WINSORIZED MODIFIED ONE STEP M-ESTIMATOR IN ALEXANDER-GOVERN TEST

TOBI KINGSLEY OCHUKO

MASTER OF SCIENCE (STATISTICS)
UNIVERSITI UTARA MALAYSIA
2016

WINSORIZED MODIFIED ONE STEP M-ESTIMATOR IN ALEXANDER-GOVERN TEST

A Thesis Submitted to College of Arts & Sciences in partial fulfilment of requirements for the degree Master of Science (Statistics) Universiti Utara Malaysia

By

Tobi Kingsley Ochuko

Awang Had Salleh
Graduate School
of Arts And Sciences

Universiti Utara Malaysia

PERAKUAN KERJA TESIS / DISERTASI
(Certification of thesis / dissertation)

Kami, yang bertandatangan, memperakukan bahawa
(We, the undersigned, certify that)

TOBI KINGSLEY OCHUKO (813614)

calon untuk Ijazah **MASTER**
(candidate for the degree of)

telah mengemukakan tesis / disertasi yang bertajuk:
(has presented his/her thesis / dissertation of the following title):

"WINSORIZED MODIFIED ONE STEP M-ESTIMATOR IN ALEXANDER-GOVERN TEST"

seperti yang tercatat di muka surat tajuk dan kulit tesis / disertasi.
(as it appears on the title page and front cover of the thesis / dissertation).

Bahawa tesis/disertasi tersebut boleh diterima dari segi bentuk serta kandungan dan meliputi bidang ilmu dengan memuaskan, sebagaimana yang ditunjukkan oleh calon dalam ujian lisan yang diadakan pada : ***3 Februari 2016.***
That the said thesis/dissertation is acceptable in form and content and displays a satisfactory knowledge of the field of study as demonstrated by the candidate through an oral examination held on: ***February 3, 2016.***

Pengerusi Viva: *(Chairman for VIVA)* **Assoc Prof Dr. Azizan Saaban** Tandatangan *(Signature)*

Pemeriksa Luar: *(External Examiner)* **Dr. Teh Sin Yin** Tandatangan *(Signature)*

Pemeriksa Dalam: *(Internal Examiner)* **Dr. Shamshuritawati Sharif** Tandatangan *(Signature)*

Nama Penyelia/Penyelia-penyelia: *(Name of Supervisor/Supervisors)* **Dr. Suhaida Abdullah** Tandatangan *(Signature)*

Nama Penyelia/Penyelia-penyelia: *(Name of Supervisor/Supervisors)* **Dr. Zakiyah Zain** Tandatangan *(Signature)*

Tarikh:
(Date) ***February 3, 2016***

Permission to Use

In presenting this thesis in fulfilment of the requirements for a postgraduate degree from Universiti Utara Malaysia, I agree that the Universiti Library may make it freely available for inspection. I further agree that permission for the copying of this thesis in any manner, in whole or in part, for scholarly purpose may be granted by my supervisor(s) or, in their absence, by the Dean of Awang Had Salleh Graduate School of Arts and Sciences. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to Universiti Utara Malaysia for any scholarly use which may be made of any material from my thesis.

Requests for permission to copy or to make other use of materials in this thesis, in whole or in part, should be addressed to:

Dean of Awang Had Salleh Graduate School of Arts and Sciences
UUM College of Arts and Sciences
Universiti Utara Malaysia
06010 UUM Sintok

Abstrak

Kajian ini tertumpu kepada ujian kumpulan bebas bagi membandingkan dua atau lebih min menggunakan kaedah berparameter iaitu ujian Alexander-Govern (*AG*). Ujian ini menggunakan min sebagai sukatan kecenderungan memusat dan dianggap sebagai alternatif yang lebih baik berbanding *ANOVA*, ujian Welch dan ujian James. Walaupun ujian *AG* mempunyai kawalan yang baik terhadap kadar ralat Jenis I dan menghasilkan kuasa yang tinggi pada varians heterogen, ujian ini tidak teguh pada data yang tidak normal. Justeru, min terpangkas telah dicadangkan dalam ujian tersebut untuk menangani masalah ketaknormalan dan kemudiannya, satu penganggar yang lebih teguh dikenali sebagai penganggar *M* satu langkah terubahsuai telah diperkenalkan. Penganggar berkenaan adalah tidak dipengaruhi oleh bilangan kumpulan, namun telah gagal untuk menghasilkan kawalan yang baik terhadap kawalan ralat Jenis I, dalam keadaan kepencongan dan kurtosis yang ekstrim. Kajian ini mencadangkan penganggar *MOM* ter*Winsor* (*WMOM*) sebagai sukatan kecenderungan memusat dalam usaha untuk meneguhkan ujian *AG*. Ujian *AG* yang ditambah baik ini, *AGWMOM* mampu menyingkirkan kewujudan data terpencil daripada taburan data. Satu kajian simulasi terhadap 5,000 set data telah dilaksanakan untuk membandingkan prestasi ujian: *AG*, *AGMOM* (ujian *AG* menggunakan penganggar *MOM*), *AGWMOM*, ujian-*t* dan *ANOVA*. Keputusan menunjukkan bahawa ujian *AGWMOM* telah meningkatkan bilangan kondisi teguh pada taburan terpencong dengan hujung normal dan taburan terpencong dengan hujung berat berbanding ujian yang lain.

Sebagai tambahan, ujian ini telah menghasilkan kuasa yang tinggi dalam kebanyakan kondisi pada empat kumpulan dengan saiz sampel tidak seimbang. Dapatan kajian mendorong untuk ujian ini menjadi paling sesuai apabila taburan data adalah berhujung berat.

Kata kunci: ujian Alexander-Govern, penganggar *MOM*, kadar ralat Jenis I, Kuasa ujian, ujian *AGWMOM*

Abstract

This research centres on independent group test of comparing two or more means by using the parametric method, namely the Alexander-Govern (*AG*) test. It uses mean as its central tendency measure and is considered as a better alternative to the *ANOVA*, the Welch test and the James test. Although the *AG* test has a good control of Type I error rate and produces a high power under variance heterogeneity, it is not robust to non-normal data. Thus, trimmed mean was proposed in the test to handle the problem of non-normality and later, a more robust estimator called modified one step *M* (*MOM*) estimator was introduced. These estimators are not influenced by the number of groups, but failed to give a good control of Type I error rate, under extreme conditions of skewness and kurtosis. This research proposes the Winsorized *MOM* (*WMOM*) estimator as a measure of central tendency in attempt to robustify the AG test. This enhanced *AG* test, *AGWMOM* is able to remove the appearance of outliers from the data distribution. A simulation study of 5,000 data sets was conducted to compare the performance of the tests: *AG*, *AGMOM* (*AG* test using *MOM* estimator), *AGWMOM*, *t*-test and *ANOVA*. The results show that the *AGWMOM* test has improved the number of robust conditions under skewed normal tailed and skewed heavy tailed distributions compared to the other tests. Additionally, the test produced high power in most conditions under four groups with unbalanced sample size. It leads that this test is convenient specifically when the data distribution is heavy tailed.

Keywords: Alexander-Govern test, *MOM* estimator, Type I error rate, power of test, *AGWMOM* test

Acknowledgement

This research is firstly acknowledged and dedicated to God Almighty, the creator of heaven and earth, the Alpha and the Omega, the Ancient of days, the Rock of ages, the King of Glory, the Great I am, that is I am, The Great Shepherd, The lover of my soul, The Man of War, the Lion of the tribe of Judah, the Counselor, the Author and Finisher of my Faith, the lily of the valley, the merciful father, the giver of life, the great Judge, the Everlasting Father, the King of kings, the LORD of Lords, My helper, the Author of wisdom, knowledge and understanding. I give God Almighty all the glory, honor, power, majesty, dominion, praises, thanksgiving, blessings, adoration and worship for everything.

Many thanks and appreciation goes to my wonderful, amiable and dynamic supervisors in person of Dr. Suhaida Binti Abdullah and Dr. Zakiyah Binti Zain for the strict and thorough supervision of the research. I am very grateful Ma's and I appreciate it. Many thanks and appreciation goes to my internal examiner in person of Dr. Shamshuritawati Sharif for her thorough and wonderful counsel in the course of carrying out this research. I appreciate it very much.

I also want to express my appreciation and gratitude to the Dean of the School of Quantitative Sciences, in person of Assc. Prof. Syed Yahaya, thank you very much Ma, for your wonderful counsel, support and encouragement, in the course of writing this Thesis. I am very grateful Ma and I appreciate it.

I also want to acknowledge and thank my wonderful, blessed, ever-caring, special, beloved and ever-dynamic parents, in person of Mr. and Mrs. D. K.O. Tobi. Dear parents, I am very grateful for everything you have done for me and I appreciate both of you very greatly. I also want to acknowledge my wonderful, blessed, unique and beloved brothers and sisters, in person of Dr. Bright, Dr. Racheal, Mrs. Deborah, Mr. Anthony, and Mrs. Flourence. I appreciate you all very greatly.

I also want to express my love, appreciation and gratitude to my wonderful, caring, blessed, dearly beloved and unique fiancée, in person of Sister Oluwabusayo. I love and appreciate you very greatly.

Lastly, I appreciate all the lecturers and staffs of the School of Quantitative Sciences.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AG	Alexander-Govern Test
MOM	Modified One Step *M*-estimator
AGMOM	Modified One Step *M*-estimator in the Alexander-Govern Test
WMOM	Winsorized Modified One Step *M*-estimator
AGWMOM	Winsorized Modified One Step *M*-estimator in the Alexander-Govern Test
ANOVA	Analysis of Variance

List of Appendices

CHAPTER ONE

INTRODUCTION

1.1 Background of the Study

This research makes comparison of the performances of the Type I error rate and power of five different tests. These tests are (i) Alexander-Govern test (*AG* test), (ii) Modified One Step *M*-estimator (*MOM*) estimator in the Alexander-Govern test (*AGMOM test*), (iii) Winsorized Modified One Step *M*-estimator (*WMOM*) estimator in the Alexander-Govern test (*AGWMOM* test), (iv) *t-test* (v) Analysis of Variances (*ANOVA*). Each test is performed under two, four and six groups conditions, with the combination of both balanced and unbalanced sample sizes, equal and unequal variances respectively, with each of the *g- and h-* distributions. The *g-* and *h-* distribution is used to determine the level of skewness and kurtosis respectively in a data distribution.

The best among the five tests will produce the best control of Type I error rate and also produce high power, under skewed heavy tailed distribution. The independent group tests such as the *ANOVA* have been applied in different field of life, for example in medicine, economics, sociology and agriculture, as discussed by Pardo, Pardo, Vincente and Esteban (1997). Three main assumptions have to be fulfilled before the *ANOVA* can work effectively, namely: (i) homogeneity of the variance (ii) normality of the data and (iii) independent observations of the data distribution.

The *ANOVA* is a classical method of analysis for comparing the differences between three or more means. It is used for testing the equality of the measure of the central tendency of a distribution, and is robust to small deviations from a normal distribution, mainly when the sample size is large enough to guarantee normality, as mentioned by Wilcox (1997, 2003).

Yusof, Abdullah, Yahaya and Othman (2011) in their research discovered that the two major problems affecting the *ANOVA* is the appearance of non-normality and heterogeneity of the variance in a data distribution. Due to this, the Type I error rate is increased and there is a reduction in the power of the test. When the distribution of the data is heavy tailed, the standard error of the mean can be greatly increased (Wilcox & Keselman, 2002). Due to this, the standard error of the *ANOVA* becomes larger than it ought to be and the power of the test is reduced. In order to obtain a good test, the Type I error rate should be well controlled and the power of the test must not be reduced. This implies that neither should Type I error rate be increased nor should there be a decrease in the power of the test.

The *ANOVA* is very sensitive to the homogeneity of variance assumption and when there is violation, the outcome of the analysis could be unreliable; whereby the *p*-value may become too conservative or large. Therefore, it is very crucial to test for the homogeneity of the variance and to check for the equality of the variance assumptions by using the correct test, so as to increase the genuineness of the results (Brown & Forsythe, 1974; Wilcox, Charlin, & Thompson, 1986).

The problem of heterogeneity of variance has been discussed by few researchers and some alternatives have been proposed. Welch (1951) introduced the Welch test that is used for testing the hypothesis of two populations with equal means. It is mentioned in different literatures as a better alternative to the *ANOVA* (Keselman, 1982; Wilcox *et al.*, 1986; Algina, Oshima & Lin, 1994; Lix, Keselman, & Keselman, 1996).

The Welch test gives a good control of Type I error rate when the variances are not equal. It becomes a common alternative to parametric procedure that deals with heteroscedasticity. However, for a small sample size, the Welch test fails to give a good control of Type I error rate, as the group size increases (Wilcox, 1988). James (1951) introduced the James test as a better solution for *ANOVA* under heterogeneity of variance. The James test is used for weighing means sample and is discussed in different literatures as a better alternative to using the *ANOVA* (Lix *et al.*, 1996; Oshima & Algina, 1992; Wilcox, 1988).

When the sample size is small, and the data distribution is non-normal, the James test fails to give a good control of Type I error rate. Both the Welch test and the James test are used for analyzing a data distribution that is non-normal with unequal variance (Brunner, Dette, & Munk, 1997; Kohr & Games, 1974; Krishnamoorthy, Lu, & Matthew, 2007; Wilcox & Keselman, 2003).

The Alexander-Govern test was introduced in 1994 to deal with heterogeneity of variance under the condition of normality, but is a test that is not robust to non-normal data. Schneider and Penfield (1997) and Myers (1998) accepted that performance of the Alexander-Govern test is better compared to the James test and

Welch test respectively. Myers (1998), suggested that the Alexander-Govern test provides a good solution to the problem of variance heterogeneity. The *AG* test can excellently put under control the Type I error rate when there is heterogeneity in the variances, under a normal data distribution.

It is a well-known fact that the common mean is a very good estimator under normal distribution but it is extremely sensitive to the presence of outliers. The Alexander-Govern test was originally developed using the common mean as its central tendency measure, hence, directly affecting its performance when dealing with non-normal data. As a result, it fails to provide a remarkable control over the probability of Type I error rate for a non-normal data and the power of the test is reduced.

Lix and Keselman (1998) introduced a better alternative to the common mean with the use trimmed mean in a few robust test statistics that improved the performance of the tests for a non-normal data. The use of trimmed mean and Winsorized variance are better alternatives to the common mean and variance respectively. This is attributed to some good properties, such as having a remarkable control of Type I error rate and the power of the test is increased, when there is a violation under the assumptions of homogeneity of the variance and when the distribution of the data is normal (Wilcox, 1995).

Trimmed mean is obtained by taking the average of the middle data only after removing a certain percentage of the largest and the smallest data value, while its variance is estimated by using the Winsorized variance. Trimming is the process of removing a fixed amount of extreme value in percentage, from both tails of a distribution during the process of analyzing data (Abdullah, Othman, Yahaya & & Yusof, (2011). Suppose in an experiment consisting of two groups, an individual may choose not to consider the two largest scores and the two smallest scores from each of the groups, such that the outliers present in either group would be removed. In using trimmed mean as a robust measure and scale in a data distribution, some limitations exist when the data are trimmed symmetrically without considering the nature of the distribution.

In general, the amount of trimming is performed regardless of the distribution of the data. There will be a great mistake in removing a data distribution where outliers are not located, mainly in a normal data distribution, because in doing so it will lead to loss of information. Meanwhile, in cases of skewed data distribution, the trimming process performed on the data must not be equal at the right and left tail of the data distribution. Another weakness in using trimmed mean is that it cannot give a good control of Type I error rate when the number of groups is more than two; i.e for four groups and above, it could no longer control the error rate in the test, especially when applied in Alexander-Govern test as its central tendency measure (Lix & Keselman, 1995).

One of the suggested estimators as a better alternative to the trimmed mean is known as the *MOM*, which is able to detect the appearance of outliers in a data distribution (Yusof, Abdullah, Yahaya & Othman, 2011). The *MOM* estimator empirically trims only the extreme data sets (Othman, Keselman, Padmanabhan, Wilcox, & Fradette, 2004). However, the main disadvantage of using the *MOM* estimator as a measure of the central tendency, in Alexander-Govern test, is that it cannot control the error rate in the test under extreme condition of skewness and kurtosis.

1.2 Problem Statement

In testing the equality of means between independent groups, two major issues need to be satisfied which are normal distribution and equal variances. The problems with unequal variances were discovered in statistical literature by Behrens (1929) and Fisher (1935). A few studies offered some alternative approaches in handling these problems. The most common approaches that made used of robust statistical tests are those of the Welch (1951) and the James (1951).

The Alexander-Govern (1994) test is a better alternative to Welch test, James test and the *ANOVA*. This is due to its simplicity in calculation (Schneider & Penfield, 1997). According to Schneider and Penfield (1997), Lix and Keselman, (1998) and Myers (1998), the Alexander-Govern test gives a good control of Type I error rate for a normal data under variance heterogeneity, but this test is not robust for a non-normal data. This results showed that the Type I error rate became out-of-control when data distribution was not normal.

The main reason why it cannot work correctly under non-normal data is because it uses common mean as its central tendency measure. The common mean is affected by the appearance of outliers when there is a deviation from normality. Lix and Keselman (1998) introduced the trimmed mean as a better alternative to the mean for a non-normal data. The trimmed mean has been used by different researchers in the past to give a good control of Type I error rate for a non-normal data (Keselman, Wilcox, Taylor, & Kowalchuk, 2000; Luh & Guo, 2005; Luh, 1999).

In applying trimmed mean in a data distribution, it possesses some disadvantages. Firstly, the percentage of trimming is placed at prior, resulting in the elimination process. Secondly, in trimming process, it should be done carefully, to minimize loss of information. Thirdly, it cannot handle large size of extreme value (Yahaya, Othman, & Keselman, 2006).

According to Abdullah, Yahaya and Othman (2007), an alternative to the use of trimmed means in Alexander-Govern test is a highly robust estimator, referred to as the *MOM* estimator. It was observed that when the distribution of the data is skewed, the *MOM* estimator kept under control the Type I error rate. The *MOM* estimator empirically trims extreme data set depending on the kind of the data set, be it a normal or skewed data distribution. When it was applied in Alexander-Govern test, it gave a remarkable result in putting under control Type I error rate, for a normal or highly skewed data distribution, but it failed to produce a remarkable control over the probability of Type I error rate under extreme condition of skewness and kurtosis (Othman *et al.*, 2004).

In a condition where the degree of skewness and kurtosis is exceptionally high, another preferred option is Winsorized mean, as introduced by Hasings, Monsteller, Tukey and Winsor (1947). Unlike the trimmed mean where data are trimmed from both tails of the distribution, the Winsorization process does not affect the sample size (Dixon & Tukey, 1968; Tukey & McLaughlin, 1963).

Ochuko, Abdullah, Zain and Yahaya (2015) described the Winsorization process as making a replacement or an exchange for the outlier detected value with a preceeding value closest to it. Winsorization has several advantages more than using the trimming procedure in a data distribution. Firstly in Winsorization, it makes a replacement or an exchange for an outlier detected value with the closest value to the outlier. Secondly, the sample size of the data remains unaltered. Thirdly, Winsorization helps to prevent loss of information. Fourthly, Winsorization helps to make the sample sizes of the data to be the same unlike using the trimming procedures.

1.3 Objective of the Research

The objective of this research is to produce a good statistical test in comparing the mean of independent groups when the assumptions of variance homogeneity and normality are violated.

The specific objectives are:

i. To modify the *AG* test using the new estimator namely the Winsorized *MOM* estimator.

ii. To evaluate the robustness of the modified *AG* test in terms of Type I error rate and power.

iii. To compare the performance of the *AG* test, *AGMOM* test, *AGWMOM* test, *t*-test and the *ANOVA*.

iv. To evaluate the reliability and efficiency of the test using real data.

1.4 The Scope of the Research

This research deals with the modification of the Alexander-Govern test, by using the Winsorized *MOM* estimator as its central tendency measure, under variance heterogeneity, to produce a good control of Type I error rate and high power under extreme conditions of skewness and kurtosis.

1.5 Significance of the Research

This research will appraise existing tests used when the assumptions of variance homogeneity and normality are violated.

To students: This research will expose and educate them on the importance of applied statistics in our world today, in overcoming insurmountable problems of the past with the aid of using simulated programmes, such as the Statistical Analysis Software (*SAS*) software programming package, for the analysis of simulated data, to making life easier and convenient in addressing problems, relating to comparing the scores among independent groups, with the goal of giving accurate results for the analysis of the independent groups.

To future researchers:

This research will be of great benefit to researchers by providing new findings in solving problems relating to comparing the scores among independent group test, on how the Winsorized *MOM* estimator was applied in the Alexander-Govern test, to overcome its weakness under non-normality in the presence of variance heterogeneity and as result, giving remarkable control of Type I error rate and to produce high power for the test, under skewed heavy tailed distribution.

1.6 Organization of the remaining Chapters

In Chapter One, the background of the study was elaborated, focusing on the independent group tests, such as the *ANOVA*, its application in different fields of life. The assumptions that must be fulfilled before the *ANOVA* can perform effectively have been described. The two main factors affecting the *ANOVA* in the control of Type I error rate and increase power were highlighted.

Other better alternatives to the *ANOVA* were mentioned, such as the Welch test and the James test. The constraints that affects the Welch test and the James test in the control of Type I error rate and high power was indicated in this chapter. The Alexander-Govern test is a better alternative to the Welch test, the James test and the *ANOVA* was mentioned and the reasons were listed.

The Alexander-Govern test is not robust to non-normal data under variance heterogeneity. As a result, the test fails to give good control for Type I error rate and high power under this condition. The problem statement in this research was identified. The objective of the research was listed. The scope of the research was elaborated. The significance of this research, namely to students and future research was explained in this chapter.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

The independent group test is all about making comparison of the equality of independent groups, either with the use of parametric or non-parametric method. In using parametric method, the *ANOVA* is seriously affected by the presence of heterogeneity of the variance and non-normality in a data distribution. The performance worsens when there are differences in the group sizes. The ability of this classical group test in controlling the Type I error rate reduces, thereby leading to an increase in the rejection of the null hypothesis testing for equal means in the distribution, especially under small sample sizes (Kulinskaya, Staudte, & Gao, 2003).

