_n$ denote a random sample of size n from a distribution that has a probability density function(pdf) $f(x; \theta)$, $\theta \in \Omega$. Let $T = t(X_1, X_2, \ldots, X_n)$ be a statistic whose pdf is $f_T(t; \theta)$. Then T is a sufficient statistic for θ if and only if the ratio

$$\frac{f(x_1; \theta), f(x_2; \theta), \ldots, f(x_n; \theta)}{f_T[t(x_1, x_2, \ldots, x_n); \theta]} = H(x_1, x_2, \ldots, x_n)$$

where $H(x_1, x_2, \ldots, x_n)$ does not depend upon $\theta \in \Omega$. see chapter 5 of (Casella and Berger, 2002).

3.5.2 Completeness

Let $f(t|\theta)$ be a family of pdfs and probability mass functions (pmfs) for a statistic $T(X)$. The family of probability distributions is called complete if $E_\theta g(T) = 0$ for all θ implies $P_\theta(g(T) = 0)$=1 for all θ. Equivalently, $T(X)$ is called a complete statistic (Casella and Berger, 2002).

Theorem 3.5.1 *Let $\mathcal{F}$ be the class of continuous distributions for a real or vector of random variables Z and let $V = g(Z)$ be a measurable function with F_V denoting the class of continuous distributions of V. If Z is complete with respect to the class F, then V is complete with respect to F_V.*

Hence under the null hypothesis H_0, the order statistics for the combined sample $X_{(1)} < X_{(2)} < \ldots < X_{(n)}$ are sufficient and complete.

3.6 Ancillary Statistic

Suppose $S = s(X)$ is a statistic taking values in the set T. If the distribution of S does not depend on θ for every $P_\theta(\theta \in \Theta)$, then S is said to be an ancillary statistic for θ. These are statistics with distributions free of the parameters and seemingly contains no information about those parameters. For example the variance S^2 of a random sample from $N(\theta, 1)$ has a distribution that does not depend upon θ and so it is an ancillary statistic.

Theorem 3.6.1 (Basu) *If T is a boundedly complete sufficient statistic for $\mathcal{P}$ and S is ancillary, then T and S are independently distributed on every $\theta \in \Theta$ or any boundedly complete sufficient statistic is independent of an ancillary statistic.*

3.7 Adaptation and Gauss Markov Model

We developed an adaptive procedure for GMM under exchangeable errors. We consider the model,

$$y = \mu + C\Delta + \epsilon \tag{3.5}$$

The usual normality restriction on the ϵ is unwind however we assumed exchangeability of ϵ. Under H_0, we consider a vector of distribution function, $\mathcal{F}$, which is centered. We are investigating centered designs whose distributions are unknown or which may vary from center (level) to center (level)

$$\mathcal{F} = \begin{bmatrix} F^{(1)} \\ F^{(2)} \\ \vdots \\ F^{(n)} \end{bmatrix} \tag{3.6}$$

We are developing an adaptive procedure that adapt within the centers (levels), because consequentially the assumption of exchangeability is applicable with centers (levels) (Okyere, 2011). Now consider the $i-th$ block of the GMM with m levels and n_i sample sizes for the development of our scheme,

$$y_{ij} = \mu + c_{ij}\Delta + \epsilon_{ij} \tag{3.7}$$

for $j = 1, ..., n_i$, and $i = 1, ..., m$. where y_{ij} is the combined response samples, c_{ij} are elements of the design matrix C_i which are 0's and 1's and Δ fixed effect parameters, ϵ_{ij} are independent and identically distributed with distribution F.

3.8 Adaptive Scheme

Let $X_{11}, X_{12}, \ldots, X_{1n_1}, X_{21}, X_{22}, \ldots, X_{2n_2}, \cdots X_{n1}, X_{n2}, \ldots, X_{nn_k}$ be random samples from continuous distribution function $f(t)$ with some amount of variations denoted by δ among the samples, that is, $f(t-\delta)$. This thesis would want to test the hypothesis that there is no difference in the sample means.

$$H_0 : \delta = 0 \; vrs \; H_1 : \delta \neq 0$$

Two assumptions are thus implied;

1. The distributions have equal variances under both the null and alternative hypothesis.
2. The $X_i's$ are assumed to be independent of each other.

The adaptive test is distribution free based on linear rank statistics or the scores of the ranks of the combined samples of the treatment groups. In this thesis the aim of the adaptive test is to identify the distribution function from which the data was drawn and to provide scores based on the Winsorised Wilcoxons to be used for the hypothesis testing among the treatment means.

3.9 Selector Statistics and Scores Function

The distribution of data is readily unknown and to avoid presumptuous assertions, data is examined and classified by considering skewness and tail weight from a class of continuous distribution. In this work, nine classification categories known as the Winsorised scores are used. For all class of continuous distribution functions, $\mathcal{F}$, there could exist heavy, moderate or light tail weight and could be right skewed, left skewed or symmetric.

3.9.1 The Selector Statistic

Hogg et al. (1975) proposed a statistic for determining the skewness of a data which is defined by Hao and Houser (2011) as the ratio of the distance between the upper end and the mid mean to the distance between the lower end and the mid mean of the data. Denoted by Q_1, it is mathematically written as:

$$Q_1 = \frac{\overline{U}_{5\%} - \overline{M}_{50\%}}{\overline{M}_{50\%} - \overline{L}_{5\%}}, \tag{3.8}$$

where $\overline{U}_{5\%}$, $\overline{M}_{50\%}$ and $\overline{L}_{5\%}$ are the averages for the upper 5%, middle 50% and the lower 5% of the $X_{(j)}^{'s}$ the ordered combined samples respectively

The measure of tail weight proposed by Hogg et al. (1975) and defined by Hao and Houser (2011) as the ratio of the distance between the averages of the upper end and the lower end to the distance between the averages of the upper half and the lower half of the data. Denoted by Q_2, it is mathematically written as:

$$Q_2 = \frac{\overline{U}_{5\%} - \overline{L}_{5\%}}{\overline{U}_{50\%} - \overline{L}_{50\%}}, \tag{3.9}$$

where $\overline{U}_{50\%}$, and $\overline{L}_{50\%}$ are the averages of the upper 50% and lower 50% of the $X'^{s}_{(j)}$; $N = n_1 + \cdots + n_k$ observations of the combined samples.
These two statistics are together called selector statistic, $S = (Q_1, Q_2)$.

3.9.2 HFR Model Selection Scheme

Based on the selector statistic, the HFR method classified the data as belonging to one of four models. These were:

- Light-tailed symmetric model
- Heavier-tailed model
- Very heavy-tailed model
- Right-skewed model

The classification scheme is as specified in Table 3.3.

Table 3.3: HFR Model Selection Scheme

Q_1 value	Q_2 value	Remarks
≥2	[2,7]	Heavier-tail model
>2	>7	Very heavy tail model
≤2	≤7	Right-skewed model
≤2	≤2	Light-tailed symmetric distribution

3.9.3 Scores functions associated with HFR Model Selection Scheme

For heavier-tailed models, ($Q_1 \leq 2$ and $2 \leq Q_2 \leq 7$), Hogg et al. (1975) proposed the Mann-Whitney-Wilcoxon scores denoted W to be used to compute the linear rank statistics;

$$S = \sum_{j=1}^{n} a_W(R_j), \tag{3.10}$$

where

$$a_W(R_j) = R_j \text{ for } j = 1, 2, \ldots, N \tag{3.11}$$

and R_jN denotes the rank of the j^{th} observation in the second sample of the combined sample.

It is asymptotically most powerful when the data follow a logistic distribution defined by;

$$f(x|a,b) = \frac{1}{b} \times \frac{exp[-(\frac{x-a}{b})]}{[1 + exp - (\frac{x-a}{b})]^2}$$

where $-\infty < x < \infty$, $-\infty < a < \infty$ and $b > 0$

For very heavy-tailed models $Q_2 > 7$, the scores for the median test denoted M were used and defined as;

$$S = \sum_{j=1}^{n} a_M(R_j) \tag{3.12}$$

where

$$a_M(R_j) = \begin{cases} 1 & \text{if } R_j > \frac{(N+1)}{2} \\ 0 & \text{otherwise} \end{cases} \tag{3.13}$$

These scores are asymptotically most powerful when data follow Laplace or double exponential distribution with its probability density function defined as;

$$f(x) = \frac{1}{2\alpha} \begin{cases} exp(-(\frac{\mu-x}{\alpha})) & \text{if } x < \mu \\ exp(-(\frac{x-\mu}{\alpha})) & \text{if } x \geq \mu \end{cases}$$

with μ defined as the location parameter and α defined as the scale parameter. For light-tailed symmetric model denoted L, ($Q_1 \leq 2$ and $1 \leq Q_2 < 2$), the scores used to compute the test statistic was given as;

$$S = \sum_{j=1}^{n} a_L(R_j) \tag{3.14}$$

where

$$a_L(R_j) = \begin{cases} R_j - \left[\frac{(N+1)}{4}\right] - \frac{1}{2} & \text{if } R_j \leq \left[\frac{(N+1)}{4}\right] \\ R_j - N + \left[\frac{(N+1)}{4}\right] - \frac{1}{2} & \text{if } R_j \geq N - \left[\frac{(N+1)}{4}\right] + 1 \\ 0 & \text{otherwise} \end{cases} \tag{3.15}$$

with mean and variance of L under the null hypothesis H_0 given as $\mu_L = 0$ and $\sigma_L^2 = \frac{mn\left[\frac{N+1}{4}\right]\left(4\left[\frac{N+1}{4}\right]^2-1\right)}{6N(N-1)}$.

For right-skewed distribution $Q_1 > 2$ and $Q_2 \leq 7$, the scores denoted S corresponding to the test statistic was used;

$$S = \sum_{j=1}^{n} a_S(R_j) \tag{3.16}$$

were used with $a_S(R_j)$ defined as;

$$a_S(R_j) = \begin{cases} R_j - \left[\frac{N+1}{2}\right] - 1 & \text{if } R_j \leq \left[\frac{N+1}{2}\right] \\ 0 & \text{otherwise} \end{cases} \tag{3.17}$$

with the mean and variance of the test statistics of S under the null hypothesis defined as $\mu_S = \frac{-nK(K+1)}{2N}$ and $\sigma_S^2 = \frac{nmK(K+1)}{12(N-1)N^2}\{-3K^2 + (4N-3)K + 2N\}$.

HFR adaptive test could be summarised as;

$$A = \begin{cases} \text{WMW(W)} & \text{if } Q_1 \leq 2 \text{ and } 2 \leq Q_2 \leq 7 \\ \text{Median test(M)} & \text{if } Q_2 > 7 \\ \text{Light-tailed scores (L)} & \text{if } Q_1 \leq 2 \text{ and } 1 \leq Q_2 < 2 \\ \text{Right skewed scores (S)} & \text{if } Q_1 > 2 \text{ and } Q_2 \leq 7. \end{cases}$$

The figure below gives a concise explanation of which scores to be used based on the values of Q_1 and Q_2.

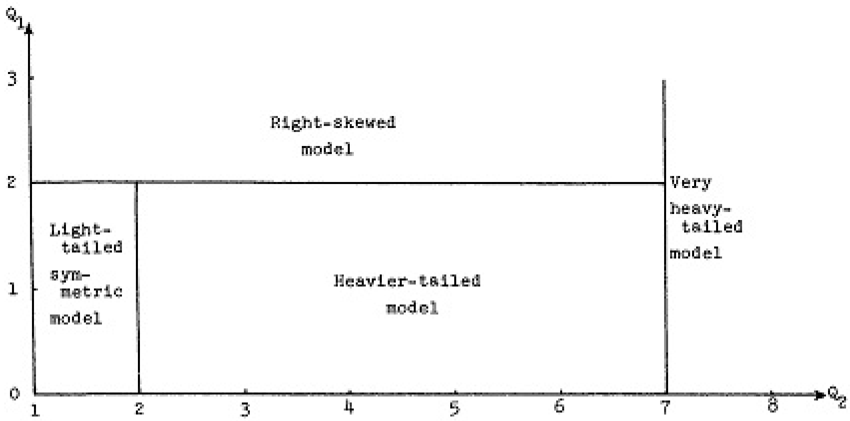

Figure 3.1: HFR model selection Criteria: Skewness Q_1 and Tailweight Q_2

This method assumed that data could not have been skewed to the left. However, an additional category of scores were created for left skewed models, $Q_1 < \frac{1}{2}$ and $Q_2 < 7$, with test statistic defined as;

$$S = \sum_{j=1}^{n} a_{LS}(R_j) \tag{3.18}$$

where

$$a_{LS}(R_j) = \begin{cases} 0 \text{ if } R_j \leq \frac{N+1}{2} \\ R_j - \frac{N+1}{2} \text{ if } R_j > \frac{N+1}{2}. \end{cases} \tag{3.19}$$

The scores selected are ancillary since they provide no information about the unknown distribution F.

3.9.4 HFR Adaptive Test and the Level of Significance

This section presents the theorem and proof that the adaptive test maintains its level of significance for data of the order statistics of the combined sample $(X_1's, \cdots, X_k's)$.

Theorem 3.9.1 (Lemma) *1. Let $\mathbb{F}$ denote the class of continuous distribution functions under consideration. Suppose that each of m tests based on the statistics $T_1, T_2, \cdots, T_m$ is distribution free over the class $\mathbb{F}$, i.e $\mathcal{P}_{H_0}(T_h \in C_h/\mathcal{F}) = \alpha$ for each $\mathcal{F} \in \mathbb{F}$ and $h = 1, \cdots, m$.*

2. Let S be some statistic that independent of $T_1, T_2, \ldots, T_m$ under H_0; for each $\mathcal{F} \in \mathbb{F}$. Suppose that S is used to decide which test T_h to conduct (S is called a selector statistic). Let U_s denotes the set of all values of S with the following decomposition; $U_s = D_1 \cup D_2 \cup \ldots \cup D_m, D_h \cap D_k = \emptyset$ for $h \neq k$, so that $S \in D_h$ corresponds to the decision to use the test T_h. The overall testing procedure is then defined by: If $S \in D_h$ then reject H_0 if $T_h \in C_h$ This two-stage adaptive test is, under H_0, distribution-free over the class of $\mathbb{F} \in \mathcal{F}$.

Proof 3.9.1

$$\begin{aligned}
\mathcal{P}_{H_0}(rejectH_0/\mathcal{F}) &= \mathcal{P}_{H_0}\left(\bigcup_{h=1}^{m}(S \in A_h \wedge T_h \in C_h/\mathcal{F})\right) \\
&= \sum_{h=1}^{m} \mathcal{P}_{H_0}(S \in A_h \wedge T_h \in C_h/\mathcal{F}) \\
&= \sum_{h=1}^{n} \mathcal{P}_{H_0}(S \in A_h/\mathcal{F}) \cdot \mathcal{P}_{H_0}(T_h \in C_h/\mathcal{F}) \\
&= \alpha \cdot \sum_{h=1}^{m} \mathcal{P}_{H_0}(S \in A_h/\mathcal{F}) \\
&= \alpha \cdot 1 \\
&= \alpha
\end{aligned}$$

3.10 Winsorization

In the process of winsorization, a fixed number of extreme scores are replaced with the closest score in the tail of the distribution. The rational is that outliers may provide some useful information concerning the magnitude of scores in the distribution, but at the same time may unduly influence the results of the analysis unless some adjustment is made (Sheskin, 2011). For illustration, consider the following observation;

$$0, 1, 18, 19, 23, 26, 26, 28, 33, 35, 98, 654$$

has a mean score of 80.08, which is not representative of the data set.Let us substitute a score of 18, for both 0 and 1 (the 2 least observed scores) and a score of 35 for 98 and 654 (the 2 highest scores). Thus the winsorized distribution scores becomes;

$$18, 18, 18, 19, 23, 26, 26, 28, 33, 35, 35, 35$$

with a mean score of 26.17, which is a far better representation of the data compared to the previous score of 80.08. A winsorized distribution may be symmetric or asymmetric. Let r represents the number of scores that are trimmed or winsorized in the right tail and l represents the number of scores trimmed in the left tail. If $r = l$, then the winsorization process is is said to be symmetric otherwise asymmetric (?).

