

The Modification and Evaluation of the Alexander-Govern Test in Terms of Power

Tobi Kingsley Ochuko¹, Suhaida Abdullah¹, Zakariah Binti Zain¹ & Sharipah Soaad Syed Yahaya¹

¹ College of Arts and Sciences, Universiti Utara Malaysia, Malaysia

Correspondence: Tobi Kingsley Ochuko, College of Arts and Sciences, Universiti Utara Malaysia, Malaysia.
E-mail: tobikingsley@rocketmail.com

Received: July 19, 2015

Accepted: September 2, 2015

Online Published: October 13, 2015

doi:10.5539/mas.v9n13p1

URL: <http://dx.doi.org/10.5539/mas.v9n13p1>

Abstract

This study centres on the comparison of independent group tests in terms of power, by using parametric method, such as the Alexander-Govern test. The Alexander-Govern (*AG*) test uses mean as its central tendency measure. It is a better alternative compared to the Welch test, the James test and the *ANOVA*, because it produces high power and gives good control of Type I error rates for a normal data under variance heterogeneity. But this test is not robust for a non-normal data. When trimmed mean was applied on the test as its central tendency measure under non-normality, the test was only robust for two group condition, but as the number of groups increased more than two groups, the test was no more robust. As a result, a highly robust estimator known as the *MOM* estimator was applied on the test, as its central tendency measure. This test is not affected by the number of groups, but could not control Type I error rates under skewed heavy tailed distribution. In this study, the Winsorized *MOM* estimator was applied in the *AG* test, as its central tendency measure. A simulation of 5,000 data sets were generated and analysed on the test, using the *SAS* package. The result of the analysis, shows that with the pairing of unbalanced sample size of (15:15:20:30) with equal variance of (1:1:1:1) and the pairing of unbalanced sample size of (15:15:20:30) with unequal variance of (1:1:1:36) with effect size index ($f = 0.8$), the *AGWMOM* test only produced a high power value of 0.9562 and 0.8336 compared to the *AG* test, the *AGMOM* test and the *ANOVA* respectively and the test is considered to be sufficient.

Keywords: Alexander-Govern test, robust, trimmed mean, *MOM* estimator, Type I error rates, *AGWMOM* test and power

1. Introduction

The independent group tests in statistics are about comparing the equality of independent groups via parametric or non-parametric approach. In parametric approach, the analysis of variance (*ANOVA*) is favourable by researchers due to its powerful detection for any differences in the means. The ability of the *ANOVA* to control Type I error rates 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; Scheffé, 1959). Ironically, in real life data, the heterogeneity of variances is very obvious; for instance in behavioural sciences, it is a common practice for researchers to work with population variances that are not equal in a data distribution (Erceg-Hurn & Mirosevich, 2008; Golinski & Cribbie, 2009; Grissom, 2000; Keselman, Kowalchuk, Agina, Lix, & Wilcox, 2000).

According to Yusof, Abdullah, Yahaya and Othman (2011) discovered that the *ANOVA* is badly affected by the appearance of heterogeneity of the variance and non-normality in a data distribution. Because of this, the Type I error rates is increased and the power of the test starts to reduce. As discussed by Wilcox and Keselman (2002) the standard error of the mean can be seriously affected when the distribution of the data is heavy tailed. As a result, the standard error of the *ANOVA* becomes bigger than it supposed to, and the power of the test is decreased. In order to obtain a good test, Type I error rates must be well controlled and also the power of the test must not be decreased. That is, neither should there be, an increase in the Type I error rates nor should there be a lost in the power efficiency of the test.

To give solution to the presence of heterogeneity of the variance, reliable alternative techniques such as the James (1951) and the Welch (1951) have been provided. Past researchers have shown that the original James tests and the Welch tests are commonly robust when there is heterogeneity of the variance and at the same time, it also has a good performance when there is a big difference in the sample sizes (Kohr & Games, 1974; Krishnamoorthy, Lu & Matthew, 2007). Another alternative technique to handle the problem of variance heterogeneity is proposed by Alexander-Govern (1994). Their method produced a remarkable control of Type I error rates and high power efficiency under normality condition. It is considered as a better alternative to the *ANOVA*, due to its easy of computation. These robust methods give a better solution to the classical test statistics, that is the *ANOVA* when the assumptions of homogeneity of the variance is violated, but these tests have failed to handle the problem of non-normality in a data distribution.

2. 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 did remarkably well compared to the *ANOVA* in the control of Type I error rates and produces high power efficiency for a normal data under variance heterogeneity (Alexander-Govern, 1994). Additionally, Schneider and Penfield (1997) reported that the Alexander-Govern test is a good alternative to the *ANOVA* under variance heterogeneity compared to the Welch test and the James test due to its simplicity in calculation and possessing good control of Type I error rates for a normal data distribution. It also produces high level of power under most experimental situations, referring to different levels of examination, when the test was applied in a data distribution, in order to identify its effectiveness in a data distribution. However, under the condition of heterogeneity of the variances, it was recommended for only normal data, but not robust for non-normal data, as discussed by Myer (1998). Other approach that is usually chosen by statistical practitioners when there is non-normal data 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 defines the given distribution for which the test is to be used. It is used to estimate 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, these approaches are not as sensitive as using parametric test when the basic assumptions of the parametric test are fulfilled. Hence, larger differences are required before a rejection of the null hypothesis is performed. In other situation, non-parametric approaches also need a large number of sample sizes to prevent loss of information. Examples of non-parametric test are, Friedman test, Mann-Whitney U test, Wilcoxon Signed-Rank 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 weakness observed in using the non-parametric tests, researchers discovered the use of robust estimators as a better alternative when dealing with non-normal data. Robust estimator that is commonly used in improving the independent group test is the trimmed means. 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 of Type I error rates, 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 loss of information.

An alternative to the use of trimmed means is a highly robust estimator, which is referred as the modified one step M-estimator. As discussed by Othman et al. (2004) the *MOM* estimator empirically trims only the extreme data set, depending on the type of the data distribution. For a skewed data distribution, the amount of trimming must not be the same at both tails of the distribution. For instance, if the distribution is skewed to the right tail, more of the right tail of the distribution should be trimmed. In using any estimator that uses trimming, one major concern, is the process of trimming itself. Trimmed means, trims data symmetrically without any regard on the nature of the distribution. While the *MOM* estimator only trims data that is observed as outliers. If both tails of the distribution are detected as outliers, the data distribution would be trimmed symmetrically, otherwise if it is one side of the distribution that is detected as outlier, it would be trimmed asymmetrically, meaning that only one tail of the data set would be trimmed.

3. 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's detection. As an example, in considering two groups when performing an experiment, an individual may decide to neglect the two highest and lowest scores from each group in the data set, with the aim of removing outliers present in either group. As

mentioned previously, trimming process need 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. In this case, if the sample size is small, trimming will make the sample size even smaller. For the Winsorization process, the two smallest and largest values are replaced or exchanged with a preceding value closest to the outlier detection values (Yusof, Abdullah & Yhaya, 2011).

In Winsorization process, the effects of outliers are removed from a data distribution by making a replacement or an exchange for the outliers values detected. It helps to prevent loss of information and hence, the data is preserved unlike the trimmed mean procedure.

4. Method

4.1 The Alexander-Govern Test

The Alexander-Govern test is a test introduced by Alexander-Govern (1994) and it uses mean as its central tendency measure. For a normal data, it possesses good control of Type I error rates and gives high power under variance heterogeneity, but is not robust to non-normal data. This test is used for comparing two or more groups and its test statistic is obtained by using the following procedure.