In a situation with unequal population variances, it will adversely affect the outcome and authenticity of the *ANOVA* mainly when the sample group sizes are not equal (Glass & Sanders, 1972; Harwell, Rubinstein, Hayes, & Olds, 1992; Kohr & Games, 1974; Scheffe, 1959). Ironically, in real life data, the heterogeneity of variance is a very common situation; for instance in behavioral sciences, it is a common practice for researchers to work with unequal variance in a data distribution (Erceg-Hurn & Mirosevich, 2008; Golinski & Cribbie, 2009; Grissom, 2000; Keselman, Kowalchuk, Algina, Lix & Wilcox, 2003).

To give solution to the presence of non-normal with heterogeneity of the variance, reliable alternative techniques such as the James (1951) and the Welch (1951) have been provided. The Alexander-Govern test is a better alternative to the Welch test, the James test and the *ANOVA*, because of its' simplicity in calculation and giving excellent control of Type I error rates for a normal data, but the test is not robust to non-normal data.

Wilcox (2003) stated that for every procedure which is based on mean, it will give poor performance when the normality distribution is deviated. From previous researches, it is observed that different approaches have been suggested in analyzing data distributions that are non-normal with heterogeneity of variances (Brunner, Dette, & Munk, 1997; Wilcox & Keselman, 2003; Cribbie, Wilcox, Bewell, & Keselman, 2007).

2.2 Robust Statistics

Robust statistics majorly deals with the spotting out of outliers in a given data distribution and reducing the appearance of the outliers as much as possible in the data distribution, so that the good observations are far more than the outliers located in the given data set. Robust statistics mainly use parametric models that permit deviations from models assumptions (Huber, 1981; Barnett & Lewis, 1994). Outliers could also be defined as observations or subsets of observations seen in a data distribution that are not consistent with the other data sets in the given data distribution (Barnett & Lewis, 1994).

According to Lix and Keselman (1998) the empirical rate of Type I error for stringent criteria of robustness must fall within the interval of $0.042 \leq \alpha \leq 0.058$ to judge the robustness of a given test at α level of significance. As stated by Bradley's (1978) the lenient criteria of robustness of a given test must fall within the interval of ($0.025-0.075$), in order to judge the robustness of a given test at α level of significance. In this research both lenient and stringent criteria of robustness was used to judge the robustness of the tests.

2.3 Dealing with Non-Normal Data

A non-normal data is a condition whereby a data is not normally distributed. Investigation under empirical test reveals that the Alexander-Govern test performed remarkably well compared to the *ANOVA* in controlling Type I error rate and power in the condition of variance heterogeneity and normality (Alexander & Govern, 1994). Additionally, Schneider and Penfield (1997) reported that the Alexander-Govern test is a good alternative to the *ANOVA* for variance heterogeneity compared to the Welch test and the James test due to its simplicity in calculation and having a good control for Type I error rate.

It also produces a high power under most experimental situations, referring to different levels of examination, when the test was applied in a data distribution, in order to evaluate its effectiveness in a data distribution. However, under the condition of heterogeneity of variances, it was recommended for only normal data but not robust to non-normal data, as discussed by Myer (1998).

With the use of non-normal data, transformation might be a favorable technique. Transformation is a special approach for transforming a data set that is non-normal in form and also having the appearance of variance heterogeneity in the data distribution. By so doing, the present scores in the distribution becomes normal and having equal variance. Despite the fact that it has the ability of transforming skewed data, it possesses some disadvantages in its usage.

Wilcox (2002) noted that using transformation on the square root of the mean and likewise with the log of the mean eliminates the influences on a real data set. Transformation also cannot remove the impact of outliers in a data distribution. In a situation where the extent of transformation is complex in a given distribution, it suffers the constraint of normalizing the data that is skewed. Other approach that is usually chosen by statistician practitioners when there is non-normal data distribution is using non-parametric methods.

According to Marascuilo and McSweeney (1977), a non-parametric test makes no exact assumption in relation to one or more of the population parameters that define the given distribution, for which the test is to be used. It is used to eliminate a nominal and ranked order data and can be described as an assumption free test or otherwise referred to as a distribution free test. However, non-parametric tests are not as sensitive as parametric tests when the basic assumptions of the parametric tests are fulfilled.

Hence, larger differences are required before a rejection of the null hypothesis is performed. In other situations, non-parametric approaches also need a large number of sample sizes to prevent the loss of information. Examples of non-parametric test are Friedman test, Mann-Whitney U test, Wilcoxon Signed-Ranked test, Fisher Exact Probability test, Kruskal-Wallis test, Cochran Q test, McNemar test and the Chi-square test as mentioned by Daniel (1990).

In considering the weaknesses observed in using the non-parametric tests, researchers have discovered the use of robust estimators as a better alternative when dealing with non-normal data. Robust estimator that is commonly chosen in improving the independent group test is the trimmed mean. This estimator has been successfully used to improve the Alexander-Govern test under non-normal distribution (Guo & Luh, 2000; Lix & Keselman, 1995; Luh, 1999). Although trimmed mean possesses a remarkable control over the probability of Type I error rate, the trimming process is performed irrespective of the nature of the distribution.

Whether outliers are present or not in a data distribution, the percentage of trimming is set at prior, thereby resulting in the elimination process done without regarding the shape of the data distribution. Therefore, it might lead to further loss of information. An alternative to the use of trimmed means is a highly robust estimator, which is referred to as the modified one-step *M*-estimator (Wilcox & Keselman, 2003).

Othman *et al.* (2004) stated that the *MOM* estimator empirically trims extreme data set only by depending solely on the nature of the data set. In a situation of skewed data, the amount of trimming should not be the same at both tails of the distribution. For example, if the data is skewed to the right, more data on the right should be trimmed from the distribution.

In using any estimator that is based on trimming, one major need to be placed under consideration, is the process of trimming itself. As mentioned previously, the trimmed mean trims data symmetrically without any consideration on the nature of the distribution. Meanwhile, the *MOM* estimator only trims data that is suspected as outliers. If both tails of a data set are detected as outliers, then the data distribution would be trimmed symmetrically, otherwise if it is one side of the data set that is detected as outliers, it would be trimmed asymmetrically, meaning that only one tail of the data set would be trimmed.

2.4 *MOM* Estimator

Scholars such as Wilcox and Keselman (2003a, 2003b) introduced the modified one step *M*-estimator to correct the problem associated with trimmed means, where the proportion of outliers is more than the percentage on trimming to be applied on the data set that is associated with the power of the test.

More trimming or other measure of location that is in a very small extent not affected by a large number of outliers is required. Also, when a data distribution is highly skewed to the right, it is very reasonable to trim more of the data set from the right tail than the left tail of the distribution. Wilcox and Keselman (2003a, 2003b) modified the one-step-*M* estimator, which is defined using the formula below:

$$\hat{\theta} = \frac{1.28 MADN_j (i_2 - i_1) + \sum_{i=i_1+1}^{n_j - i_2} Y(i) j}{n_j - i_1 - i_2} \quad (2.1)$$

where

n_j = the sample sizes of the data distribution,

i_1 = the number of X_{ij} observations when $\frac{|X_i - M|}{MAD_n} < -K$,

$i_2 =$ is the number of X_{ij} observations when $\frac{|X_i - M|}{MAD_n} > K$,

In eliminating $1.28MADN_j\,(i_2 - i_1)$, where $MADN_j = MAD_j / 0.6745$, *the* MAD_j is the median of the values of: $\left|Y_{ij} - \hat{M}_j\right|, \ldots, \left|Y_{nj}\, j - \hat{M}_j\right|$. Note that $\hat{M}_j$ is defined as the median of the j^{th} group, i_1 is the number of observations where $Y_{ij} - \hat{M}_j > 2.24MADN_j$.

Therefore, the modified *M*-estimator proposed by Wilcox and Keselman (2003) is defined as:

$$\hat{\theta}_j = \sum_{i=i_1+1}^{n_j - i_2} \frac{Y(i)j}{n_j - i_1 - i_2}. \tag{2.2}$$

A one-step *M* estimator is defined as:

$$\hat{\theta}_{OS} = \frac{1.28(MAD_n)(i_2 - i_1) + \sum_{i-i_1+1}^{n-i_2} X_{(i)}}{n - i_1 - i_2} \quad , \tag{2.3}$$

given $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$, as the observations expressed from the least value to the largest value. The expression $1.28\,(MAD_n)\,(i_2 - i_1)$ *in* $\hat{\theta}_{OS}$, arises for technical reasons; overlooking it, results to the popularly known estimator, otherwise called the modified one step *M*-estimator (*MOM*).

Outliers in a data distribution can be detected by using the formula below:

$$\frac{|X_j - M|}{MAD_n} > K \quad , \tag{2.4}$$

or when $\frac{|X_j - M|}{MAD_n} < -K$, (2.5)

where X_j represents the ordered random sample observations, M is the median of the ordered random samples and MAD_n is the median absolute deviation about the median. The value of K is 2.24. This value was proposed by Wilcox and Keselman (2003) for detecting the presence of outliers in a data set, because it has very small standard error, when sampling from a normal distribution.

The *MOM* estimator is also defined as:

$$MOM = \frac{1}{n_j - i_1 - i_2} \sum_{i=i_1+1} X_{(i)} \quad (2.6)$$

According to Abdullah, Yahaya and Othman (2007), the formula for the *MOM* estimator is defined in Equation (2.6),

where:

$X_{(i)}$ = the i^{th} sample ordered observations for the *j* group,

and $K = 2.24$

Substituting K equals to 2.24, to obtain a small reasonable standard error, for a normal data distribution, either i_1 or i_2 is used, which is defined as follows:

i_1 is the number of observations, when $\frac{|X_i - M|}{MAD_n} < -2.24$ (2.7)

i_2 is the number of observations,

when $\frac{|X_i - M|}{MAD_n} > 2.24$. (2.8)

The *MOM* estimator is obtained by using either equation (2.7) or equation (2.8) as expressed above, which is used to detect the presence of outliers from the ordered data set. In using equation (2.8), if the absolute value of the observed ordered data set subtracted from the median, divided by the value of the MAD_n is greater than *K*, then that observed value is considered to be an outlier.

The estimate of location for the *MOM* estimator is defined as the average of the values remaining after all the outliers, if there are any present in the data set, are removed. The value of 2.24 is motivated in part with the aim of getting a reasonable small standard error when taking samples from a normal distribution. The one step *M*–estimator is more satisfactory in obtaining a relatively small standard error, but the *MOM* estimator has some advantages over the one-step *M*-estimator, as follows.

The *MOM* estimator is very flexible in relation to the number of observations that should be removed as outliers from the distribution.

1. The *MOM* estimator can deal relatively large number of outliers.
2 It results in using the common mean, when no outliers are found in a data distribution.
3. It permits different amount of trimming from the left tail to the right tail of the distribution.

2.5 Trimming and Winsorization Methods

Trimming process is a technique used in removing a certain amount of data either by setting the percentage of trimming at prior or identify the amount of trimming using any procedure of outlier detection (Yusof, Abdullah & Yahaya, 2011). As an example, in considering two groups when performing an experiment, a person may decide to neglect the two highest and the two lowest scores from each group in the data set, with the aim of removing outliers present in either group. As mentioned earlier, trimming process needs to be done carefully because every single data value brings valuable information. The obvious disadvantage of trimming process is that it reduces the number of data to be considered. Therefore, if the sample size is small, trimming will make the sample size even smaller.

Winsorization is another approach in dealing with the influence of outliers. The term Winsorization was discovered by Hasings, Monsteler, Tukey and Winsor (1947) as a change in a data distribution by restricting extreme values with the aim of reducing the appearance of outliers from the distribution. In Winsorization process, the value of the outlier detected are replaced or exchanged with a preceding value, closest to it. As a result, the sample size of the data set is not affected.

To illustrate the Winsorization process, consider the following data set given as: 1, 2, 3, 4, 5, 6, 7 and 8 (Tukey & McLaughlin, 1963; Staudte & Sheather, 2011). The total sample size of this data set is eight. The mean of the distribution is 4.5. Therefore, 20% trimmed mean of the sample size of the data set is 1.6 and the approximated value is 2.

This indicates that two data would be discarded or trimmed from the left tail and the right tail of the data distribution. Hence, the distribution becomes; 3, 4, 5 and 6 respectively. As a result of this, the sample size is affected. It is reduced to four instead of eight (50% reduction) with the use of trimmed mean on the distribution.

For the Winsorization process, the two smallest and greatest values are replaced or exchanged with a preceding value closest to the outlier detected values. Hence, the Winsorized distribution becomes: 3, 3, 3, 4, 5, 6, 6, 6. Winsorization is a process that involves making a replacement or an exchange for the outliers values detected. Hence, the sample sizes in the data distribution remains the same. It helps to prevent loss of information and hence, the data is preserved, unlike the trimmed mean procedure.

2.6 The Alexander-Govern Test and its Test Statistic

The Alexander-Govern test is proposed by Alexander-Govern (1994). This test uses mean as its central tendency measure. It gives a remarkable control of Type I error rate and high power for a normal data, under variance heterogeneity. But this test is not robust for a non-normal data and fails to give a good control of Type I error rate under this condition. The test is used for comparing two or more groups. The test statistic for the Alexander-Govern (AG) test is obtained by using the following procedures.

The procedure in obtaining the test statistic for the Alexander-Govern test starts by first ordering the data set, having population $j\,(j=1,\ldots,J)$. For each of the data set, the mean is calculated using:

$$\bar{X}_j = \frac{\sum_j X_{ij}}{n_j} \qquad (2.9)$$

where X_{ij} represent the observed ordered random samples and n_j denote the sample size of the j observations. The mean is used as a measure of the central tendency in the Alexander-Govern (1994) procedure. After obtaining the mean, the estimate of the usual unbiased variance is calculated using:

$$s_j^2 = \frac{\sum (X_{ij} - \bar{X}_j)^2}{(n_j - 1)} \qquad (2.10)$$

where, $\bar{X}_j$ is used for estimating μ_j for the population j . The standard error of the mean is calculated for each group using:

$$S_{ej} = \sqrt{\frac{s_j^2}{n_j}} \qquad (2.11)$$

The weight (w_j) for the group sizes with j population of the ordered sample data is defined such that summation of the weight $(\sum w_j)$ should be equal to 1. So, the weight for each of the group is calculated using the formula below:

$$w_j = \frac{1/S^2_{ej}}{\sum_j 1/S^2_{ej}} \qquad (2.12)$$

The null hypotheses testing for the Alexander-Govern (1994) method, for the equality of the mean and under variance heterogeneity are expressed as:

$H_0 : \mu_1 = \ldots = \mu_j$

$H_A : \mu_i \neq \mu_j$, for at least $i \neq j$

The alternative hypothesis, (H_A) contradicts the claim or statement made by the null hypothesis. The grand mean for all the groups is calculated using:

$$\hat{\mu}=\sum_{j=1}^{J} w_j \bar{X}_j \qquad (2.13)$$

where, $\bar{x}_j$ and w_j are defined in Equations (2.9) and (2.12) respectively.

The t statistic for each group is calculated by using:

$$t_j = \frac{\bar{X}_j - \hat{\mu}}{S_{ej}} \qquad (2.14)$$

where, $\bar{x}_j, \hat{\mu}$ and S_{ej} are given in Equations (2.9), (2.13) and (2.11) respectively.

The t statistic is distributed as a t variable, having $(n_j - 1)$ degrees of freedom, for each of the independent groups in the order data set. The t statistic obtained for each of the group is converted to a standard normal deviates (z_j) with the use of Hill's (1970) normalization approximation in the Alexander-Govern (1994) approach. The formula is defined below:

$$Z_j = c + \frac{[c^3 + 3c]}{b} - \frac{[4c^7 + 33c^5 + 240c^3 + 855c]}{[10b^2 + 8bc^4 + 1000b]} \qquad (2.15)$$

where $c = [a \times \log_e (1 + \frac{t_j^2}{v_j})]^{1/2}$ (2.16)

$v_j = n_j - 1, a = v_j - 0.5$ and $b = 48a^2$ (2.17)

The test statistic for the Alexander-Govern (AG) approach is defined below:

$$A = \sum_{j=1}^{J} Z_j^2 \qquad (2.18)$$

A chi-square distribution table is used to obtain the p-value for the Alexander-Govern test at $\sigma = 0.05$ level of significance. If the p-value is less than 0.05, it is concluded that the test is significant otherwise, it is not.

2.7 Summary

In comparing independent groups, the classical group test such as the *ANOVA* is seriously affected by the appearance of heterogeneity of the variance and non-normality in a data distribution. Reliable parametric alternatives such as the James test, the Welch test, and the Alexander-Govern test have been proposed to solve the problem of variance heterogeneity. The Alexander-Govern test is considered as a better alternative to the Welch test and the James test because it is easy to compute, it produces a high level of power and possesses a remarkable control over the probability of Type I error rate. However, it has a weakness being that, it is not robust to non-normality under variance heterogeneity.

When trimmed mean was applied in the Alexander-Govern test, it was only robust for two group cases, but when there was an increment in the group sizes above two, the test was no longer robust and hence, could not give a good control of Type I error rate. A highly robust estimator known as the *MOM* estimator was applied on the test, as a substitute for its measure of central tendency. This estimator is not affected by the number of groups. It gave an excellent control of Type I error rate under a skewed distribution. But it failed to give a good control of Type I error rate, under skewed heavy tailed distribution.

Therefore, in this research, the Winsorized *MOM* estimator was applied in Alexander-Govern test to overcome its weakness under non-normality in the presence of variance heterogeneity, in an extreme condition of skewness and kurtosis and it gave the test an excellent control of Type I error rate and high power.

CHAPTER THREE

RESEARCH METHODOLOGY

3.1 Introduction

This chapter explains how the Winsorized modified one step *M*-estimator (*WMOM*) was used as a replacement for the central tendency measure in Alexander-Govern test, to overcome the weakness of the test for non-normality under variance heterogeneity, in an extreme condition of skewness and kurtosis. In this chapter, the test statistic for the Winsorized modified one step *M*-estimator is explained in detailed.

There are five different variables that were used in this research, which are: balanced (equal) and unbalanced (unequal) sample size, variance ratio, group sizes i.e $J = 2$, 4 and 6, types of distribution and nature of pairing. The research design shows the combination and pairing of both balanced (equal) and unbalanced (unequal) sample sizes with both equal and unequal variance, for both positive and negative pairing condition with each of the *g*- and *h*- distribution for two, four and six groups conditions.

In the research design, 84 conditions of pairing were used for the five different tests, which are: the Alexander-Govern (*AG*) test, the modified one step *M*-estimator in the Alexander-Govern (*AGMOM*), the Winsorized modified one step *M*-estimator in the Alexander-Govern (*AGWMOM*) test, the *t*-test and the *ANOVA* respectively. Lastly, the statistical power of a test is defined as the probability of not accepting the null hypothesis when it is false. The power of a test is affected by three main factors, which are: sample size, level of significance and the effect size. The Effect Size Index is divided into three types, namely: small, middle and large. In this research the Large Effect Size Index was used to produce high power for each of the tests.

3.2 The Modified Alexander-Govern Test

The *WMOM* estimator is applied on the data distribution where the outlier detected value is replaced or exchanged with a preceding value closest to the position the outlier is located. The *WMOM* estimator is obtained by averaging the Winsorized data distribution. It is expressed by using formula (3.1):

$$WMOM = \bar{X}_{WMOMj} = \frac{\sum_{j=1}^{J} X_{WMOMj}}{n} \tag{3.1}$$

The *WMOM* estimator is used as a replacement for the common mean as the central tendency measure in the Alexander-Govern test, for the following reasons:

i. to eliminate the appearance of outliers from the data distribution.

ii. to make the Alexander-Govern test to be robust to non-normality.

The Winsorized sample variance is defined as:

$$S^2_{WMOMj} = \frac{\sum_{j=1}^{J} (X_j - \bar{X}_{WMOMj})^2}{n-1}, \tag{3.2}$$

where, X_j is the random ordered observed sample and $\bar{X}_{WMOMj}$ is the Winsorized *MOM* estimator for the Winsorized data distribution. The standard error of *WMOM* is obtained by using the bootstrapping method. The bootstrapping algorithm for estimating the standard errors is defined below.

Firstly, we chose B independent bootstrap samples defined as: $x^{*1}, x^{*2}, ..., x^{*B}$, where each of these random samples comprises of n data values selected with replacement from x defined below:

$$x^* = (x_1, x_2, ..., x_n) \tag{3.3a}$$

$$F \rightarrow (x_1^*, x_2^*, ..., x_n^*) \tag{3.3b}$$

The indication of the symbol (*) shows that x^* is not the real data set of x but it refers to a randomized or resampled version of x. Where s is used for estimating $t(\hat{F})$ and $\hat{F}$ is the empirical distribution for the probability of $\frac{1}{n}$ on each of the observed values of $x_i, i = 1, 2, ..., n$.

In estimating the standard error of the bootstrap samples, the number of B falls within the range of (25 – 200). According to Efron and Tibshirani (1998) bootstrap sample size of 50 is sufficient enough to give a reasonable estimate of the standard error of the *MOM* estimator. In this research, the same sample size was used to estimate the standard error of the *MOM* estimator.

Secondly, the bootstrap replications equating to each of the bootstrap samples is defined below:

$$\hat{\theta}^*(b) = s(x^{*b}) b = 1, 2, ..., B. \tag{3.4}$$

$s(x^*)$ is the mean of the bootstrap data distribution were evaluated.

Thirdly, we estimate the bootstrap estimate of $se_F(\hat{\theta})$ from the sample standard deviation of the bootstrap replications as defined below:

$$\hat{se}_B = \{\sum_{b=1}^{B}[\hat{\theta}^*(b) - \hat{\theta}^*(\cdot)]^2/(B-1)\}^{1/2}, \tag{3.5}$$

where $\hat{\theta}^*(\cdot) = \sum_{b=1}^{B} \hat{\theta}^*(b)/B$

and $\hat{\theta}^* = s(x^*)$

The weight w_j for the Winsorized data distribution for each group is expressed as below:

$$w_j = \frac{1/S^2_{eWMOMj}}{\sum_{j=1}^{J} 1/S^2_{eWMOMj}}, \tag{3.6}$$

where $\sum_{j=1}^{J} 1/S^2_{eWMOMj}$ is the sum of the inverse of the square of the standard error for all the groups in the ordered data distribution, from the real data distribution. Where $1/S^2_{eWMOMj}$ is the reciprocal of the standard error of the Winsorized data distribution and S^2_{eWMOMj} is the standard error of the Winsorized data distributon and is defined using the formula below:

$$S^2_{eWMOMj} = \sqrt{\frac{s_j^2{}_{WMOMj}}{n_j}} \tag{3.7}$$

The variance weighted estimate of the total mean for the Winsorized data distribution for all the groups is expressed as:

$$\hat{\mu}_j = \sum_{j=1}^{J} w_j \bar{X}_{WMOMj} \tag{3.8}$$

where w_j is given in equation (3.6) and $\bar{X}_{WMOMj}$ is given in equation (3.1).