3.10.1 The Winsorised Scores

Hogg et al. (1975) measures of skewness and tail weight depict the ratio of the difference in averages of the combined sample, that is, of the form $\overline{U}_{\alpha_1} - \overline{V}_{\alpha_2}$, where α_1 and α_2 are some fractions to be trimmed from the combined ordered data.

Now, let

$$m(\alpha_1,\, \alpha_2) = \frac{1}{l} \sum_{j=b_1+1}^{n-b_2} X_{(j)} \tag{3.20}$$

where $X'_{(j)}s$ are the ordered statistic of the combined sample, $b_1 = [n\alpha_1]$, $b_2 = [n\alpha_2]$, $[x]$ denotes the smallest integer greater than x and $l = n - b_1 - b_2$. This redefines the measures of skewness and tail weight proposed by Hogg et al. (1975). Thus, the measure of skewness and tail weight are now denoted as Q_1^* and Q_2^* respectively and defined as;

$$Q_1^* = \frac{(m(0.95, 0) - m(0.25, 0.25))}{(m(0.25, 0.25) - m(0, 0.95))}$$

and

$$Q_2^* = \frac{(m(0.95, 0) - m(0, 0.95))}{(m(0.5, 0) - (0, 0.25))}$$

Therefore, to adapt on the error measurements of model (3.2), under H_0 the combined error measurements are used to obtain the measures of skewness and tail weight. This thesis makes use of benchmarks proposed by Al-shomrani (2003). The cutoff values for measures of skewness and tail weight depend on the sample size n, but as $n \to \infty$, the measures converges to that proposed by Hogg et al. (1975).

For Q_1^*, the

$$\text{lower cutoff } = 0.36 + \frac{0.68}{n}$$

$$\text{upper cutoff } = 2.73 - \frac{3.72}{n}$$

and for Q_2^*, when the sample size is less than 25

$$\text{lower cutoff } = 2.17 - \frac{3.01}{n}$$

$$\text{upper cutoff } = 2.63 - \frac{3.94}{n}$$

but when the sample size is at least 25, then the lower and upper cutoff are respectively defined as;

$$\text{lower cutoff } = 2.24 - \frac{4.68}{n}$$

and

$$\text{upper cutoff } = 2.63 - \frac{9.37}{n}.$$

These cutoff points are used in the selection of a rank score associated with the unknown distribution. In this thesis, the rank test used is given by;

$$T_\varphi = \sum_{j=1}^{n} \varphi\left[\frac{R(Z_j)}{n+1}\right] I(Z_j = Y_j) \tag{3.21}$$

where $\varphi(j) = \varphi\left(\frac{j}{n+1}\right)$, $a_\varphi(1)$, $a_\varphi(2)\ldots$ $a_\varphi(n)$ are scores and φ satisfies the following conditions:

- φ is a non-decreasing function and square integrable on $(0, 1)$.
- φ is differentiable on $(0, 1)$ and since φ is square integrable, then $\int_0^1 \varphi(u)du = 0$ and $\int_0^1 \varphi^2(u)du = 1$.

Thus from the model given by (3.2), then under H_0, e_i has density f and distribution F, with the optimal score $\varphi_f(u)$ given by

$$\varphi_f(u) = \frac{-f'(F^{-1}(u))}{f(F^{-1}(u))}$$

This thesis employs nine winsorised scores. Hettmansperger (1984) classified these scores under four generic cases and they include;

1.
$$\varphi_I(u) = \begin{cases} s_3, & u > s_1 \\ s_3 + \frac{s_3 - s_2}{s_1}(u - s_1), & \text{otherwise.} \end{cases}$$

2.
$$\varphi_{II}(u) = \begin{cases} \frac{-s_3}{s_1}(u - s_1), & u < s_1 \\ \frac{-s_4}{s_2 - 1}(u - 1) + s_4, & u > s_2 \\ 0, & \text{otherwise.} \end{cases}$$

3.
$$\varphi_{III}(u) = \begin{cases} s_2, & u < s_1 \\ s_3 + \frac{s_2 - s_3}{s_1 - 1}(u - 1), & \text{otherwise.} \end{cases}$$

4.
$$\varphi_{IV}(u) = \begin{cases} s_3, & u < s_1 \\ s_4, & u > s_2 \\ s_3 + \frac{s_4 - s_3}{s_2 - s_1}(u - s_1), & \text{otherwise.} \end{cases}$$

where s_1, s_2, s_3, s_4 and s_5 are parameters and $\varphi_i(j) = \varphi_i\left(\frac{j}{n+1}\right)$.

Figure 3.2 represents the plot of the scores of the adaptation (Okyere, 2011).

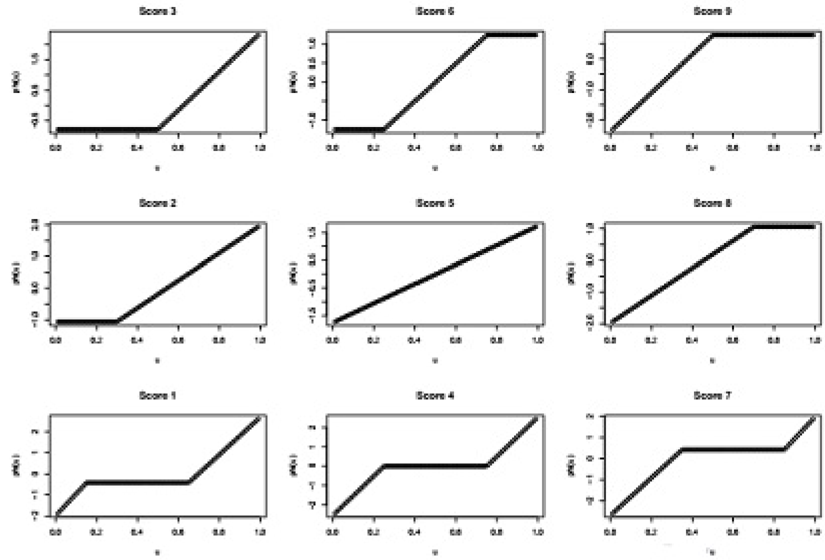

Figure 3.2: Plots of the Nine Winsorised Scores

Table 3.4 below provides the benchmarks needed for the nine winsorised wilcoxon scores proposed by (Hettmansperger, 1984).

Table 3.4: Benchmarks for Winsorised Scores

Skewness	Tail weight	Score function
Left	Light	$\varphi_1 = \varphi_{III}$ with parameters ($s_1 = 0.1$,$s_2 = -1$ and $s_3 = 2.0$)
Left	Medium	$\varphi_2 = \varphi_{III}$ with parameters ($s_1 = 0.3$,$s_2 = -1$ and $s_3 = 2.0$)
Left	Heavy	$\varphi_3 = \varphi_{III}$ with parameters ($s_1 = 0.5$,$s_2 = -1$ and $s_3 = 2.0$)
Symmetric	Light	$\varphi_4 = \varphi_{II}$ with parameters ($s_1 = 0.25$,$s_2 = 0.75$, $s_3 = -1$, $s_4 = 1.0$ and $s_5 = 0.0$)
Symmetric	Medium	Wilcoxon scores $\varphi_5 = \sqrt{12}[u - \frac{1}{2}]$
Symmetric	Heavy	$\varphi_6 = \varphi_{IV}$ with parameters ($s_1 = 0.25$,$s_2 = 0.75$, $s_3 = -1$ and $s_4 = 1.0$)
Right	Light	$\varphi_7 = \varphi_{II}$ with parameters ($s_1 = 0.9$,$s_2 = -2$ and $s_3 = 1.0$ $s_4 = 1$, and $s_5 = 0$)
Right	Medium	$\varphi_8 = \varphi_I$ with parameters ($s_1 = 0.7$,$s_2 = -2$ and $s_3 = 1.0$)
Right	Heavy	$\varphi_9 = \varphi_I$ with parameters ($s_1 = 0.5$,$s_2 = -2$ and $s_3 = 1.0$)

.

The summary of the adaptive test is thus based on the selector statistics $S =$

$\{Q_1^*, Q_2^*\}$ corresponding to the regions A_k, for $k = 1, 2, \ldots, 9$. That is;

$$A_1 = \left\{Q_1^* < \hat{Q}_{1l}^*, \, Q_2^* > \hat{Q}_{2u}^*\right\}$$
$$A_2 = \left\{\hat{Q}_{1l}^* < Q_1^* < \hat{Q}_{1u}^*, \, Q_2^* > \hat{Q}_{2u}^*\right\}$$
$$A_3 = \left\{Q_1^* > \hat{Q}_{1u}^*, \, Q_2^* > \hat{Q}_{2u}^*\right\}$$
$$A_4 = \left\{Q_1^* < \hat{Q}_{1l}^*, \, \hat{Q}_{2l}^* < Q_2^* < \hat{Q}_{2u}^*\right\}$$
$$A_5 = \left\{\hat{Q}_{1l}^* < Q_1^* < \hat{Q}_{1u}^*, \, \hat{Q}_{2l}^* < Q_2^* > \hat{Q}_{2u}^*\right\}$$
$$A_6 = \left\{Q_1^* > \hat{Q}_{1u}^*, \, \hat{Q}_{2l}^* < Q_2^* < \hat{Q}_{2u}^*\right\}$$
$$A_7 = \left\{Q_1^* < \hat{Q}_{1l}^*, \, Q_2^* < \hat{Q}_{2l}^*\right\}$$
$$A_8 = \left\{\hat{Q}_{1l}^* < Q_1^* < \hat{Q}_{1u}^*, \, Q_2^* < \hat{Q}_{2l}^*\right\}$$
$$A_9 = \left\{Q_1^* > \hat{Q}_{1u}^*, \, Q_2^* < \hat{Q}_{2l}^*\right\}$$

where $\hat{Q}_{1l}^*$, $\hat{Q}_{1u}^*$, $\hat{Q}_{2l}^*$ and $\hat{Q}_{2u}^*$ are the benchmarks from the ordered samples or residuals (Al-shomrani, 2003). The regions are however identified from the corresponding parameters associated with the scores as shown in Table 3.4. An example of the scores are shown on Figure 3.3 based on a sample size of 50 with the regions displayed.

Supposing A_k and φ_k are the regions and scores selected respectively, then the adaptive test, $AD(S, \varphi)$ is $AD(S, \varphi) = T_{\varphi_k}, \quad S \in A_k$, where

$$T_{\varphi_k}(\Delta) = \sum_{j=1}^{n_2} a_{\varphi_k}(R(x_j - \Delta)) \tag{3.22}$$

is the test statistic on the ranks and score φ_k, associated with region A_k and hence distribution free. Under the null hypothesis H_0, $E[T_{\varphi_k}(\Delta)]$ is zero (Okyere, 2011).

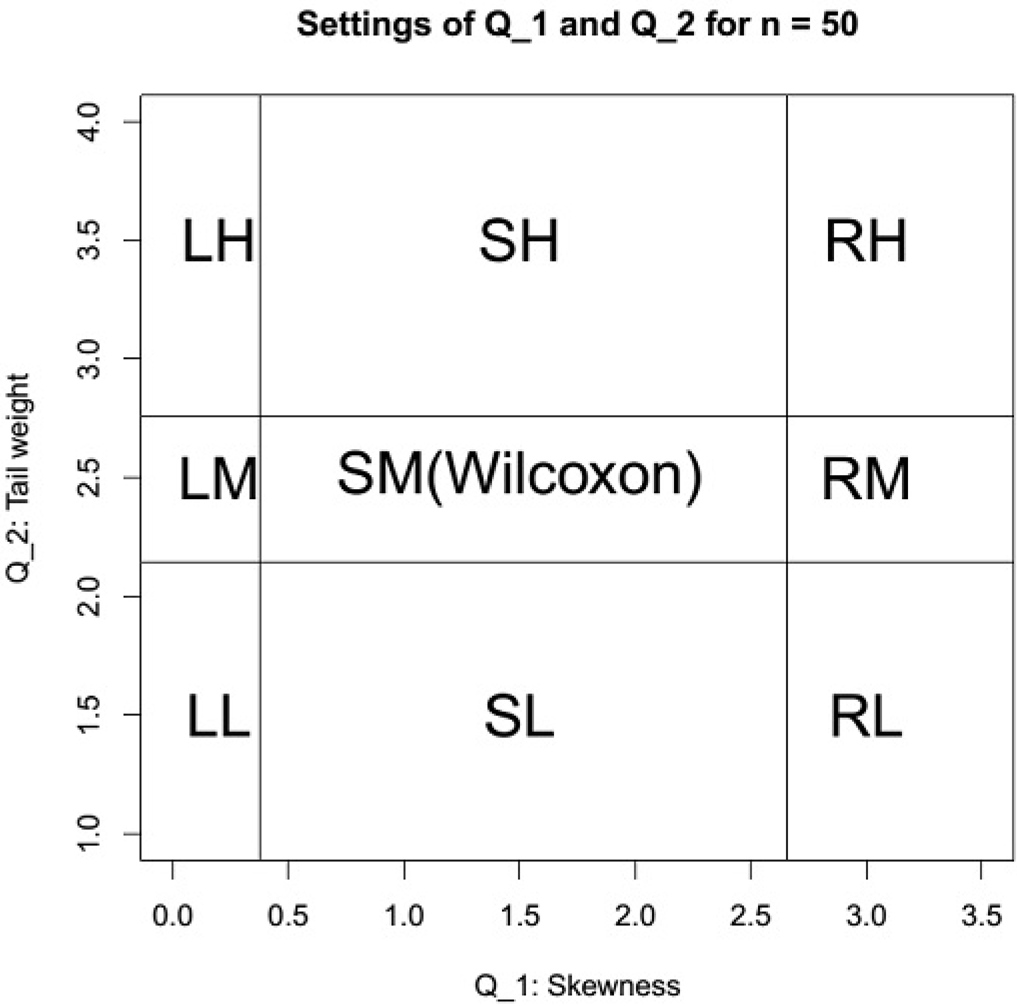

Figure 3.3: Plot of benchmarks with n=50

Hence,

$$
\begin{aligned}
E_0(T_{\varphi_k}) &= \sum_{j=1}^{n_2} E_0[a_{\varphi_k}(R(x_j))] \\
&= \sum_{j=1}^{n_2} \sum_{i=1}^{n} a_{\varphi_k}(i).\frac{1}{n} \\
&= 0
\end{aligned}
$$

since the ranks of the $x_j's$ are uniform on the integers $1, 2, \ldots, n$ and $\sum_{i=1}^{n} a_{\varphi_k}(i) = 0$.

To evaluate the variance, since $E_0(T_{\varphi_k}) = 0$, then

$$
\begin{aligned}
Var_0(T_\varphi) &= E_0(T^2_{\varphi_k}) \\
&= \sum_{j=1}^{n_2}\sum_{j'=1}^{n} E_0[a_{\varphi_k}(R(x_j))a_{\varphi_k}(R(x_{j'}))] \\
&= \left\{\frac{n_2}{n} - \frac{n_2(n_2-1)}{n(n-1)}\right\} \\
&= \frac{n_1 n_2}{n(n-1)} S_a^2
\end{aligned}
$$

since $S_a = \sum_{j=1}^{n} a_{\varphi_k}(R(x_j))$. for a detailed proof see (Hettmansperger, 1984).

