

Graphical Representation of the Power Rates for the Winsorized Modified Alexander-Govern Test

Tobi Kingsley Ochuko*, Suhaida Abdullah, Zakiah Zain, Sharipah Syed Soaad Yahaya

College of Arts and Sciences, School of Quantitative Sciences, Universiti Utara Malaysia, Kedah, Malaysia

Abstract Aims and Objectives: This research deals with the comparison of the power rates of five different tests, namely: the Alexander-Govern (AG) test, the modified one step M-estimator in the Alexander-Govern (AGMOM) test, the Winsorized modified one step M-estimator in the Alexander-Govern (AGWMOM) test, the t -test and the ANOVA, for two, four and six group conditions, positively and negatively with each of the g - and h - distribution. To see of the five tests which one of them will produce the highest power for $g = 0$ and $h = 0$, $g = 0$ and $h = 0.5$, $g = 0.5$ and $h = 0$ and $g = 0.5$ and $h = 0.5$ respectively. **Method:** A line graph is used to show the graphical representation of the trends of the five listed tests, with respect to the effect size index (d) for two groups and (f) for more than two groups conditions. To see those tests that produced the minimum power value of 0.5 and also those tests that produced a power value of 0.8 and are considered to be sufficient and high. **Results:** The power values of the five different tests, under normal distribution for two group condition only, shows that the power values of the five tests is above the minimum power value of 0.5. **Conclusions:** The AGWMOM test produced the highest power values under skewed heavy tailed distribution, for four group condition only, with values of 0.9562 and 0.8336, compared to the other four tests, and the power of the test is regarded as sufficient and high.

Keywords Power rates, Alexander-Govern (AG) test, the AGMOM test and the AGWMOM test

1. Introduction

This research centres on the comparison of the power rates of five different tests, namely: the Alexander-Govern test (AG), the modified one step M-estimator (MOM), the Winsorized modified one step M-estimator in the Alexander-Govern test (AGWMOM), the t -test and the ANOVA, with the g - and h - distribution, for two, four and six group conditions respectively. To see of the five tests, which one of them will produce the minimum power value of 0.5 and also, a power value of 0.8 and the test is considered to be sufficient and high.

The ANOVA is a classical method of analysis, that have been is applied in different fields of life, such as in medicine, economics, sociology and agriculture, as stated by [24]. Three assumptions have to be fulfilled before the method can work effectively, they are: (i) homogeneity of the variance (ii) normal distribution of the data (iii) independent observations.

According to [33] the two major problems confronting the ANOVA, is the appearance of non-normality and variance heterogeneity in a data distribution. As a result, the Type I error rates is increased and the power of the test is

reduced. To obtain a good test, Type I error rates should be controlled and likewise the power of the test should not be reduced. This implies that the Type I error rates should not be increased and there should be no loss in the power of the test.

The ANOVA is used 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 data distribution, and is robust to small deviations from normality mainly when the sample size is large enough to guarantee normality, as explained by [29, 30].

The ANOVA is very sensitive to the assumption of homogeneity of variance, such that when there is a violation, the result of the analysis could be unreliable: then the p -value becomes too conservative or may be large. Therefore, it is very important 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 authenticity of the results [5, 31].

The problem of variance heterogeneity has been addressed by few scholars and some alternatives were proposed. [27] Proposed the Welch test that is used for testing the hypothesis of two populations with equal means. It was discussed in different literatures as an alternative to the ANOVA [4, 12, 16, 31]. The Welch test gives a good control of Type I error rates for unequal variances.

It is an alternative to the parametric method that deals with heteroscedasticity. However, when the sample size is

* Corresponding author:

tobikingsley@rocketmail.com (Tobi Kingsley Ochuko)

Published online at <http://journal.sapub.org/statistics>

Copyright © 2016 Scientific & Academic Publishing. All Rights Reserved

small and as the group sizes increases, the Welch test fails to give a good control of Type I error rates [28]. [10] Proposed a better alternative to the *ANOVA*, namely the James test. The James test is used for weighing sample means, as mentioned by different scholars [16, 22, 28].

The James test fails to give a good control of Type I error rates for a small sample size and when the data distribution is non-normal. Both the Welch test and the James test are used for analysing a data distribution that is non-normal with variance heterogeneity [6, 13, 14, 30].