The t statistic for the Winsorized data distribution for each of the group is obtained using the formula below:

$$t_j = \frac{\bar{X}_{WMOMj} - \hat{\mu}_j}{S_{eWMOMj}} \tag{3.9}$$

where $\bar{x}_{WMOMj}, \hat{\mu}_j$ and S_e is the Winsorized *MOM* estimator, the total mean for the Winsorized data distribution and the standard error of the Winsorized data distribution respectively. In the Alexander-Govern (1994) approach, the t_j value is transformed to a standard normal with the use of Hill's (1970) normalization approximation technique and the hypothesis testing of the Winsorized data distribution, where S^2_{WMOMj} is the usual estimate of the Winsorized sample variance of the *WMOM* estimator for μ_j is defined below:

$H_0 : \mu_1 = \ldots = \mu_j$

$H_A : \mu_i \neq \mu_j$, for at least $i \neq j$

The alternative hypothesis, (H_A) contradicts the claim or statement made by the null hypothesis. The grand mean for all the groups is calculated using:

Thus, the normalization approximation formula for the Alexander-Govern method, using the Winsorized data distribution is expressed as the original AG test (see Section 2.6.1).

The test statistic of the Winsorized Modified One Step *M*- estimator in the Alexander-Govern test (*AGWMOM*) test for all the groups in the ordered data sample is expressed as:

$$AGWMOM = Z^2{}_{WMOMj} \tag{3.10}$$

The test statistic for the $AGWMOM$ test follows a chi-square distribution at $\alpha = 0.05$ level of significance, having (J – 1) chi-square degrees of freedom. The *p*-value can be determined using a standard chi-square distribution table. If the *p*- value of the $AGWMOM$ test is less than 0.05, then we can say that the test is significant, otherwise, the test is referred to as not significant.

3.3 Variables Investigated in this Research

There are five different types of variables that were used in this research, namely: sample size, variance ratio (equal and unequal variance), group sizes, types of distribution, and nature of pairing. All these variables were manipulated to investigate the strength and the weakness of the original Alexander-Govern test (AG) test, the Modified One Step *M*-estimator in Alexander-Govern test ($AGMOM$) test, the

Winsorized Modified One Step *M*-estimator in Alexander-Govern test (*AGWMOM*) test, the *t-test* and the *ANOVA*.

3.3.1 Balanced and Unbalanced Sample Size

Researchers such as Yusof, Abdullah, Yahaya and Othman (2011) made used of equal and unequal sample sizes (balanced and unbalanced) condition in their research findings. For a balanced sample size, they used sample sizes of N = 60, where $n_1 = 15$, $n_2 = 15$, $n_3 = 15$ and $n_4 = 15$, and for N=80, where $n_1 = 20$, $n_2 = 20$, $n_3 = 20 n_4 = 20$ and $n_5 = 20$.Under an unbalanced sample size condition, where N = 60, they selected $n_1 = 12$, $n_2 = 14$, $n_3 = 16$ and $n_4 = 18$, for N=80, they selected $n_1 = 10$, $n_2 = 20$, $n_3 = 20$ and $n_4 = 30$. The selection of the sample sizes chosen by these researchers gave them a remarkable control of Type I error rate.

On the other hand, Othman *et al.* (2004) and Keselman *et al.* (2007) in their researches, used unbalanced sample size of N = 25, where $n_1 = 10$, and $n_2 = 15$, for N = 30, where $n_1 = 10$ and $n_2 = 20$, for N = 40, where $n_1 = 15$ and $n_2 = 25$, for N= 70, where $n_1 = 10$, $n_2 = 15$, $n_3 = 20$ and $n_4 = 25$, for N = 90, where $n_1 = 15$, $n_2 = 20$, $n_3 = 25$ and $n_4 = 30$. The selection of the sample sizes chosen by these researchers gave them an excellent control of Type I error rate.

Abdullah, Yahaya and Othman (2007) used unbalanced sample sizes of N = 40, where $n_1 = 15$ and $n_2 = 25$, for N = 80, where $n_1 = 10, n_2 = 15, n_3 = 25$ and $n_4 = 30$. The selection of the sample sizes chosen by these researchers are considered to be moderate quantities of sample sizes and it gave them a good control of Type I error rate and a high power for their test in the course of their analysis. Furthermore, Yusof, Abdullah and Yahaya (2012) also used unbalanced samples in their research, such as for N = 60, where $n_1 = 12, n_2 = 14, n_3 = 16$ and $n_4 = 18$, for N = 80, where $n_1 = 10, n_2 = 20, n_3 = 20$ and $n_4 = 30$. The selection of the sample sizes used by these researchers gave them a remarkable control of Type I error rate.

In this research, both balanced (equal sample sizes) and unbalanced sample sizes (unequal sample sizes) were selected. Under a balanced sample size condition, for N = 40, the sample sizes used are $n_1 = 20$ and $n_2 = 20$, under two groups condition. For the case of four groups, where N = 80, n_1, n_2, n_3 and n_4 were all set equal to 20. Under a balanced sample size for six group condition, the total N = 120, where $n_1 = 20, n_2 = 20, n_3 = 20, n_4 = 20, n_5 = 20$ *and* $n_6 = 20$.

For unbalanced sample size condition, the sample sizes used are $n_1 = 16$, and $n_2 = 24$, for four group case, where $n_1 = 15, n_2 = 15, n_3 = 20$ and $n_4 = 30$. For six group condition, the sample sizes used are $n_1 = 2, n_2 = 4, n_3 = 4, n_4 = 16, n_5 = 32$, $n_6 = 62$.

The selection of both balanced and unbalanced sample sizes in this research, has assisted us to see the performance of our newly proposed method under these two conditions of sample sizes. The quantities of sample sizes chosen in this research are referred to as moderate amount of sample sizes to make comparison on the effect of power on the number of groups on this new method (Abdullah, Yahaya & Othman, 2008).

3.3.2 Variance Ratios

Othman *et al.* (2004) and Keselman *et al.* (2007) in their research analysis used unequal variance ratios of 36:1, 8:1 as well as 36:1:1:1 and 8:1:1:1. The variance ratios of 36:1 and 36:1:1:1 are considered as extreme conditions of variance heterogeneity, while the variance ratio of 8:1 and 8:1:1:1 are considered as less extreme conditions of variance heterogeneity. These researchers agreed that the variance ratios of 36:1 and 36:1:1:1 are considered large enough to give researchers acceptable results for their data analysis.

Meanwhile, Abdullah, Yahaya and Othman (2008) used different extreme conditions of variance heterogeneity of 1:36, 1:1:1:36, 1:1:1:1:1:36, 1:4:16:36, and 1:4:4:16:16:36 respectively. It was observed that the variance ratios of 1:1:1:36 produced a higher power compared to the variance ratio of 1:4:16:36. The difference in the values of the power between these two variance ratios was 0.77, which gave a power of 0.8.

The variance ratios of 1:4:16:36 produced a smaller power compared to the 0.8, for a large effect size. The variance ratio of 1:1:1:1:1:36, produced a higher power compared to the variance ratio of 1:4:4:16:16:36, where the difference in the power values was as large as 0.8 (0.8098). The selection of variance ratio can have a great influence on the power of a test. The variance ratio of 1:1:1:1:1:36 makes the test to be more powerful.

In this research, the variance ratios that were used are: (1:1), (1:36) and (1:1) (36:1), (1:1:1:1) and (1:1:1:36) and (1:1:1:1) (36:1:1:1), (1:1:1:1) (1:4:16:36) and (1:1:1:1)

(36:16:4:1) and (1:1:1:1:1:1) (1:4:4:16:16:36) and (1:1:1:1:1:1) (36:16:16:4:4:1) under variance heterogeneity. The selection of the variance ratio in this research has helped to show how well this new method can perform under extreme condition of variance heterogeneity.

3.3.3 Group Sizes

Scholars such as Othman *et al.* (2004) and Keselman *et al.* (2007) used group size of $J = 2$ and the results shows that all the method they used were robust. All the values of their method fell within the Bradley's (1978) stringent condition of robustness between the interval of 0.025 and 0.075. In the case of $J = 4$, all the methods they used fell within the Bradley's condition of robustness except for one which was not robust under very severe condition of non-normality.

According to Yusof, Othman and Yahaya (2008) in their research for $J = 2$, all the values for the method they used fell within the stringent criterion of robustness. In the case of $J = 4$, all the method used fell within the Bradley's condition of robustness with one of the methods used having the closest value to the nominal value of 0.05.

In the work done by Abdullah, Yahaya and Othman (2007), they used group sizes of $J = 2$, $J = 4$ and $J = 6$, and a standard stringent criteria of robustness was considered within the interval 0.042 and 0.058 to judge the robustness of the three methods they used in their analysis, namely the Alexander-Govern test with common mean, the Alexander-Govern test with trimmed mean, and the Alexander-Govern test with *MOM* estimator. It was discovered that the Alexander-Govern test using common mean and the *MOM* estimator as a measure of the central tendency for the test, were robust for all the groups, the trimmed mean was robust only for two groups.

All these findings showed that the number of group gives significant impact to the performances of the test. Similarly, in this research, different groups sizes of J =2, J = 4 and J = 6 were used to investigate the performance of this new method.

3.3.4 Types of Distribution

In this research, four different types of distribution, namely: standard normal distribution, symmetric heavy tailed distribution, skewed normal distribution and skewed heavy tailed distribution, were used to examine the effects of Type I error rate on the types of distribution. These four different types of distribution, represents different levels of skewness and kurtosis, by using the *g*- and *h*- distribution.

The term heavy tailed distribution in probability distribution theory could be described as the tail that is not exponentially bounded. As a result, the tails are heavier than the exponential distribution. While a skewed normal tail distribution describes the measure of the symmetry of the probability of the real-valued random variable about the distribution. Skewness in a data distribution is defined as the curve that is seen distorted or when it is skewed either to the right tail or to the left tail of a given data sets. The word "kurtosis" is defined as the measure of the peak of a distribution and it shows how high the distribution is close to the mean. The distribution of the data is said to be symmetric when g = 0 and h = 0 as discussed by Abdullah, Yahaya and Othman, (2007).

According to Yusof, Abdullah, Yahaya and Othman (2011), the *g*- and *h*- distribution is modified from the normal distribution, where *g* is a constant that controls the value of skewness in the distribution, and *h* is a constant that controls the value of kurtosis in the distribution. As the value of the *g* and *h* increases, the level of skewness and kurtosis increases accordingly.

The observations of the *g*- and *h*- distribution are obtained by transforming the standard normal variates using the formula below:

$$Y_{ij} = \begin{Bmatrix} \dfrac{\exp(gZ_{ij}) - 1}{g} \exp(hZ^2_{ij}/2) \; for\ g \neq 0 \\ Z_{ij} \exp(hZ^2_{ij}/2) \qquad for\ g = 0 \end{Bmatrix}, \tag{3.11}$$

where Z_{ij} is the standard normal distribution with *i* and *j* population.

The values of the *g*- and *h*- distribution that were used in this research are: $g = 0$ and $h = 0$ (standard normal distribution), $g = 0$ and $h = 0.5$ (symmetric heavy tailed distribution), $g = 0.5$ and $h = 0$ (skewed normal tailed distribution), and $g = 0.5$ and $h = 0.5$ (skewed heavy tailed distribution) as discussed by Abdullah, Yahaya and Othman (2007). The characteristics of the *g*- and *h*- distribution are presented in Table 3.1.

Table 3.1

The characteristics of the g- and h- distribution

g- (Nonnegative content)	*h*- (Nonnegative content)	Skewness	Kurtosis	Types of distribution
0	0	0	3	Standard normal
0	0.5	0	11986.20	Symmetric heavy tailed
0.5	0	1.81	18393.60	Skewed normal tailed
0.5	0.5	120.10	18393.60	Skewed heavy tailed

Source: Wilcox (1997)

3.3.5 Types of Pairing

In this research, the robustness of the Alexander-Govern test, the Modified One Step *M*-estimator in the Alexander-Govern test and the Winsorized Modified One Step *M*-estimator in the Alexander-Govern test were determined by using two types of pairing: positive pairing and negative pairing. Positive pairing is a kind of pairing that occurs, whereby a smaller sample size is being paired with a smaller variance while a larger sample size is being paired with a larger variance. While negative pairing is the kind of pairing that occurs, whereby a smaller sample size is being paired with a larger variance, and a larger sample size is being paired with a smaller variance (Othman *et al.*, 2004 & Keselman *et al.*, 2007).

These conditions of pairing are selected, since they can mainly give conservative or conventional results for positive pairings and liberal or substantial results for negative pairings accordingly (Keselman *et al.*, 2007). In a balanced condition for two groups (see Table 3.2), for four groups (see Table 3.3) and for six groups (see Table 3.4) is a condition where a balanced sample size is combined with a balanced variance ratio. It is also referred to as a perfect condition with the combination of both balanced sample size with equal variance.

3.4 Research Design

The Alexander-Govern test is a test that uses mean as a measure of its central tendency, but is not robust for non-normal data under variance heterogeneity. For the design of this research, both balanced and unbalanced sample sizes were paired with equal and unequal variance for two groups (J = 2), four groups (J = 4), and for six groups (J = 6), positively and negatively with each of the *g*- and *h*- distribution.

For each of the tests namely: the *AG test*, the *AGMOM test*, *the AGWMOM* test, the *t-test* and the *ANOVA*, data set of 5,000 were simulated in the research design. The values of the skewness and kurtosis for the *g*- and *h*- distributions are theoretical values and a computer generated values, based on 5,000 observations were simulated for these values (Wilcox, 1997). 5,000 data sets were used in this research to give us a satisfactory result for each of the tests. To obtain the pseudo random variates, *SAS* generator *RANNOR (SAS Institute,* 1999) was used with a nominal level of $\alpha = 0.05$ for the analysis of the tests in this research.

The robustness of the *WMOM* estimator with respect to the Type I error rate and the power of the test, was obtained by manipulating the five listed variables as mentioned previously in Section 3.3. The research design used in this research, for two, four and six groups, shows the combination and pairing of both balanced and unbalanced samples with equal and unequal variance, with each of the *g*- and *h*- distribution, positively and negatively. Each of this pairing condition is denoted by C1 to C84 respectively. The research design is illustrates in tabular form, for two groups (see Table 3.2), four groups (see Table 3.3) and six groups (see Table 3.4) groups respectively.

Table 3.2

Research Design for Two Groups Condition with N = 40

The *g*- and *h*-distribution	**Balanced and Unbalanced sample size**	**Variance ratio**	**Nature of Pairing**	**Notations for the Conditions**
$g = 0$ and $h = 0$	20:20	1:1	Balanced condition	C1
		1:36	Positive Pairing	C2
	16:24	1:1		C3
		1:36	Positive Pairing	C4
		36:1	Negative Pairing	C5
$g = 0$ and $h = 0.5$	20:20	1:1	Balanced condition	C6
		1:36	Positive Pairing	C7
	16:24	1:1		C8
		1:36	Positive Pairing	C9
		36:1	Negative Pairing	C10
$g = 0.5$ and $h = 0$	20:20	1:1	Balanced condition	C11
		1:36	Positive Pairing	C12
	16:24	1:1		C13
		1:36	Positive Pairing	C14
		36:1	Negative Pairing	C15
$g = 0.5$ and $h = 0.5$	20:20	1:1	Balanced condition	C16
		1:36	Positive Pairing	C17
	16:24	1:1		C18
		1:36	Positive Pairing	C19
		36:1	Negative Pairing	C20

Table 3.3

Research Design for Four Groups Condition with N = 80

The g- and h-distribution	Balanced and Unbalanced sample size	Variance ratio	Nature of Pairing	Notations for the Nature of Pairing
$g = 0$ and $h = 0$	20:20:20:20	1:1:1:1	Balanced condition	C21
		1:1:1:36	Positive Pairing	C22
		1:4:16:36	Positive Pairing	C23
	15:15:20:30	1:1:1:1		C24
		1:1:1:36	Positive Pairing	C25
		36:1:1:1	Negative Pairing	C26
		1:4:16:36	Positive Pairing	C27
		36:16:4:1	Negative Pairing	C28
$g = 0$ and $h = 0.5$	20:20:20:20	1:1:1:1	Balanced condition	C29
		1:1:1:36	Positive Pairing	C30
		1:4:16:36	Positive Pairing	C31
	15:15:20:30	1:1:1:1		C32
		1:1:1:36	Positive Pairing	C33
		36:1:1:1	Negative Pairing	C34
		1:4:16:36	Positive Pairing	C35
		36:16:4:1	Negative Pairing	C36
$g = 0.5$ and $h = 0$	20:20:20:20	1:1:1:1	Balanced condition	C37
		1:1:1:36	Positive Pairing	C38
		1:4:16:36	Positive Pairing	C39
	15:15:20:30	1:1:1:1		C40
		1:1:1:36	Positive Pairing	C41
		36:1:1:1	Negative Pairing	C42
		1:4:16:36	Positive Pairing	C43
		36:16:4:1	Negative Pairing	C44
$g = 0.5$ and $h = 0.5$	20:20:20:20	1:1:1:1	Balanced condition	C45
		1:1:1:36	Positive Pairing	C46
		1:4:16:36	Positive Pairing	C47
	15:15:20:30	1:1:1:1		C48
		1:1:1:36	Positive Pairing	C49
		36:1:1:1	Negative Pairing	C50
		1:4:16:36	Positive Pairing	C51
		36:16:4:1	Negative Pairing	C52

Table 3.4

Research Design for Six Groups Condition with N = 120

The *g*- and *h*-distribution	**Balanced and Unbalanced sample size**	**Variance ratio**	**Nature of Pairing**	**Notations for the Nature of Pairing**
$g = 0$ and $h = 0$	20:20:20:20:20:20	1:1:1:1:1:1	Balanced condition	C53
		1:1:1:1:1:36	Positive Pairing	C54
		1:4:4:16:16:36	Positive Pairing	C55
	2:4:4:16:32:62	1:1:1:1:1:1		C56
		1:1:1:1:1:36	Positive Pairing	C57
		36:1:1:1:1:1	Negative Pairing	C58
		1:4:4:16:16:36	Positive Pairing	C59
		36:16:16:4:4:1	Negative Pairing	C60
$g = 0$ and $h = 0.5$	20:20:20:20:20:20	1:1:1:1:1:1	Balanced condition	C61
		1:1:1:1:1:36	Positive Pairing	C62
		1:4:4:16:16:36	Positive Pairing	C63
	2:4:4:16:32:62	1:1:1:1:1:1		C64
		1:1:1:1:1:36	Positive Pairing	C65
		36:1:1:1:1:1	Negative Pairing	C66
		1:4:4:16:16:36	Positive Pairing	C67
		36:16:16:4:4:1	Negative Pairing	C68
$g = 0.5$ and $h = 0$	20:20:20:20:20:20	1:1:1:1:1:1	Balanced sample size	C69
		1:1:1:1:1:36	Positive Pairing	C70
		1:4:4:16:16:36	Positive Pairing	C71
	2:4:4:16:32:62	1:1:1:1:1:1		C72
		1:1:1:1:1:36	Positive Pairing	C73
		36:1:1:1:1:1	Negative Pairing	C74
		1:4:4:16:16:36	Positive Pairing	C75
		36:16:16:4:4:1	Negative Pairing	C76
$g = 0.5$ and $h = 0.5$	20:20:20:20:20:20	1:1:1:1:1:1	Balanced condition	C77
		1:1:1:1:1:36	Positive Pairing	C78
		1:4:4:16:16:36	Positive Pairing	C79
	2:4:4:16:32:62	1:1:1:1:1:1		C80
		1:1:1:1:1:36	Positive Pairing	C81
		36:1:1:1:1:1	Negative Pairing	C82
		1:4:4:16:16:36	Positive Pairing	C83
		36:16:16:4:4:1	Negative Pairing	C84

The research design was used to determine the robustness of the modified Alexander-Govern test. By using this research design, the best procedure was obtained for the tests. According to Lix and Keselman (1998), the empirical rate of Type I error must be within the interval of $0.042 \leq \hat{\alpha} \leq 0.058$ that is use judge the robustness of a given test at α level of significance. The interval of the values selected in this research gave a strict condition for the robustness of the tests, with the aim of producing minimum error rate with deviation from model assumptions.

Abdullah, Yahaya and Othman, (2007) used the interval of 0.042 and 0.058 for evaluating the robustness of the test in their analysis. The interval selected by these researchers, shows that a test is said to be robust when its' Type I error rate is within the stringent criterion of robustness. Otherwise, if the test falls outside the stringent criteria of robustness, then the Type I error rate is out of control. According to Bradley's (1978) the lenient criteria of robustness should be within the interval of (0.025 – 0.075). This interval of robustness is also selected in this research, to see those tests that can give excellent control of Type I error rate.

3.5 Statistical Power Analysis

The statistical power of a test is defined as the probability that it will definitely result in significant outcomes (Cohen, 1988). It could also be described as the capacity of a test to recognize any effect when the effect size occurs. Cohen (1988) explains that the effect size is the extent at which a phenomenon is observed in the population. As a result, the null hypothesis becomes false in the population. When making hypothesis testing, the probability of accepting the null hypothesis when it is false, is referred to as Type II error which is represented as β. In addition, the power of a test could be defined as the probability of not accepting the null hypothesis when it is false, and it is represented as $1-\beta$ (Cohen, 1988).

The power of a test is affected by three variables, namely: (i) sample size, (ii) level of significance and (iii) effect size.

The sample size: In detecting the power of a test, the selection of the sample size chosen by the researcher is very important. The selection of the sample sizes directly affects the power of a test. For a small sample size selected, it will result to a very small amount of the power of the test. When the sample size is large, it will definitely result to a large amount of the power of the test. Hence, the selection of the sample size chosen by the researcher will directly affect the power output of the test. The power of a test is directly proportional to the quantity of the sample sizes selected (Abdullah, Yahaya & Othman, 2008).