The technique to obtain the test statistic for the Alexander-Govern test begins by ordering the data distribution, with population j ($j = 1, \dots, J$). For each of the data distribution, the mean is obtained using:

$$\bar{X} = \frac{\sum_j X_{ij}}{n_j} \quad (1)$$

where

X_{ij} represents the observed ordered random samples of the data set and n_j is the sample sizes of the observations. The mean represents the measure of central tendency in the Alexander-Govern (1994) method. After the mean is obtained, the estimate of the usual unbiased variance can be calculated by using:

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

where

\bar{X}_j is used for estimating μ_j , for the population (j). The standard error of the mean is obtained for each of the groups by using:

$$S_{ej} = \left[\frac{s_j^2}{n_j} \right]^{1/2} \quad (3)$$

The weight (w_j) for the group sizes with j population of the ordered sample data is defined where $\sum w_j$ must be equal to 1. Therefore, the weight (w_j) for each of the group is obtained by using the formula below:

$$w_j = \frac{1/S_{ej}^2}{\sum_j 1/S_{ej}^2} \quad (4)$$

The null hypothesis testing for the Alexander-Govern (1994) procedure, for the equality of the mean, under variance heterogeneity is defined as:

$$\begin{aligned} H_0: \mu_1 &= \dots \mu_J \\ H_A: \mu_1 &\neq \dots \mu_j, j=1, \dots, J \end{aligned} \quad (5)$$

The alternative hypothesis disclaims the statement made by the null hypothesis. The variance weighted estimate of the total means for all the groups in the ordered data set, is obtained by using:

$$\hat{\mu} = \sum_{j=1}^J w_j \bar{X}_j \quad (6)$$

where

w_j represent the weight for each of the group in the data distribution and \bar{X}_j is the mean of each of the groups in the ordered data set. The t statistic for each group is obtained by using:

$$t_j = \frac{\bar{X}_j - \hat{\mu}}{S_{ej}} \quad (7)$$

where

\bar{X}_j is the mean for each of the groups. $\hat{\mu}$ is the grand mean for all the groups under analysis and S_{ej} denotes the standard error of the mean for each of the groups in the population j .

The t statistic is distributed as a t variable with $n_j - 1$ degrees of freedom for ν .

where

ν is the degree of freedom for each of the groups in the ordered data set. The t statistic is obtained for each of the groups and is transformed to a standard normal deviates by using the Hill's (1970) normalization approximation in the Alexander-Govern (1994) method. The formula is defined as:

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

where

$$c = [a * \log_e (1 + \frac{t_j^2}{\nu_j})]^{1/2} \quad (9)$$

where

$$\nu_j = n_j - 1, a = \nu_j - 0.5, b = 48a^2, \quad (10)$$

The test statistic for the Alexander-Govern method is defined as:

$$A = \sum_{j=1}^J Z_j^2 \quad (11)$$

The level of significance used for the Alexander-Govern, test statistic is $\alpha = 0.05$, with $j - 1$ chi-square degree of freedom. The critical value is obtained by using the chi-square distribution table. Thereafter, we compare our chi-square value with the test statistic value of the Alexander-Govern test. Supposed if \underline{A} obtained is greater than the value of the test from the chi-square distribution table, we reject H_0 and conclude that the means of the different groups are different from each other, otherwise we do not reject H_0 .

5. The Modified Alexander-Govern Test

Given an ordered data sample of X_1, X_2, \dots, X_n , with sample size n and group sizes j .

Firstly, the median of the data set is obtained by selecting the middle value from the observations. The MAD estimator is the median of the set of absolute values of the differences between each of the score and the median. It is the median of $|X_j - M|, \dots, |X_n - M|$. Thereafter, the median absolute deviation about the median (MAD_n) estimator is obtained by using:

$$MAD_n = \frac{MAD}{0.6745}, \quad (12)$$

According to Wilcox and Keselman (2003) the constant value of 0.6745 is used to rescale the MAD estimator, with the aim of making the denominator estimates σ when sampling from a normal distribution.

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

$$\frac{|X_j - M|}{MAD_n} > K, \quad (13)$$

$$\text{Or when } \frac{|X_j - M|}{MAD_n} < K, \quad (14)$$

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 a very small standard error, when sampling from a normal distribution.

To detect the presence of outliers from the ordered data set, we use either equation (13) or equation (14), which is also referred to as the *MOM* estimator. The Winsorized *MOM* estimator would be applied on the data distribution when the outlier detected is replaced or exchanged with a preceding value closest to where the outlier is located.

In this study, the mean as a measure of the central tendency in Alexander-Govern test is modified by substituting it with the Winsorized modified one step M- estimator (*WMOM*), as the central tendency measure of the test. The Winsorized *MOM* estimator is applied on the data set, where outliers found are exchanged or replaced with a preceding value closest to the location where the outlier are observed.

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

1. To remove the presence of outliers from the data distribution
2. To make the Alexander-Govern test to be robust to non-normal data.

The Winsorized *MOM* can be defined as the averaging of the Winsorized data distribution. It is expressed by using:

$$WMOM = \bar{X}_{WMOMj} = \frac{\sum_{j=1}^J X_{WMOMj}}{n}, \quad (15)$$

where

\bar{X}_{WMOMj} is the mean of the Winsorized data distribution, X_{WMOMj} is the ordered sample data of the Winsorized data distribution and n is the sample size of the Winsorized data distribution.

The Winsorized sample variance is expressed as:

$$S^2_{WMOMj} = \frac{\sum_{j=1}^J (X_j - \bar{X}_{WMOMj})^2}{n-1}, \quad (16)$$

where

S^2_{WMOMj} is the Winsorized sample variance for the Winsorized data distribution and \bar{X}_{WMOMj} is the Winsorized *MOM* estimator for the Winsorized data distribution.

The standard error of the Winsorized *MOM* estimator is obtained with the use of the bootstrapping methods, in order to estimate the standard error and the bootstrapping algorithm for estimating the standard errors is expressed below:

Firstly, we select B independent bootstrap samples defined as: $x^{*1}, x^{*2}, \dots, x^{*B}$, where each of these random samples comprises of n data values chosen with replacement from x expressed below:

$$\begin{aligned} x^* &= (x_1^*, x_2^*, \dots, x_n^*) \\ \hat{F} &\rightarrow (x_1^*, x_2^*, \dots, x_n^*) \end{aligned} \quad (17)$$

The appearance of the star indicates that x^* is not the real data set of x instead it represents a randomized or

resampled version of x .

In estimating the standard error of the bootstrap samples, the number of B lies within the interval of (25-200). In this study, we made use of 50 amounts of the bootstrap samples to give a reasonable estimate of the standard of the MOM estimator (Efron & Tibshirani, 1998).

Secondly, we evaluate the bootstrap replications equating to each of the bootstrap samples expressed as:

$$\hat{\theta}^*(b) = s(x^{*b}), b=1, 2, \dots, B. \quad (18)$$

Thirdly, we estimate the standard error of $se_F(\hat{\theta})$ from the sample standard deviations of the bootstrap (B) replications as expressed below:

$$se_B = \left\{ \sum_{b=1}^B [\hat{\theta}^*(b) - \hat{\theta}(\cdot)]^2 / (B-1) \right\}^{1/2}, \quad (19)$$

where

$$\hat{\theta}(\cdot) = \sum_{b=1}^B \hat{\theta}^*(b) / B. \quad (20)$$

The weight w_j for the Winsorized data set for each group is defined as:

$$w_j = \frac{1 / S_{eWMOMj}^2}{\sum_{j=1}^J 1 / S_{eWMOMj}^2} \quad (21)$$

where

$\sum_{j=1}^J 1 / S_{eWMOMj}^2$ is the sum of the reciprocal of the square of the standard error for all the groups in the ordered data distribution.

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

$$\hat{\mu}_j = \sum_{j=1}^J w_j \bar{X}_{WMOMj} \quad (22)$$

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

$$t_j = \frac{\bar{X}_{WMOMj} - \hat{\mu}}{S_{eWMOMj}} \quad (23)$$

where

S_{eWMOMj} is the Winsorized sample standard error from the Winsorized data distribution for each of the

independent group of \bar{X}_{WMOMj} .