The adaptive test is thus distribution free since the selector statistics S is obtained from only the order statistic and the test statistic T_{φ_k} is based on the ranks. Hence for a region A_k, the corresponding decision rule at α is to reject H_0 if:

$$
\left| \frac{T_\varphi}{\sqrt{Var_0(T_{\varphi_k})}} \right| \geq z_{\frac{\alpha}{2}}. \tag{3.23}
$$

For example the wilcoxon scores defined by $\varphi(u) = \sqrt{12}[u - 0.5]$ is regarded the optimal rank scores for the logistic distribution and the sign scores which is defined as $\varphi(u) = \text{sgn}[u - 0.5]$ as the optimal score for a double exponential distribution.

3.11 Exchangeability of Random Variables

Error measurements or observations are said to be exchangeable if they are considered independent, identically distributed (i.i.d), or if they are jointly normal with identical covariances (Good, 2002). Considering the two-sample location problem where $X_i's$ are random sample, i.i.d with continuous distribution function $F(x)$ and $Y_i's$ also being random sample, i.i.d with distribution function $F(x - \Delta)$ with the hypothesis defined as $H_0 : \Delta = 0 \quad$ against $H_1 : \Delta \neq 0$, then the exact test for H can be obtained by transforming the variable by subtracting

0 from each of the $X_i's$ and Δ from each of the $Y_i's$. Thus, "a set of observations (random variables) $\mathbf{X}$ will be said to be transformably exchangeable if there exists a transformation(measurable transformation) T, such that $T\mathbf{X}$ is exchangeable "(Good, 2002).

Theorem 3.11.1 *An infinite sequence of random variables* $(Y_1, Y_2, \ldots, Y_n, \ldots)$ *is said to be infinitely exchangeable under probability measure* P, *if the joint probability of every finite subsequence* $(Y_{n_1}, Y_{n_2}, \ldots, Y_{n_k})$ *satisfies* $(Y_{n_1}, Y_{n_2}, \ldots, Y_{n_k}) \stackrel{d}{=} (Y_{\tau(n_1)}, Y_{\tau(n_2)}, \ldots, Y_{\tau(n_k)})$ *for all permutations* τ *defined on the set* $\{1, 2, 3, \ldots, k\}$

An infinite sequence of random variables is said to be infinitely exchangeable, if every finite sequence of its variables (events) is exchangeable (Mahmoud, 2008). Mahmoud (2008) explained exchangeability of random variables with this illustration. Consider the white-blue pólya-eggenberger urn scheme in which balls are sampled with replacements, and whenever a ball colour appears in the sample, an extra ball of the same colour is added. Now supposing two balls were used initially, each of a different colour. Let $\hat{W}_i$ be an indicator of picking a white ball in the i^{th} draw for all $i \geq 1$. That is;

$$\hat{W}_i = \begin{cases} 1 & \text{if the ball in the } i^{th} \text{ draw is white} \\ 0 & \text{otherwise} \end{cases}$$

Thus, the probabilities of all sequences with only 1 blue colour is given by;

$$P(\hat{W}_1 = 1, \hat{W}_2 = 1, \hat{W}_3 = 0) = \frac{1}{2} \times \frac{2}{3} \times \frac{1}{4} = \frac{1}{12}$$

Likewise,

$$P(\hat{W}_1 = 1, \hat{W}_2 = 0, \hat{W}_3 = 1) = \frac{1}{2} \times \frac{1}{3} \times \frac{2}{4} = \frac{1}{12}$$

and

$$P(\hat{W}_1 = 0, \hat{W}_2 = 1, \hat{W}_3 = 1) = \frac{1}{2} \times \frac{1}{3} \times \frac{2}{4} = \frac{1}{12}$$

Also, the probabilities of all sequences of obtaining only one white ball in the first three picks is given by;

$$\begin{aligned} P(\hat{W}_1 = 1, \hat{W}_2 = 0, \hat{W}_3 = 0) &= P(\hat{W}_1 = 0, \hat{W}_2 = 1, \hat{W}_3 = 0) \\ &= P(\hat{W}_1 = 0, \hat{W}_2 = 0, \hat{W}_3 = 1) \\ &= \frac{1}{12} \end{aligned}$$

and finally the probabilities of all sequences of the same colour in all three picks is;

$$\begin{aligned} P(\hat{W}_1 = 1, \hat{W}_2 = 1, \hat{W}_3 = 1) &= P(\hat{W}_1 = 0, \hat{W}_2 = 0, \hat{W}_3 = 0) \\ &= \frac{1}{4} \end{aligned}$$

Generally, for $\hat{W}_1$, $\hat{W}_2$, $\ldots$, $\hat{W}_n$ being an arbitrarily large but fixed $n \geq 2$ and suppose that there is a total of k among the indicators that are 1,and the rest 0 occurring at at positions $1 \leq \tau_1 < \tau_2 < \ldots, \tau_k \leq n$. Then, the probability of this event is

$$\begin{aligned} &\mathcal{P}[\hat{W}_1 = 0, \cdots \hat{W}_{\tau_1 - 1} = 0, \hat{W}_{\tau_1} = 1, \hat{W}_{\tau_1} = 1, \hat{W}_{\tau_1+1} = 0, \cdots, \hat{W}_{\tau_k - 1} = 0, \hat{W}_{\tau_k} = 1, \hat{W}_n = 0] \\ &= \frac{1}{2} \times \frac{2}{3} \times \cdots \times \frac{\tau_1 - 1}{\tau_1} \times \frac{1}{\tau_1 + 1} \times \frac{\tau_1}{\tau_1 + 2} \times \frac{\tau_1 + 1}{\tau_1 + 3} \times \cdots \times \frac{\tau_2 - 2}{\tau_2} \\ &\quad \times \frac{2}{\tau_2 + 1} \times \frac{\tau_2 - 1}{\tau_2 + 2} \times \cdots \times \frac{\tau_k - k}{\tau_k} \times \frac{k}{\tau_{k+1}} \times \frac{\tau_k - k + 1}{\tau_k + 2} \times \cdots \times \frac{n - k}{n + 1} \\ &= \frac{k!(n-k)!}{(n+1)!} \end{aligned} \tag{3.24}$$

Conversely

$$\mathcal{P}[\hat{W}_1 = 1, \hat{W}_2 = 1, \cdots \hat{W}_k = 1, \hat{W}_{k+1} = 0, \hat{W}_{k+2} = 0, \cdots, \hat{W}_n = 0]$$
$$= \frac{1}{2} \times \frac{2}{3} \times \frac{3}{4} \times \cdots \times \frac{k}{k+1} \times \frac{1}{k+2} \times \frac{2}{k+3} \times \cdots \times \frac{n-k}{n+1}$$
$$= \frac{k!(n-k)!}{(n+1)!}. \tag{3.25}$$

Comparing equations (3.24) and (3.25), the probability of drawing k white balls in n draws is independent of where in the sequence the white balls were drawn (Mahmoud, 2008). The characteristics of interest is that there are k white balls and that all the sequences with the same number of balls have the same probability.

Theorem 3.11.2 (De Finetti's Theorem) *Let $X_1, X_2, \ldots, X_n, \ldots$ be an infinite sequence of random variables. Suppose that, for any n, $X_1, X_2, \ldots, X_n$ is exchangeable: $\mathcal{P}(x_1, x_2, \ldots, x_n) = \mathcal{P}(x_{\tau_1}, x_{\tau_2}, \ldots, x_{\tau_n})$ for all permutations τ of $\{1, 2, \ldots, n\}$. Then*

$$\mathcal{P}(x_1, x_2, \ldots, x_n) = \int \{\Pi_{i=1}^{n} \mathcal{P}(x_i/\theta)\} \mathcal{P}(\theta)\, d\theta.$$

for some parameter θ, some prior distribution of θ and some sampling model $\mathcal{P}(x/\theta)$. The prior and sampling model depend on the form of the belief model $\mathcal{P}(x_1, x_2, \ldots, x_n)$. θ is the parameter that describes the conditions under which the random variables are generated.

The implication here is that any probability measure describing an exchangeable sequence that is infinite can be expressed as a mixture of independent and identically distributed (iid) probability measures.

Theorem 3.11.3 (Lemma) *Independent and identically distributed random variables are exchangeable, conversely, exchangeability does not imply that random variables are independent and identically distributed.*

Thus, for the k-sample data, under H_0, the error terms e_i for $1 \leq i \leq n$ of equation (3.2) are exchangeable. That is, the random variables $e_1, e_2, \ldots, e_n$ are exchangeable iff for every permutation $\tau \in \mathcal{E}^n$, the joint distribution of $(e_{\tau_1}, e_{\tau_2}, \ldots, e_{\tau_n})$ is identical to the joint distribution of $e_1, e_2, \ldots, e_n$.

Proof 3.11.1 *Suppose $E_1, E_2, \ldots, E_n$ are conditionally i.i.d given some unknown parameter θ. Then for any permutation τ of $\{1, 2, \ldots, n\}$ and any set of values $(e_1, e_2, \ldots, e_n) \in \mathcal{E}^n$. Then*

$$
\begin{aligned}
\mathcal{P}(e_1, e_2, \ldots, e_n) &= \int (e_1, e_2, \ldots, e_n/\theta)\mathcal{P}(\theta)\, d\theta \\
&= \int \left[\mathcal{P}(e_1/\theta) \times \mathcal{P}(e_2/\theta) \ldots \mathcal{P}(e_n/\theta)\right] \mathcal{P}(\theta)\, d\theta \\
&= \int \left\{\Pi_{i=1}^{n} \mathcal{P}(e_i/\theta)\right\} \mathcal{P}(\theta)\, d\theta \\
&= \int \left\{\Pi_{i=1}^{n} \mathcal{P}(e_{\tau_i}/\theta)\right\} \mathcal{P}(\theta)\, d\theta \\
&= \mathcal{P}(e_{\tau_1}, e_{\tau_2}, \ldots, e_{\tau_n}).
\end{aligned}
$$

In Bernardo (1996), it is stated that it is of great importance and sufficient to assess for any n, the form of the joint probability density of $P(y_1, y_2, \ldots, y_n)$ in order to predict a future observable quantity given a sequence of'similar' observations. Furthermore, in probability theory if the observations or measurement errors are regarded as exchangeable, then they are deemed as being drawn from a random sample from the same model and there exist a prior probability distribution over the parameter of that model, hence requiring a Bayesian approach, however the representation theorem does not specify the model nor the required prior distribution.

3.12 Estimation of the Scale parameter

"For a specific distribution, the optimum scores is selected such that the asymptotic efficacy C_φ is as large as possible or equivalently such that the asymptotic variance of $\hat{\Delta}_\varphi$ is small as possible "(Hettmansperger and McKean, 2011).

The scale parameter τ_φ is defined as;

$$\begin{aligned}\tau_\varphi^{-1} &= \int_0^1 \varphi(u)\varphi_f(u)du \\ &= \int_0^1 \varphi(u)\left\{\frac{-f'(F^{-1}(u))}{f(F^{-1}(u))}\right\}du\end{aligned}$$

$\varphi_f(u)$ is referred to as the optimal score function.

If $\hat{\Delta}$ is an estimator whose variance achieves the Cramer-Rao lower bound ($\forall\,\Delta$), it is called efficient. That is,

$$Var(\hat{\Delta}) \geq \frac{\left(\frac{\mathrm{d}}{\mathrm{d}\Delta}E(\hat{\Delta})\right)^2}{nI(\Delta)}. \tag{3.26}$$

Thus, for the j^{th} observation in the $k-th$ sample, select scores with efficacy as large as possible or with asymptotic variance τ_φ as small as possible. The proof is as shown below;

Proof 3.12.1

$$\begin{aligned}\tau_{\varphi_j}^{-1} &= \int_0^1 \varphi(u)\varphi_f(u)du \\ &= \int_0^1 \varphi(u)\left\{\frac{-f'(F^{-1}(u))}{f(F^{-1}(u))}\right\}du \\ &= \frac{\int_0^1 \varphi(u)\varphi_f(u)du}{\sqrt{\int_0^1 \varphi_f^2(u)du}\sqrt{\int_0^1 \varphi^2(u)du}}\sqrt{\int_0^1 \varphi_f(u)du} \\ &= \left\{\frac{\int_0^1 \varphi(u)\varphi_f(u)du}{\sqrt{\int_0^1 \varphi_f^2(u)du} \times 1}\right\}\sqrt{\int_0^1 \varphi_f(u)du} \\ \therefore \tau_{\varphi_j}^{-1} &= \rho\sqrt{\int_0^1 \varphi_f^2(u)du} \\ &= \rho\sqrt{I(f)}\end{aligned}$$

where ρ is the correlation coefficient and $\int_0^1 \varphi_f^2 du$ is the Fisher information denoted by $I(f)$. Hence, by the Cramér-Rao lower bound, the smallest asymptotic variance obtainable is asymptotically efficient. Thus, to maximise τ_φ the score function is chosen such that $\rho = 1$ and $\varphi(u) = \varphi_f(u)$ (Hettmansperger and McKean, 2011).

Since $\hat{\Delta}_\varphi$ is location and scale equivariant, only the form $f(x)$ is needed. Therefore $\tau_\varphi = \frac{1}{\sqrt{I(f)}}$. The resulting estimate $\hat{\Delta}_\varphi$ is asymptotically efficient, implying that τ_t is a consistent estimator for τ. Hence for an estimator τ, the average of these estimators of the data is evaluated resulting in;

$$\tau = \frac{1}{j}\sum_{t=1}^{j} \tau_t$$

which is consistent for τ (Rashid et al., 2012).

3.12.1 Overall Test Statistic of Adaptation on Sample

In this subsection, the overall test statistic on a sample from j observed data is developed. Since the distribution of the errors or residuals are unknown, the "selector statistic" are used. Under H_0, it has been established that at observed data point, by definition 3.11.1, the error terms are exchangeable, thus the order statistics of the combined sample are sufficient and complete (Okyere, 2011). Denoted by φ_{kj}, the score that is selected at the j^{th} observed data that falls in region k, then the test statistic is

$$T_{\varphi_{kj}} = \sum_{i=1}^{n} a_j(R_j(X_i^{(j)})),$$

where $T_{\varphi_{kj}}$ is asymptotically standard normal and distribution free. Thus, a pooled test statistic is used to obtain the overall test. Under H_0, the overall test statistic, T is

$$
\begin{aligned}
T &= \sum_{t=1}^{q} T_{\varphi_{kj}} \\
&= \sum_{t=1}^{q} \sum_{i=1}^{n} a_j(R_j(X_i^{(j)}))
\end{aligned}
$$

which is also an asymptotic distribution $N(0, q)$ (Okyere, 2011). Hence, for the test $H_0 : \Delta = 0$ vrs $H_1 : \Delta \neq 0$, we fail to reject H_0 if;

$$
T = \left| \frac{\sum\limits_{t=1}^{q} X_t}{\sqrt{t}} \right| \leq z_{\frac{\alpha}{2}},
$$

since the asymptotic test is distribution free at each time point, the overall test statistic is also distribution free (Okyere, 2011).

3.13 Rank Based Estimation

Consider the function,

$$
\|v\| = \sum_{j=1}^{n} a(R(v_i))v_i,
$$

where $a(j)$'s are the scores, such that, $a(1) \leq a(2) \leq \ldots \leq a(n)$ and $\sum a(j) = 0$. Assume also that, $a(j) = -a(n+1-j)$. The rank-based procedures are used due to the fact that they are robust and also because the overall dispersion function denoted $D(\Delta)$ is convex. Thus, it is worth noting that the adaptation is performed at each observed data points, since it has been established that at each observed data point, under H_0 for the model (3.2), the error measurements are exchangeable.

Theorem 3.13.1 *Suppose* $a_j(1) \leq a_j(2) \leq \ldots \leq a_j(n)$, *and* $a(j) = -a(n+1-j)$, *then the function* $\| \cdot \|_\varphi$ *is a pseudo-norm.*

Next we define a pseudo-norm.

Theorem 3.13.2 (Pseudo-norm) *An operation $\|\cdot\|_\varphi$ is called a pseudo-norm if it satisfies the following four conditions:*

(i) $\|u+v\|_\varphi \leq \|u\|_\varphi + \|v\|_\varphi \ \forall \ u, v \in \mathbb{R}^n$.