Murphy and Myors (1998) stated that the power of a test must be above 0.5 and can be considered sufficient when the value is 0.8 and above. When the power of a test is 0.8, it shows that success which is the probability of not accepting the null hypothesis is four times as certain as failure. When the power of a test is 0.9, it shows that the success is nine times as certain as failure.

The level of significance: It is the process of neglecting the null hypothesis when it is actually true, and is otherwise referred to as Type I error. The level of significance is expressed as α. To obtain the power of a test, the value of α selected is very crucial (Abdullah, Yahaya & Othman, 2008). The level of significance selected for this research is α=0.05. When the value of α to be chosen is too small, it will definitely result to a smaller amount of the power of the test.

Effect size: In statistics, it is observed that the probability of the null hypotheses, that is the *p*-value, decreases as the effect size increases and the sample size increases accordingly. The effect size shows the differences between the maximum and minimum means between two groups, divided by the standard deviation inside the population (Cohen, 1998).

3.5.1 The Effect Size Index

In this research, the effect size that was used for two groups ($J = 2$), four groups ($J = 4$) and six groups ($J = 6$) and their pattern of variability is explained below:

3.5.2 The Effect Size Index for $J = 2$

Abdullah, Yahaya and Othman (2008) stated that when considering two population groups, the effect size index, is the effect size that we are aiming at detecting. By definition, the effect size index (*d*) is expressed as:

$$d = \frac{m_A - m_B}{\sigma} \tag{3.12}$$

where:

$|m_A - m_B|$ is the absolute value of the difference between the maximum and minimum means between the two groups.

$\sigma =$ the standard deviation of the population.

According to Murphy and Myors (1998) the effect size is said to be small when $d = 0.2$, it is said to be medium when $d = 0.5$ and it is said to be large when $d = 0.8$.

3.5.3 The Effect Size Index for *J* = 4 or More

According to Cohen (1988) when *k* > *2*, where *k* represents the number of means in the distribution, and it implies that the number of means is increased above two. The association between the number of means and the range of the standardized mean relies precisely on how much the means are dispersed over the range in the distribution. The spread of the means (*f*) is expressed as:

$$f = \frac{\sigma_m}{\sigma} \tag{3.13}$$

where:

σ_m is the original scale units of the standard deviation

Under this situation, d is no longer an effect size, but it represent the largest and the smallest means or otherwise, the range of the standardize means.

By definition, d is expressed as:

$$d = \left| \frac{m_A - m_B}{\sigma} \right|$$

(3.14)

Given:

m_A is the biggest value of *K* means, and

m_B is the smallest value of K means

The f index relies on the specification in the patterns of differences of the means. Cohen (1998) also mentioned that there are three forms of variability that shows the association between f and d as expressed below.

Form 1:

Small variability, where f is expressed as:

$$f = d\sqrt{\frac{1}{2k}} \tag{3.15}$$

f is the effect size index for more than two groups i.e for four groups and above, while d is the effect size index for two groups condition.

Form 2:

Medium variability, where f is expressed as:

$$f = \frac{d}{2}\sqrt{\frac{k+1}{3(k-1)}} \tag{3.16}$$

Form 3:

Large variability, where f is expressed as:

$$f = \frac{1}{2}d \text{ , when } k \text{ is said to be even;} \tag{3.17}$$

$$f = d\frac{\sqrt{k^2-1}}{2k} \text{ , when } k \text{ is odd} \tag{3.18}$$

*Note *that k* is the number of means in the population.

When f is 0.1, the effect size index is said to be small, when f is 0.25, the effect size index is said to be medium, when f is 0.4, the effect size index is said to be large accordingly (Abdullah, Yahaya & Othman, 2008).

Table 3.5

Pattern of Variability of the Effect Size Index for 4 Groups and 6 Groups

The Effect Size Index	**For $J = 4$**	**For $J = 6$**
Small	$-\frac{1}{2}d, 0, 0, \frac{1}{2}d$	$-\frac{1}{2}d, 0, 0, 0, 0, \frac{1}{2}d$
Medium	$-\frac{1}{2}d, -\frac{1}{4}d, \frac{1}{4}d, \frac{1}{2}d$	$-\frac{1}{2}d, -\frac{1}{3}d, -\frac{1}{6}d, \frac{1}{6}d, \frac{1}{3}d, \frac{1}{2}d$
Large	$-\frac{1}{2}d, -\frac{1}{2}d, \frac{1}{2}d, \frac{1}{2}d$	$-\frac{1}{2}d, -\frac{1}{2}d, -\frac{1}{2}d, \frac{1}{2}d, \frac{1}{2}d, \frac{1}{2}d$

Source: Cohen (1988).

It should be noted that the effect size index is not considered in analyzing for the Type I error rate for the *AGMOM* test, the *AGWMOM* test, the *t*-test and the *ANOVA*, in this research. The effect size index (*d*) is only used in analyzing for the power of the tests. The effect size index used in analyzing the power of the tests for two groups condition in this research are: d = 0.2, 0.3, 0.4, 0.5, 0.6, 0.7 and 0.8 respectively. For four groups condition, d = 0.29, 0.57, 0.80, 0.80, 1.00, 1.20, 1.40 and 1.60 respectively. Under six groups condition, d = 0.34, 0.69, 0.88, 0.80, 1.00, 1.20, 1.40 and 1.60 respectively.

Case 1: The pairing of equal sample size with equal variance, the population means are calculated as thus:

The effect size index for two groups is obtained using the formula below:

$$d = \frac{\mu_A - \mu_B}{\sigma} \tag{3.19}$$

where

μ_A is the first population mean,

μ_B is the second population mean and

σ is the standard deviation of the population means.

When $d = 0.2$, $\mu_A = 1$, and $\sigma = 1$

substituting into Equation (3.18);

$0.2 = \frac{1 - \mu_B}{1}$

$0.2 = 1 - \mu_B$

$\mu_B = 1 - 0.2$

Therefore, $\mu_B = 0.8$.

Then, the population means for two groups conditions, with the pairing of equal sample size with equal variance are

$(\mu_A, \mu_B) = (1, 0.8)$

Case 2: Pairing of equal sample size with equal variance, for both positive and negative pairing.

Where $d = 0.2$, the standard deviation is obtained for the unequal variances using the formula below:

$$\sigma = \sqrt{\frac{v_1 + v_2}{2}}, \tag{3.20}$$

where, v_1 *and* v_2 represents the variance ratio for two groups condition.

For example, when v_1 = 1 and v_2 = 36,

$$\sigma = \sqrt{\frac{1+36}{2}}$$

$$\sigma = 4.3012.$$

By using Equation (3.18) the second population mean can be obtained.

For example, when d = 0.2, μ_A = 1 and δ = 4.3012, substituting into Equation (3.18)

$$0.2 = \frac{1 - \mu_B}{4.3012}$$

$$0.2(4.3012) = 1 - \mu_B$$

$$\mu_B = 1 - 0.8602$$

$$\mu_B = 0.1398$$

Thus,

$$(\mu_A, \mu_B) = (1, 0.1398).$$

In the research design for analyzing the Type I error rates of the tests, the effect size index *(d)* is not considered.

Case 3: The pairing of unequal sample with unequal variance i.e. (16, 24), (1, 36) or (16, 24), (1, 36).

For $v = (1:36)$ with $s = (16, 24)$, the standard deviation is obtained by using Equation (3.19):

$$\sigma = \sqrt{\frac{(n_1 \times \sigma^2{}_1) + (n_2 \times \sigma^2{}_2)}{n_1 + n_2}} \tag{3.21}$$

Substituting, $n_1 = 16$, $n_2 = 24$, $\sigma^2{}_1 = 1$ and $\sigma^2{}_2 = 36$ into Equation (3.19),

$$\sigma = \sqrt{\frac{(16)(1) + (24)(36)}{16 + 24}}$$

$= 4.6904$

Substituting $d = 0.2$, $\mu_A = 1$ and $\sigma = 4.6904$ into Equation (3.18), the mean for the second population is given by

$0.2 = \frac{1 - \mu_B}{4.6904}$

$(0.2)\ (4.6904) = 1 - \mu_B$

$0.9381 = 1 - \mu_B$

$\mu_B = 1 - 0.9381$

$\mu_B = 0.0619$

Hence, $(\mu_A, \mu_B) = (1, 0.0619)$

Under four and six groups condition the effect size index (*f*) that was used for this research are: $f = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7$ and 0.8

When $f = 0.1$, the value of d is calculated using the formula for small effect size index, in Equation (3.14).

For four groups case, where k is defined as the number of groups in the distribution, Substituting $k = 4$, and $f = 0.1$ into Equation (3.14),

$$0.1 = d\sqrt{\frac{1}{2(4)}}$$

$$0.1 = 0.35d$$

$$d = 0.1/0.35$$

$$d = 0.29.$$

The population mean is obtained by using the formula for calculating the large pattern of variability for four groups in Table 3.5 as:

$$-\frac{1}{2}d, -\frac{1}{2}d, \frac{1}{2}d, \frac{1}{2}d$$

By substituting $d = 0.29$ into the formula,

$$-\frac{1}{2}(0.29), -\frac{1}{2}(0.29), \frac{1}{2}(0.29), \frac{1}{2}(0.29)$$

= - 0.145, - 0.145, 0.145, and 0.145. These are the values of the population means for four groups when $d = 0.29$.

When $f = 0.5$, it falls under large effect size index, since k of 4 is an even number, we use Equation (3.16) to calculate d:

$$0.5 = \frac{1}{2}(d)$$

$$d = 1.0$$

From Table 3.5, the population means for four groups is obtained by:

$$-\frac{1}{2}d, -\frac{1}{2}d, \frac{1}{2}d, \frac{1}{2}d = --\frac{1}{2}(1.0), -\frac{1}{2}(1.0), \frac{1}{2}(1.0), \frac{1}{2}$$

μ = - 0.50, - 0.50, 0.50, and 0.50.

For six groups condition, where $f = 0.3$ and $k = 6$, the value of d is obtained using Equation (3.15):

$$0.3 = \frac{d}{2}\sqrt{\frac{6+1}{3(6-1)}}$$

$0.6 = 0.68d$

$d = 0.6/0.68$

$= 0.88$

From Table 3.5, when d = 0.88, the population means for the large pattern of variability is obtained using:

$$-\frac{1}{2}d, -\frac{1}{2}d, -\frac{1}{2}d, \frac{1}{2}d, \frac{1}{2}d, \frac{1}{2}d$$

$$\mu = -\frac{1}{2}(0.88), -\frac{1}{2}(0.88), -\frac{1}{2}(0.88), \frac{1}{2}(0.88), \frac{1}{2}(0.88), \frac{1}{2}(0.88)$$

$$\mu = -0.44, -0.44, -0.44, 0.44, 0.44, 0.44$$

In conclusion, the power of a test is affected by the quantity of the sample sizes chosen; the higher the sample sizes selected, the higher would be the power of the test. The sample size selected in a data distribution is directly proportional to the power of a test.

CHAPTER FOUR

RESULTS AND ANALYSIS

4.1 Introduction

This chapter examines the performance of the Type I error rate and power for each of the tests, namely: the *AG* test, the *AGMOM* test, the *AGWMOM* test, the *t*-test and the *ANOVA*, for four different distributions, under two, four and six group conditions. So as to see of the five different tests mentioned above, which among them will give a good control of Type I error rate and high power.

4.2 The Type I Error Rate

The Type I error rate of the five different tests that were used in this research must fall under three criteria of robustness. Which are (i) those tests that fall within the stringent criteria of robustness, (ii) those tests that fall within the lenient criteria of robustness and (iii) those tests that do not fall on neither stringent criteria of robustness nor lenient criteria of robustness and are considered not to be robust.

This research considers stringent criteria of robustness, within the interval of (0.042 – 0.058), to judge the robustness of the tests (Lix & Keselman, 1998) and also considers the lenient criteria of robustness to judge the robustness of the tests that are within the interval of (0.025 – 0.075) as stated by Bradley's (1978). These intervals of robustness are selected in this research, to see those tests that can give excellent control of Type I error rate.

All the values presented in the Tables, are bolded and italized, bolded and unbolded. The bolded and italized values are those values that are strictly within the stringent criteria of robustness. The bolded values represent those values that are within the lenient criteria of robustness, but do not fulfill the stringent criteria of robustness. The unbolded values are those values that are considered not to be robust. This implies that they neither within the stringent criteria of robustness nor within the lenient criteria of robustness.

4.2.1 Two Groups Case

Under two groups condition, the Type I error rate is compared for each of the tests, namely, the *AG* test, the *AGMOM* test, the *AGWMOM* test and the *t-test* with each of the four different distributions, in order to see those test that are within the stringent criteria of robustness and also those that are within the lenient criteria of robustness. As a result, the tests are said to have remarkable control of Type I error rate.

4.2.1.1 Normal Distribution (*g*= 0; *h*= 0)

In Table 4.1, under a normal distribution, for two group condition, all the Type I error rate for the *AG* test, the *AGMOM* test and the *AGWMOM* test performed well where these tests are robust in all conditions regardless of the sample sizes and the variance ratio. The *t*-test also performed quite good where it provides good control of Type I error rate in all the conditions except for negative pairing condition. The test is considered not robust where the Type I error rate with value of 0.1078 is outside the robust criteria.

Among all the tests, the *AG* test has better result compared to the *AGMOM* test and the *AGWMOM* test because its' Type I error rate fall within the stringent criteria of robustness.

Table 4.1

Comparison of the Type I Error Rate for the AG, AGMOM, AGWMOM, and t-test *Under Normal Distribution for Two Groups Condition*

Sample Size	Equal and Unequal Variance	*AG*	*AGMOM*	*AGWMOM*	*t*-test
20:20	1:1	***0.0508***	**0.0414**	**0.0392**	***0.0528***
	1:36	***0.0562***	***0.0528***	***0.0496***	**0.0710**
16:24	1:1	***0.0484***	***0.0430***	0.0386	***0.0570***
	1:36	***0.0570***	***0.0552***	***0.0496***	**0.0618**
	36:1	***0.0498***	***0.0450***	***0.0438***	0.1078

4.2.1.2 Symmetric Heavy Tailed Distribution ($g = 0$ and $h = 0.5$)

In Table 4.2, under a skewed normal tailed distribution, the Type I error rate for the *AG* test, the *AGMOM* test and the *AGWMOM* all fall within the lenient criteria of robustness and gave the best control of Type I error rate over the *t*-test. For both positive and negative pairing condition, the *t*-test produced Type I error rate with value 0.0138 and 0.0814 that fall outside the criteria of robustness.

Table 4.2

*Comparison of the Type I Error Rate for the AG, AGMOM, AGWMOM, and t-*test, *Under a Symmetric Heavy Tailed Distribution for Two Groups Condition*

Sample Size	**Equal and Unequal Variance**	***AG***	***AGMOM***	***AGWMOM***	***t-test***
20:20	1:1	**0.0336**	**0.0262**	**0.0346**	**0.0356**
	1:36	**0.0340**	**0.0358**	**0.0392**	**0.0402**
16:24	1:1	**0.0304**	**0.0266**	**0.0352**	***0.0430***
	1:36	**0.0394**	**0.0340**	**0.0412**	0.0138
	36:1	**0.0312**	**0.0294**	**0.0346**	0.0814

4.2.1.3 Skewed Normal Tailed Distribution (g = 0.5 and h = 0)

In Table 4.3, under a skewed normal tailed distribution, for two groups conditions, the *AG* test, the *AGMOM* test and the *AGWMOM* test gave an excellent control of Type I error rate over the *t*-test, because these tests are robust in all conditions. While the *t*-test only robust with stringent criteria when variance are equal regardless the number of sample size. It is also found to be robust under condition of positive nature of pairing.

Table 4.3

Comparison of the Type I Error Rate for the AG, AGMOM, AGWMOM, and t-test *Under a Skewed Normal Tailed Distribution for Two Groups Condition*

Sample Size	**Equal and Unequal Variance**	***AG***	***AGMOM***	***AGWMOM***	***t-test***
20:20	1:1	***0.0508***	***0.0420***	**0.0364**	***0.0474***
	1:36	***0.0562***	***0.0534***	***0.0558***	0.0882
16:24	1:1	***0.0480***	***0.0434***	**0.0386**	***0.0570***
	1:36	***0.0570***	***0.0560***	**0.0588**	**0.0380**
	36:1	***0.0498***	***0.0504***	***0.0450***	0.1538

4.2.1.4 Skewed Heavy Tailed Distribution (g =0.5 and h = 0.5)

In Table 4.4, under a skewed heavy tailed distribution, the *AGMOM* and *AGWMOM* test gave a remarkable control of Type I error rate compared to the *AG* test and the *t*-test, because the test falls within the robust criteria in all the conditions. The *AG* test has Type I error rate that are outside the criteria of robustness, which are under balanced sample size with unequal variance and unbalanced sample size with positive pairing condition. The Type I error rate of the *t*-test has a value of 0.0138 and 0.0878, for both positive and negative condition of pairings accordingly, that falls outside the criteria of robustness.

Table 4.4

Comparison of the Type I Error Rate for the AG, AGMOM, AGWMOM, and t-test *Under a Skewed Heavy Tailed Distribution for Two Groups Condition*

Sample Size	**Equal and Unequal Variance**	***AG***	***AGMOM***	***AGWMOM***	***t-test***
20:20	1:1	**0.0336**	**0.0258**	**0.0314**	**0.0288**
	1:36	0.3400	**0.0374**	***0.0470***	***0.0430***
16:24	1:1	**0.0274**	**0.0272**	**0.0352**	**0.0370**
	1:36	0.3940	**0.0378**	***0.0422***	0.0138
	36:1	**0.0312**	**0.0332**	**0.0298**	0.0878

4.2.2 Four Groups Condition

Under four groups condition, the Type I error rate is compared for each of the tests, namely: the *AG* test, the *AGMOM* test, the *AGWMOM* test and the *ANOVA* for each of the four different distributions, to see which of the test is more robust and have an excellent control of Type I error rate.

4.2.2.1 Normal Distribution ($g = 0$ and $h = 0$)

In Table 4.5, under a normal distribution, for four group conditions, the *AG* test, the *AGMOM* test and the *AGWMOM* test gave an excellent control of Type I error rate compared to the *ANOVA*. Under balanced sample the *ANOVA* falls outside the criteria of robustness when the variances are unequal. Under unbalanced sample size the test is considered not robust when the nature of pairing is negative.

Table 4.5

Comparison of the Type I error rates for the AG, AGMOM, AGWMOM, and the ANOVA Under a Normal Distribution for Four Groups condition

sample size	**Equal and Unequal Variance**	***AG***	***AGMOM***	***AGWMOM***	***ANOVA***
20:20:20:20	1:1:1:1	***0.0518***	**0.0404**	**0.0386**	***0.0518***
	1:1:1:36	***0.0522***	***0.0428***	**0.0408**	0.1096
	1:4:16:36	***0.0544***	***0.0500***	***0.0468***	0.0798
15:15:20:30	1:1:1:1	***0.0504***	***0.0478***	***0.0458***	***0.0500***
	1:1:1:36	***0.0514***	***0.0482***	***0.0458***	**0.0334**
	36:1:1:1	***0.0504***	***0.0486***	***0.0446***	0.1696
	1:4:16:36	***0.0520***	***0.0492***	***0.0464***	**0.0320**
	36:16:4:1	***0.0516***	***0.0514***	***0.0468***	0.1446

4.2.2.2 Symmetric Heavy Tailed Distribution ($g = 0$ and $h = 0.5$)

In Table 4.6, under a symmetric heavy tailed distribution, the *AGWMOM* test gave an outstanding control of Type I error rate compared to the other three tests. Only one condition of the test did not fall within the criteria of robustness. The *AG* test was found to be robust only under balanced sample size condition. For the *AGMOM* test, it is robust only under two conditions with variance ratio of 1:4:16:36 that are balanced sample size and unbalanced sample size with positive pairing. The *ANOVA* still can be considered robust as long as the variances are equal regardless of the sample size. It is also robust under balanced sample size with variance ratio 1:4:16:36.

Table 4.6

Comparison of the Type I Error Rate for the AG, AGMOM, AGWMOM, and the ANOVA Under a Symmetric Heavy Tailed Distribution, for Four Groups Condition

Sample Size	**Equal and Unequal Variance**	***AG***	***AGMOM***	***AGWMOM***	***ANOVA***
20:20:20:20	1:1:1:1	**0.0280**	0.0218	**0.0282**	**0.0336**
	1:1:1:36	**0.0282**	0.0230	**0.0310**	0.0782
	1:4:16:36	**0.0282**	**0.0260**	**0.0330**	***0.0484***
15:15:20:30	1:1:1:1	0.0240	0.0192	**0.0660**	**0.0344**
	1:1:1:36	0.0238	0.0212	0.0772	0.0182
	36:1:1:1	0.0208	0.0192	**0.0664**	0.1328
	1:4:16:36	0.0230	**0.0258**	**0.0298**	0.0178
	36:16:4:1	0.0238	0.0234	**0.0286**	0.1130

4.2.2.3 Skewed Normal Tailed Distribution ($g = 0.5$ and $h = 0$)

In Table 4.7, under a skewed normal tailed distribution, the *AGWMOM* test gave an excellent control of Type I error rate compared to the *AG* test, the *AGMOM* test and the *ANOVA* because all the conditions of the tests falls within the criteria of robustness. Under equal variance condition all the tests are found to be robust. For the *AG* test, it is robust under unequal variance with values 1:1:1:36 regardless of sample size and nature of pairings. While the *AGMOM* test, is not robust under the condition of positive pairing. The *ANOVA* is not robust under unequal variance with balanced sample size. It is also not robust when the natures of pairings are negative.

Table 4.7

Comparison of the Type I Error Rate for the AG, AGMOM, AGWMOM, and the ANOVA Under a Skewed Normal Tailed Distribution for Four Groups Condition

Sample Size	**Equal and Unequal Variances**	***AG***	***AGMOM***	***AGWMOM***	***ANOVA***
20:20:20:20	1:1:1:1	**0.0620**	***0.0436***	***0.0452***	***0.0550***
	1:1:1:36	**0.0620**	***0.0460***	**0.0272**	0.1714
	1:4:16:36	0.0756	***0.0546***	**0.0262**	0.1098
15:15:20:30	1:1:1:1	**0.0596**	***0.0460***	***0.0466***	***0.0508***
	1:1:1:36	**0.0272**	0.0148	***0.0520***	**0.0756**
	36:1:1:1	**0.0602**	***0.0482***	***0.0520***	0.2330
	1:4:16:36	0.0228	0.0102	***0.0550***	***0.0444***
	36:16:4:1	**0.0646**	***0.0560***	***0.0462***	0.1954

4.2.2.4 Skewed Heavy Tailed Distribution (g = 0.5 and h = 0.5)

In Table 4.8, under a skewed heavy tailed distribution, the *AGWMOM* test is discovered to be robust in all the condition tested that makes this test produced a remarkable control of Type I error rate compared to the *AG* test, the *AGMOM* test and the *ANOVA*. The Type I error rate for the *AG* test produced robust values for all balanced conditions. When the sample size is unbalanced the test becomes not robust even when the variances are equal.