For the Alexander-Govern (1994) technique, the t_j value is transformed to standard normal by using the Hill's (1970) normalization approximation and the hypothesis testing of Winsorized data distribution, where S_{WMOMj}^2 is the usual estimate of the Winsorized sample variance of the $WMOM$ estimator for μ_j is defined as:

$$\begin{aligned} H_0: \mu_1 &= \dots = \mu_J \\ H_0: \mu_1 &\neq \dots \neq \mu_J, \text{ for } j = (j=1, \dots, J). \end{aligned} \quad (24)$$

The normalization approximation formula for the Alexander-Govern technique, using the Winsorized data distribution, is defined as:

$$Z_{WMOMj} = c + \frac{[c^3 + 3c]}{b} - \frac{[4c^7 + 33c^5 + 240c^3 + 855c]}{[10b^2 + 8bc^4 + 1000b]}$$

where

$$c = [a * I_n (1 + \frac{t_j^2}{v_j})]^{1/2} \quad (25)$$

where

$$v_j = n_j - 1, a = v_j - 0.5, b = 48a^2 \quad (26)$$

where

a, b and c are unknown parameters, v_j is the degree of freedom for each of the groups in the ordered sample data with sample size n .

The test statistic $AGWMOM$ for all the groups in the ordered sample data is defined as:

$$AGWMOM = \sum_{j=1}^J Z_{WMOMj}^2 \quad (27)$$

The test statistic for the $AGWMOM$ test follows a chi-square distribution at $\alpha=0.05$ with $J - 1$ chi-square degree of freedom. The chi-square distribution table is used to determine the critical value.

6. The Variables Used in This Study

In this study, there are five types of variables that were used. They are: equal and unequal sample sizes, equal and unequal variance, group sizes, types of distribution and the nature of pairing. All these variables listed above were manipulated to show the strength and the weakness of the five different tests, namely: the AG test, the $AGMOM$ test, the $AGWMOM$ test, the t -test and the $ANOVA$.

Table 1. The properties of the g- and h- distribution

| g-(Non-negative content) | h-(Non-negative content) | Skewness | Kurtosis | Types of distribution |
|--------------------------|--------------------------|----------|----------|------------------------|
| 0 | 0 | 0 | 3 | Normal distribution |
| 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).

7. The Design for the Study

The Alexander-Govern test is a test introduced by Alexander-Govern (1994). This test uses mean as its central tendency measure and has good control of Type I error rates and high power under variance heterogeneity for a normal distribution, but the test is not robust for a non-normal data. The test is also used for comparing two or more groups. In the research design for this study, both equal and unequal sample sizes were paired with equal and unequal variance, for two groups ($J = 2$), for four groups ($J = 4$) and for six groups ($J = 6$), positively and negatively with each of the g- and h- distribution.

5,000 data sets were simulated for each of the five different tests, namely: the AG test, the $AGMOM$ test, the $AGWMOM$ test, the t -test and the $ANOVA$. The pseudo random variates were obtained by using *SAS* generator *RANNOR* (*SAS Institute*, 1999) with a nominal level of $\alpha=0.05$ for the analysis of the power of the tests. 50

bootstrap samples were used in this study to estimate the standard error of the *MOM* estimator. As stated by Efron and Tibshirani (1998) 50 amounts of the bootstrap samples is the sufficient amount to produce a good estimate of $se_F(\hat{\theta})$.

Table 2. The Research Design for Two Groups

| The g- and h-distribution | Balanced and Unbalanced sample size | Variance ratio | Nature of Pairing | Notation for the condition of pairing |
|---------------------------|-------------------------------------|----------------|--------------------|---------------------------------------|
| 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. The Research Design for Four Group Condition

| 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 an h = 0 | 20:20:20:20 | 1:1:1:1 | Balanced condition | C37 |
| | | 1:1:1:36 | Positive Pairing | C38 |
| | | 1:4:16:36 | Negative Pairing | C39 |
| | 15:15:20:30 | 1:1:1:1 | - | C40 |

| | | | | |
|---------------------|-------------|-----------|--------------------|-----|
| g = 0.5 and h = 0.5 | 20:20:20:20 | 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 |
| | | 1:1:1:1 | Balanced condition | C45 |
| | | 1:1:1:36 | Positive Pairing | C46 |
| | 15:15:20:30 | 1:4:16:36 | Positive Pairing | C47 |
| | | 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 4. The Research Design for Six Groups Condition

| 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 |
| | 2:4:4:16:16:36 | 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 |
| | | 1:1:1:1:1:1 | Balanced condition | C61 |
| g = 0 and h = 0.5 | 20:20:20:20:20:20 | 1:1:1:1:1:36 | Positive Pairing | C62 |
| | | 1:4:4:16:16:36 | Positive Pairing | C63 |
| | | 1:1:1:1:1:1 | - | C64 |
| | 2:4:4:16:32:62 | 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 |
| | | 1:1:1:1:1:1 | Balanced condition | C69 |
| | | 1:1:1:1:1:36 | Positive Pairing | C70 |
| g = 0.5 and h = 0 | 20:20:20:20:20:20 | 1:4:4:16:16:36 | Positive Pairing | C71 |
| | | 1:1:1:1:1:1 | - | C72 |
| | | 1:1:1:1:1:36 | Positive Pairing | C73 |
| | 2:4:4:16:32:62 | 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 |

8. The Power of the Tests

The term power of a test is defined statistically as the probability that it will definitely result in significant outcomes. It could also be described as the capability of a test to recognize any effect, when the effect size occurs. According to Cohen (1988) explained that the effect size is the extent at which a phenomenon is

observed in the population, which implies that the null hypothesis becomes false in the population. In statistics the power of a test is affected by three factors, which are: the level of significance (α), the sample size of the data distribution (n_j), and the effect size (ES). The effect size is divided into three categories, namely: small, medium and large respectively (Abdullah, Yahaya & Othman, 2008).

In this research, the large pattern of variability was used to analyse the power of the tests, so as to produce a high power for the test.

9. Pattern of Variability

| The Effect Size Index | For J = 4 | For J = 6 |
|-----------------------|--|---|
| Small | $-\frac{1}{2}d, 0, 0, \frac{1}{2}d$ | $-\frac{1}{2}d, 0, 0, 0, 0, \frac{1}{2}d$ |
| Middle | $-\frac{1}{2}d, -\frac{1}{4}d, \frac{1}{4}d, \frac{1}{2}d$ | $-\frac{1}{2}d, -\frac{1}{3}d, -\frac{1}{6}d, \frac{1}{6}d, \frac{1}{3}d, \frac{1}{2}d$ |
| 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).

For small variability, f is expressed as:

$$f = d \sqrt{\frac{1}{2k}}, \quad (28)$$

For medium variability, f is expressed as:

$$f = \frac{d}{2} \sqrt{\frac{k+1}{3(k-1)}}, \quad (29)$$

For large variability, f is expressed as:

$$f = \frac{1}{2}d, \text{ in a condition, where } k \text{ is said to be even} \quad (30)$$

$$f = d \frac{\sqrt{k^2 - 1}}{2k}, \text{ in a situation, where } k \text{ is said to be odd} \quad (31)$$

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).

The power of each of the tests is represented graphically where the y-axis corresponds to the power of each of the tests and the horizontal axis represents the effect size index (d) for two groups conditions and (f) for more than two groups condition. The graph is used to show the trend of the power of each of the tests in relation to the effect size index. According to Murphy and Myers (1998) the power of a test must be above 0.5, and it can be considered to be sufficient, 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 index (d and f) respectively.

For the analysis of the power of the tests, 84 conditions of pairing were considered i.e C1 – C84 for each of the tests, to see those tests that produced low power, sufficient and high power respectively.