(ii) $\|\alpha u\|_\varphi = |\alpha|\|u\|_\varphi \ \forall \ \alpha \in \mathbb{R},\ u \in \mathbb{R}^n$.

(iii) $\|u\|_\varphi \geq 0 \ \forall \ u \in \mathbb{R}^n$.

(iv) $\|u\|_\varphi = 0$ *if and only if* $u_1 = u_2 = u_3 = \ldots = u_n$.

The shift parameter Δ is estimated using the following pseudo-norm,

$$\|v\|_\varphi = \sum_{i=1}^{n} a[R(v_i)]v_i, v_i \in \mathbb{R}^n, \tag{3.27}$$

where $R(v_i)$ denotes the rank of v_i among the $v_1, v_2, \ldots, v_n$ and the scores at each observed data point generated as

$$a_j[i] = \varphi_j\left[\frac{i}{(n+1)}\right]$$

for $\varphi_j(u)$ a non-decreasing bounded square-integrable function defined on $(0,1)$ such that standardizing the square-integrable function yields $\int_0^1 \varphi_j(u)du = 0$ and $\int_0^1 \varphi_j^2(u)du = 1$, $a(i)$ is the score such that $a(1) \leq \ldots \leq a(n)$ and $\sum a(i) = 0$. For example, the Wilcoxon pseudo-norm is generated by the linear score function $\varphi(u) = \sqrt{12}\left(u - \frac{1}{2}\right)$ and the sign score is generated by $\varphi(u) = \operatorname{sgn}\left(u - \frac{1}{2}\right)$.

3.13.1 Jaeckel Dispersion Function

The geometry of rank-based estimation is similar to that is similar to that of least squares. In rank based regression however, we replace euclidean distance with another measure of distance, the Jaeckel's dispersion function defined by the rank based estimator of the shift parameter Δ denoted by $\hat{\Delta}$ is given by;

$$\hat{\Delta}_\varphi = \text{Argmin}\|\mathbf{Z} - \mathbf{C}\Delta\|_\varphi \tag{3.28}$$

Denoting the negative of the gradient of $\|\mathbf{Z} - \mathbf{C}\Delta\|$ by $S_\varphi(\Delta)$, then based on equation (3.27),

$$S_\varphi(\Delta) = \sum_{j=1}^{n_2} a_\varphi(R(X_j - \Delta)) \tag{3.29}$$

where $\hat{\Delta}_\varphi$ approximately solves the equation $S_\varphi(\hat{\Delta}_\varphi) \doteq 0$ (Hettmansperger and McKean, 2011). Thus, for each observed data, under the null hypothesis, the gradient of the rank test statistic is

$$S_\varphi = \sum_{j=1}^{n_2} a_\varphi(R(X_j)). \tag{3.30}$$

Since the test statistic only depends on the ranks of the combined sample it is distribution free under the null hypothesis. Thus

$$E_0(S_\varphi) = 0 \tag{3.31}$$

and

$$\sigma_\varphi^2 = V_0(S_\varphi) = \frac{n_1 n_2}{n(n-1)} \sum_{i=1}^{n} a^2(i), \tag{3.32}$$

where the variance can also be expressed as

$$\sigma_\varphi^2 = \frac{n_1 n_2}{n-1}\{\sum_{i=1}^{n} a^2(i)\frac{1}{n}\} \doteq \frac{n_1 n_2}{n-1},$$

and the approximation is due to the fact that the term in braces is a Riemann sum of $\int \varphi^2(u)du = 1$ and hence, converges to 1 (Hettmansperger and McKean, 2011). Note that T_φ and S_φ are used interchangeably, in this work. Kloke and McKean (2013) however used equation (3.28) termed the dispersion function proposed by (Jaeckel, 1972) to find an estimate of Δ based on ranks, and subsequently showed that;

$$\begin{aligned} \|\mathbf{Z} - \mathbf{C}\Delta\|_\varphi &= (\mathbf{Z} - \mathbf{C}\Delta)^T a(R(\mathbf{Z} - \mathbf{C}\Delta)) \\ &= \sum_{i=1}^{n} (Z_i - c_i^T\Delta) a(R(Z_i - c_i^T\Delta)) \end{aligned} \tag{3.33}$$

where $a(j) = \varphi(\frac{j}{n+1})$, the gradient defined as

$$S_{\varphi}(\Delta) = -\nabla\|\mathbf{Z} - \mathbf{C}\Delta\| = \mathbf{C}^{T}a(R(\mathbf{Z} - \mathbf{C}\Delta))$$

and the estimator $\hat{\Delta}_{\varphi}$ solves $S_{\varphi}(\Delta) \doteq 0$.

3.13.2 Ordinary Least Squares Estimation

Assume the linear model;

$$\mathbf{Y} = \alpha\mathbf{1}_n + \mathbf{X}\beta + \mathbf{e} \tag{3.34}$$

where $\mathbf{Y} = [Y_1, Y_2, \ldots, Y_n]^T$ is an $n \times 1$ response vector, $\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^T \\ \vdots \\ \mathbf{x}_n^T \end{bmatrix}$ is an $n \times p$ design matrix centred and of full rank, $\mathbf{1}_n$ is an $n \times 1$ error vector of ones, $\mathbf{e} = [e_1, e_2, \ldots, e_n]^T$ is an $n \times 1$ vector of i.i.d errors. Then under the hypothesis $H_0 : \beta = 0$ against $H_1 : \beta \neq 0$, the ordinary least squares(OLS) estimator for β is $\hat{\beta}_{OLS}$ and is given by;

$$\hat{\beta} = \text{Argmin}\|\mathbf{Y} - \mathbf{X}\beta\|_2^2 \tag{3.35}$$

where $\|u\|_2^2 = (\sum_{i=1}^{n} u_i^2)$. Hence, $\hat{\mathbf{Y}}_{OLS} = \mathbf{X}\beta_{OLS}$ is the closest vector in the euclidean distance to $\mathbf{Y}$. Since $\mathbf{X}$ is of full rank, the solution of the estimator is given as

$$\hat{\beta}_{OLS} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y},$$

and $\hat{\beta}_{OLS}$ is asymptotically distributed as

$$\hat{\beta}_{OLS} \dot{\sim} N_p(\beta, \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}).$$

The estimator $\hat{\beta}_{OLS}$ is not robust enough since an outlier can have significant affect on the estimator (Hettmansperger and McKean, 2011).

Another form of the parametric case in the ordinary least squares(OLS) approach of estimating $\hat{\beta}$ of β is to minimize the length

$$\|\varepsilon\| = [\varepsilon^T \varepsilon]^{\frac{1}{2}}$$

of the vector of residuals defined by $\varepsilon = Y - X\hat{\beta}$ (Staudte and Sheather, 1990). In addition, for an $n \times p$ matrix X, the column space is defined as $\mathcal{R}(X) = \{\theta : \theta = X\beta, \beta \in \mathbb{R}^p\}$ which is a subspace of $\mathbb{R}^n$ (Rao et al., 2008). Now, since

$$X^T(Y - X\hat{\beta}) = 0,$$

that is, $X^T X\hat{\beta} = X^T Y$, the equation

$$X^T X\beta = X^T Y$$

provides a solution and $X\beta$ is unique for all solutions of β. This satisfies the notion that $\hat{\theta}$ exist (Rao et al., 2008). Then by the independence of the residuals, the estimator

$$\hat{\beta} \sim N[\beta, \quad \sigma^2(X^T X)^{-1}]$$

and

$$Cov(\hat{\beta}) = \sigma^2(X^T X)^{-1},$$

which reaches the Cramér-Rao bound for the model and thus optimal for all unbiased estimators.

3.13.3 Pitman Regular

Theorem 3.13.3 *An estimating function $S_\varphi(\Delta)$ is pitman regular if the following conditions hold:*

1. *$S_\varphi(\Delta)$ is non increasing in Δ.*
2. *$\overline{S}_\varphi(\Delta) = \frac{S_\varphi(\Delta)}{n^\gamma}$, for some $\gamma > 0$ there exists a function $\mu(\Delta)$ such that $\mu(0) = 0$, $\mu'(\Delta)$ is continuous at 0, $\mu'(0) > 0$ and either $\overline{S}_\varphi(0) \xrightarrow{p_\theta} \mu(\Delta)$ or*

$E_\Delta[\overline{S}_\varphi(0)] = \mu(\Delta)$.

3. $\sup_{|b|\leq B} |\sqrt{n}\overline{S}(\frac{b}{\sqrt{n}}) - \sqrt{n}\overline{S}(0) + \mu'(0)b| \xrightarrow{p} 0$ *for any* $B > 0$.

4. *There is a constant* $\sigma(0)$ *such that* $\sqrt{n}\{\frac{\overline{S}(0)}{\sigma(0)}\} \xrightarrow{D_0} N(0,1)$ *and* $c = \frac{\mu'(0)}{\sigma(0)}$ *called the efficacy of* $S_\varphi(\Delta)$.

3.13.4 Asymptotic Distribution and Efficacy of $\hat{\Delta}_\varphi$

To develop the asymptotic null distribution of S_φ, using equation (3.30), it follows then that from equations (3.2) and (3.3) the linear rank statistic is

$$S_\varphi = \sum_{i=1}^{n} c_i a(R(Z_i)) = \sum_{i=1}^{n} (c_i - \overline{c}) a(\frac{n}{n+1} F_n(Z_i)), \tag{3.36}$$

where F_n is the empirical distribution function of $Z_1, \ldots, Z_n$. The score function is monotonic and square integrable. Now, let T_φ be the random variable defined by

$$T_\varphi = \sum_{i=1}^{n} (c_i - \overline{c})\varphi(F(Z_i)). \tag{3.37}$$

Hence, comparing equations (3.36) and (3.37), it implies that T_φ is an approximation of S_φ (Hettmansperger and McKean, 2011). Consequently, under H_0 the distribution of T_φ is approximately normal and has the same distribution as S_φ on condition that the second moment of their difference goes to 0 (Hettmansperger and McKean, 2011). That is,

$$Var\left[\frac{T_\varphi - S_\varphi}{\sigma_\varphi}\right] \to 0.$$

A proof of this is found in the appendix of Hettmansperger and McKean (2011). Also, assume that $f(x)$ has a finite Fisher information. This means that f is absolutely continuous, non decreasing and square integrable $\varphi(u)$ such that $0 \leq I(f) = \int_0^1 \varphi_f^2(u)du < \infty$ and $\lim(\frac{n_i}{n}) = \lambda_i$, $0 < \lambda_i < 1$, $i = 1, 2$ and $\lambda_1 + \lambda_2 = 1$ (Hettmansperger and McKean, 2011).

The square integrable function is defined as

$$\varphi_f(u) = \frac{-f'(F^{-1}(u))}{f(F^{-1}(u))}$$

where f is uniformly bounded.

Defining the scale parameter τ_φ as

$$\begin{aligned}\tau_\varphi^{-1} &= \int_0^1 \varphi(u)\varphi_f(u)du \\ &= \int_0^1 \varphi(u)\left\{\frac{-f'(F^{-1}(u))}{f(F^{-1}(u))}\right\}du\end{aligned} \tag{3.38}$$

Since the error measurements are independently distributed, then for the asymptotic representation of $\hat{\Delta}_\varphi$, the gradient $S_\varphi(\Delta)$ should satisfy the four conditions under 3.13.3 for the observed data (Hettmansperger and McKean, 2011). Hence $S_\varphi(\Delta)$ is non-increasing which satisfies the first condition.

Thus, from equation (3.29),

$$S_\varphi(\Delta) = \sum_{i=1}^{n_2} \varphi\left[\frac{n_1}{n+1}F_{n_1}(Y_i - \Delta) + \frac{n_2}{n+1}F_{n_2}(Y_i)\right] \tag{3.39}$$

where F_{n_1} and F_{n_2} are the empirical distribution functions of $X_1, X_2, \ldots, X_{n_1}$ and $Y_1, Y_2, \ldots, Y_{n_2}$ respectively (Hettmansperger and McKean, 2011). The second condition of 3.13.3 is satisfied if $E\left[\frac{1}{n}S_\varphi(\Delta)\right] = \mu(\Delta)$ and the $\mu' > 0$. Now from equation (3.39),

$$\begin{aligned}E\left[\frac{1}{n}S_\varphi(\Delta)\right] &\rightarrow \lambda_2 \int_{-\infty}^{\infty} \varphi\left[\lambda_1 F(x) + \lambda_2 F(x-\Delta)\right] f(x-\Delta)dx \\ &= \lambda_2 \int_{-\infty}^{\infty} \varphi\left[\lambda_1 F(x+\Delta) + \lambda_2 F(x)\right] f(x)dx \\ &= \mu_\varphi(\Delta) > 0\end{aligned}$$

Differentiating $\mu_\varphi(\Delta)$ and evaluating at $\Delta = 0$, to obtain an asymptotic efficacy results. This is illustrated below;

$$\begin{aligned}
\mu'_\varphi(0)|_{\Delta=0} &= \lambda_1\lambda_2 \int_{-\infty}^{\infty} \varphi'[F(t)]f^2(t)\mathrm{d}t \\
&= \lambda_1\lambda_2 \int_{-\infty}^{\infty} \varphi[F(t)] \left(\frac{-f'(t)}{f(t)}\right) f(t)\mathrm{d}t \\
&= \lambda_1\lambda_2 \int_0^1 \varphi(u)\varphi_f(u)\mathrm{d}u \\
&= \lambda_1\lambda_2\tau_\varphi^{-1} > 0
\end{aligned}$$

For proper evaluation and proof of $E\left[\frac{1}{n}S_\varphi(\Delta)\right]$ see Hettmansperger and McKean (2011). The second condition of Pitman regular is thus satisfied.

For the third condition, a rigorous proof is obtained from Theorem $A.2.8$ and the general rank regression statistics developed in section $A.2.2$ of the appendix in Hettmansperger and McKean (2011). Thus, the asymptotic linearity of $S_\varphi(\Delta)$ is given by;

$$\frac{1}{\sqrt{n}}S_\varphi\left(\frac{\delta}{\sqrt{n}}\right) = \frac{1}{\sqrt{n}}S_\varphi(0) - \tau_\varphi^{-1}\lambda_1\lambda_2\delta + O_p(1) \tag{3.40}$$

uniformly for $|\delta| \leq B$, where $B > 0$ and τ_φ is as defined in equation (3.38).

Finally, from condition 4 of 3.13.3,

$$\frac{1}{\sqrt{n}}\frac{S_\varphi(0)}{\sqrt{\lambda_1\lambda_2}} \sim N(0,1).$$

Thus, the efficacy of the test based on S_φ found in Hettmansperger and McKean (2011) is given by;

$$c_\varphi = \frac{\tau_\varphi^{-1}\lambda_1\lambda_2}{\sqrt{\lambda_1\lambda_2}} = \tau_\varphi^{-1}\sqrt{\lambda_1\lambda_2}. \tag{3.41}$$

since the asymptotic efficacy is given by $c_\varphi = \frac{\mu'(\Delta)}{\sigma(0)}$.

Theorem 3.13.4 *Suppose $S_\varphi(\Delta)$ is pitman regular with efficacy c_φ, then $\sqrt{n}(\hat{\Delta} - \Delta)$ converges in distribution to $Z \sim N\left(0, \frac{1}{c_\varphi^2}\right)$.*

Thus, since the estimate $\hat{\Delta}_\varphi$ solves the equation $S_\varphi(\hat{\Delta}_\varphi) \doteq 0$, then, based on the Pitman regularity and theorem 3.13.4, the asymptotic distribution of $\hat{\Delta}_\varphi$ is given by

$$\sqrt{n}(\hat{\Delta} - \Delta) \xrightarrow{D} N(0,\, \tau_{\varphi}^{2}(\lambda_1\lambda_2)^{-1})$$

.