However, it still can consider robust for negative pairing with variance ratio 36:1:1:1 and positive pairing with variance ratio of 1:4:16:36. The *AGMOM* test is not robust under equal variance regardless of the sample sizes. Yet, it is robust under positive and negative pairing conditions. For the *ANOVA,* it is robust with equal variance ratio regardless of the sample sizes. The Type I error rate of the *ANOVA* for both positive and negative pairing condition falls outside the criteria of robustness.

Table 4.8

Comparison of the Type I Error Rate for the AG, AGMOM, AGWMOM, and the ANOVA Under a Skewed Heavy Tailed Distribution for Four Groups Condition

Sample Size	**Equal and Unequal Variance**	***AG***	***AGMOM***	***AGWMOM***	***ANOVA***
20:20:20:20	1:1:1:1	**0.0322**	0.0206	**0.0398**	**0.0290**
	1:1:1:36	**0.0320**	0.0220	**0.0326**	0.0880
	1:4:16:36	**0.0336**	**0.0250**	**0.0336**	***0.0512***
15:15:20:30	1:1:1:1	0.3000	0.0190	**0.0274**	**0.0336**
	1:1:1:36	0.3960	**0.0256**	*0.0474*	**0.0240**
	36:1:1:1	**0.0272**	**0.0260**	*0.0466*	0.1394
	1:4:16:36	**0.0360**	**0.0266**	**0.0320**	0.0164
	36:16:4:1	0.0166	**0.0256**	**0.0384**	0.1130

4.2.3 Six Groups

For six groups condition, the results of the Type I error rate are presented as following.

4.2.3.1 Normal Distribution ($g = 0$ and $h = 0$)

In Table 4.9, under a normal distribution, for six group conditions, the *AG* test, the *AGMOM* test and the *AGWMOM* test have the best control of Type I error rate compared to the *ANOVA*. The Type I error rate of the three tests falls within the criteria of robustness in all the balanced conditions. It is no doubt that the *ANOVA* is very good only under perfect condition, which is normal distribution with balanced sample size and equal variance. However, this test is not robust when the variances are not equal. Under unbalanced conditions all the tests are not robust except for *ANOVA* which is still robust as long as the variances are equal.

Table 4.9

Comparison of the Type I Error Rate for the AG, AGMOM, AGWMOM, and the ANOVA, Under a Normal Distribution for Six Groups Condition

Sample Size	**Equal and Unequal Variances**	***AG***	***AG MOM***	***AGW MOM***	***ANOV A***
20:20:20: 20:20:20	1:1:1:1:1:1	***0.0522***	***0.0440***	**0.0402**	***0.0530***
	1:1:1:1:1:36	***0.0522***	***0.0444***	**0.0406**	0.1260
	1:4:4:16:16:36	***0.0572***	***0.0488***	***0.0464***	0.0810
2:4:4:16: 32:62	1:1:1:1:1:1	0.1522	0.1864	0.1796	**0.0640**
	1:1:1:1:1:36	0.1434	0.1698	0.1724	0.0002
	36:1:1:1:1:1	0.1192	0.1432	0.1378	0.5992
	1:4:4:16:16:36	0.0920	0.0872	0.0926	0.0020
	36:16:16:4:4:1	0.1148	0.1454	0.1362	0.6878

4.2.3.2 Symmetric Heavy Tailed Distribution ($g = 0$ and $h = 0.5$)

In Table 4.10, under a symmetric heavy tailed distribution, the *AG* test gave a remarkable control of Type I error rate compared to the *AGMOM* test, the *AGWMOM* test and the *ANOVA*. The *AG* test is seen to be robust in all conditions under balanced sample size. It is also robust under unbalanced sample size with positive nature of pairing. The *AGMOM* test is not robust in all the conditions of balanced sample sizes. This test is robust under positive pairing condition, and also the test is found to be robust under negative with variance of 36:16:16:4:4:1. The *AGWMOM* test has its Type I error rate fall within the interval of robustness, in all the balanced condition but the test is not robust under unbalanced samples. The *ANOVA* falls within the criteria of robustness, under balanced condition, except for variance ratio of (1:1:1:1:1:36). Under unbalanced sample size, this test is only robust for negative pairing with variance value of (36:16:16:4:4:1).

Table 4.10

Comparison of the Type I Error Rate for the AG, AGMOM, AGWMOM, and the ANOVA Under a Symmetric Heavy Tailed Distribution for Six Groups Condition

Sample Size	**Equal and Unequal Variance**	***AG***	***AGMOM***	***AGWMOM***	***ANOVA***
20:20:20:20:20:20	1:1:1:1:1:1	**0.0260**	0.1092	**0.0266**	**0.0350**
	1:1:1:1:1:36	**0.0258**	0.0186	**0.0256**	0.0922
	1:4:4:16:16:36	**0.0248**	0.0216	**0.0288**	***0.0520***
2:4:4:16:32:62	1:1:1:1:1:1	0.0794	0.1092	0.1092	0.0988
	1:1:1:1:1:36	**0.0656**	***0.0450***	0.0896	0.0040
	36:1:1:1:1:1	0.0796	0.0896	0.0982	0.3890
	1:4:4:16:16:36	**0.0348**	***0.0486***	***0.0442***	0.0130
	36:16:16:4:4:1	0.0898	***0.0456***	0.1008	***0.4732***

4.2.3.3 Skewed Normal Tailed Distribution ($g = 0.5$ and $h = 0$)

In Table 4.11, under a skewed normal tailed distribution, the *AGMOM* test and the *AGWMOM* test is more robust compared to the *AG* test and the *ANOVA*. Under balanced condition, all the Type I error rate of the *AGMOM* test and the *AGWMOM* test fall within the interval of robustness. The *AG* test is robust only under equal variance and unequal variance with value of (1:1:1:1:1:36).The *ANOVA* has its Type I error rate fall within the interval criteria of robustness, only under equal variances regardless of the sample sizes.

Table 4.11

Comparison of the Type I Error Rate for the AG, AGMOM, AGWMOM, and the ANOVA Under a Skewed Normal Tailed Distribution for Six Groups Condition

Sample Size	**Equal and Unequal Variances**	***AG***	***AGMOM***	***AGWMOM***	***ANOVA***
20:20:20:20:20:20	1:1:1:1:1:1	**0.0650**	***0.0498***	***0.0456***	***0.0544***
	1:1:1:1:1:36	**0.0728**	***0.0508***	***0.0440***	0.2070
	1:4:4:16:16:36	0.0860	***0.0576***	***0.0514***	0.1184
2:4:4:16:32:62	1:1:1:1:1:1	0.2080	0.1944	0.2118	**0.0670**
	1:1:1:1:1:36	0.2734	0.1692	0.2188	0.0060
	36:1:1:1:1:1	0.1678	0.1600	0.1740	0.5692
	1:4:4:16:16:36	0.2514	0.0880	0.1430	0.0034
	36:16:16:4:4:1	0.1418	0.1636	0.1620	0.6722

4.2.3.4 Skewed Heavy Tailed Distribution (g = 0.5 and h = 0.5)

In Table 4.12, under a skewed heavy tailed distribution with balanced sample size, the *AGWMOM* produced the most convincing results where it is robust in all the conditions. While the *ANOVA* is robust for two conditions and the *AG* test is robust under one condition only. The *AG* test is not robust in all the balanced condition. Under the unbalanced sample sizes the *AG* test and the *ANOVA* is not robust in all the conditions. While the *AGMOM* is robust under one condition and the *AGWMOM* test is robust under two conditions.

Table 4.12

Comparison of the Type I Error Rate for the AG, AGMOM, AGWMOM, and the ANOVA Under a Skewed Heavy Tailed Distribution for Six Groups Condition

Sample Size	Equal and Unequal Variances	*AG*	*AGMOM*	*AGWMOM*	*ANOVA*
20:20:20:	1:1:1:1:1:1	**0.0370**	0.0208	**0.0286**	**0.0330**
	1:1:1:1:1:36	0.0186	0.0186	**0.0292**	0.1028
20:20:20	1:4:4:16:16:36	0.0200	0.0246	**0.0300**	***0.0574***
	1:1:1:1:1:1	0.1212	0.1136	**0.0320**	0.0970
	1:1:1:1:1:36	0.1236	0.0964	0.1028	0.0100
2:4:4: 16:32:62	36:1:1:1:1:1	0.1108	0.0898	0.1036	0.3336
	1:4:4:16:16:36	0.0888	***0.0478***	***0.0524***	0.0200
	36:16:16:4:4:1	0.1044	0.0962	0.1046	0.4090

4.2.4 Overall Conclusion on the Type I Error Rate

As the overall conclusion of the Type I error rate, the number of conditions is counted to see how many conditions of the tests can be considered as stringent robust (SR), only lenient robust (LR – SR) and not robust (NR). It should be noted that the total conditions of each test under each distribution is twenty-one (21). Table 4.13 shows the number of conditions according to the types of distribution.

Table 4.13

Number of Conditions Based on the Type I Error Rates

Distribution	Robustness	*AG*	*AGMOM*	*AGWMOM*	*t-test/ANOVA*
Normal Distribution (g = 0 and h = 0)	SR	16	14	10	5
	LR – SR	0	2	6	5
	NR	5	5	5	11
	Total	21	21	21	21
Symmetric Heavy Tailed Distribution (g = 0 and h = 0.5)	SR	0	3	1	4
	LR – SR	13	7	15	5
	NR	8	11	5	12
	Total	21	21	21	21
Skewed Normal Tailed Distribution (g = 0.5 and h = 0)	SR	5	14	11	6
	LR – SR	8	0	5	3
	NR	8	7	5	12
	Total	21	21	21	21
Skewed Heavy Tailed Distribution (g = 0.5 and h = 0.5)	SR	0	1	5	3
	LR – SR	9	10	9	6
	NR	12	10	7	12
	Total	21	21	21	21
	Grand Total	84	84	84	84

Note: SR = Stringent Robust, LR = Lenient Robust and NR = Not Robust

In Table 4.13, under a normal distribution, for all the group sizes, the *AG* test, the *AGMOM* test and the *AGWMOM* test are more robust compared to the *t*-test and the *ANOVA*. The *AG* test has 16 of its condition fall within the stringent criteria of robustness. None of the conditions of the *AG* test is within the lenient criteria of robustness. The *ANOVA* also performed quite good in the control of Type I error rate with 10 conditions of the tests that are regarded as robust. Among the four tests, the *ANOVA* has the highest number of conditions with the total number of 11 conditions that are considered not robust. The *AG* test has the best control of Type I error rate with 16 conditions of the test that are considered to be robust compared to the other four tests.

Under symmetric heavy tailed distribution, the *AGWMOM* test has the best control of Type I error rate with a total of 16 conditions of the test that are regarded as robust compared to the other three tests. The *ANOVA* has the highest number conditions that are considered not robust compared to the other three tests.

Under a skewed normal tailed distribution, the *AGWMOM* test has the best control of Type I error rate, with a total of 16 conditions of the test that are referred to as robust, compared to the other three tests. None of the conditions of the *AG* test falls within the stringent criteria of robustness. The *ANOVA* has the highest number of conditions that are referred to as not robust compared to the other three tests.

Under skewed heavy tailed distribution, the *AGWMOM* test has the best control of Type I error rate with 14 conditions of the test that are referred to as robust, compared to the *AG* with 9 conditions, *AGMOM* with 11 conditions and *ANOVA* with 9 conditions. The *AGWMOM* test has the highest number of conditions that falls within the stringent criteria of robustness, compared to the other three tests. The *AGMOM* test has the highest number of conditions that are said to be robust, under lenient criteria of robustness, compared to the other three tests.

As the distribution changes from normal to non-normal distribution, the number of conditions of the *ANOVA* for both stringent and lenient criteria of robustness remains the same. That is a total of nine conditions of the *ANOVA* are considered to be robust when the distribution is non-normal. Both the *AG* test and the *ANOVA* have equal number of conditions that are considered not robust, compared to the other two tests.

In overall, the *AGWMOM* test gave the best control of Type I error rate under non-normality, compared to the *AG* test, the *AGMOM* test and the *ANOVA*, because has it always has the highest number of conditions under robust criteria.

4.3 The Power Rate of the Test

In this section, the power of the tests is explained for each of the tests, for each of the four different types of distribution, with the pairing of the sample sizes and variances, positively and negatively, for two, four and six groups respectively.

The power rate of the tests is represented graphically where the *y*-axis corresponds to the power of the tests and the horizontal axis represents the effect size index *d* for two groups case and *f* for more than two group case. The graph is used to show the trend of the power of the tests in relation to the effect size index. According to Murphy and Myors (1998) the power of a test is considered sufficient when it is 0.5. It can be considered to be high when its value is 0.8 and above.

From the graph, it reveals those tests that have low power, sufficient and high power with respect to the effect size indexes (*d* and *f*). In this research there are 84 conditions, denoted by C1 to C84, which can be referenced from Table 3.2, Table 3.3 and Table 3.4, in the research design for two, four and six groups in Chapter 3.

4.3.1 Two Groups Condition under Normal Distribution

In Figure 4.1, the power of the four tests, namely, the *AG* test, the *AGMOM* test, the *AGWMOM* test and the *t*-test is increasing as the effect size index is increasing. The power of the four tests is regarded as sufficient, since their power values are above 0.5. The *AG* test has the highest power in C3 and C4. While the *t*-test, has the highest power in C1, C2 and C5. In C5, despite the fact that the *t*-test has the highest power with value of 0.8004, as reference from appendix B2, the test is regarded as not robust, because its' Type I error rate is outside the criteria of robustness. As a result, the power of the test is referred to as not very good.

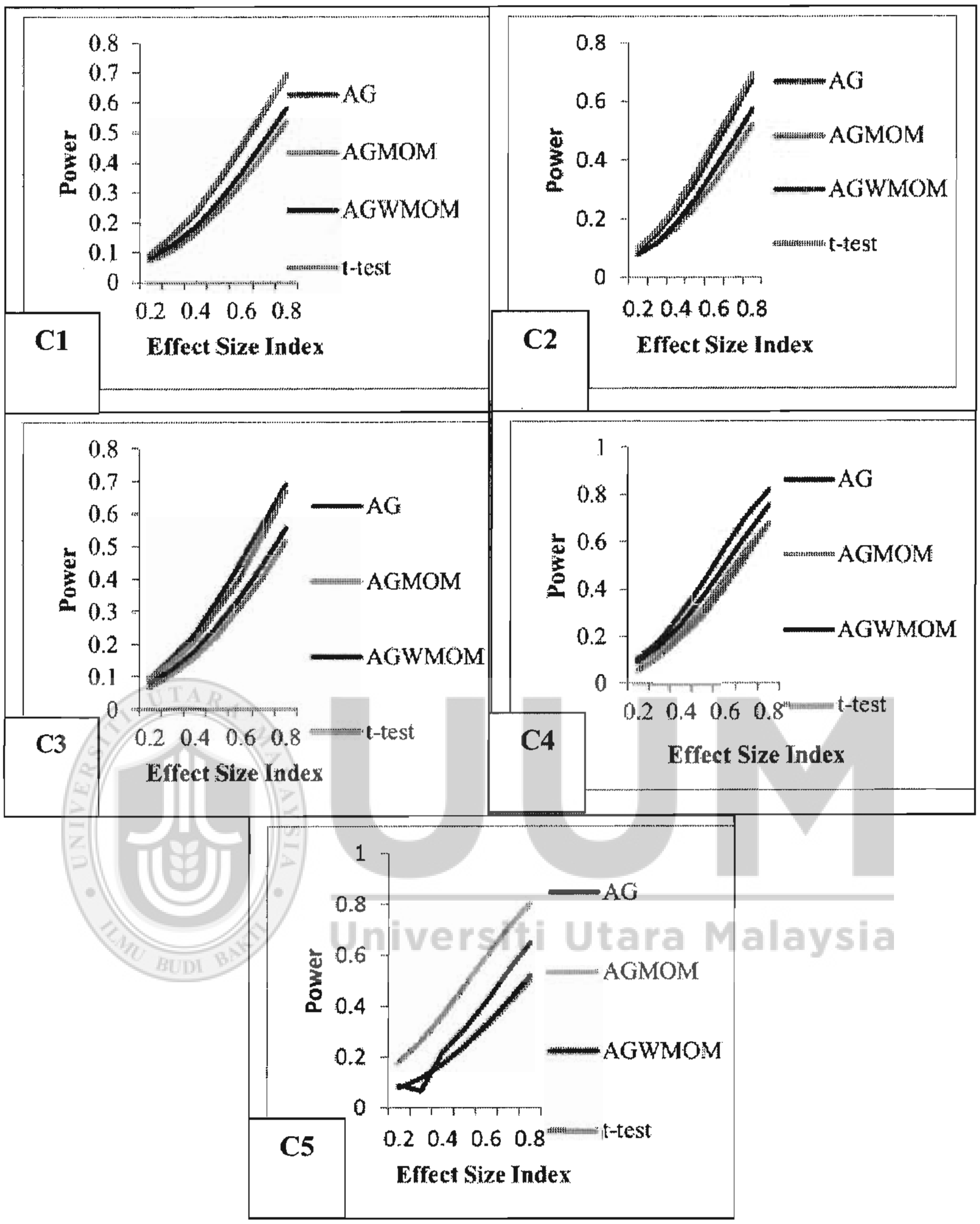

Figure 4.1. Power versus Effect Size Index, for two groups condition under a normal distribution.

4.3.2 Two Groups Condition under Symmetric Heavy Tailed Distribution

In Figure 4.2, from C6 to C10, the power of the four tests is increasing as the effect size index is increasing. In C6, C7, C9 and C10, the power of the four tests are referred to as not sufficient, because their power values is not up to 0.5. Only in C8, there are two tests that achieve 0.5 i.e the *AGMOM* and the *AGWMOM*. The *AGWMOM* has the highest power compared to the other three tests, under symmetric heavy tailed distribution, for two group condition.

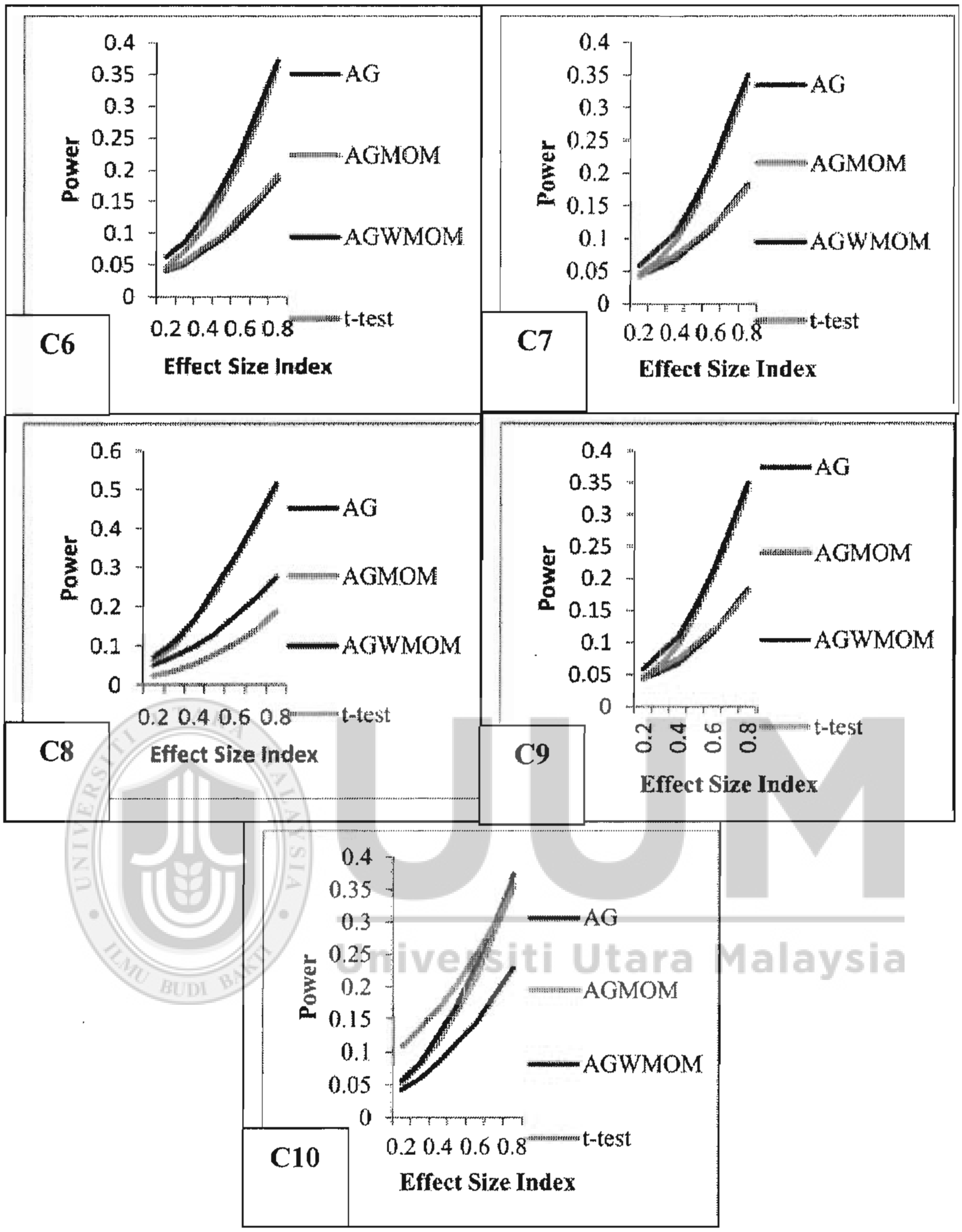

Figure 4.2. Power versus Effect Size Index, for a symmetric heavy tailed distribution, for two groups condition

4.3.3 Two Groups Condition under a Skewed Normal Tailed Distribution

In Figure 4.3, the power of the four tests is displayed according to conditions C11 to C15. The power is increasing as the effect size index is increasing. All the tests are considered not having high power since they did not achieve 0.8 in all conditions, except for the *AG* test which obtained power value of 0.8540 under C15. In C11 and C13, the *AG* and *t*-test are considered having sufficient power, when both tests achieve a power value of 0.5. In C14, only the *AGMOM* test and *AGWMOM* test reach the sufficient value of power.