10. The Power of the Tests for Two Groups Conditions, Under Normal Distribution

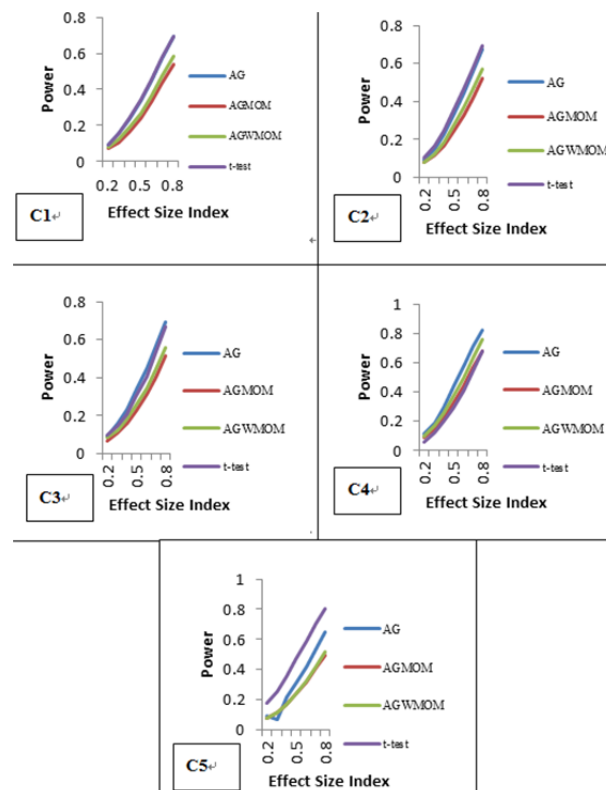


Figure 1. Power against Effect Size Index, for two group condition, under a normal distribution

In Figure 1 above, under a normal distribution, for two group conditions, from C1 to C5, 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 in like manner.

11. The Power of the Tests for Two Groups Conditions, under Symmetric Heavy Tailed Distribution

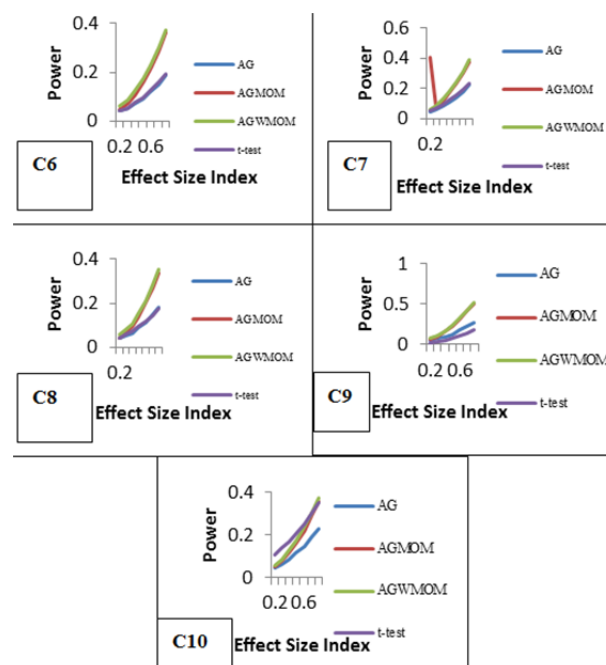


Figure 2. Power versus Effect Size Index, under a symmetric heavy tailed distribution, for two group condition

In Figure 2, under a symmetric heavy tailed distribution, for two group condition, in C6, the power of the four tests is below 0.5. The *AGWMOM* test has the highest power. The power of the four tests is regarded as insufficient. In C7, the power of the four tests is not up to 0.5. The *AGWMOM* test has the highest power. In C8, the power of the four tests is below 0.5. The *AGWMOM* test has the highest power. The power of the four tests is regarded as insufficient. In C9, the power of the *AGMOM* test and the *AGWMOM* test is below 0.5. The *AGWMOM* test has the highest power. In C10, the power of the four tests is below 0.5 and is considered to be very low.

11. The Power of the Tests for Two Groups Conditions, Under Skewed Normal Tailed Distribution

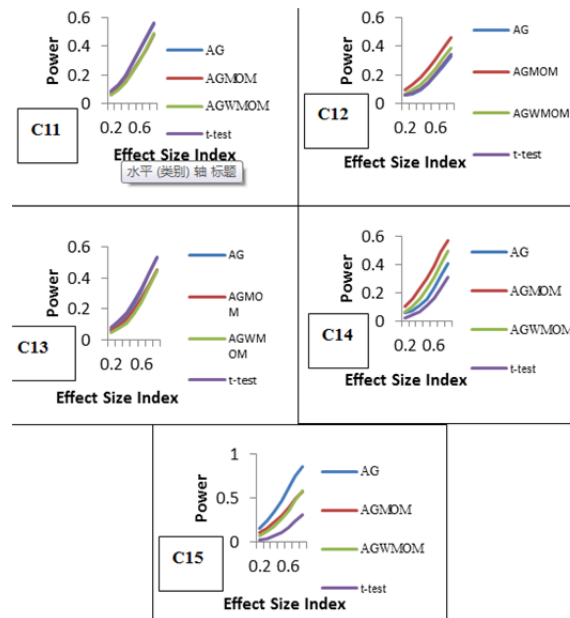


Figure 3. Power versus Effect Size Index, for two group condition, under a skewed normal tailed distribution

In Figure 3, under a skewed normal tailed distribution, for two group condition, in C11, the power of the *AG* test and the *t-test* is above 0.5. The *t-test* has the highest power. The power of the four tests is below 0.8 and is referred to as insufficient. In C12, the power of the four tests is not up to 0.5. The *AGMOM* test has the highest power. The power of the four tests is said to be insufficient, because their values are less than 0.8. In C13, the *AG* test and the *t-test* have their power values above 0.5 and the power of the four tests are regarded as insufficient. In C14, the *AGMOM* test has the highest power and is also above 0.5. The power of the four tests are said to be insufficient, because their power values are not up to 0.8.

In C15, the power of the *AG* test, the *AGMOM* test and the *AGWMOM* test is more than 0.5 and the *AG* test has the highest power, because its value is above 0.8. The power of the *AG* test is considered to be sufficient.

11. The Power of the Tests for Two Groups Conditions, Under Skewed Heavy Tailed Distribution

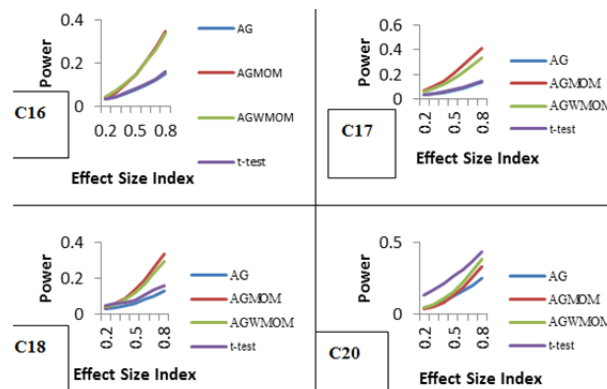


Figure 4. Power against Effect Size Index, for two group condition, when $g = 0.5$ and $h = 0.5$

In Figure 4, for $g = 0.5$ and $h = 0.5$, for C16, the power of the four tests is below 0.5. The *AGMOM* test has the highest power. The power of the four tests is below 0.8 and is regarded as insufficient. In C17, the power of the four tests is considered to be very small and not up to 0.5. The *AGMOM* test has the highest power. In C18, only the power of the *AGMOM* test is above 0.5. The power of the four tests is said to be insufficient, because their power values is less than 0.8. In C20, the power of the four tests is less than 0.5 and also considered to be insufficient.

12. The Power of the Tests for Four Groups Conditions, Under Normal Distribution

In Figure 5 below, under a normal distribution for four group conditions, in C21, the Type I error rates of the *AG* test, the *AGMOM* test, the *AGWMOM* test and the *ANOVA* are close to the nominal level of $\alpha = 0.05$ and the four tests have a maximum power value of 1.0000. In C22, only the Type I error rates of the *ANOVA* is below the nominal level of significance. The *AG* test has the highest power. The power of the other three tests is above 0.8 and is regarded as sufficient and high. In C23, only the Type I error rates of the *ANOVA* is above 0.05. The *AG* test has the highest power. The power value of the four tests is below the 0.5 bench mark. In C24, the power of the four tests is above 0.5 and also above 0.8 and are referred to as high and sufficient. In C25, only the power of the *ANOVA* is not up to 0.5. The power of the *AG* test, the *AGMOM* test and the *AGWMOM* test are above 0.5 and also more than 0.8 and are said to sufficient and high. In C26, only the Type I error rates of the *ANOVA* is above 0.05 level of significance. The *AG* test has the highest power. The power of the other three tests is greater than 0.8 and considered to be high and sufficient. In C27, the power of the four tests is less than 0.5 and also not up to 0.8. As result, the four tests are referred to as insufficient with low power. In C28, the power of the four tests is less than 0.5. The power of the four tests are said to be low and insufficient, because their power values is less than 0.8.