3.13.5 Asymptotic Relative Efficiency

For any two consistent test statistics P and Q, of any hypothesis H_0, the asymptotic relative efficiency is the ratio of sample sizes required to obtain identical power against the same alternative H_1, taking the limit as the sample size n tends to infinity and as H_1 tends to H_0 (Hao and Houser, 2011). This implies that the asymptotic relative efficiency (ARE) lies in the interval $(0, \infty)$ when the tests are positive ie $ARE_{P,Q} \in (0, \infty)$. Also, when $ARE_{P,Q} \in (0, 1)$ then the test statistic P is regarded less efficient than Q, the test P is however considered efficient as the test Q when the $ARE_{P,Q} = 1$, lastly the test P is more efficient than the test Q when the $ARE_{P,Q} \in (1, +\infty)$.

Alternatively, let T_p and T_Q be two linear rank statistics based on the score generating functions P and Q. Then the asymptotic relative efficiency (ARE) is given by;

$$ARE(T_P, T_Q/f) = \frac{AE(T_P/f)}{AE(T_Q/f)} \tag{3.42}$$

where $AE(T_P/f)$ and $AE(T_Q/f)$ are the asymptotic efficacies of P and Q respectively (Kössler, 2010).

In this thesis the asymptotic relative efficiency is based on the mean squared errors of the score functions.

Theorem 3.13.5 *The asymptotic relative efficiency between two estimates or two tests based on the score functions $\varphi_1(u)$ and $\varphi_2(u)$ of one score function* *relative to other is defined by;*

$$e(\varphi_1, \varphi_2) = \frac{{c_{\varphi_1}}^2}{{c_{\varphi_2}}^2} = \frac{\tau_{\varphi_2}^2}{\tau_{\varphi_1}^2}$$

where c_1 *and* c_2 *are respectively the efficacies of the two estimates and* τ_{φ_i}, $i = 1, 2$ *are the scale parameters of the two score functions.*

Chapter 4

Analysis, Results and Discussion

4.1 Introduction

This chapter presents results from the simulation studies from some known continuous distribution and an application of proposed procedure on a real data to ascertain the relative efficiency of the adaptive test proposed in this thesis to the parametric counterpart. The chapter includes four section; simulation under Pure Hogg (run under H_0), fitting models to residuals from an Ordinary Least Square (OLS), fitting models to residuals from Wilcoxon Scores, and application on real data. The R-package, *Rfit* by Kloke and McKean (2012) is used for the study. Some other functions were written and implemented in R for the study. Table 4.1 presents the score functions associated with the nine Wilcoxon winsorized scores categorization of continuous distributions.

Table 4.1: Score functions for the Nine Wilcoxon Winsorized Scores categories

Skewness	**Tail weight**	**Score**	**Score function**
Left	Light	LL	afriscores3
Left	Medium	LM	afriscores4
Left	Heavy	LH	bentscores3
Symmetric	Light	SL	bentscores2
Symmetric	Medium	SM	wscores
Symmetric	Heavy	SH	bentscores4
Right	Light	RL	afriscores2
Right	Medium	RM	afriscores1
Right	Heavy	RH	bentscores1

The score functions *wscores*, *bentscores*1, *bentscores*2, *bentscores*3 and *bent−scores*4 are functions in the Rfit package. However, $afriscores1$, $afriscores2$, $afriscores3$ and $afriscores4$ were self-written and implemented in Rfit. The functions are as proposed by Hettmansperger (1984). The continuous distributions considered in this work include the normal distribution, contaminated normal distribution (5%, 10%, 15% and 20%), laplace distribution, truncated logistic distribution and mixture of distributions for balanced ANOVA models.

4.2 Simulation under Pure Hogg

In this section, a simulation study of adaptive test were performed. Under Pure Hogg, adaptation is directly performed on the simulated data (the observed samples). The adaptive test is compared with the F-statistics. Simulation results for normal, double exponential (laplace), contaminated normal, the truncated logistic and a mixture of distributions for a balanced ANOVA models are presented. 10,000 simulations were carried out for 5, 10, 15, and 20 observations each, assigned to three (3) levels with inter and intra-level correlation coefficient $\rho = 0$. These distributions were used owing to their properties and the adaptive scheme proposed. For example, the normal distribution is symmetric and has moderate tails and the double exponential distribution may be symmetric but heavy tailed.

4.2.1 Normal distribution

Simulation results are presented in Table 4.2 for Normal distribution with $\mu = 0$ and $\sigma = 1$;

Table 4.2: Simulation Results of Adaptive Test and Parametric Test for Normal distribution

Sample size	F-test			Adaptive Test			
(n_1, n_2, n_3)	Value	p-value	σ	Score	Value	p-value	τ
(5,5,5)	8.919	0.00423	1.102	SM	1.36112	0.29325	2.3005
(10,10,10)	37.36	1.67e-08	0.600	SM	6.4436	0.00515	2.1062
(15,15,15)	9.354	0.00437	0.898	SM	14.82175	1e-05	1.8637
(20,20,20)	25.12	51e-08	0.894	SM	31.24179	0.0000	1.6138

The selector statistics for the adaptive test identified the normal distribution with $\mu = 0$ and $\sigma = 1$ as a symmetric skewed and medium tailed distribution. From the variance returned in 4.2, it is obvious that the parametric F-test outperforms the adaptive test at all the level sample sizes considered. However, with exception of sample size (5,5,5), the two tests suggested a rejection to null hypothesis of no difference in level means. The ARE of the F-test over the adaptive test if the data under consideration is from a normal distribution is between 25% and 55%. It was observed that the ARE increased as sample sizes of the levels increased.

4.2.2 Laplace Distribution

The laplace distribution also known as double exponential distribution. This distribution is characterized by location θ (any real number) and scale λ (has to be greater than a 0) parameters. The probability density function of $Laplace(\theta, \lambda)$ is:

$$f(x|\theta, \lambda) = \frac{1}{2\lambda} exp\left(-\frac{|x-\theta|}{\lambda}\right).$$

The cumulative density function looks even more impressive, yet rather easy to integrate because of the absolute value in the formula:

$$F(x|\theta, \lambda) = \frac{1}{2} exp\left(-\frac{|x-\theta|}{\lambda}\right), when\, (x \leq \theta).$$

and

$$F(x|\theta, \lambda) = 1 - \frac{1}{2} exp\left(-\frac{|\theta-x|}{\lambda}\right), when\, (x > \theta).$$

Unlike the exponential, the laplace is defined $-\infty < x < \infty$. If $\theta = 0$, then the probability density function for Laplace on $x > 0$ is equal to 1/2 of the probability of the exponential. The expected value of a Laplace distribution is

$$E(x) = \theta$$

As in the case of other symmetrical distributions, such as the Normal and the logistic distributions, Laplace's location is the same as its mean, median, and mode. The variance is:

$$Var(x) = 2\lambda^2.$$

Using the Laplace distribution with rate = 1, 10,000 simulations were carried out. Simulation results are presented in Table 4.2;

Table 4.3: Simulation Results of Adaptive Test and Parametric Test for Laplace Distribution

Sample size	F-test			Adaptive Test			
(n_1, n_2, n_3)	Value	p-value	σ	Score	Value	p-value	τ
(5,5,5)	2.167	0.157	1.1912	SH	0.66953	0.53007	0.962598
(10,10,10)	0.368	0.696	1.3730	SH	2.99766	0.06673	1.202722
(15,15,15)	0.929	0.403	1.4426	SH	1.35735	0.2684	1.040336
(20,20,20)	0.783	0.462	1.3142	SH	1.35488	0.26616	1.265793

The Laplace distribution with rate=2 was identified by the adaptive test as a symmetric skewed and heavy tailed distribution. From Table 4.3, the variance returned suggest that the adaptive test performed better at all the level sample sizes considered than the F-test. However, the two tests failed to reject the null hypothesis of no difference in level means at all sample sizes considered. The ARE of the adaptive test over the F-test if the data under consideration is from a Laplace distribution is between 4% and 20%. It was observed that the ARE decreased as sample sizes of the levels increased.

4.2.3 Truncated Logistic Distribution

The Logistic distribution is characterized by location μ and scale $= \sigma$ has distribution function,

$$F(x) = \frac{1}{1 + e^{-(x-\mu)/\sigma}}$$

and density

$$f(x) = \frac{1}{\sigma} \frac{e^{(x-\mu)/\sigma}}{(1 + e^{(x-\mu)/\sigma})^2}$$

It is a long-tailed distribution with mean=μ and variance=$\pi^2/3\sigma^2$.

Using the Truncated Logistic distribution, 10,000 simulations were carried out. Simulation results are presented in Table 4.2;

Table 4.4: Simulation Results of Adaptive Test and Parametric Test for Truncated Logistic Distribution

Sample size	F-test			Adaptive Test			
(n_1, n_2, n_3)	Value	p-value	σ	Score	Value	p-value	τ
(5,5,5)	2.728	0.106	0.0690	SL	8.78345	0.00447	0.145689
(10,10,10)	2.379	0.112	0.0765	SL	0.58683	0.56303	0.2340154
(15,15,15)	0.735	0.486	0.06937	SL	5.57869	0.0071	0.2100686
(20,20,20)	2.403	0.0995	0.08595	SL	0.40197	0.67088	0.2830328

The truncated logistic distribution with $\mu = 0$ and $\sigma = 1$ was identified by the adaptive test as a symmetric skewed and light tailed distribution. From Table 4.4, the variance returned suggest that the adaptive test underperformed at all the level sample sizes considered than the F-test. However, the two tests failed to reject the null hypothesis of no difference in level means at all sample sizes considered except at sample size 5, where the Adaptive test rejected H_0 whereas the F-test resulted otherwise. The ARE of the adaptive test over the F-test if the data under consideration is from a truncated logistic distribution is between 30% and 48%. It was observed that the ARE decreased as sample sizes of the levels increased.

4.2.4 Contaminated Normal Distribution

Let Z be random samples drawn from normal distributions, $I_{1-\epsilon}$ be a discrete random variable defined by;

$$I_{1-\epsilon} = \begin{cases} 1 \text{ with prob. } 1-\epsilon \\ 0 \text{ with prob } \epsilon \ , \end{cases}$$

and assume that Z and $I_{1-\epsilon}$ are independent (Hogg et al., 2005).

Let $Q = ZI_{1-\epsilon} + \sigma_c Z(1 - I_{1-\epsilon})$, then by the independence of Z and $I_{1-\epsilon}$, the cdf of Q is given by;

$$\begin{aligned}
F_Q(q) =& Pr(Q \leq q) \\
=& Pr[Q \leq q, I_{1-\epsilon} = 1] + Pr[Q \leq q, I_{1-\epsilon} = 0] \\
=& Pr[Q \leq q / I_{1-\epsilon} = 1] Pr[I_{1-\epsilon} = 1] + Pr[Q \leq q / I_{1-\epsilon} = 0] Pr[I_{1-\epsilon} = 0] \\
=& Pr[Q \leq q](1-\epsilon) + Pr[Q \leq q/\sigma_c]\epsilon \\
=& \Phi(q)(1-\epsilon) + \Phi\left(\frac{q}{\sigma_c}\right)\epsilon
\end{aligned}$$

where σ_c is the standard deviation of contamination, $I_{1-\epsilon}$ is the characteristics function, ϵ is the percentage of contamination. Thus, the pdf of the contaminated distribution is given by;

$$\frac{\mathrm{d}F_Q(q)}{\mathrm{d}q} = (1-\epsilon)f_Q(q) + \frac{\epsilon}{\sigma_c} f_Q\left(\frac{q}{\sigma_c}\right).$$

Simulation results are presented in Table 4.5;

Table 4.5: Simulation Results of Adaptive Test and Parametric Test for Contaminated Normal Distribution

Sample size	Level	F-test			Adaptive Test			
(n_1, n_2, n_3)	%	Value	p-value	σ	Score	Value	p-value	τ
(5,5,5)	5%	1.002	0.396	9.672	SH	2.20517	0.15289	6.111384
	10%	1.764	0.213	4.797	SH	4.96343	0.02687	3.038976
	15%	0.579	0.575	6.145	SH	1.18599	0.33884	4.818696
	20%	3.317	0.0713	11.40	SH	0.95542	0.41207	6.154036
(10,10,10)	5%	2.246	0.125	7.171	SM	1.76329	0.19066	8.800966
	10%	1.472	0.247	7.79	SM	2.22532	0.12747	8.875872
	15%	1.281	0.294	7.279	SM	0.05517	0.94643	7.649184
	20%	3.133	0.0597	6.277	SH	0.82296	0.44985	5.836568
(15,15,15)	5%	0.044	0.957	8.097	SH	0.46592	0.63076	4.821166
	10%	1.09	0.345	8.359	SH	1.19274	0.31345	5.927768
	15%	0.266	0.768	7.116	SH	0.56405	0.57315	5.865964
	20%	0.939	0.399	6.808	SH	1.96386	0.15299	5.401612
(20,20,20)	5%	2.783	0.0703	7.374	SH	1.0307	0.36331	5.998714
	10%	2.053	0.138	5.791	SH	0.10909	0.89684	4.705272
	15%	0.095	0.91	6.211	SH	0.68746	0.50697	4.953914
	20%	0.809	0.45	5.694	SH	2.84496	0.06642	5.708576

The Normal distribution at 5% contamination was identified by the adaptive test as a symmetric skewed and heavy tailed distribution. From Table 4.5, the variability in the model fit suggest that the adaptive test outperformed at all the sample sizes considered than the F-test. However, the two tests failed to reject the null hypothesis of no difference in level means at all sample sizes considered. The ARE of the Adaptive test over the F-test if the data under consideration is from a 5% contaminated normal distribution is between 20% and 40%. It was observed that the ARE decreased as sample sizes of the levels increased.

At 10% contamination, sample size of 5 was identified as symmetric skewed and heavy tailed distribution whereas that for the other sample sizes were identified as symmetric skewed and heavy tailed distribution. From Table 4.5, the variance returned by the models suggest that the Adaptive Test performed better than the F-test. The ARE of the Adaptive test over the F-test is between 20% to 50%.

At 15% contamination, sample size of 5 was identified as symmetric skewed and heavy tailed distribution whereas that for the other sample sizes were identified as symmetric skewed and heavy tailed distribution. From Table 4.5, the variance returned by the models suggest that the Adaptive Test performed better than the F-test at sample sizes 5, 10 and 15. The ARE of the Adaptive test over the F-test is between 17% to 25%.

At 20% contamination, the distribution was identified as symmetric skewed and heavy tailed distribution. From Table 4.5, the variance returned by the models suggest that the Adaptive Test performed better than the F-test at sample sizes 5, 10 and 15. The ARE of the Adaptive test over the F-test is between 7% to 48%. However, at sample size 20, the F-test performed better than the Adaptive test with ARE of 2%.

4.2.5 Mixture of Distributions

We considered the situation where the data for the various levels are from different distributions. 10,000 simulations were performed where data for level one samples were simulated from weibull distribution (shape = 2 and scale = 1), data for level two samples were from truncated logistic distribution (location = 0 and scale = 1) and data for level three samples were from laplace distribution (rate = 1). Table 4.6 presents the results from the simulated studies.

Table 4.6: Simulation Results of Adaptive Test and Parametric Test for Mixture of Distributions

Sample size	F-test			Adaptive Test			
(n_1, n_2, n_3)	Value	p-value	σ	Score	Value	p-value	τ
(5,5,5)	2.94	0.0914	0.9452	SH	6.27413	0.01364	0.4236017
(10,10,10)	1.88	0.172	1.217	SH	2.22578	0.12742	0.6232764
(15,15,15)	4.973	0.0115	0.7351	SH	5.26962	0.00908	0.6460359
(20,20,20)	0.867	0.426	1.0393	SH	10.00631	0.00019	0.9048917

The Adaptive test identified the data as a symmetric skewed and heavy tailed distribution. The test decision on rejection or otherwise of H_0 at sample sizes (5,5,5) and (20,20,20) differs among the two tests at 5% level of significance. The Adaptive test identified significance difference in means whiles the F-test concluded otherwise. However, at sample sizes (10,10,10) and (15,15,15), both test yielded same decision results. The variances for the two tests suggest that the Adaptive test performs better than the F-test. The ARE of the Adaptive test over the F-test is between 40% and 64%. There seems to be a decline in efficiency of the Adaptive test over F-test as the sample sizes increase.