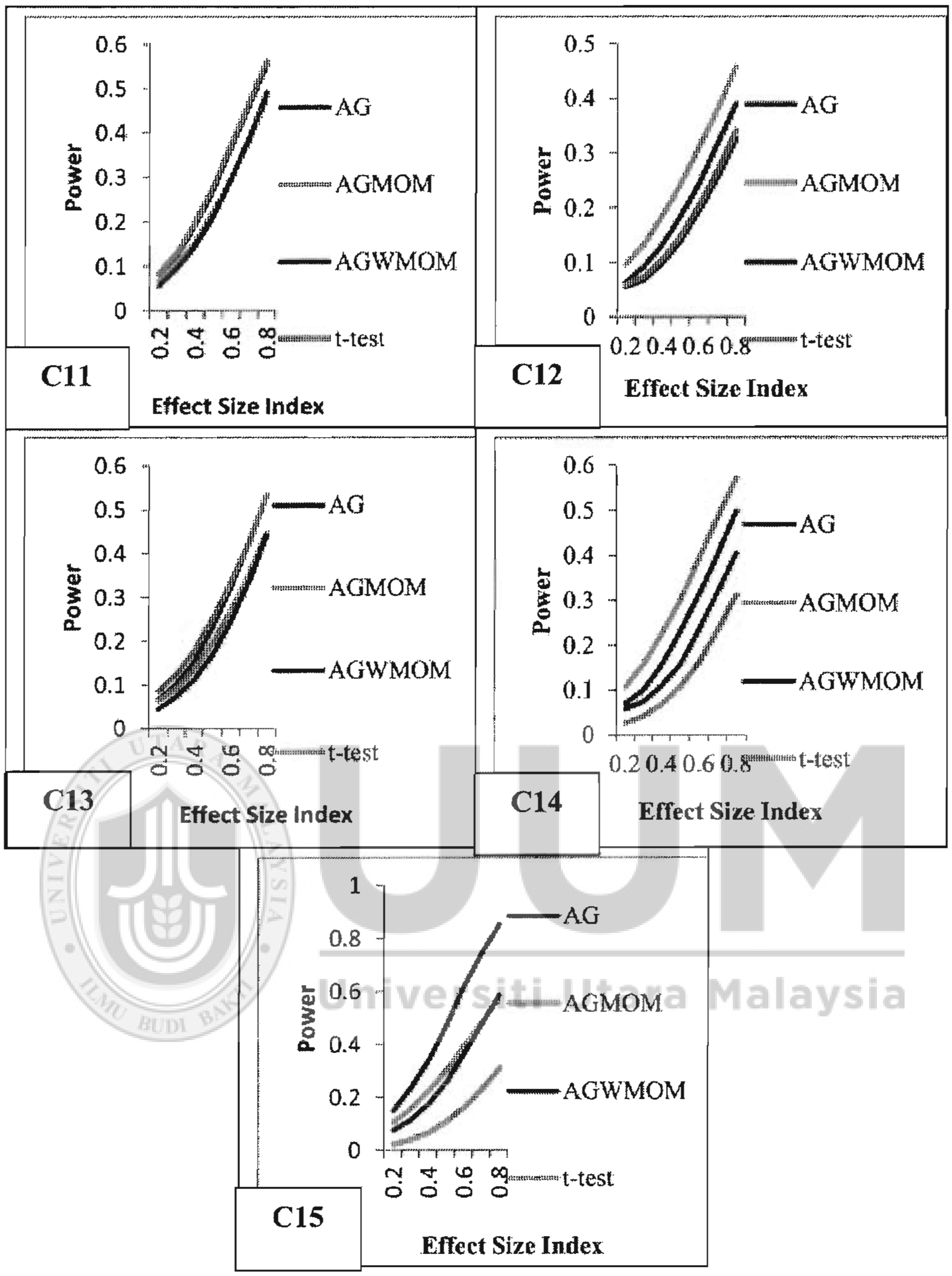

Figure 4.3. Power versus Effect Size Index, for two groups condition under a skewed normal tailed distribution

4.3.4 Two Groups Condition under a Skewed Heavy Tailed Distribution

In Figure 4.4, the power of the four tests is increasing as the effect size index is increasing. In C16 to C20, the power of the four tests is not up to 0.5 and are said to be low and insufficient. In C16, C17, C18 and C19, the *AGMOM* test has the highest power. In C20, the *t*-test has the highest power.

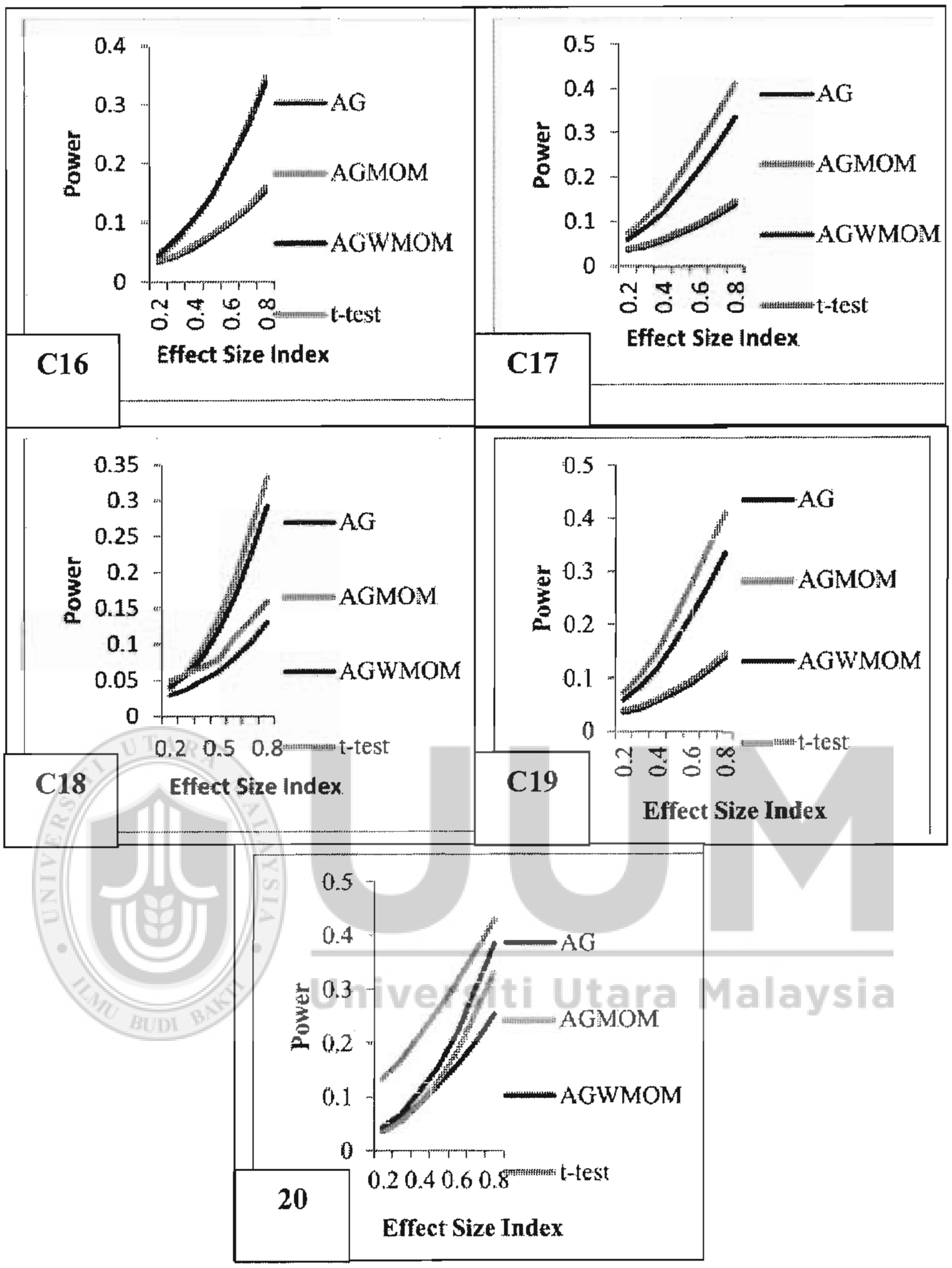

Figure 4.4. Power versus Effect Size Index, for two groups condition, for g = 0.5 and h = 0.5

4.4 The Power of the Four Tests, For Four Groups Condition, Under Four Different Distributions

The power of the *AG* test, the *AGMOM* test, the *AGWMOM* test and the *ANOVA* are examined in four different distributions under four group conditions.

4.4.1 Four Groups Condition under a Normal Distribution

The power values of all the compared tests are displayed in figure 4.5. The power of the *AG* test, the *AGMOM* test, the *AGWMOM* test and the *ANOVA* is increasing as the effect size index is increasing in like manner. All the four tests have sufficient and high power in C21. The *AG* test, *AGMOM* test and the *AGWMOM* test are regarded as having high and sufficient power in C22, C25 and C26. The four tests are observed to have a very low power in C23, C24, C27 and C28 respectively.

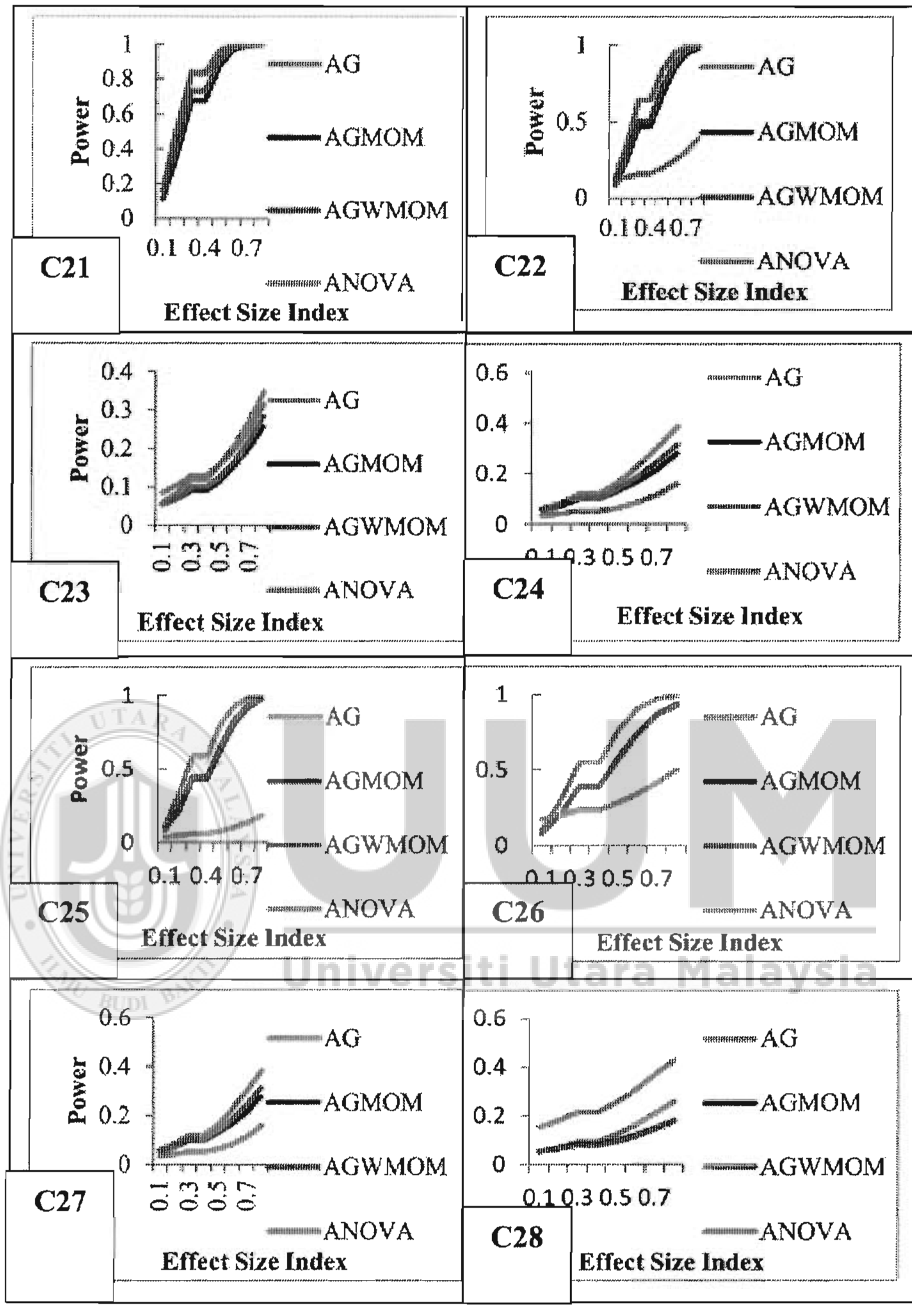

Figure 4.5. Power versus Effect Size Index for Four Groups Under a normal distribution

4.4.2 Four Groups Condition under Symmetric Heavy Tailed Distribution

In Figure 4.6, it can be noticed that the power of the four tests is increasing as the effect size index is increasing in all conditions except in C32. In C32, the power values of the tests are found not to be consistent with the effect size index, due to fact that all the tests are not robust under this condition (refer to Table 4.6). In C29, 30, 31, 33 and 34, the *AGMOM* and the *AGWMOM* have high power where their power values achieve 0.8. In C33, the *AGWMOM* test is observed not to be robust. In C35 and C36, the power of the all the tests is observed to be very low and insufficient.

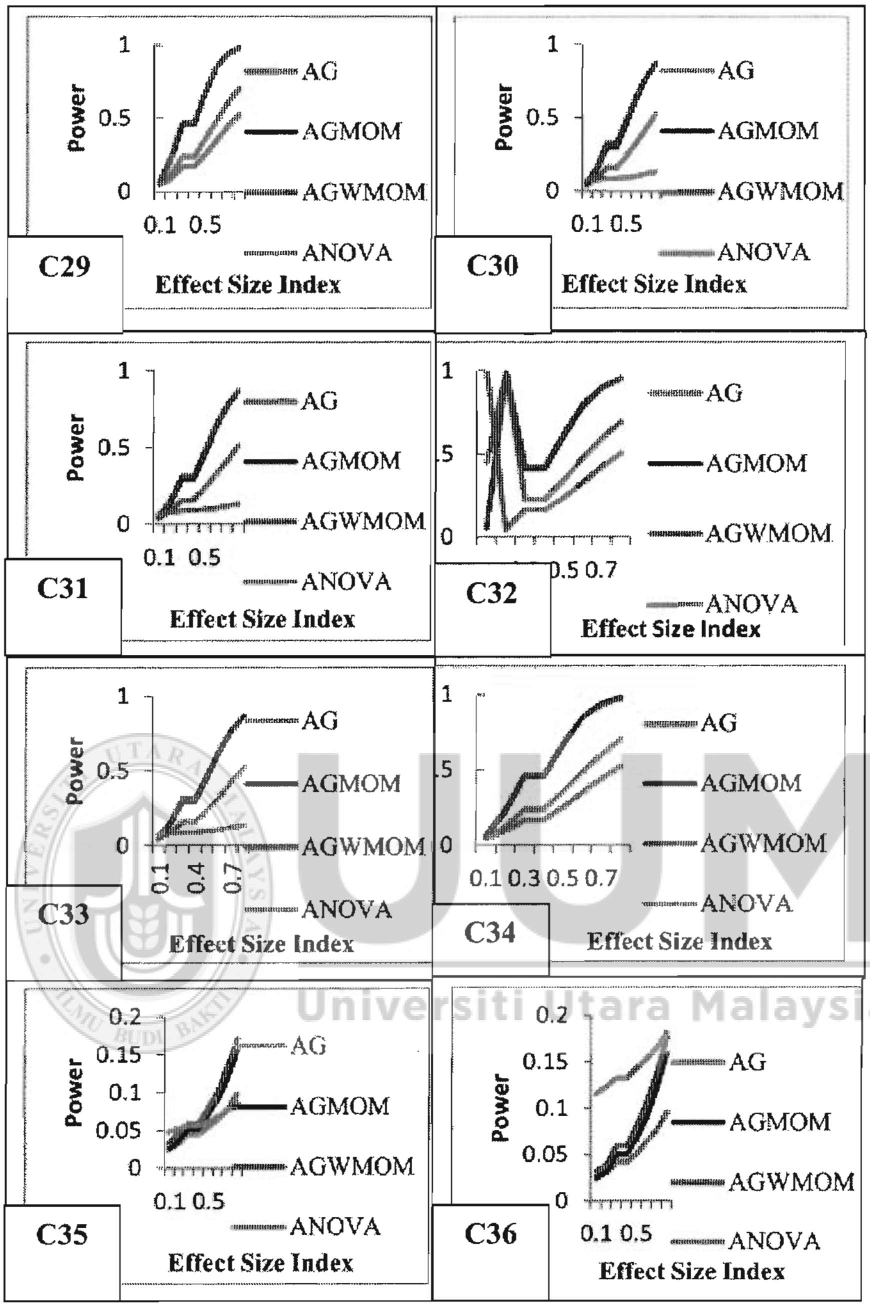

Figure 4. 6. Power against Effect Size Index, for four groups condition, for g = 0 and h = 0.5

4.4.3 Four Groups Condition under a Skewed Normal Tailed Distribution

The four compared test in Figure 4.7, is increasing as the effect size index is increasing. All the four tests have high and sufficient power in C37. The *AG*, *AGMOM* and the *AGWMOM* test produced sufficient and high power in C38, C40 and C42. The *AGWMOM* test and the *ANOVA* have sufficient power in C39. Only the *AG* test has sufficient power in C43. The four tests have very low power in C41 and C44.

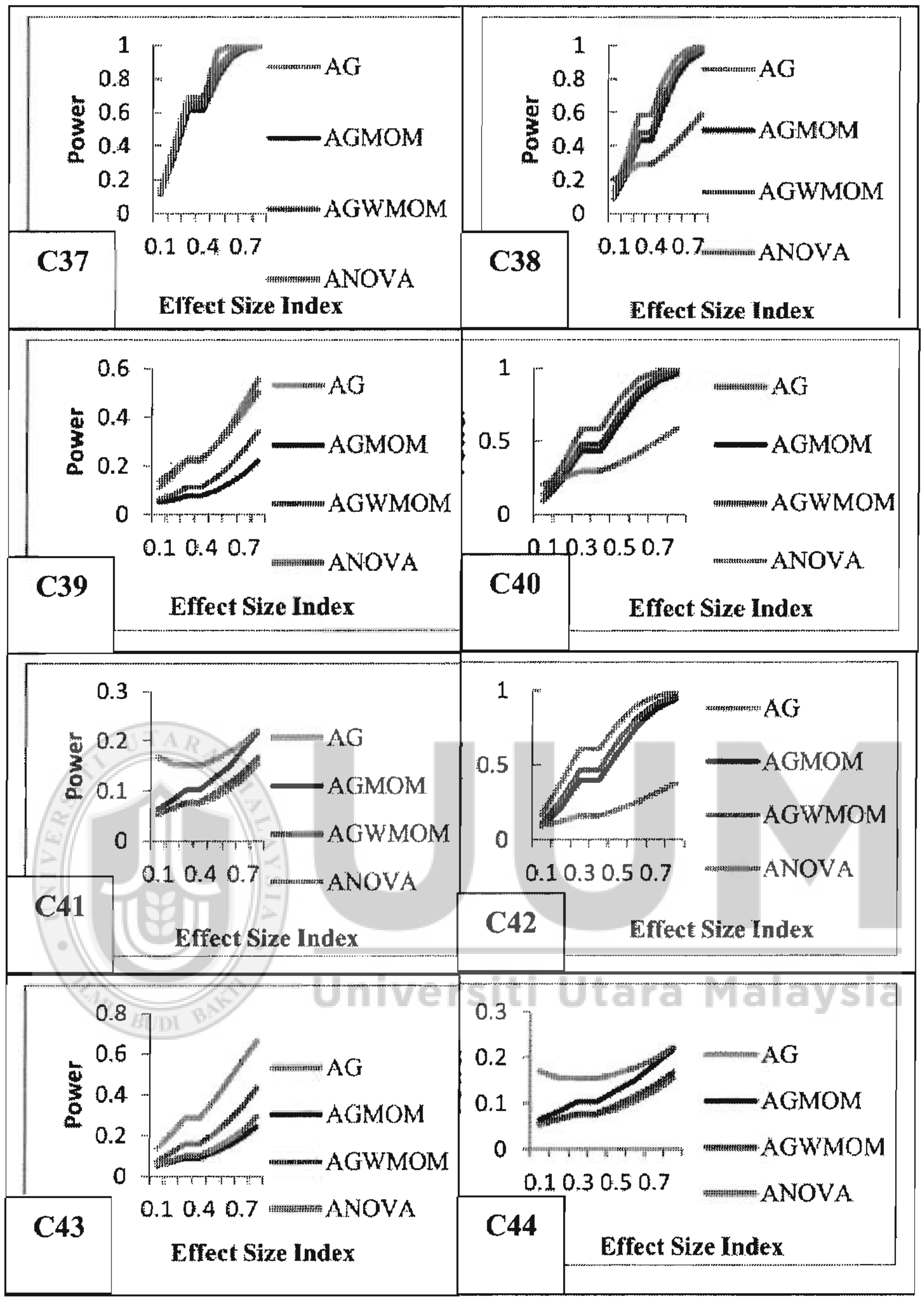

Figure 4.7. Power versus Effect Size Index, for four groups condition, Under a skewed normal tailed distribution

4.4.2 Four Groups Condition, Under a Skewed Heavy Tailed Distribution

The power values of the four tests are increasing as the effect size index is increasing, except in C51, where the *ANOVA* is found to be decreasing as the effect size index is increasing. The *AGMOM* and the *AGWMOM* test produced a high and sufficient power values in C45, C46, C48, C49 and C50. The power of the four tests is considered to be very low in C47, C51 and C52. In C51, the power value of the *ANOVA* is regarded as very low and is the test not robust.