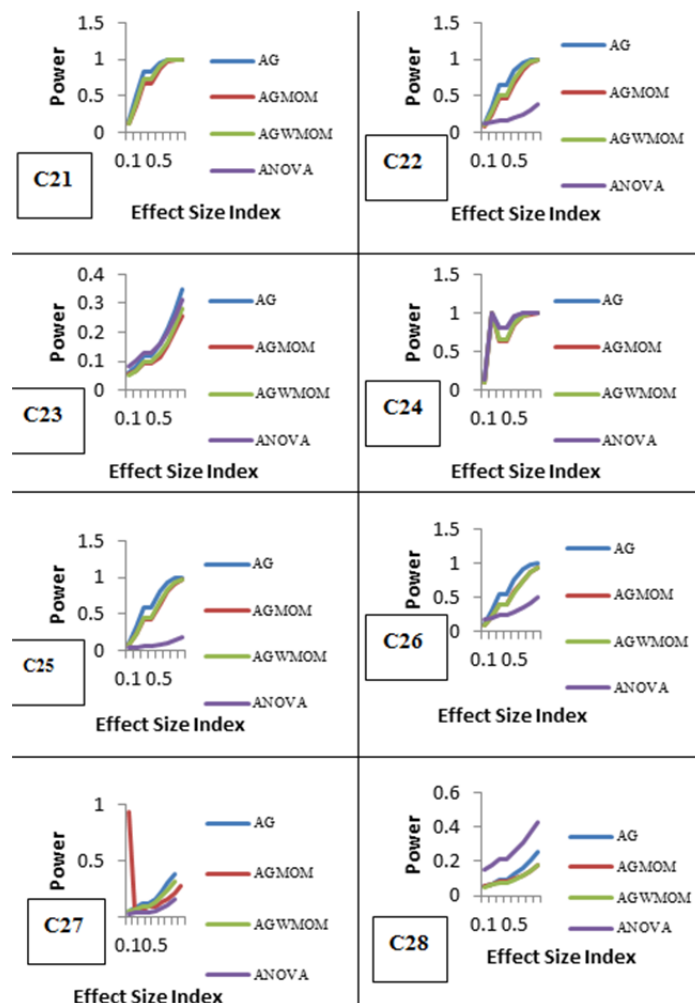


Figure 5. Power versus Effect Size Index, under a normal distribution, for four group condition

13. The Power of the Tests for Four Groups Conditions, Under a Symmetric Heavy Tailed Distribution

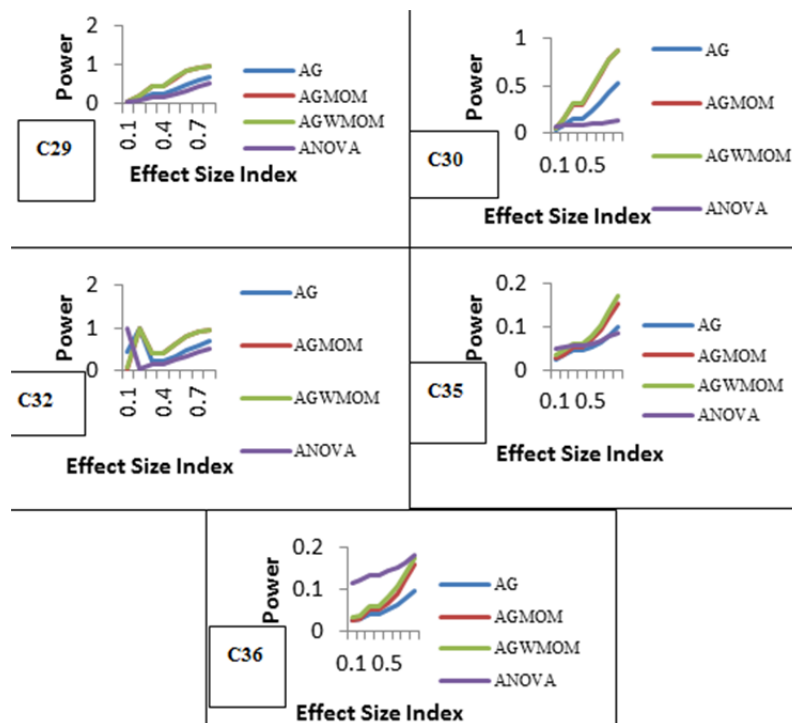


Figure 6. Power against Effect Size Index, under four group condition, for a symmetric heavy tailed distribution

In Figure 6, under a symmetric heavy tailed distribution, for four group conditions, for C29, the power of the four tests is more than 0.5. The *AGMOM* test has the highest power. The *AGMOM* test and the *AGWMOM* test have their power values more than 0.8 and are said to be high and sufficient. In C30, only the power of the *ANOVA* is less than 0.5. The *AGWMOM* test has the highest power. Both the power of the *AGMOM* test and the *AGWMOM* test is above 0.8 and is regarded as sufficient and high. In C32, the *AGWMOM* test has the highest power. The power of the four tests is above 0.5. Both the power of the *AGMOM* test and the *AGWMOM* test is more than 0.8 and are said to be high and sufficient. In C35, the power of the four tests is said to be very small and not up to 0.5. In C36, the power of the four tests is considered to be very small and insufficient.

14. The Power of the Tests for Four Groups Conditions, Under a Skewed Normal Tailed Distribution

In Figure 7 below, in C37, the *t-test* has the highest power with a maximum value of 1. The power of the four tests is above 0.5 and likewise, the power of the four tests is more than 0.8 and is regarded as high and sufficient. In C38, only the Type I error rates of the *ANOVA* is not robust. The *AG* test has the highest power and is considered to be high and sufficient.

In C39, both the power of the *AG* test and the *ANOVA* is above 0.5. The power of the four tests is below 0.8 and are said to be insufficient. In C40, the power of the four tests is more than 0.5. The *AG* test has the highest power. The power of the *AG* test, the *AGMOM* test and the *AGWMOM* test is greater than 0.8 and is referred to as sufficient and high. In C41, both the *AG* test and the *ANOVA* have their Type I error rates considered as not robust. The *AG* test has the highest power. In C42, only the Type I error rates of the *ANOVA* is above the nominal level of 0.05. The power of the other three tests is more than 0.8 and considered to be sufficient and high. In C43, only the power of the *AG* test is more than 0.5. The power of the four tests are said to be insufficient. In C44, only the Type I error rates of the *ANOVA* is above 0.05 level of significance. The power of the four tests is below the 0.5 bench mark.