4.3 Fitting Models to Least Squares Residuals

This section presents simulated results from adapting and performing F-test on least squares residuals.

4.3.1 Normal distribution

10,000 simulation results are presented in Table 4.7;

Table 4.7: Simulation Results of Adaptive Test and Parametric Test of Least Squares Residuals for data from the Normal distribution

Sample size	F-test			Adaptive Test			
(n_1, n_2, n_3)	Value	p-value	σ	Score	Value	p-value	τ
(5,5,5)	0	1	3.499	SM	0.00797	0.99207	6.330272
(10,10,10)	0	1	3.126	SM	0.01473	0.98531	4.772586
(15,15,15)	0	1	3.357	SM	0.04802	0.95317	4.772586
(20,20,20)	0	1	3.138	SM	0.04556	0.9555	4.115804

The adaptive test identified the least squares residuals of the data from the normal distribution with $\mu = 0$ and $\sigma = 1$ as a symmetric skewed and medium tailed distribution. From Table 4.7, the parametric F-test outperforms the adaptive test at all the level sample sizes considered by returning small variances compared to those for the adaptive test. However, the decision from the two tests are similar for all sample sizes considered. The ARE of the F-test over the adaptive test if the data under consideration is from a normal distribution is between 23% and 55%. It was observed that the ARE decreased as sample sizes of the levels increased.

4.3.2 Laplace Distribution

Simulation results are presented in Table 4.8. The Laplace distribution with rate=1 was identified by the adaptive test as a symmetric skewed and heavy tailed distribution. From Table 4.8, the variance returned suggest that the adaptive test performed better at all the level sample sizes considered than the F-test. However, the two models failed to reject the null hypothesis of no difference in level means at all sample sizes considered. The ARE of the adaptive test over the F-test if the

Table 4.8: Simulation Results of Adaptive Test and Parametric Test for Laplace Distribution

Sample size	F-test			Adaptive Test			
(n_1, n_2, n_3)	Value	p-value	σ	Score	Value	p-value	τ
(5,5,5)	0	1	0.5199	SH	0.17005	0.84562	0.4820709
(10,10,10)	0	1	0.5405	SH	0.40064	0.67381	0.4326355
(15,15,15)	0	1	0.5903	SH	0.26173	0.77096	0.387311
(20,20,20)	0	1	0.6081	SH	2.05907	0.13696	0.4558181

data under consideration is from a Laplace distribution is between 8% and 34%. It was observed that the ARE increased as sample sizes of the levels increased.

4.3.3 Truncated Logistic Distribution

Simulation results are presented in Table 4.9.

Table 4.9: Simulation Results of Adaptive Test and Parametric Test for Truncated Logistic Distribution

Sample size	F-test			Adaptive Test			
(n_1, n_2, n_3)	Value	p-value	σ	Score	Value	p-value	τ
(5,5,5)	0	1	0.2445	SM	0.10068	0.90497	0.3207393
(10,10,10)	0	1	0.3298	SL	0.10387	0.9017	0.4851264
(15,15,15)	0	1	0.2346	SL	0.05094	0.9504	0.5292146
(20,20,20)	0	1	0.3163	SL	0.06224	0.93973	0.5813575

The Truncated Logistic distribution with $\mu = 0$ and $\sigma = 1$ was identified by the adaptive test as a symmetric skewed and light tailed distribution for sample sizes 10, 15 and 20 and as a symmetric skewed and medium tailed distribution for sample size 5. From Table 4.9, the variance returned suggest that the adaptive test underperformed at all the level sample sizes considered than the F-test. The ARE of the F-test over the Adaptive test if the data under consideration is from a truncated logistic distribution is between 24% and 57%. It was observed that the ARE increased as sample sizes of the levels increased.

4.3.4 Contaminated Normal Distribution

Simulation results are presented in Table 4.10. The Normal distribution at 5% contamination was identified by the adaptive test as a symmetric skewed and

Table 4.10: Simulation Results of Adaptive Test and Parametric Test for Contaminated Normal Distribution

Sample size	Level	F-test			Adaptive Test			
(n_1, n_2, n_3)	%	Value	p-value	σ	Score	Value	p-value	τ
(5,5,5)	5%	0	1	4.161	SH	0.09876	0.90669	3.420167
	10%	0	1	2.4866	SH	0.22523	0.80163	2.078935
	15%	0	1	2.65198	SH	0.55388	0.58873	2.604724
	20%	0	1	2.5589	SH	0.2209	0.80499	1.993167
(10,10,10)	5%	0	1	2.48837	SM	0.01711	0.98305	3.712071
	10%	0	1	2.7761	SM	0.04056	0.96031	4.040014
	15%	0	1	2.6079	SM	0.23305	0.79369	2.615836
	20%	0	1	2.9339	SM	0.02253	0.97774	2.97184
(15,15,15)	5%	0	1	2.83143	SM	0.13192	0.87677	3.116345
	10%	0	1	2.89603	SM	0.02491	0.97541	3.129687
	15%	0	1	2.2261	SM	0.13855	0.87102	2.32268
	20%	0	1	3.12250	SM	0.04162	0.95927	3.49266
(20,20,20)	5%	0	1	2.92147	SM	0.16957	0.84445	3.113383
	10%	0	1	2.92147	SM	0.01925	0.98094	3.298762
	15%	0	1	2.73203	SM	0.27948	0.75721	2.861447
	20%	0	1	3.04729	SM	0.02595	0.9744	3.176907

medium tailed distribution for sample sizes 10, 15 and 20. However, at sample size 5, it was identified as symmetric skewed and heavy tailed distribution. From Table 4.10, the variance returned suggest that the adaptive test outperformed the F-test at sample size 5. However, the two models failed to reject the null hypothesis of no difference in level means at all sample sizes considered. The ARE of the adaptive test over the F-test at sample size 5 is 18%. But at sample sizes 10, 15 and 20, the F-test performed better than the Adaptive test, with an ARE of between 6% and 33%. It is worthy of note that the ARE decreases as the sample sizes increase.

At 10% contamination, the distribution was identified by the adaptive test as a symmetric skewed and medium tailed distribution at sample sizes 10, 15 and 20. However, at sample size 5, it was identified as symmetric skewed and heavy tailed distribution. From Table 4.10, the variance returned suggest that the adaptive test outperformed the F-test at sample size 5. However, the two models failed to reject the null hypothesis of no difference in level means at all sample sizes considered. The ARE of the adaptive test over the F-test at sample size 5 is 16%.

But at sample sizes 10, 15 and 20, the F-test performed better than the Adaptive test, with an ARE of between 6% and 33%. The ARE decreases as the sample sizes increase.

At 15% contamination, the distribution was identified as a symmetric skewed and medium tailed distribution at sample sizes 10, 15 and 20. However, at sample size 5, it was identified as symmetric skewed and heavy tailed distribution. From Table 4.10, the variance returned suggest that the adaptive test outperformed the F-test at sample size 5. However, the two models failed to reject the null hypothesis of no difference in level means at all sample sizes considered. The ARE of the adaptive test over the F-test at sample size 5 is 8%. Meanwhile, at sample sizes 10, 15 and 20, the F-test performed better than the Adaptive test, with an ARE of between 4% and 8%. The ARE decreases as the sample sizes increase.

At 20% contamination, the distribution was identified as a symmetric skewed and medium tailed distribution at sample sizes 10, 15 and 20. However, at sample size 5, it was identified as symmetric skewed and heavy tailed distribution. From Table 4.10, the variance returned suggest that the adaptive test outperformed the F-test at sample size 5. However, the two models failed to reject the null hypothesis of no difference in level means at all sample sizes considered. The ARE of the adaptive test over the F-test at sample size 5 is 22%. Meanwhile, at sample sizes 10, 15 and 20, the F-test performed better than the Adaptive test, with an ARE of between 1% and 11%.

4.3.5 Mixture of Distributions

10,000 simulations were performed where data for level one samples were simulated from the Weibull distribution (shape = 2 and scale = 1), data for level two samples were from the Truncated Logistic distribution (location = 0 and scale =

1) and data for level three samples were from the Laplace distribution (rate = 1). Table 4.11 presents the results from the simulated studies.

Table 4.11: Simulation Results of Adaptive Test and Parametric Test for Mixture of Distributions

Sample size	F-test			Adaptive Test			
(n_1, n_2, n_3)	Value	p-value	σ	Score	Value	p-value	τ
(5,5,5)	0	1	0.79347	SH	2.0458	0.17199	0.3439972
(10,10,10)	0	1	1.2284	SH	0.05751	0.94423	0.3619935
(15,15,15)	0	1	0.8486	SH	0.22933	0.79606	0.5834737
(20,20,20)	0	1	1.0733	SH	1.19387	0.31051	0.4278501

The adaptive test identified the data as a symmetric skewed and heavy tailed distribution. The variances for the two tests from Table 4.11 suggest that the Adaptive test performs better than the F-test. The ARE of the adaptive test over the F-test is between 32% and 70%.

4.4 Fitting Models to Wilcoxon Winsorised Scores Residuals

This section presents simulated results from adapting and performing F-test on Wilcoxon Winsorised Scores residuals.

4.4.1 Normal distribution

10,000 simulation results are presented in Table 4.12. The adaptive test identified

Table 4.12: Simulation Results of Adaptive Test and Parametric Test of the Wilcoxon Winsorised Scores Residuals for data from the Normal distribution

Sample size	F-test			Adaptive Test			
(n_1, n_2, n_3)	Value	p-value	σ	Score	Value	p-value	τ
(5,5,5)	0.022	0.978	2.15267	SH	0.05218	0.94938	2.004418
(10,10,10)	0.002	0.998	1.47513	SM	0	1	2.007322
(15,15,15)	0.03	0.971	1.7506	SM	0	1	2.112924
(20,20,20)	0.196	0.822	1.8105	SM	0	1	2.279802

the Wilcoxon Winsorised scores residuals of the data from the Normal distribution with $\mu = 0$ and $\sigma = 2$ as a symmetric skewed and medium tailed distribution for sample sizes 10, 15, and 20. However, at sample size 5, it was identified as a symmetric skewed and heavy tailed distribution. From Table 4.12, the parametric F-test performs quiet better than the Adaptive test at sample sizes 10, 15 and 20, by returning small variances. The ARE of the F-test over the Adaptive test if the data under consideration is from a normal distribution is between 17% and 36%. However, at sample size 5, the Adaptive test performed better than the F-test, with an ARE of 8%.

4.4.2 Laplace Distribution

Simulation results are presented in Table 4.13.

Table 4.13: Simulation Results of Adaptive Test and Parametric Test of the Wilcoxon Winsorised Scores Residuals for Laplace Distribution

Sample size	F-test			Adaptive Test			
(n_1, n_2, n_3)	Value	p-value	σ	Score	Value	p-value	τ
(5,5,5)	0.315	0.736	0.89549	SH	0	1	0.7514154
(10,10,10)	0.167	0.847	0.7075	SH	0	1	0.3416036
(15,15,15)	0.84	0.439	0.99207	SH	0	1	0.6948067
(20,20,20)	0.279	0.758	0.73287	SH	0	1	0.5607453

The Laplace distribution with $rate = 1$ was identified by the adaptive test as a symmetric skewed and heavy tailed distribution for all the sample sizes considered. From Table 4.13, the variance returned suggest that the Adaptive test performed better at all the level sample sizes considered than the F-test. However, the two models failed to reject the null hypothesis of no difference in level means at all sample sizes considered. The ARE of the Adaptive test over the F-test if the data under consideration is from a Laplace distribution is between 16% and 52%.

4.4.3 Truncated Logistic Distribution

Simulation results are presented in Table 4.14.

The Truncated Logistic distribution with $\mu = 0$ and $\sigma = 1$ was identified by the adaptive test as a symmetric skewed and light tailed distribution for sample

Table 4.14: Simulation Results of Adaptive Test and Parametric Test of the Wilcoxon Winsorised Scores Residuals for Truncated Logistic Distribution

Sample size	F-test			Adaptive Test			
(n_1, n_2, n_3)	Value	p-value	σ	Score	Value	p-value	τ
(5,5,5)	0.325	0.729	0.8666	SH	0	1	0.5888089
(10,10,10)	0.129	0.88	0.7385	SL	0.49609	0.61435	0.6480697
(15,15,15)	0.238	0.789	0.82885	SM	0.36055	0.69943	0.9515803
(20,20,20)	0.128	0.88	0.80436	SM	0.30459	0.73862	0.8185455

sizes 10, as a symmetric skewed and medium tailed distribution for sample size 15 and 20 and as a symmetric skewed and heavy tailed distribution for sample size 5. From Table 4.14, the variance returned suggest that the Adaptive test underperformed than the F-test at level sample sizes 15 and 20. The ARE of the F-test over the Adaptive test is between 2% and 13%. However, at sample sizes 5 and 10, the Adaptive test performed better than the F-test, with an ARE of 32% and 12% respectively.

4.4.4 Contaminated Normal Distribution

Simulation results are presented in Table 4.15. The Normal distribution at 5%

Table 4.15: Simulation Results of Adaptive Test and Parametric Test of the Wilcoxon Winsorised Scores Residuals for Contaminated Normal Distribution

Sample size	Level	F-test			Adaptive Test			
(n_1, n_2, n_3)	%	Value	p-value	σ	Score	Value	p-value	τ
(5,5,5)	5%	0.49	0.624	2.2280	SH	0	1	2.077739
	10%	0.179	0.839	2.40583	SH	0	1	2.22602
	15%	0.082	0.921	2.82913	SH	0	1	2.208968
	20%	0.61	0.559	2.60576	SH	0	1	2.92846
(10,10,10)	5%	0.052	0.949	2.91719	SM	0	1	2.93733
	10%	0.028	0.972	2.3394	SM	0	1	2.363568
	15%	0.016	0.984	3.01380	SM	0	1	3.378316
	20%	0.001	0.999	2.11447	SM	0	1	2.34367
(15,15,15)	5%	0.063	0.939	3.5523	SM	0	1	4.03708
	10%	0.027	0.973	3.00366	SM	0	1	3.254252
	15%	0.082	0.921	2.315599	SM	0	1	2.505954
	20%	0.003	0.89	3.202187	SM	0	1	3.693798
(20,20,20)	5%	0.023	0.977	2.7344	SM	0	1	2.836659
	10%	0.003	0.997	2.78998	SM	0	1	2.844048
	15%	0.013	0.987	2.60576	SM	0	1	2.909152
	20%	0.117	0.89	2.45438	SM	0	1	2.539937

contamination was identified by the adaptive test as a symmetric skewed and medium tailed distribution for sample sizes 10, 15 and 20. However, at sample size 5, it was identified as symmetric skewed and heavy tailed distribution. From Table 4.15, the variance returned suggest that the adaptive test outperformed the F-test at sample size 5 . However, the two models failed to reject the null hypothesis of no difference in level means at all sample sizes considered. The ARE of the Adaptive test over the F-test at sample size 5 is about 7%. But at sample sizes 10, 15 and 20, the F-test performed better than the Adaptive test, with an ARE of between 1% and 12%.

At 10% contamination, the distribution was identified by the adaptive test as a symmetric skewed and medium tailed distribution at sample sizes 10, 15 and 20. However, at sample size 5, it was identified as symmetric skewed and heavy tailed distribution. From Table 4.15, the variance returned suggest that the adaptive test outperformed the F-test at sample size 5. Again, the two models failed to reject the null hypothesis of no difference in level means at all sample sizes considered. The ARE of the Adaptive test over the F-test at sample size 5 is 8%. But at sample sizes 10, 15 and 20, the F-test performed better than the Adaptive test, with an ARE of between 1% and 8%.