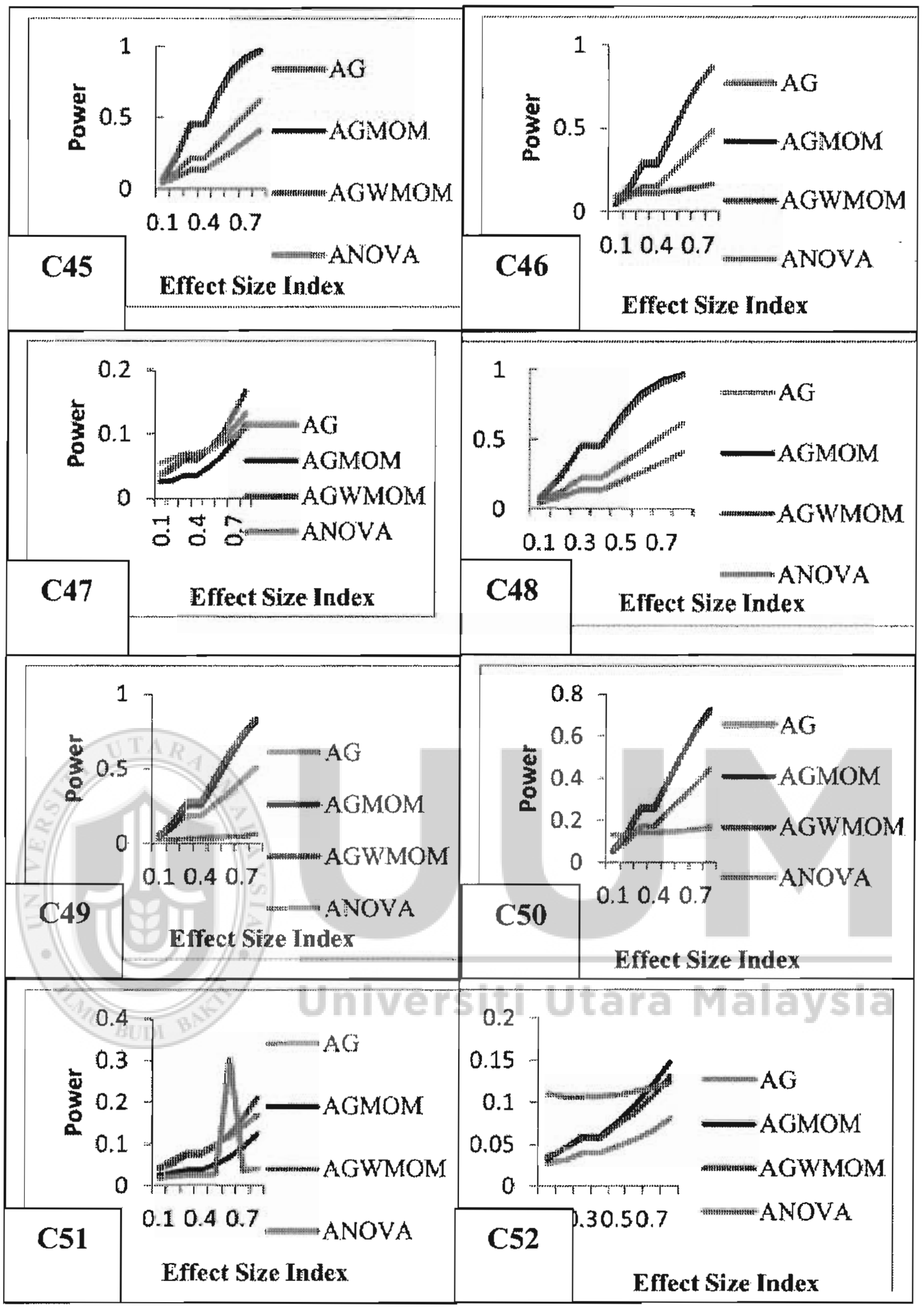

Figure 4.8. Power versus Effect Size Index, under a skewed heavy tailed distribution, for four groups condition

4.5 The Power of the Tests, For Six Groups Condition, Under Four Different Distributions

The power of the four tests, namely, the *AG* test, the *AGMOM* test, the *AGWMOM* test and the *ANOVA* will be investigated under four different distributions, to see of the four tests which one of them will have its power value above the 0.5 and also the test that can produce a power value of 0.8 and above, that can be considered as sufficient and high, under six groups condition.

4.5.1 Six Groups Condition, Under a Normal Distribution

In Figure 4.9, all the compared tests are increasing as the effect size index is increasing accordingly. All the four tests have high and sufficient power in C53 and C55. In C56 and C58, the four tests are observed to have sufficient power. Only the *ANOVA* have a sufficient power in C60. The *AG*, *AGMOM* and the *AGWMOM* test have sufficient power in C57. The power of the four tests is considered to be very low in C59.

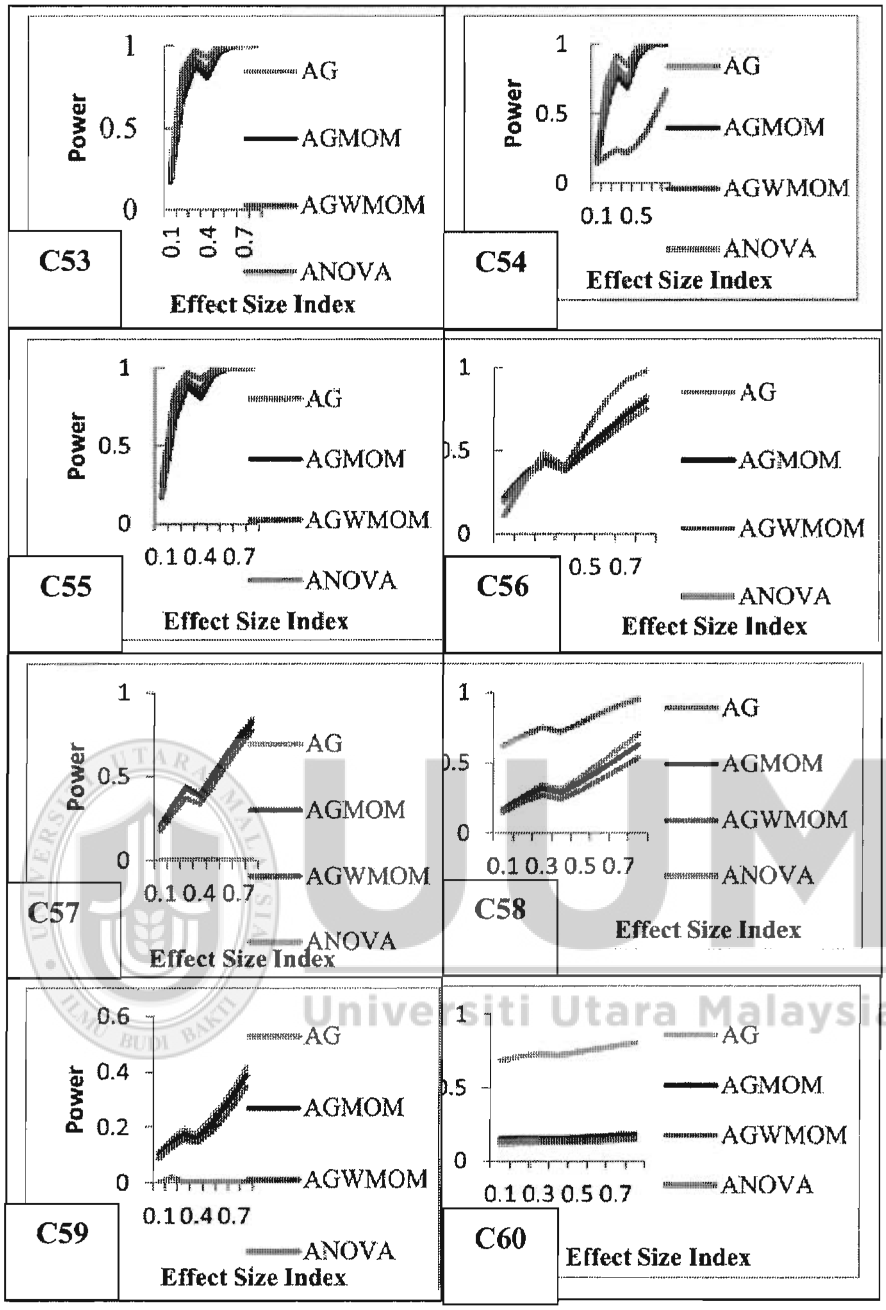

Figure 4.9. Power versus Effect Size Index, for six groups condition under a normal distribution

4.5.2 Six Groups Condition, Under a Symmetric Heavy Tailed Distribution

In Figure 4.10, the power of the four tests is increasing as the effect size index is increasing, except in C68. In C8, the power of the tests are observed not to be consistent with the effect size index, because the *AG* test and the *AWMOM* test are seen not to be robust under this condition (see Table 4.10). In C61, all the four tests achieve a sufficient power. The *AG* test, the *AGMOM* test and the *AGWMOM* test have sufficient power in C62, C64 and C65. In C63, C66 and C67, the power values of the four tests are below 0.5 and are regarded as not sufficient.

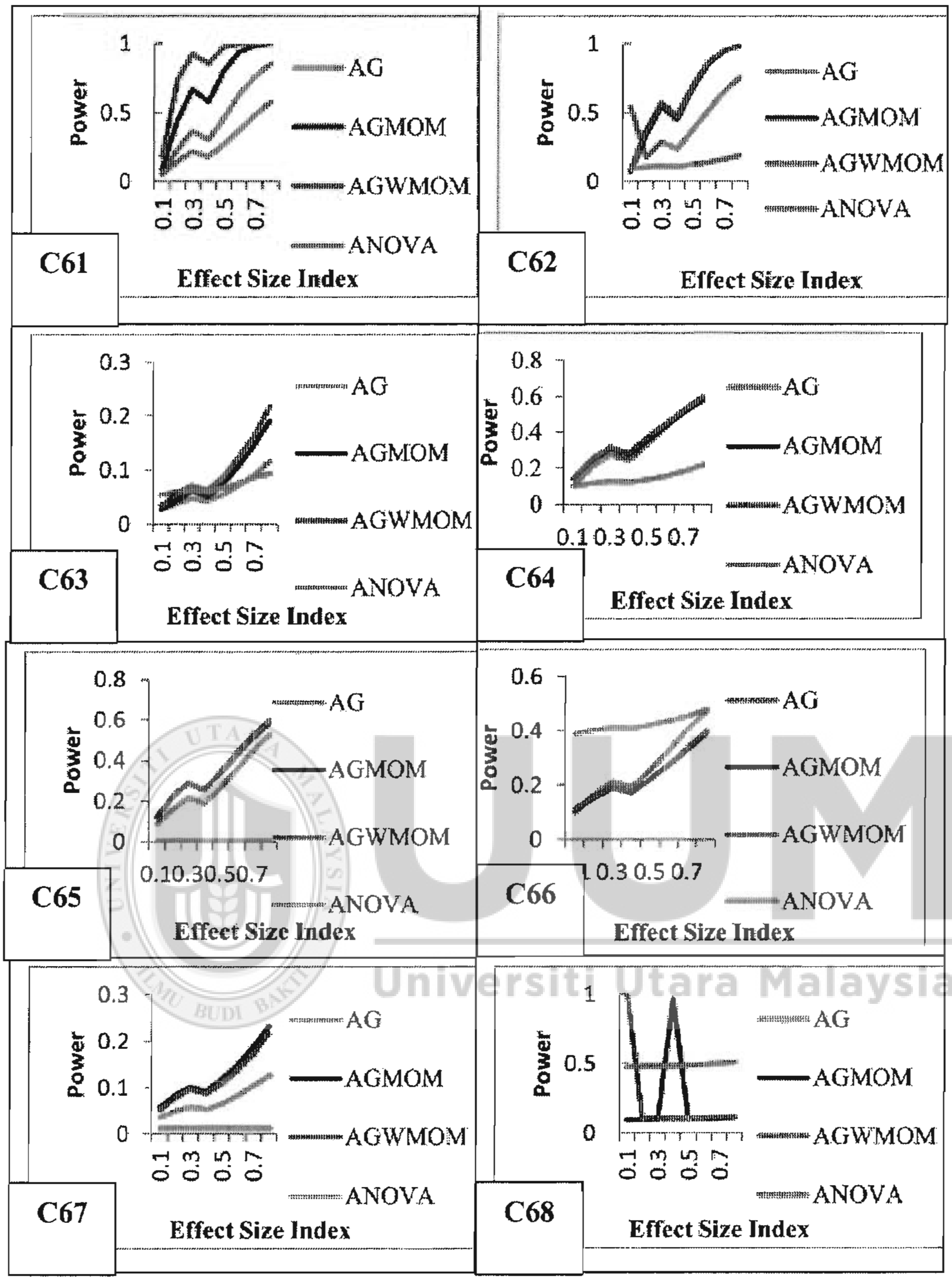

Figure 4.10. Power versus Effect Size Index, for six groups condition, under a symmetric heavy tailed distribution

4.5.3 Six Groups Condition, Under a Skewed Normal Tailed Distribution

In Figure 4.11, the power of the four tests is increasing as the effect size index is increasing accordingly. The power of the four tests is regarded as sufficient and high in C69 and C70. The four tests have sufficient power only in C74. In C75, the *AG* test, the *AGMOM* test and the *AGWMOM* test have sufficient power. Only the *ANOVA* has sufficient power in C76. The power of the four tests is referred to as very low in C73.

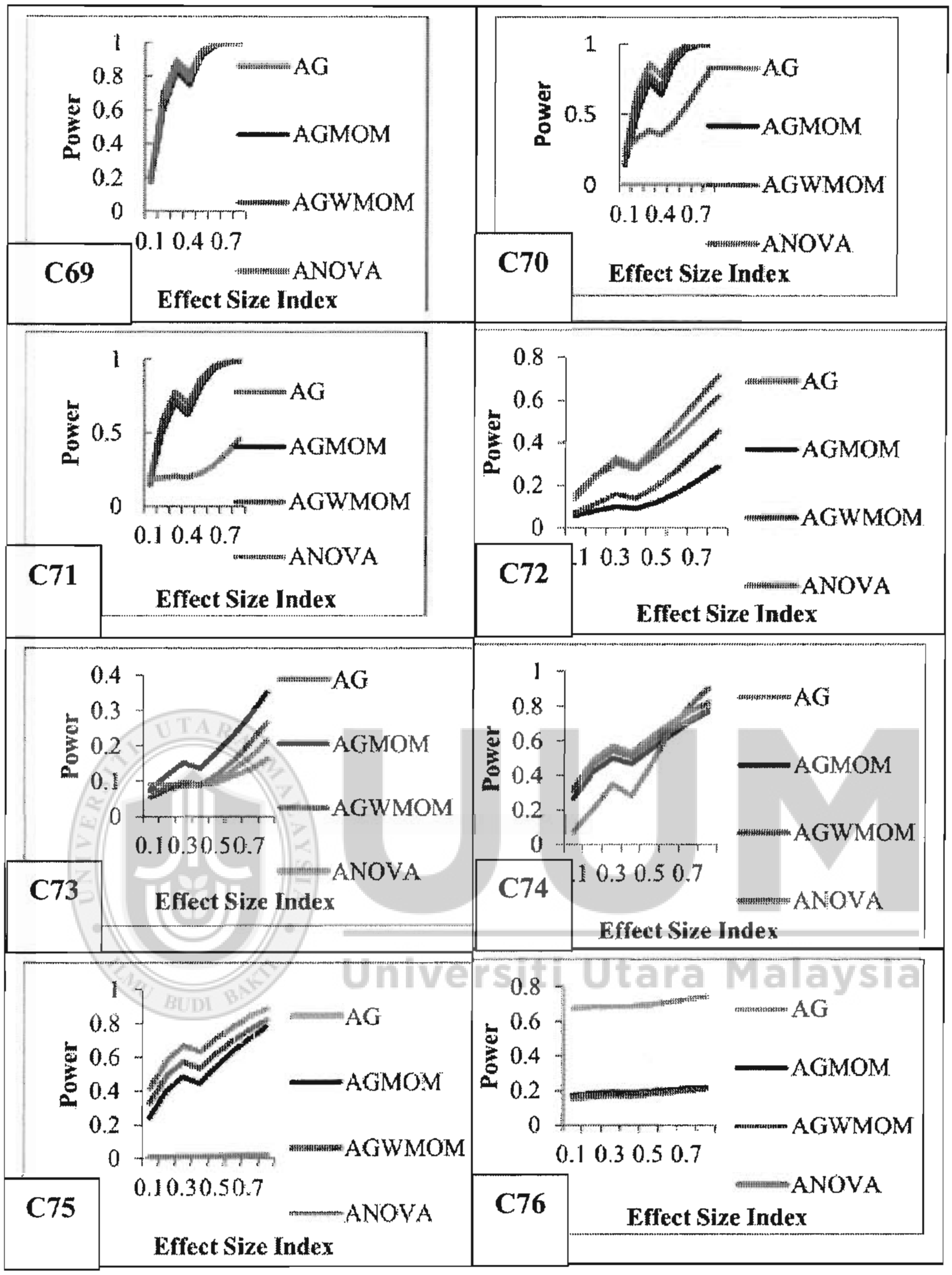

Figure 4.11. Power versus Effect Size Index under a skewed normal tailed distribution, for six groups condition

4.5.4 Six Groups Condition, Under a Skewed Heavy Tailed Distribution

The power of the compared tests in Figure 4.12 is increasing as the effect size index is increasing. The *AGMOM* test and *AGWMOM* test have sufficient and high power in C77 and C78. The *AG*, *AGMOM* and *AGWMOM* test have sufficient power in C77, C78, C80 and C81. Only the *AG* test has sufficient power in C82. The four tests have a very low power in C79, 83 and C84.

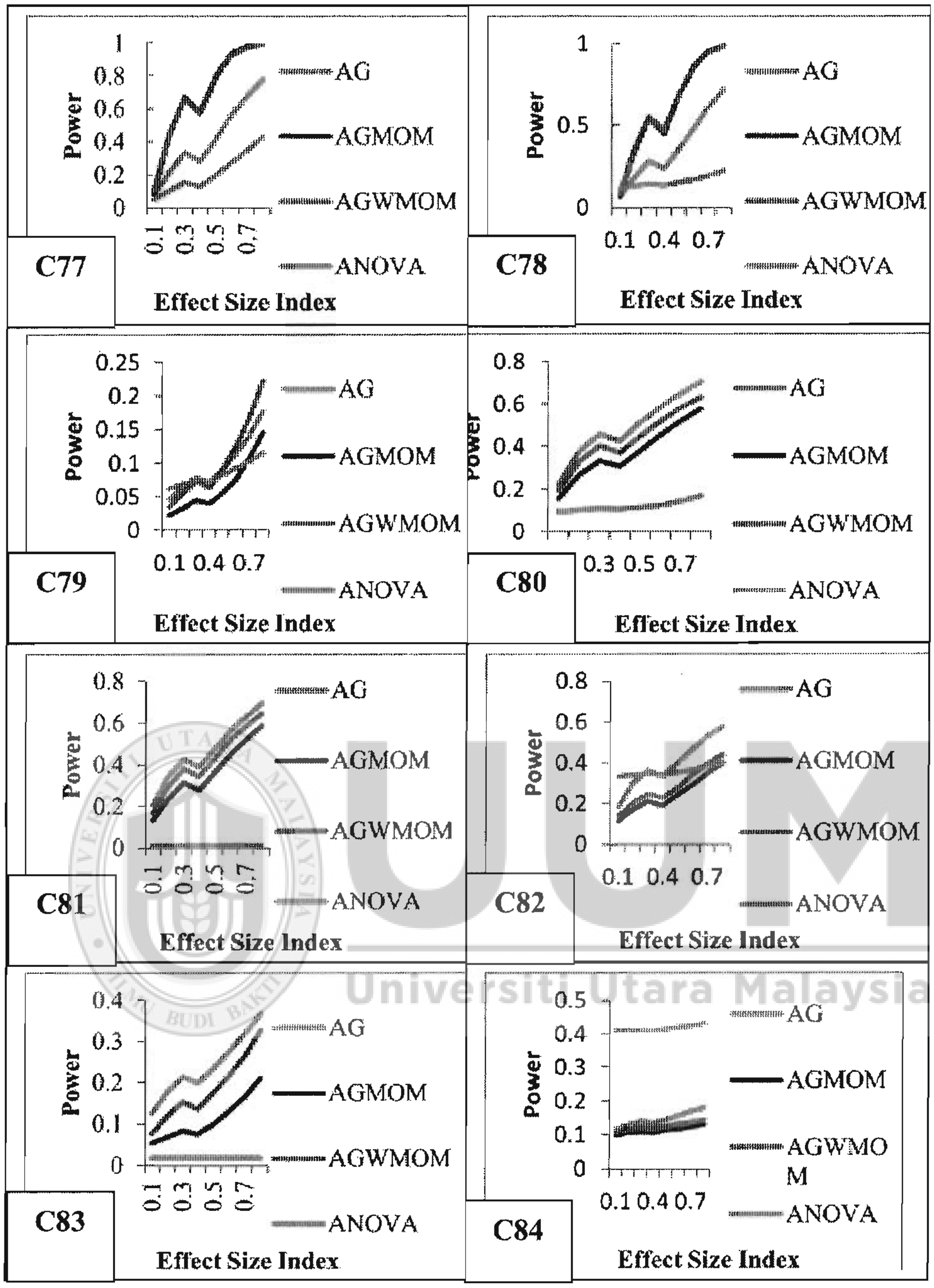

Figure 4.12. Power versus Effect Size Index, for six groups condition, under a skewed heavy tailed distribution

4.6 Evaluating the Capacity of the Test Using Real Data

To fulfill the fourth objective of this research, a real life data was used which was extracted from Keselman *et al.* (2007) that comprises of three independent groups, namely group young, middle and old, see appendix O. The test of homogeneity of the variance was used for the three independent groups, using the Levene's test to determine if the three independent groups are different from each other or not as the reaction time changes.

In this section, test of homogeneity of variances, descriptive statistics, test statistic of AG test and $AGWMOM$ test, and test of normality are performed to show the advantages of each test.

Table 4.13

Test of Homogeneity of Variances

Test of Homogeneity of Variances
Reaction

Levene Statistic	df1	df2	p-value
1.821	2	43	.174

$\alpha = 0.05$

H_O: If there is no difference between the groups

H_1: If there is difference between the groups

If the p-value from the test of homogeneity is less than 0.05, we can reject H_O otherwise failed to reject H_O . When the p-value is > 0.05, we accept H_O and reject H_1. The p-value from the test of homogeneity of the variance, is > 0.05, i.e 0.174 > 0.05, implies that we accept H_O and conclude that there is no difference between the groups as the reaction time changes.