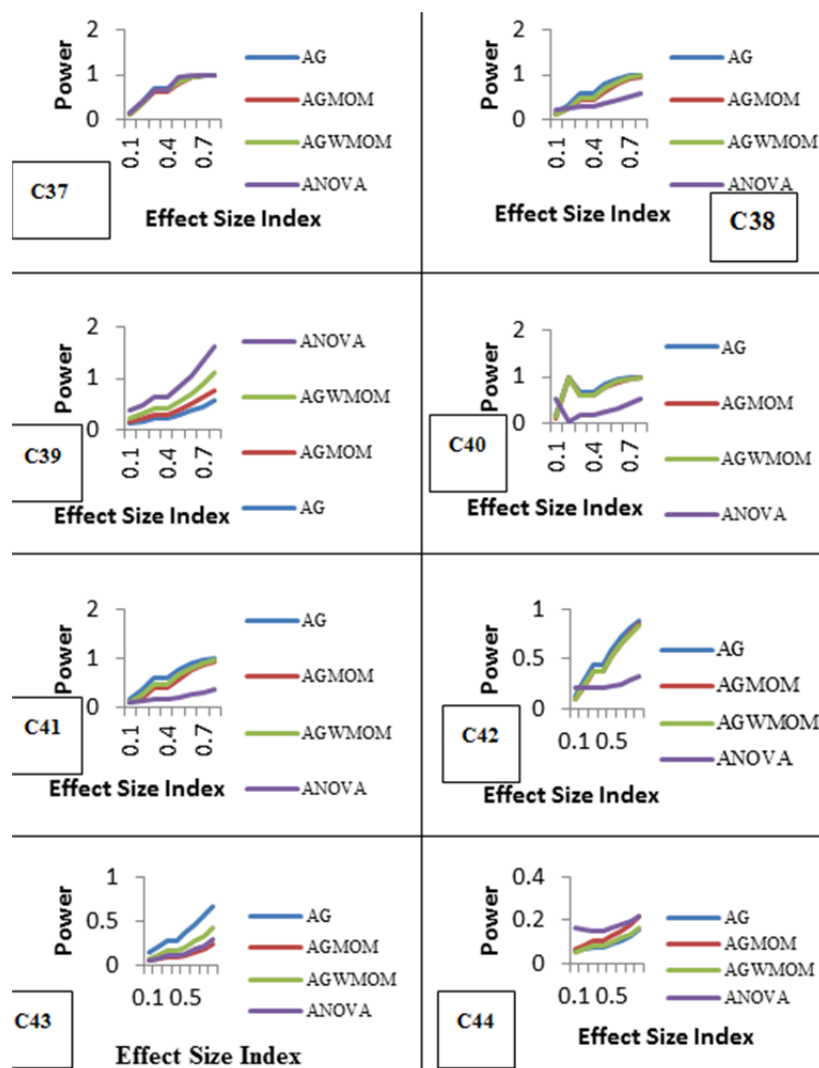
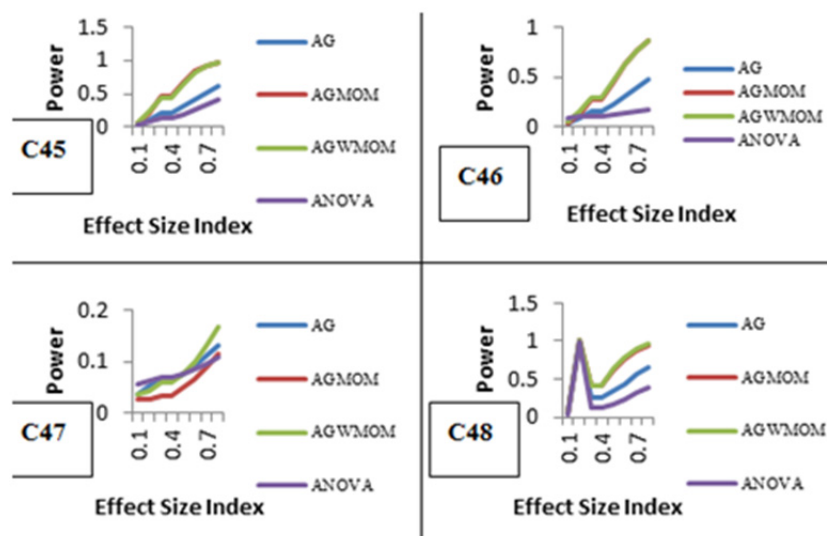


Figure 7. Power versus Effect Size Index, for four group condition, for $g = 0.5$ and $h = 0$

15. The Power of the Tests for Four Groups Conditions, Under a Skewed Heavy Tailed Distribution



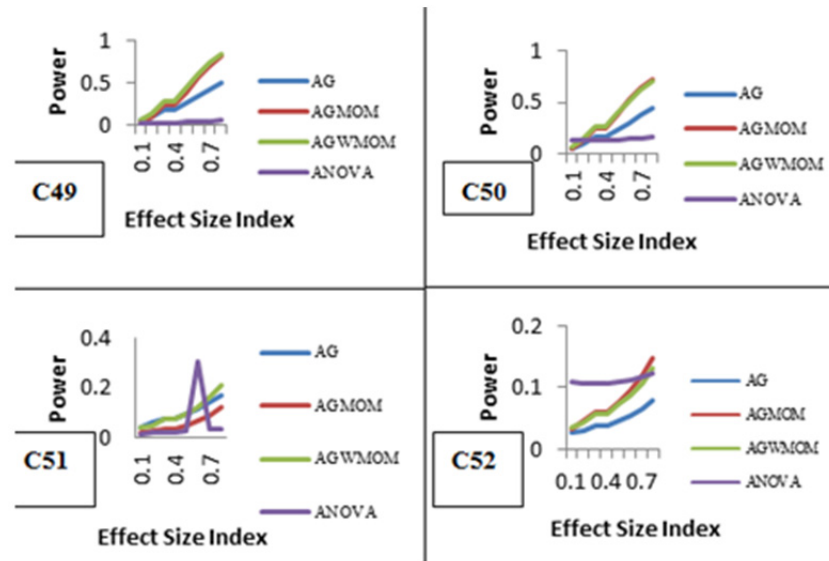


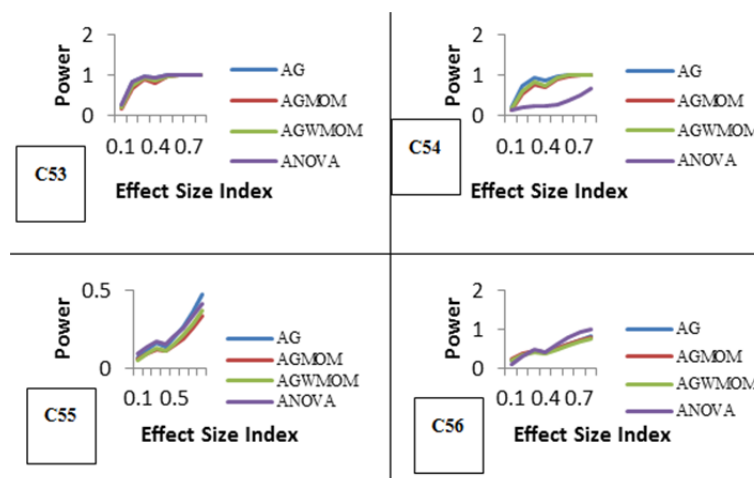
Figure 8. Power against Effect Size Index, for four group condition, when $g = 0.5$ and $h = 0.5$

In Figure 8, for C45, the *AGMOM* test has the highest power. Both the *AGMOM* test and the *AGWMOM* test have their power values above 0.8 and are said to be high and sufficient. In C46, the *AGWMOM* test has the highest power and is greater than 0.8 and considered to be sufficient and high. In C47, the *AGWMOM* test has the highest power.

In C48, only the power of the *ANOVA* is observed to be below 0.5. The *AGWMOM* test has the highest power. The power of the *AGMOM* and *AGWMOM* test is more than 0.8 and is regarded as sufficient and high. In C49, the Type I error rates of the *AGMOM* test and the *ANOVA* are considered not robust. The *AGWMOM* test has the highest power and is regarded high and sufficient, because its power value is above 0.8. In C50, the *AGMOM* test has the highest power. The power of the *AGMOM* and the *AGWMOM* test are above 0.5. In C51, the Type I error rates of the four tests is below the nominal level of 0.05. The power of the four tests is below the 0.5 bench mark.

In C52, the Type I error rates of the *AGWMOM* test and the *ANOVA* are considered not to be robust. The *AGMOM* test has the highest power.