At 15% contamination, the distribution was identified as a symmetric skewed and medium tailed distribution at sample sizes 10, 15 and 20. However, at sample size 5, it was identified as symmetric skewed and heavy tailed distribution. From Table 4.15, the variance returned suggest that the adaptive test outperformed the F-test at sample size 5 . However, the two models failed to reject the null hypothesis of no difference in level means at all sample sizes considered. The ARE of the Adaptive test over the F-test at sample size 5 is 22%. Meanwhile, at sample sizes 10, 15 and 20, the F-test performed better than the Adaptive test, with an ARE of between 8% and 11%.

At 20% contamination, the distribution was identified as a symmetric skewed and medium tailed distribution at sample sizes 10, 15 and 20. However, at sample size 5, it was identified as symmetric skewed and heavy tailed distribution. From Table 4.15, the variance returned suggest that the adaptive test performed better than the F-test at sample size 5 . However, the two models failed to reject the null hypothesis of no difference in level means at all sample sizes considered. The ARE of the Adaptive test over the F-test at sample size 5 is 13%. Meanwhile, at sample sizes 10, 15 and 20, the F-test performed better than the Adaptive test, with an ARE of between 3% and 13%.

4.4.5 Mixture of Distributions

10,000 simulations were performed where data for level one samples were simulated from weibull distribution (shape = 2 and scale = 1), data for level two samples were from truncated logistic distribution (location = 0 and scale = 1) and data for level three samples were from Laplace distribution (rate = 1). Table 4.16 presents the results from the simulated studies.

Table 4.16: Simulation Results of Adaptive Test and Parametric Test of the Wilcoxon Winsorised Scores Residuals for Mixture of Distributions

Sample size	F-test			Adaptive Test			
(n_1, n_2, n_3)	Value	p-value	σ	Score	Value	p-value	τ
(5,5,5)	0.738	0.498	1.104355	LH	4.50916	0.03463	0.7247588
(10,10,10)	0.092	0.912	0.481363	LM	0.01351	0.98658	0.3602497
(15,15,15)	0.334	0.718	0.87647	SH	0	1	0.5245058
(20,20,20)	0.04	0.961	0.73702	SH	3e-05	0.9997	0.4968895

From Table 4.16, the adaptive test identified the data at sample size 5 as a left skewed and heavy tailed distribution, at sample size 10 as left skewed and medium tailed distribution and at sample sizes 15 and 20 as symmetric skewed and heavy tailed distributions. The variances for the two tests from Table 4.16 suggest that the Adaptive test performs better than the F-test at all sample sizes. The ARE of the Adaptive test over the F-test is between 33% and 40%.

4.5 Application

4.5.1 Apple Orchard Grafting Experiment

Pearce (1992) conducted an experiment to investigate five types of root-stock in apple orchard grafting. The following data represent the extension growth (cm) after four years.

X1 = extension growth for type I

X2 = extension growth for type II

X3 = extension growth for type III

X4 = extension growth for type IV

X5 = extension growth for type V

Table 4.17: Output from the Apple Orchard Grafting Experiment

Type	Extension growth (cm)							
X1	2569	2928	2865	3844	3027	2336	3211	3037
X2	2074	2885	3378	3906	2782	3018	3383	3447
X3	2505	2315	2667	2390	3021	3085	3308	3231
X4	2838	2351	3001	2439	2199	3318	3601	3291
X5	1532	2252	3083	2330	2079	3366	2416	3100

The analysis performed is presented in Table (4.18);

Table 4.18: Results of Adaptive Test and Parametric Test for Apple Orchard Grafting Experiment

Sample size	F-test			Adaptive Test			
(n_1, n_2, n_3)	Value	p-value	σ	Score	Value	p-value	τ
(8,8,8)	1.49	0.226	510.2362	SM	1.08381	0.37948	618.4604

From Table 4.18, the Adaptive test identified the data as symmetric skewed and medium tailed distribution. The F-test reported the least variance depicting it as the most efficient for the data. The ARE of F-test over the Adaptive test is 17.5%. However, both test conclusions favour the null hypothesis, H_0. This results agree with the findings from the simulation studies, where F-test performed better than the Adaptive Test for data from symmetric skewed and moderate tailed distribution.

4.5.2 Automatic Valve Shutoff Mechanism Experiment

This example was extracted from Montgomery (2001). The response time in milliseconds was determined for three different types of circuits that could be used in an automatic valve mechanism. The results are shown in the Table 4.19

Table 4.19: Output from the Automatic Valve Shutoff Mechanism Experiment

Circuit Type	Response Time				
1	9	12	10	8	15
2	20	21	23	17	30
3	6	5	8	16	7

The analysis performed is presented in Table (4.19);

Table 4.20: Results of Adaptive Test and Parametric Test for Automatic Valve Shutoff Mechanism Experiment

Sample size	F-test			Adaptive Test			
(n_1, n_2, n_3)	Value	p-value	σ	Score	Value	p-value	τ
(5,5,5)	16.08	0.000402	4.11096	RL	16.08075	4e-04	3.325721

From Table 4.19, the Adaptive test identified the data as right skewed and light tailed distribution. Although, both test at 1% and 5% level of significance concluded against the null hypothesis, H_0, the Adaptive test reported the least variance depicting it as the most efficient for the data. The ARE of Adaptive test over the F-test is 19.1%. This results agree with the findings from the simulation studies, where Adaptive test performed better than the F-Test for data from non-symmetric skewed and varying tailed distribution.

Chapter 5

Summary and Conclusion

5.1 Introduction

This chapter presents the summary of the findings, discussions and conclusion of the study.

5.2 Summary

The asymptotic properties of statistical estimators greatly rely on the central limit theorem, however, in practice sample sizes are not often large. The F-test employs the assumption of normality of the data, and could this test be the optimal test to conduct in all situations? In this thesis, an adaptive testing procedure is developed for hypothesis testing of one-way ANOVA models and the efficiency compared to the traditional F-test. The adaptive procedure proposed used the linear rank test, and thus, non-parametric. The procedure uses the data to ascertain which statistical test is most efficient. It is conducted in two phases. In the first phase, a selection statistic is computed from the estimate of skewness and tail-weight. In the second phase, the selector statistic is used to determine the appropriate score function for the analysis. This procedure has been proven to have several advantages. Notably, it can increase the power of the test if the error distribution is skewed and makes narrow confidence internals, are robust for both validity and efficiency and automatically downweight outliers, which has the effect of making the results less sensitive to observations that do not agree with the model (O'Gorman, 2004).

The procedure for the adaptive test is as follows;

- Let $X_{11}, X_{12}, \ldots, X_{1n_1}, X_{21}, X_{22}, \ldots, X_{2n_2}, \cdots X_{n1}, X_{n2}, \ldots, X_{nn_k}$ be the ordered combined random samples from continuous distribution function $f(t)$ with some amount of variations denoted by δ among the samples, that is, $f(t-\delta)$. The hypothesis that there is no difference in the sample means, that is, $H_0 : \delta = 0$ is tested against $H_1 : \delta \neq 0$

- Data is examined and classified by considering skewness and tail weight from a class of continuous distribution. The measure of skewness (Q_1) according to Hogg et al. (1975) is define as;

$$Q_1 = \frac{\overline{U}_{5\%} - \overline{M}_{50\%}}{\overline{M}_{50\%} - \overline{L}_{5\%}}, \tag{5.1}$$

  where $\overline{U}_{5\%}$, $\overline{M}_{50\%}$ and $\overline{L}_{5\%}$ are the averages for the upper 5%, middle 50% and the lower 5% of the $X'^{s}_{(j)}$ the ordered combined samples respectively.
  The measure of tailweight, Q_2, according to Hogg et al. (1975) defined as;

$$Q_2 = \frac{\overline{U}_{5\%} - \overline{L}_{5\%}}{\overline{U}_{50\%} - \overline{L}_{50\%}}, \tag{5.2}$$

  where $\overline{U}_{50\%}$, and $\overline{L}_{50\%}$ are the averages of the upper 50% and lower 50% of the $X'^{s}_{(j)}$; $N = n_1 + \cdots + n_k$ observations of the combined samples.
  These two statistics are together called selector statistic, $S = (Q_1, Q_2)$.

- Then specify cutoff points for the measures of skewness and tailweight. The benchmarks proposed by Al-shomrani (2003) is used. The cutoff values depend on the sample size n, but as $n \to \infty$, the measures converges to that proposed by Hogg et al. (1975).
  For Q_1^*, the

$$\text{lower cutoff } (clq1) = 0.36 + \frac{0.68}{n}$$

$$\text{upper cutoff } (cuq1) = 2.73 - \frac{3.72}{n}$$

and for Q_2^*, when the sample size is less than 25

$$\text{lower cutoff}(clq2) = 2.17 - \frac{3.01}{n}$$

$$\text{upper cutoff}(cuq2) = 2.63 - \frac{3.94}{n}$$

but when the sample size is at least 25, then the lower and upper cutoff are respectively defined as;

$$\text{lower cutoff}(clq2) = 2.24 - \frac{4.68}{n}$$

and

$$\text{upper cutoff}(cuq2) = 2.63 - \frac{9.37}{n}.$$

- The distributional categorization is evaluated as displayed in Table 5.1

Table 5.1: The Nine distributional categorization of data

Skewness	Tailweight	Distribution
$Q_1 \leq clq1$	$Q_2 \leq clq2$	Left skewed light tailed
$Q_1 \leq clq1$	$Q_2 > clq2$ and $Q_2 \leq cuq2$	Left skewed medium tailed
$Q_1 \leq clq1$	$Q_2 > cuq2$	Left skewed heavy tailed
$Q_1 \leq clq2$	$Q_2 \leq clq2$	Symmetric skewed light tailed
$Q_1 \leq clq2$	$Q_2 > clq2$ and $Q_2 \leq cuq2$	Symmetric skewed medium tailed
$Q_1 \leq clq2$	$Q_2 > cuq2$	Symmetric skewed heavy tailed
$Q_1 \leq clq2$	$Q_2 \leq clq2$	Right skewed light tailed
$Q_1 \leq clq2$	$Q_2 > clq2$ and $Q_2 \leq cuq2$	Right skewed medium tailed
$Q_1 \leq clq2$	$Q_2 > cuq2$	Right skewed heavy tailed

- The cutoff points are used in the selection of a rank score associated with the unknown distribution.The rank test used is given by;

$$T_\varphi = \sum_{j=1}^{n} \varphi \left[\frac{R(Z_j)}{n+1} \right] \tag{5.3}$$

where $\varphi(j) = \varphi\left(\frac{j}{n+1}\right)$, $a_\varphi(1),\ a_\varphi(2)\ldots\ \ a_\varphi(n)$ are scores.

- Nine winsorised scores classified by Hettmansperger (1984) under four generic cases are used.They include;

1.

$$\varphi_I(u) = \begin{cases} s_3, & u > s_1 \\ s_3 + \frac{s_3 - s_2}{s_1}(u - s_1), & \text{otherwise.} \end{cases}$$

2.

$$\varphi_{II}(u) = \begin{cases} \frac{-s_3}{s_1}(u - s_1), & u < s_1 \\ \frac{-s_4}{s_2 - 1}(u - 1) + s_4, & u > s_2 \\ 0, & \text{otherwise.} \end{cases}$$

3.

$$\varphi_{III}(u) = \begin{cases} s_2, & u < s_1 \\ s_3 + \frac{s_2 - s_3}{s_1 - 1}(u - 1), & \text{otherwise.} \end{cases}$$

4.

$$\varphi_{IV}(u) = \begin{cases} s_3, & u < s_1 \\ s_4, & u > s_2 \\ s_3 + \frac{s_4 - s_3}{s_2 - s_1}(u - s_1), & \text{otherwise.} \end{cases}$$

where s_1, s_2, s_3, s_4 and s_5 are parameters and $\varphi_i(j) = \varphi_i\left(\frac{j}{n+1}\right)$.

- The associated score functions and parameters are presented in Table 5.2

Table 5.2: Benchmarks for Winsorised Scores

Skewness	Tail weight	Score function
Left	Light	$\varphi_1 = \varphi_{III}$ with parameters ($s_1 = 0.1$,$s_2 = -1$ and $s_3 = 2.0$)
Left	Medium	$\varphi_2 = \varphi_{III}$ with parameters ($s_1 = 0.3$,$s_2 = -1$ and $s_3 = 2.0$)
Left	Heavy	$\varphi_3 = \varphi_{III}$ with parameters ($s_1 = 0.5$,$s_2 = -1$ and $s_3 = 2.0$)
Symmetric	Light	$\varphi_4 = \varphi_{II}$ with parameters ($s_1 = 0.25$,$s_2 = 0.75$, $s_3 = -1$, $s_4 = 1.0$ and $s_5 = 0.0$)
Symmetric	Medium	Wilcoxon scores $\varphi_5 = \sqrt{12}[u - \frac{1}{2}]$
Symmetric	Heavy	$\varphi_6 = \varphi_{IV}$ with parameters ($s_1 = 0.25$,$s_2 = 0.75$, $s_3 = -1$ and $s_4 = 1.0$)
Right	Light	$\varphi_7 = \varphi_{II}$ with parameters ($s_1 = 0.9$,$s_2 = -2$ and $s_3 = 1.0$ $s_4 = 1$, and $s_5 = 0$)
Right	Medium	$\varphi_8 = \varphi_I$ with parameters ($s_1 = 0.7$,$s_2 = -2$ and $s_3 = 1.0$)
Right	Heavy	$\varphi_9 = \varphi_I$ with parameters ($s_1 = 0.5$,$s_2 = -2$ and $s_3 = 1.0$)

The thesis employed the above procedure in a simulation study. Simulation under Pure Hogg (run under H_0), fitting models to residuals from an Ordinary Least

Square (OLS), fitting models to residuals from Wilcoxon Scores, and application on real data were conducted to ascertain the set objectives. The R-package, *Rfit* by (Kloke and McKean, 2012) was used for the study. Some other functions were written and implemented in R for the study. Simulation results for normal, double exponential, contaminated normal, the truncated logistic and a mixture of distributions for a balanced ANOVA models are presented. 10,000 simulations were carried out for 5, 10, 15, and 20 observations each, assigned to three (3) levels with intralevel correlation coefficient $\rho = 0$. These distributions were used owing to their properties and the adaptive scheme proposed. For example, the normal distribution is symmetric and has moderate tails and the double exponential distribution may be symmetric but heavy tailed.

The findings from the simulation studies revealed that the parametric F-test for oneway ANOVA model performed better than the non-parametric adaptive test proposed for symmetric skewed and moderate tailed distributions and then for symmetric skewed and light tailed distributions with ARE between 2% and 55%. The normal distribution and the truncated logistic distribution were identified to these distributional characterization. However, the adaptive test outperformed the F-test in symmetric and non-symmetric skewed with varying tail weights distribution with ARE between 5% and 70%. Simulations from the Laplace distribution, the contaminated normal distribution and mixture of distribution confirmed the superiority of the Adaptive test over the F-test. At small sample sizes, the Adaptive test exhibited its robustness in cases where there were outliers in the data, however, the F-test displayed some weakness in performance. This results is in agreement with Hill et al. (1988) when they used lung cancer data to demonstrate the dominance of their adaptive procedures over the parametric and rank based procedures when the size of each sample was at least 20. Also, O' Gorman (1997) in his evaluation of the power and significance level of the adaptive procedure, conducted Monte Carlo simulations to compare with procedures such as the

F-test, Kruskal-Wallis test and the normal scores. He concluded that all the tests maintained their level of significance for dataset with at least 24 observations, but the adaptive tests proved to be more powerful for distributions that were skewed when the total number of observations were at least 24. This study also revealed that, even for symmetric skewed and moderate tailed distribution and symmetric skewed and light tailed distribution, at very small sample sizes, the adaptive test performed appreciable well compared to the performance of the F-test in symmetric and non-symmetric skewed with varying tail weights distributions. Miao and Gastwirth (2009) reported results from simulation studies proved that their adaptive procedure maintained its nominal level of significance for all continuous distributions even for sample sizes as small as 10 and has almost the same power as the best signed rank test for a broad class of distribution functions. There are several advantages the researcher enjoys from using the Adaptive test (Hogg et al. (1975); Büning (1996)). The distributional characterization of the data at hand is adequately specified and the assurance of high breakdown point of the Winsorised mean is confirmed.