In Table 4.14 below, shows the descriptive statistics for the three independent groups, for the AG test.

Table 4.14

Descriptive Statistics for the Young, Middle and Old Groups using the AG test and the *AGWMOM* test

Test statistic	Descriptive statistic	Young	Middle	Old
AG test	Mean	544.0511	473.6992	571.6813
	Standard error	59.7266	144.6221	49.5377
AGWMOM test	Mean	505.8433	456.8608	551.0392
	Standard error	4.9059	12.1963	6.7518

In Table 4.14, the mean of the three independent groups, namely: the young, middle and old groups, are stated above. The standard errors for the young, middle and old groups are considered to be very high with values 59.7266, 144.6221 and 49.5377 respectively, for the three independent groups. This is as a result of the presence of outliers in the real life data for the *AG* test.

In Table 4.14, the Winsorized mean for the three independent groups, namely: the young, middle and old groups respectively are; 505.8433, 456.8608 and 551.0392 and are observed to be smaller in comparison to the mean for the young, middle and old groups respectively of the *AG* test. The standard errors for the Winsorized young, middle and old groups respectively are: 4.9059, 12.1963 and 6.7518 and are considered to be far smaller compared to the standard error for the young, middle and old groups of the *AG* test in Table 4.15. This is as a result of the elimination of the presence of outliers from the real life data that have been replaced with the preceding values closest to the outlier values from the real life data.

Table 4.15

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	Df	Sig.	Statistic	Df	Sig.
Young	.185	18	.200*	.924	18	.319
Middle	.347	11	.000	.721	11	.001
Old	.199	14	.200*	.935	14	.431

Shapiro-Wilk Test is a test that is frequently used for sample sizes that is less than 50. This test can be used to handle sample size that is more than 2000 (Shapiro & Wilk, 1965). Therefore, the Shapiro-Wilk Test is used to test for the normality of the three independent groups, which are the young, middle and old groups. For the significance level of $\alpha = 0.05$, if the significant value of any of the three independent groups is greater than 0.05, the data is considered to be normally distributed. Otherwise, if the significant value is less than 0.05, the data distribution is regarded as non-normal. The results from Table 4.16 show that the p-value for the group young and old are greater than 0.05, hence both groups are said to be normally distributed. The middle group has a p-value of 0.001 which is less than 0.05 and is regarded as non-normally distributed.

Table 4.16

The statistic test for the AG test and the AGWMOM test

Test	**Test Statistic**	***p*-Value**
Original *AG*	5.3237	0.06982
AGWMOM	30.1280	0.0000002869

In Table 4.16, the results of the test statistics show that the *AG* test has a p-value of 0.0698, that is regarded as not significant because its value is greater than 0.05, while the *AGWMOM* test produced a p-value of 0.0000002869 that is considered to be less than 0.05 and is said to be significant.

In conclusion, the *AGWMOM* test is considered to be more reliable and efficient in minimizing error as much as possible from the real life data, because the test produced a smaller standard error for the three independent groups, namely: the young, middle and old group respectively, in comparison to the *AG* test. Therefore, the performance of the *AGWMOM* test is more efficient and reliability compared to the *AG* test in evaluating the efficient and reliability of the tests using real life data.

CHAPTER FIVE

DISCUSSION AND CONCLUSION

5.1 Summary

The Alexander-Govern test is a test proposed by Alexander-Govern (1994). This test uses mean as its central tendency measure and is considered as a better alternative to the *ANOVA*, the Welch test and the James test, for producing an excellent control of Type I error rate and high power for a normal data under variance heterogeneity. But the Alexander-Govern test is not robust to non-normal data. Researchers such as Lix and Keselman (1998) proposed the trimmed mean in Alexander-Govern test to solve the problem of non-normality for the test.

Wilcox and Keselman (2003) introduced the *MOM* estimator as a measure of the central tendency for the test. Abdullah, Yahaya and Othman (2007) also used the *MOM* estimator as a measure of the central tendency in Alexander-Govern test and it gave them a remarkable control over the probability of Type I error rate for both normal and skewed data distribution. The *MOM* estimator is not influenced by the number of groups. It gave a good control over the probability of Type I error rate under normal and highly skewed condition for all group sizes. But this estimator fails to give a good control over the probability of Type I error rate in an extreme condition of skewness and kurtosis.

In this research, the *AGWMOM* test was applied in Alexander-Govern test to overcome its weakness for non-normal data in an extreme condition of skewness and kurtosis and also under variance heterogeneity.

5.2 Implication and Conclusion

In Table 4.13 as the distribution changes from normal to symmetric heavy tailed distribution, the number of conditions of the *AGWMOM* test decreased from ten to one under stringent criteria of robustness and increased from six to 16 under lenient criteria of robustness. This is as a result of a decrease in the number of conditions of the *AGWMOM* test (from 5 cases to 4 cases) that are considered not robust. When the distribution changes from normal to skewed normal tailed distribution the conditions of the *AGWMOM* test changed, from ten to eleven under stringent criteria of robustness. The number of conditions of the test also changed from six to four under lenient criteria of robustness. The number of conditions of the test that is said not to be robust changed from five to six cases.

It is observed that as the distribution changes from normal to skewed heavy tailed distribution, the conditions of the *AGWMOM* test decreases from ten to five, under stringent criteria of robustness and increased from six to thirteen under lenient criteria of robustness. This is as a result of a reduction in the conditions of the test that are regarded as not robust. This is because the *AGWMOM* test has a high rise in the number of conditions of the test under lenient criteria of robustness, from six to thirteen, that led to a decrease in the number of conditions of the test that is considered not robust.

As the distribution changes from symmetric heavy tailed to skewed heavy tailed distribution, the number of conditions of the *AGWMOM* test increased from one to five under stringent criteria of robustness and decreased from 15 to nine under lenient criteria of robustness. This is as a result of an increase in the number of conditions of the test that is regarded as not robust from five to seven.

A change in the distribution from skewed normal tailed to skewed heavy tailed, shows that the *AGWMOM* test experienced a reduction in the conditions of the test that fall under stringent criteria of robustness, from eleven to five. There is an increase in the conditions of the test from four to thirteen under lenient criteria of robustness and this brought about a reduction in the conditions of the test that are considered not robust, from six to three. It can be seen that when the distribution moves from normal to skewed heavy tailed distribution, the number of conditions of the *AGWMOM* test under stringent criteria of robustness, decreased ten to five. This led to a decrease in the number of conditions of the test that is not robust, from five to three.

Example to illustrate robustness of the *AGWMOM* from normal to skewed heavy tailed distribution is shown in Table 5.1.

Table 5.1

Number of Conditions of AGWMOM test from normal to skewed normal tailed distribution

Robustness	Normal	Skewed heavy tailed
SR	10	5
NR	5	3

5.3 Suggestion and Future Research

In this research, for both stringent and lenient criteria of robustness, the *AGWMOM* test has provided a remarkable control of Type I error rate under skewed normal tailed distribution and skewed heavy tailed distribution, compared to the *AG* test, the *AGMOM* test and the *ANOVA* respectively. But it can be observed that as the distribution changes from normal to skewed normal tailed distribution, the robustness of the test increased from ten to eleven under stringent criteria of robustness and decreased from six to four under lenient criteria of robustness.

Under skewed heavy distribution, the robustness of the *AGWMOM* test reduces under stringent criteria of robustness from ten to five and increases under lenient criteria of robustness from six to thirteen. Future research can be done, to introduce a more robust estimator that can increase the robustness of the test for both stringent and lenient criteria of robustness, as the distribution changes from normal to skewed normal tailed distribution and from normal to skewed heavy tailed distribution respectively.

REFERENCES

Abdullah, S., Yahaya, S. S. S., & Othman, A. R. (2007). Proceedings of The 9th Islamic Countries Conference on Statistical Sciences 2007. In Modified One Step M Estimator as a Central Tendency Measure for Alexander-Govern *Test* P. 834–842.

Abdullah, S, Syed Yahaya, & Othman, A. R. (2008). A Power Investifation of Alexander-Govern Test Using Modified One Step M-Estimator as the Central Tendency Measure. IASC 2008: December 5-8, Yokohama, Japan.

Alexander, R. A., & Govern, D. M. (1994). A New and Simpler Approximation for ANOVA Under Variance Heterogeneity. *Journal Educational Statistics*, *19*(2), 91–101.

Algina, J., Oshima, T. C., & Lin, W.-Y. (1994). Type I Error Rates for Welch's Test and James's Second-Order Test Under Nonnormality and Inequality of Variance When There Are Two Groups. *Journal of Educational and Behavioral Statistics*, *19*(3), 275–291.

Barnett V. and Lewis T. (1994). *Outliersin Statistical Data, 3rd edition*, 584 pp. Chichester, UK: Wiley.

Behrens W. V. (1929). An approximate Degrees of Freedom Soslution to the Multivariate Fisher Problem. *Biometrika Trust*, *68*, 807–837.

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

Brunner, E., Dette, H., & Munk, A. (1997). Box-Type Approximations in Nonparametric Factorial Designs. *Journal of the American Statistical Association*, *92*(440), 1494–1502.

Cohen, J.(1988). Statistical power analysis for the behavioural sciences. New York: Chapman & Hall.

Cribbie, R.A., Wilcox, R. R., Bewell, C., & Keselman, H. J. (2007). Tests for treatment group equality when data are non-normal and heteroscedastic. *Journal of Modern Applied Statistical Methods*, 6, 117-132.

Daniel, W. W. (1990). *Applied nonparametric statistics (2nd ed.).* Boston: PWS-Kent Publishing Company.

Dixon, W. J., & Tukey, J. W. (1968). Approximate Behavior of the Distribution of Winsorized t (Trimming/Winsorization 2). *Technometrics*, *10*(1),83–98.

Efron, B., & Tibshirani. (1998). *An introduction to the bootstrap*. NewYork: Chapman & Hall.

Erceg-Hurn, D. M., & Mirosevich, V. M. (2008). Modern robust statistical methods:an easy way to maximize the accuracy and power of your research.*The American psychologist*, *63*(7), 591–601.

Fisher, R. A. (1935). The Logic of Inductive Inference Author (s): R . A. Fisher Source : Journal of the Royal Statistical Society , Vol . 98 , No.1 (1935), pp. 39-82 Published by: Wiley for the Royal Statistical Society Stable URL: http://www.jstor.org/stable/234243. *Journal of the Royal Statistical Society*, *98*(1), 39–82.

Glass, G. V, & Sanders, J. R. (1972). Consequences of Failure to Meet Assumptions Underlying the Fixed Effects Analyses of Variance and Covariance Author (s): Gene V . Glass , Percy D. Peckham and James R. Sanders Published by: American Educational Research Association Stable URL : http:, *42*(3), 237–288.

Golinski, C., & Cribbie, R. A. (2009). The expanding role of quantitative methodologistsin advancing psychology. *Canadian Psychology/Psychologie canadienne*, *50*(2), 83–90. *Consulting and Clinical Psychology*, *68*(1), 155–165.

Guo, J., & Luh, W. (2000). An invertible transformation two-sample trimmed t -statistic under heterogeneity and nonnormality, *49*, 1–7.

Guo, J. H., & Luh, W. M. (2000). Testing Methods for the One-Way Fixed Effects ANOVA Models of Log-Normal Samples. *Journal of Applied Statistics,* 27(6), 731-738.

Hampel, F. R., Ronchetti, E.M., Rousseeuw, P.J., and Stahel, W.A. (1986). *Robust Statistics: The Approach Based on Influence Functions,* 502 pp. New York: Wiley. [Gives a survey of robust statistical techniques.].

Harwell, M. R., Rubinstein, E. N., Hayes, W. S., & Olds, C. C. (1992). Summarizing Monte Carlo Results in Methodological Research : The One- and Two-Factor Fixed Effects ANOVA Cases. *Journal of Educational Statistics*, *17*(4), 315–339.

Hasings, C., Monsteller, F., Tukey, J. W., & Winsor, C. P. (1947). Low moments for small sample: a comparative study of order statistics. *Annals of Mathematical Statistics*, *18*, 413–426.

He, X., Simpson, D.G., & Portnoy, S.L. (1990). Breakdown robustness of tests. *Journal of the Simulation and computation,* B8(4), 446 - 452.

Huber, P.J. (1964). Robust estimation of a location parameter.*The Annals of Mathematical Statistics* 35, 73-101.

James, G. S. (1951). Variances are Unknown when the ratios of the population variances, *38*(3/4), 324–329.

Keselman, H. ., Wilcox, R. ., Lix, L. M., Algina, J., & Fradette, K. (2007). Adaptive robust estimation and testing. *British Journal of Mathematical and Statistical Psychology*, *60*, 267–293.

Keselman, H.J., Wilcox, R. R., Algina, J., Othman, A.R. (2004). A Power Comparison Of Robust Test Statistics Based On Adaptive Estimators. *Journal of Modern Applied Statistical Methods.*

Keselman, H. J., Kowalchuk, R. K., Algina, J., Lix, L. M., & Wilcox, R. R. (2000). Testing treatment effects in repeated measure designs: Trimmed means and bootstrapping. *British Journal of Mathematical and Statistical Psychology*, *53*, 175–191.

Keselman, H. J., Wilcox, R. R., Taylor, J., & Kowalchuk, R. K. (2000). Tests for mean equality that do not require homogeneity of variances: Do they really work? *Communications in Statistics: Simulation and computation*, *29*, 875–895.

Keselman, H.J., Wilcox, R.R., Othman, A.R., & Fradette, K. (2002). Trimming, Transforming statistics, and bootstrapping: Circumventing the biasing effects of heteroscedasticity and nonnormality. *Journal of Modern Applied Statistical Methods,* 1, 288-309.

Keselman, J. J. C. and H. J. (1982). Parametric Alternatives to the Analysis of Variance Author (s): Jennifer J . Clinch and H . J . Keselman Source : Journal of Educational Statistics , Vol . 7 , No . 3 (Autumn , 1982), pp . 207-214 Published by : American Educational Research Associati. *Journal of Educational Statistics*, *7*(3), 207–214.

Kohr, R. L., & Games, P. A. (1974). Robustness of the analysis of variance, the Welch procedure, and a Box procedure to heterogeneous variances. *Journal of Experimental Education*, *43*, 61–69.

Krishnamoorthy, K., Lu, F., & Mathew, T. (2007). A parametric bootstrap approach for ANOVA with unequal variances: Fixed and random models. *Computational Statistics & Data Analysis*, *51*(12), 5731– 5742. doi:10.1016/j.csda.2006.09.039.

Kulinskaya, E., Staudte, R. G., & Gao, H. (2003). Power Approximations in Testing for Unequal Means in a One-Way ANOVA Weighted for Unequal Variances. *Communications in Statistics - Theory and Methods*, *32*(12), 2353–2371.

Lix, L M, & Keselman, H. J. (1998). To trim or not to trim. *Educational and Psychological Measurement*, *58*(3), 409–429.

Lix, L M, Keselman, J. C., & Keselman, H. J. (1996). Consequences of assumption violations revisited: A quantitative review of alternatives to the one-way analysis of variance F test. *Review of Educational Research*, *66*, 579–619.

Lix, Lisa M., & Keselman, H. J. (1995). Approximate degrees of freedom tests: A unified perspective on testing for mean equality. *Psychological Bulletin*, *117*(3), 547–560. doi:10.1037//0033-2909.117.3.547.

Luh, W. M. (1999). Developing trimmed mean test statistics for two-way fixed-effects ANOVA models under variance heterogeneity and nonnormality. *Journal of Experimental Education*, *67*(3), 243–265.

Luh, W. M., & Guo, J. H. (2005). Heteroscedastic test statistics for one-way analysis of variance: The trimmed means and Hall's transformation conjunction. *The Journal of Experimental Education*, *74*(1), 75–100.

Marascuilo, L. A., & McSweeney, M. (1977). *Nonparametric and distribution-free methods for the social sciences*. Monterey, CA: Brooks/Cole Publishing Company.

Md Yusof, Z., Abdullah, S., Syed Yahaya, S. S., & Othman, A. R. (2011). Type I Error Rates of Ft Statistic with Different Trimming Strategies for Two Groups Case. *Modern Applied Science*, *5*(4).

Murphy, K.R., & Myors, B. (1998). Statistical power analysis: A simple and general model for traditional and modern hypothesis tests. Mahwah, NJ: Lawrence Erlbaum.

Myers, L. (1998). Comparability of The James' Second-Order Approximation Test and The Alexander and Govern A Statistic for Non-normal Heteroscedastic Data. *Journal of Statistical Simulation Computational*, *60*, 207–222.

Oshima, T. C., & J Algina. (1992). Type I error rates for James's second- order test and Wilcoxon's Hm test under heteroscedasticity and non- normality. *British Journal of Mathematical and Statistical Psychology*, *45*, 255–263.

Othman, A. R., Keselman, H. J., Padmanabhan, A. R., Wilcox, R. R., & Fradette,K. (2004). Comparing measures of the "typical" score across treatment groups. *The British journal of mathematical and statistical psychology*, *57*(Pt 2), 215–234.

Pardo, J. A., Pardo, M. C., Vicente, M. L., & Esteban, M. D. (1997). A statistical information theory approach to compare the homogeneity ofseveral variances. *Computational Statistics & Data Analysis*, *24*(4), 411–416.

Scheffe, H. (1959). *The analysis of variance*. New York: Wiley.

SAS Institute Inc. (1999). *SAS/IML User's Guide version 8.* Cary, NC: SAS Institute Inc.

Schrader, R. M.., & Hettmansperger, T.P. (1980). Robust analysis of variance.*Biometrika,* 67, 93-101.

Schneider, P. J., & Penfield, D. A. (1997). Alexander and Govern's Approximation: Providing an alternative to ANOVA Under Variance Heterogeneity. *Journal of Experimental Education*, *65*(3), 271–287.

Shapiro, S. S., and Wilk, M.B. (1965). An analysis of variance test for normality complete samples, *Biometrika* 52, 591 – 611.

Sprent, P. (1993). Applied nonparametric statistical methods (2nd.). London: Chapman & Hall.

Staudte, R. G., & Sheather, S. J. (2011). *Robust estimation and testing*. New York: Wiley.

Syed Yahaya, S.S., Othman, A. R. & Keselman, H. J. (2006). Comparing the" Typical Scores" Across Independent Groups Based on Different Criteria for Trimming. *Metodoloski*.

Tukey, J. (1977). *Explanatory Data Analysis.* (M. Reading, Ed.). Addison-Wesley.

TUKEY, J. W., & McLAUGHLIN, D. (1963). Less Vulnerable Confidence and Significance Procedures for Location Based on a Single Sample: TRIMMING /WINSORIZATION 1. *The Indian Journal of Statistic*, *25*(3).

Welch, B. L. (1951). On the comparison of several means: An aternative approach. Biometrica, 38, 330–336.

Wilcox, R R. (1988). A new alternative to the ANOVA F and new results on James's second-order method. *British Journal of Mathematical and Statistical Psychology*, *41*, 109–117.

Wilcox, R R. (1997). *Introduction to robust estimation and hypothesis testing*. San Diego, CA: Academic Press.

Wilcox, R.R. (2002). Understanding the practical advantages of modern ANOVA methods *Journal of Clinical Child and Adolescent Psychology,* 31, 399-412.

Wilcox RR, Keselman HJ (2002). Power analysis when comparing trimmed means. J. Modern Appl. Stat. Methods, 1(1): 24-31.

Wilcox, R.R., & Keselman, H. J. (2003). Modern Robust Data Analysis Methods: Measures of Central Tendency. *Psychological Methods,* 8(3), 254-274.

Wilcox, R R, Charlin, V. L., & Thompson, K. L. (1986). New Monte Carlo results on the robustness of the ANOVA F, W, and F statistics. *Communications in Statistics-Simulation*, *15*, 933–943.

Wilcox, R. R. (1995). ANOVA: A paradigm for low power and misleading measures of effect size? *Review of Educational Research*, *65*, 51–77.

Wilcox, R.R. (1994). A one-way random effects model for trimmed. *Psychometrika*, *59*(3), 289–306.

Wilcox, R. R. (1997). Introduction to robust estimation and hypothesis testing. San Diego, CA:Academic Press.

Wilcox, Rand R. (2002). Understanding the practical advantages of modern ANOVA methods. *Journal of clinical child and adolescent psychology :53,31*(3), 399–412.

Wilcox, Rand R, & Keselman, H. J. (2003). Modern Robust Data Analysis Methods : Measures of Central Tendency.*Psychological Methods*, *8*(3), 254–274.

Wilcox, R. R. & Keselman, H.J. (2003a). Modern robust data analysis methods: Measures of central tendency. *Psychological Methods,* 8, 254 - 274.

Wilcox, R.R. ., & Keselman, H.J. (2003b). Repeated measures one-way ANOVA based on modified one-step M-estimator.*British Journal of Mathematical and Statistical Psychology,* 56, 15 - 25.

Yahaya, S. S. S., Othman, A. R., & Keselman, H. J. (2006). Comparing the "Typical Score" Across Independent Groups Based on Different Criteria for Trimming, *3*(1), 49–62.

Yang, K., Li, J., & Gao, H. (2006). *The impact of sample imbalance on identifying differently expressed genes.* BMC Bioinformatics, 79(Suppl 4), S8.

Yusof, Z., Othman, A.R. & Yahaya, S.S.S. (2008). TYPE I ERROR RATES OF T1 STATISTIC. International J. of Math. Sci. & Engg. Appls. (IJMSEA). 2(4), 305-312.

Yusof, Z., Abdullah, S. & Yahaya, S.S.S. (2011). Type I Error Rates of F_tStatistic with Different Trimming Strategies for TWO Groups Case.Modern Applied Science, Vol. 5, No. 4; August 2011.

Yusof Z, Abdullah, S, Yahaya, S.S.S, Othman, R.A (2011). Testing the equality of central tendency measures using various trimming strategies.African Journal of Mathematics and Conputer Science Research, Vol. 4(1), pp. 32-38, January 2011.

Yusof, Z., Abdullah S., Yahaya S.S.S. & Othman A.R.(2012). TYPE I ERROR AND POWER RATES OF F_t STATISTIC WITH TRIMMED MEAN. Far East Journal of Mathematical Sciences (FJMS). 69(1), 37-50.