16. The Power of the Tests for Six Groups Conditions, Under a Normal Distribution



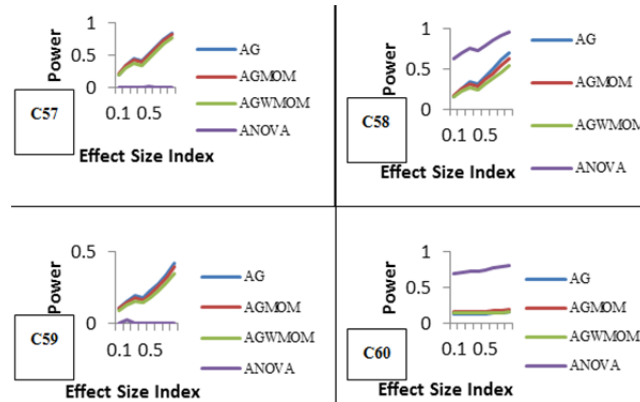


Figure 9. Power against Effect Size Index, for six group condition, under a normal distribution

In Figure 9, for C53, the power of the four tests reached a maximum level of 1. The power of the four tests is above 0.5 and also above 0.8 and are regarded as high and sufficient. In C54, the power of the four tests is above 0.5. The *AG* test has the highest power. The power of the *AG* test, the *AGMOM* test and the *AGWMOM* test are more than 0.8 and are referred to as sufficient and high. In C55, the power of the four tests is below 0.5. In C56, the power of the four tests is more than 0.5. The power of the *AG* test, the *AGMOM* test and the *ANOVA* is above 0.8 and are considered to be high and sufficient. In C57, only the power of the *ANOVA* is less than 0.5. Both the *AG* test and the *AGMOM* test have high and sufficient power. In C58, the power of the four tests is more than 0.5. The *ANOVA* has the highest power and is regarded as sufficient and high. In C59, the power of the four tests is considered to be very small and insufficient, below the bench mark of 0.5. In C60, only the power of the *ANOVA* is more than 0.5 and also more than 0.8, and is referred to as sufficient and high.

17. The Power of the Tests for Six Groups Conditions, Under a Symmetric Heavy Tailed Distribution

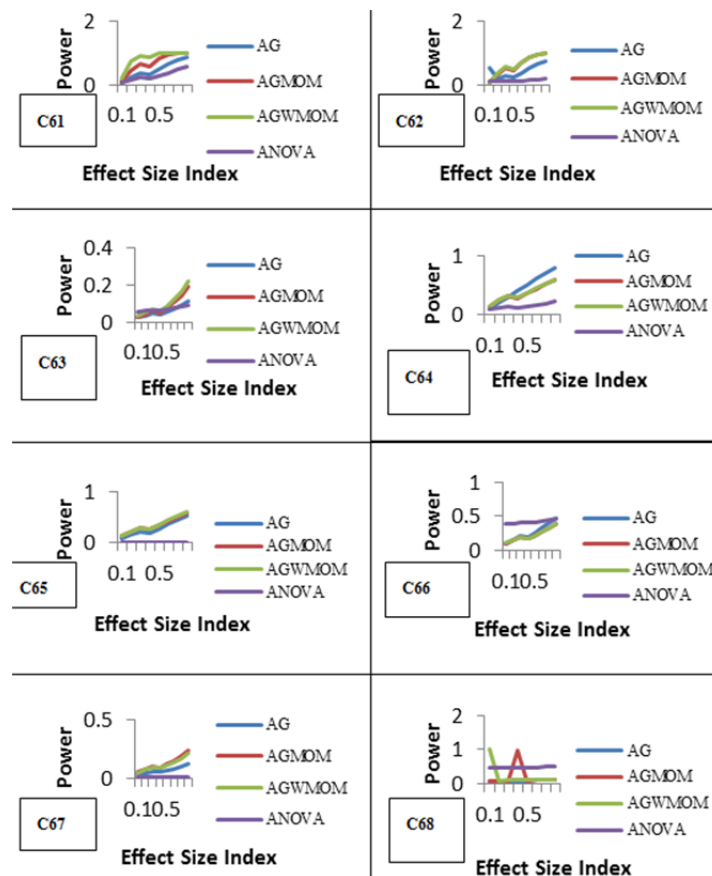


Figure 10. Power versus Effect Size Index, for six group condition, where $g = 0$ and $h = 0.5$

In Figure 10, for C61, the power of the four tests is above 0.8 and referred to as high and sufficient. In C62, the Type I error rates of the *AG* test and the *AGWMOM* test are regarded as robust. The *AGWMOM* test has the highest power and is above 0.8 and said to be high and sufficient. In C63, the power of the four tests is considered to be very small and insufficient, below the bench mark of 0.5. In C64, only the power of the *ANOVA* is below 0.5. The *AGWMOM* has the highest power and more than 0.8 and considered to be sufficient and high. In C65, only the power of the *ANOVA* is less than 0.5. It can be seen that the *AGWMOM* test has the highest power and is above 0.8 and regarded as high and sufficient. In C66, the power of the four tests is said to be small and insufficient. In C67, the power of the four tests is considered to be very small and insufficient. In C68, the Type I error rates of the four tests are outside the nominal level of 0.05. The four tests have their power values below the 0.5 bench mark.

18. The Power of the Tests for Six Groups Conditions, Under a Skewed Normal Tailed Distribution

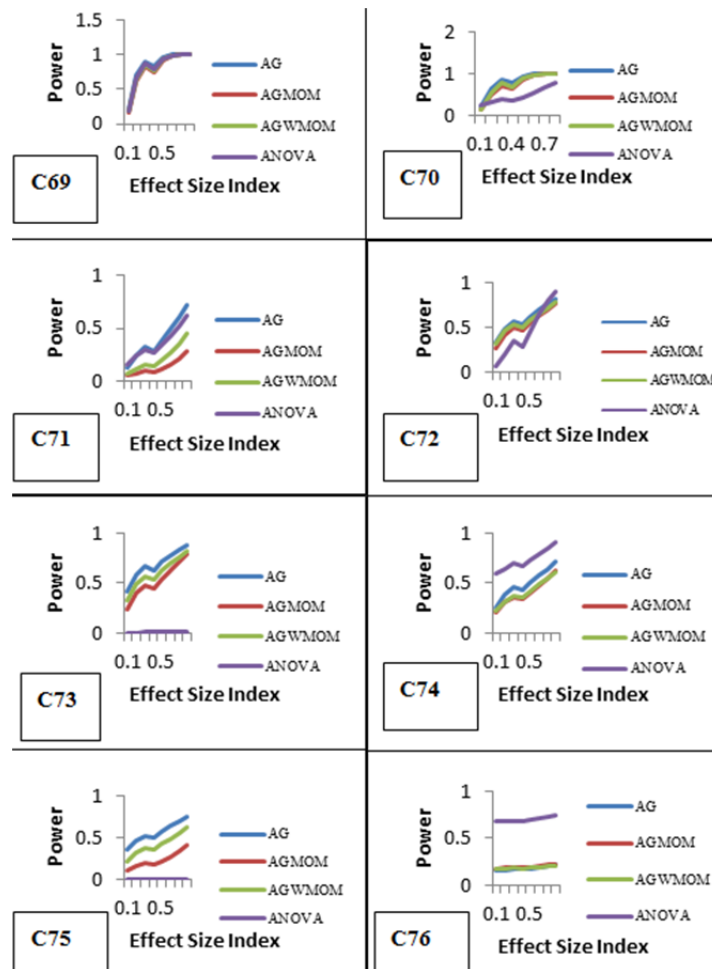


Figure 11. Power versus Effect Size Index, for six groups conditions, for $g = 0.5$ and $h = 0$

In C69, the *AG* test has the highest power. The power of the four is above 0.5 and also above 0.8 and are said to be high and sufficient. In C70, only the Type I error rates of the *ANOVA* is considered not robust. The power of the four tests is more than 0.5. The *AG* test, the *AGMOM* test and the *AGWMOM* test have their power values above 0.8 and are regarded as sufficient and high. In C71, both the power of the *AG* test and the *ANOVA* is more than 0.5. The *AG* test has the highest power. In C72, the power of the four tests is above 0.5. The *AG* test and the *ANOVA* have their power values above 0.8 and are said to be high and sufficient. In C73, only the power of the *ANOVA* is less than 0.5. The power of the *AG* test and the *AGWMOM* test is more than 0.8 and considered to be sufficient and high. In C74, the power of the four tests is more than 0.5. The *ANOVA* has the highest power. In C75, the *AG* test has the highest power. Both the *AG* test and the *AGWMOM* test have their power values above 0.5. In C76, only the power of the *ANOVA* is above the 0.5 bench mark. The power of the four tests is below 0.8 and is regarded as low and insufficient.