5.3 Conclusion

The findings of the study reveal the several advantages of the use of the adaptive test. The distributional characterization of the data at hand is known to the researcher. This information is very crucial in data analysis. The robustness of the adaptive test implies higher reliability of results from use. Hogg et al. (1975) through monte carlo simulation confirmed that the adaptive test performs powerfully over a broad class of distributions and is to be preferred over some popular non-adaptive tests including parametric ones. Although, the F-test displayed superiority in efficiency in symmetric skewed, medium and light tailed distributions, the adaptive test was more efficient in more broader class of continuous distribution. The performance of these test at small sample sizes was of much importance in this thesis because most sensitive areas of the application of oneway ANOVA models often has very low sample size usage. The adaptive test was more efficient at very small sample sizes compared to the F-test. It is important to also note that the F-test also performed appreciably well as the sample sizes increased. Based on the findings of this study, the adaptive test should be incorporated in statistical analysis of oneway ANOVA models. It should be performed alongside the parametric F-test and comparative efficiency will inform the sort of results presented for an analysis.

REFERENCES

Al-shomrani, A. A. (2003). *A Comparison of different schemes for selecting and estimating score functions based on residuals*. PhD thesis, Western Michigan University.

Andrews, D., Bickel, P., Hampel, F., Huber, P., Rogers, W., and Tukey, J. (1972). *Robust Estimates of Location: Survey and Advances*. Princeton University Press, Princeton, NJ.

Anscombe, F. J. (1967). Topics in the Investigation of Linear Relations Fitted by the Method of Least Squares. *Journal of the Royal Statistical Society*, 29.

Bandyopadhyay, U. and Dutta, D. (2007). Adaptive nonparametric tests for a single sample location problem. *Statistical Methodology*, 4(4):423–433.

Bernardo, J. M. (1996). The concept of exchangeability and its applications. *Far East Journal of Mathematical Sciences*, 4:111–122.

Botton, S. and Bon, C. (2009). *Pharmaceutical Statitistics:Practical and Clinical Application*. Marcel Dekker, Inc., 5th edition.

Büning, H. (1996). Adaptive Tests for the c-Sample Location Problem. *Communications in Statistics- Theory and Methods*, 25.

Casella, G. and Berger, R. L. (2002). *Statistical Inference*. Wadsworth Group, Thomson Learning Inc, second edition.

Freidlin, B., Miao, W., and Gastwirth, J. L. (2003). On the use of the shapiro-wilk test in two-stage adaptive inference for paired data from moderate to very heavy tailed distributions. *Biometrical Journal*, 45(7):887–900.

Good, P. I. (2002). Extensions of the concept of exchangeability and their applications. *Journal of Modern Applied Statistical Methods*, 1(2):243–247.

Hao, L. and Houser, D. (2011). Adaptive procedures for nonparametric ttest: Seven decades of advances. *Communications in Statistics- Theory and Methods.*

Hettmansperger, T. P. (1984). *Statistical Inference based on Ranks.* John Wiley.

Hettmansperger, T. P. and McKean, J. W. (2011). *Robust Nonparametric Statistical Methods.* Chapman and Hall/CRC Monographs on statistics and Applied Probability Series. CRC Press, second edition.

Hill, N. J., Padmanabhan, A. R., and Puri, M. L. (1988). Adaptive nonparametric procedures and applications. *Journal of the Royal Statistical Society*, 37(2):205–218.

Hogg, R. V. (1974). Adaptive Robust Procedures: A Partial Review and Some Suggestions for Future Applications and Theory. *Journal of the American Statistical Association*, 69(348).

Hogg, R. V., Craig, A. T., and McKean, J. W. (2005). *Introduction to Mathematical Statistics.* Pearson Prentice Hall, sixth edition.

Hogg, R. V., Fisher, Doris, M., and Randles, R. H. (1975). Two-sample Adaptive Distribution-Free Test. *Journal of the American Statistical Association*, 70(351).

Huber, P. J. (1972). Robust statistics: A review. *Annals of Mathematical Statistics*, 43:1041–1067.

Jaeckel, L. A. (1972). Estimating regression coefficients by minimizing the dispersion of the residuals. *Annals of Mathematical Statistics*, 43(5):1449–1458.

Kloke, J. D. and McKean, J. W. (2012). Rfit: Rank-based estimation for linear models,. *The R Journal*, 4(2):57–64.

Kloke, J. D. and McKean, J. W. (2013). R packages for rank based estimates. *R Journal.*

Kössler, W. (2010). Max-type rank tests, u- tests and adaptive tests for the two-sample location problem- an asymptotic power study. *Journal of Computational Statistics and Data Analysis*, 54(9):2053–2065.

Mahmoud, H. (2008). *Pólya Urn Models*. Chapman & Hall/ CRC Press Texts in Statistical Science.

Miao, W. and Gastwirth, J. L. (2009). A new two stage adaptive nonparametric test for paired differences. *Statistics and its inference*, 2:213–221.

Montgomery, D. C. (2001). *Design and Analysis of Experiments*. John Wiley and Sons, Inc., 5th edition.

O' Gorman, T. W. (1997). An adaptive test for the one-way layout. *The Canadian Journal of Statistics/ La Revue de Statistique*, 25(2):269–279.

O' Gorman, T. W. (2002). An adaptive test of significance for a subset of regression cefficients. *Statistics in Medicine*, 21(22):3527–3542.

O' Gorman, T. W. (2006). An adaptive test for a subset of coefficients in a multivariate regression model. *Biometrical Journal*, 48(5):849–859.

O'Gorman, W. T. (2004). *Applied Adaptive Statistical Methods*. ASA-SIAM Series.

Okyere, G. A. (2011). *Robust Adaptive Scheme for Linear Mixed Model*. PhD thesis, Western Michigan University.

Pearce, S. C. (1992). Data analysis in agricultural experimentation. i. contrasts of interest. *Experimental Agriculture*, 28:245–253.

Rao, C. R., Toutenburg, H., Shalabh, and Heumann, C. (2008). *Linear models and Generalizations: Least Squares and Alternatives*. Springer Series in Statistics. Springer-Verlag Berlin Heidelberg New York, third edition edition.

Rashid, M. M., McKean, J. W., and Kloke, J. D. (2012). R estimates and associated inferences for mixed models with covariates in a multi-center clinical trial. *Statistics in Biopharmaceutical Research*, 4(1):37–49.

Sheskin, D. J. (2011). *Handbook of Parametric Statistical Procedures*. Chapman Hall/CRC., 3rd edition.

Staudte, R. G. and Sheather, S. J. (1990). *Robust Estimation and Testing*. John Wiley & Sons, Inc.

Appendix A

ASYMPTOTIC RESULTS

A.1 Simple Linear Rank Statistics

Consider the sequence of random variables $Y_1, Y_2, \ldots, Y_n$ be iid with common density function $f(y)$ which follows the assumption that $f(t)$ is absolutely continuous and has finite information, i.e $0 \leq I(f) = \int_0^1 \varphi_f^2(u)du < \infty$. Let $c_1, c_2, \ldots, c_n$ denote a sequence of centred ($\bar{c} = 0$), regression coefficients and assume that;$n^{-1}\mathbf{C}^T\mathbf{C} \to \sum > 0$ as $n \to \infty$ and $\lim_{n\to\infty} \sup_{1\leq i<n} = 0$. Thus

$$\frac{\max c_i^2}{\sum_{i=1}^n c_i^2} \to 0 \tag{A.1.1}$$

and

$$\frac{1}{n}\sum_{i=1}^n c_i^2 \to \sigma_c^2,\ \sigma_c^2 > 0 \tag{A.1.2}$$

for some constant σ_c^2.

Assuming the score function $\varphi(u)$ is defined on the interval $(0, 1)$ satisfying;

$$\begin{cases} \varphi(u)\text{being non decreasing square integrable and bounded function} \\ \int_0^1 \varphi(u)du = 0\text{and} \int_0^1 \varphi^2(u)du = 1. \end{cases} \tag{A.1.3}$$

Then the linear rank statistics is defined by;

$$S = \sum_{i=1}^n c_i a(R(Y_i)) \tag{A.1.4}$$

where the scores are generated as $a(i) = \varphi(\frac{i}{n+1})$.

A.1.1 Asymptotic Distribution Theory

Following the assumptions of section A.1 then the mean and variance of the linear rank statistics S are given by $E(S) = 0$ and $Var(S) = \sum_{i=1}^{n} c_i^2 \{\frac{1}{n-1}\sum_{i=1}^{n} a^2(i)\} \doteq \sum_{i=1}^{n} c_i^2$ respectively. The approximation of the variance is due to the fact that the quantity in the braces is the Riemann sum of $\int_0^1 \varphi^2(u)du = 1$.

Now following (A.1.4), S can also be written as;

$$S = \sum_{i=1}^{n} c_i \varphi\left(\frac{n}{n+1}F_n(Y_i)\right) \tag{A.1.5}$$

where F_n is the empirical distribution of $Y_1, Y_2, \ldots, Y_n$. Hence the approximation of S denoted T is given by;

$$T = \sum_{i=1}^{n} c_i \varphi(F(Y_i)) \tag{A.1.6}$$

with the mean and variance of T following A.1.3 given as $E(T) = 0$ and $Var(T) = \sum_{i=1}^{n} c_i^2$ respectively.

Thus based on the central limit theorem and the law of large numbers $\frac{1}{\sqrt{n}}T$ is asymptotically distributed as $N(0, \sigma_c^2)$.

Since the means of S and T are the same, then it follows that S has the same asymptotic distribution as T provided the second moment of their difference is zero.

Proof A.1.1

$$\begin{aligned}
E\left[\left(\frac{1}{\sqrt{n}}S - \frac{1}{\sqrt{n}}T\right)^2\right] &= \frac{1}{n}E\left[\sum_{i=1}^{n} c_i \left(\varphi\left(\frac{n}{n+1}F_n(Y_i)\right) - \varphi\left(F(Y_i)\right)\right)^2\right] \\
&\leq \frac{n}{n-1}\left\{\frac{1}{n}\sum_{i=1}^{n} c_i^2\right\} E\left[\left(\varphi\left(\frac{n}{n+1}F_n(Y_i)\right) - \varphi(F(Y_i))\right)^2\right] \\
&\to \sigma_c^2 \cdot 0
\end{aligned}$$

This follows a derivation from pages 422 to 425 of (Hettmansperger and McKean, 2011).

Theorem A.1.1 *From the above stated assumptions,* $\frac{1}{\sqrt{n}}(T - S) \xrightarrow{p} 0$*, and* $\frac{1}{\sqrt{n}}S \xrightarrow{D} N(0, \sigma_c^2)$.

This establishes the asymptotic distribution of the simple linear rank statistics.

Appendix B

B.1 R-codes

B.1.1 Selector Statistics

```
selstat = function(xs){
  n = length(asp)
This calculates the benchmarks / cut-off proposed by Ali (2003)
  clq1 = 0.36 + (0.68/n)
  cuq1 = 2.73 - (3.72/n)
  if(n < 25){
    clq2 = 2.17 - (3.01/n)
    cuq2 = 2.63 - (3.94/n)
  } else {
    clq2 = 2.24 - (4.68/n)
    cuq2 = 2.95 - (9.37/n)
  }
  cus = c(clq1,cuq1,clq2,cuq2)
Datapoints are ordered.
  iord = order(asp)
  xs = asp[iord]
This calculated the attributes in the formular
for the measure of skewness and tailweight
  a=xs[(0.95*n):(n+1)]
  b=xs[(0.25*n):(0.75*n)]
  c=xs[1:(0.05*n)]
  d=xs[(0.5*n):(n+1)]
  e=xs[1:(0.5*n)]
```

```
  um1 = mean(a)
  lm1 = mean(c)
  mm1 = mean(b)
  um2 = mean(d)
  lm2 = mean(e)
  ulmeans = c(um1,lm1,mm1,um2,lm2)
The two selector statistics are calculated
  q1 = (um1 - mm1)/(mm1 - lm1)
  q2 = (um1 - lm1)/(um2 - lm2)
  qs = c(q1,q2)
Statement for selecting distributional characteristics
associated with (q1 and q2)
  if(q1 <= clq1){
    if(q2 <= clq2){score = 'LL'}
    if((q2 > clq2) && (q2 <= cuq2)){score = 'LM'}
    if(q2 > cuq2){score = 'LH'}
  } else if(q1 <= clq2){
    if(q2 <= clq2){score = 'SL'}
    if((q2 > clq2) && (q2 <= cuq2)){score ='SM'}
    if(q2 > cuq2){score = 'SH'}
  } else {
    if(q2 <= clq2){score = 'RL'}
    if((q2 > clq2) && (q2 <= cuq2)){score = 'RM'}
    if(q2 > cuq2){score = 'RH'}
  }
  score = score
  list(score=score, qs = qs,cus=cus,ulmeans=ulmeans)
}
```

B.1.2 Score Functions

```
afri4.phi = function(u,param){
  s1=param[1]
  s2=param[2]
  s3=param[3]
  k1=s2-s3
  k2=s1-1
  ifelse(u < s1,s2,s3+(k1/k2)*(u-1))
}
afri4.Dphi = function(u, param){
  s1=param[1]
  s2=param[2]
  s3=param[3]
  k1=s2-s3
  k2=s1-1
  ifelse(u < s2,0,k1/k2)
}
afriscores4<-new("scores",phi=afri4.phi,Dphi=afri4.Dphi,
param=c(0.3,-1,2))

afri3.phi = function(u,param){
  s1=param[1]
  s2=param[2]
  s3=param[3]
  k1=s2-s3
  k2=s1-1
  ifelse(u < s1,s2,s3+(k1/k2)*(u-1))
}
```

```
afri3.Dphi = function(u, param){
  s1=param[1]
  s2=param[2]
  s3=param[3]
  k1=s2-s3
  k2=s1-1
  ifelse(u < s2,0,k1/k2)
}
afriscores3<-new("scores",phi=afri3.phi,Dphi=afri3.Dphi,
param=c(0.1,-1,2))

afri1.phi = function(u,param){
  s1=param[1]
  s2=param[2]
  s3=param[3]
  k=s3-s2
  ifelse(u > s1,s3,s3+(k/s1)*(u-s1))
}
afri1.Dphi = function(u, param){
  s1=param[1]
  s2=param[2]
  s3=param[3]
  k=s3-s2
  ifelse(u > s1,0,k/s1)
}
afriscores1<-new("scores",phi=afri1.phi,Dphi=afri1.Dphi,
param=c(0.7,-2,1))

afri2.phi <- function (u, param)
```

```
{
  s1 = param[1]
  s2 = param[2]
  s3 = param[3]
  s4 = param[4]
  ifelse(u < s1,  -s3/s1 * (u + s1), ifelse(u > s2, -s4/(s2-1)
  * (u - 1) + s4, 0))
}
afri2.Dphi <- function (u, param)
{
  s1 = param[1]
  s2 = param[2]
  s3 = param[3]
  s4 = param[4]
  ifelse(u < s1, - s3/s1, ifelse(u > s2, -s4/(s2-1), 0))
}
afriscores2=new("scores",phi=afri2.phi,Dphi=afri2.Dphi,
param=c(0.9,-2,1,0))
```