19. The Power of the Tests for Six Groups Conditions, Under a Skewed Heavy Tailed Distribution

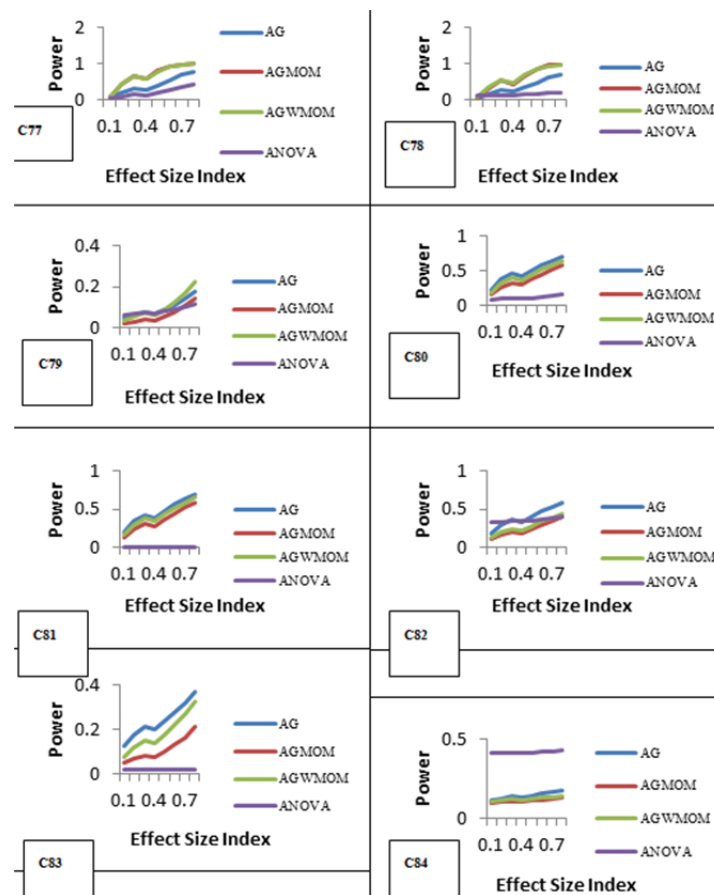


Figure 12. Power against Effect Size Index, under a skewed heavy tailed distribution, for six group condition

In Figure 12, for C77, only the power of the *ANOVA* is less than 0.5. The *AGMOM* test has the highest power. Both the power of the *AGMOM* test and the *AGWMOM* test is greater than 0.8 and are referred to as sufficient and high. In C78, the *AGMOM* test has the highest power. The power of the *AGMOM* test and the *AGWMOM* test is above 0.5 and are also above 0.8, and are said to be high and sufficient. In C79, the power of the four tests is observed to be very low. In C80, the Type I error rates for the four tests is not within the nominal level of 0.05. The *AG* test has the highest power. The power of the four tests is below 0.5. In C81, the Type I error rates of the four tests is not within the nominal interval of 0.05. The *AG* test has the highest power. Only the power value of the *ANOVA* is less than 0.5. In C82, only the power of the *AG* test is above 0.5 and is observed to have the highest power. In C83, the Type I error rates of the *AG* test and the *ANOVA* is not within the nominal level of 0.05. The *AG* test has the highest power. The power of the four tests is below the 0.5 bench mark. In C84, the power of the four tests is below 0.5. The *ANOVA* has the highest power. The Type I error rates of the four tests is above the nominal level of 0.05.

20. Discussion and Conclusion

According to Murphy and Myers (1998) the power of a test must be above 0.5 and is considered to be sufficient, when its power value is 0.8 and above.

In this research, only when $g = 0.5$ and $h = 0.5$, with the pairing of unequal sample of (15:15:20:30) with equal variance of (1:1:1:1), the *AGWMOM* test produced a high power value of 0.9562 compared to the *AG* test, the *AGMOM* test and the *ANOVA* and is considered to be sufficient. Also, with the pairing of unequal sample size of (15:15:20:30) with unequal variance of (1:1:1:36), the *AGWMOM* test produced a high power value of 0.8336 compared to the *AG* test, the *AGMOM* test and the *ANOVA* and is regarded as sufficient.

Acknowledgement

I give "God Almighty", all the thanks, glory, praises, honour, power, worship and majesty for everything. He is

the author of wisdom, knowledge and understanding.

I also want to thank and acknowledge my wonderful, blessed and beloved parents, in person of Mr. and Mrs. D. K. O. Tobi, for their constant encouragement, love, support, sacrifice, backings and goodwill. I love and appreciate them very greatly.

References

- Abdullah, S., Syed, Y., & Othman, A. R. (2008). A Power Investigation 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 of Educational Statistics*, 19(2), 91-101.
- Cohen, J. (1988). Statistical power analysis for the behavioural sciences, New York: Chapman & Hall.
- Daniel, W. W. (1990). *Applied nonparametric statistics* (2nd ed.). Boston: PES-Kent Publishing Company.
- Efron, B., & Tibshirani. (1998). *An introduction to the bootstrap*. New York: Chapman & Hall.
- Erceg-Hurn, D. M., & Mirosevich, V. M. (2008). An easy way to maximize the accuracy and power of your research. *American Psychologist*, 63, 591-601.
- Glass, G. V., Peckham, P. D., & Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the fixed-effects analysis of variance and covariance. *Review of Educational Research*, 42, 237- 288.
- Golinski, C., & Cribbie, R. A. (2009). The expanding role of quantitative methodologists in advancing psychology. *Canadian Psychology*, 50, 83-90.
- Grissom, R. J. (2000). Heterogeneity of variance in clinical data. *Journal of Consulting and Clinical Psychology*, 68, 155-165.
- Guo, J. H., & Luh, W. M. (2000). Testing Methods for One-Way Fixed Effects ANOVA Models of Log-Normal Samples. *Journal of Applied Statistics*, 27(6), 731-738.
- Harwell, M. R., Rubinstein, E. N., Haves, 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.
- Hill, G. W. (1970). Algorithm 395, Student's *t*-distribution. *Communications of the ACM*, 13, 617-619.
- James, G. S. (1951). The comparison of several groups of observations when the ratios of the population variances are unknown. *Biometrika*, 38, 324-329.
- Kohr, R. L., & Games, P. A. (1974). Robustness of the analysis of variances. *Journal of Experimental Education*, 43, 61-69.
- Krishnamoorthy, K., Lu, F., & Mathew, T. (2007). A parametric bootstrap
- 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. Doi: 10.1081/STA-12002538.p approach for ANOVA with unequal variances: Fixed and random models. *Computational Statistics & Data Analysis*, 51(12), 5731-5742. <http://dx.doi.org/10.1016/j.csda.2006.09.039>
- 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. <http://dx.doi.org/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.
- Marascuilo, L. A., & McSweeney, M. (1977). *Nonparametric and distribution-free methods for the social sciences*. Monterey, CA: Brooks/Cole Publishing Company.
- Muphy, K. R., & Myers, 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 Computational*, 60, 207-222.
- Othman, A. R., Keselman, H. J., Padmanabhan, A. R., Wilcox, R. R., & Fradette, K. (2004). Comparing

- measures of the “typical” score across treatments groups. *The British Journal of Mathematical and statistical psychology*, 57(Pt 2), 215-234. Retrieved from [http:// www.ncbi.nlm.nih.gov/pubmed/15511305](http://www.ncbi.nlm.nih.gov/pubmed/15511305)
- Scheffe, H. (1959). *The analysis of variance*. New York: Wiley.
- 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.
- Welch, B. L. (1951). On the comparison of several mean values: An alternative approach. *Biometrika*, 38, 330-336.
- Wilcox, R. R., & Keselman, H. J. (2002). Power analysis when comparing trimmed means. *J. Modern Appl.Stat. Methods*, 1(1), 24-31.
- Yusof, Md. Z., Abdullah, S., & Yahaya, S. S. S. (2011). Type I Error Rates of F_t Statistic with Different Trimming Strategies for TWO Groups Case. *Modern Applied Science*, 4.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).