The [3] introduced the Alexander-Govern test as a better alternative to the Welch test, the James test and the *ANOVA*, due to its simplicity in calculation as discussed by [26]. The usage of the robust test statistic such as the Alexander-Govern test is very helpful when the assumption of homogeneity of variance is violated. However, this test has its disadvantages. [26, 17, 20] observed that this method is only effective for a normal data and is not suitable for a non-normal data.

Their results showed that the Type I error rates became out-of-control, when the data distribution was not normal. The main reason why it cannot work effectively for a non-normal data is because it uses common mean as its measure of central tendency. Common mean is a highly sensitive measure with 0% breakdown point, even when only one data value is changed the value of the mean will be totally changed. Therefore, this measurement cannot handle any occurrence of outliers as well as a slight deviation from a normal distribution. In finding a solution to this problem, [17] suggested using robust estimator such as the trimmed mean, in different statistical tests that originally used the mean as the central measure of tendency.

Their findings proved that by substituting robust estimator such as the trimmed mean to the common mean is able to improve the performance of the tests in terms of Type I error rates, for a non-normal data. Trimmed mean was also used in Alexander-Govern test by [15, 18] and they observed that with the use of Winsorized variance and trimmed mean is able to eliminate the appearance of outliers in a skewed data distribution. This shows that with the use of trimmed mean, non-normality problem can be eliminated. Trimmed mean serves as an estimator that as a substitute for the common mean as a central tendency measure in a data distribution. Trimmed mean have been used by different scholars, due to its efficiency and reliability in putting under control Type I error rates [11, 19, 18].

In applying trimmed mean in a data distribution, it possesses some drawbacks, namely: (i) the percentage of trimming is placed at prior, resulting in the elimination process. (ii) in trimming process, it must be done carefully, to minimize loss of information (iii) it cannot handle large size of extreme value [32]. According to [1] an alternative to the use of trimmed mean in Alexander-Govern test is a highly robust estimator, known as the *MOM*. It was discovered that when the data distribution is skewed, the *MOM* estimator gave a good control of Type I error rates.

The *MOM* estimator empirically trims extreme data depending on the nature of the distribution, be it skewed or normal distribution.

When it was applied in Alexander-Govern test, it gave a remarkable control of Type I error rates, for a normal or highly skewed data distribution, but it failed to give a good control of Type I error rates under extreme condition of skewness and kurtosis [23].

In this research, the Winsorized *MOM* estimator was applied in Alexander-Govern test to overcome its weakness for non-normality, under variance heterogeneity, in an extreme condition of skewness and kurtosis, to give a remarkable control of Type I error rates and to produce high power for the test.

2. The Alexander-Govern Test and Its Test Statistic

The Alexander-Govern test was introduced by [3]. This test is used for comparing three or more groups. The mean serves as a measure of the central tendency for the test and gives a good control of Type I error rates for a normal data under variance heterogeneity. But this test is not robust to non-normal data. The test statistic for the test is defined by using the following methods.

First, the data sets are ordered, with population sizes of j ($j = 1, \dots, J$). For each of the data sets, the mean is calculated by using:

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

Where X_{ij} is the observed ordered random sample with n_j as the sample size of the observations. The mean is used as the central tendency measure in the [3]. After obtaining the mean, the usual unbiased estimate of the variance is obtained by using the formula:

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

Where \bar{X}_j is used for estimating μ_j for the population j . The standard error rate of the mean is calculated using the formula below:

$$S_{ej} = \sqrt{\frac{s^2_j}{n_j}}, \quad (3)$$

The weight (w_j) for the group sizes with population j of the observed ordered random sample is defined, such that $\sum w_j$ must be equal to 1. Then the weight (w_j) for each of the groups is calculated using the formula:

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

The null hypothesis testing for the [3] for the equality of mean, under heterogeneity of variance is expressed as:

$$H_o : \mu_1 = \mu_2 = \dots = \mu_j$$

$$H_A : \mu_1 \neq \mu_j$$

For at least $i \neq j$

The alternative hypothesis contradicts the statement made by the null hypothesis. The variance weighted estimated of the total mean for all the groups in the ordered data distribution, is calculated by using the formula below:

$$\hat{\mu} = \sum_{j=1}^J w_j \bar{X}_j, \quad (5)$$

Where, w_j , is the weight for each of the independent groups in the data distribution and \bar{X}_j is the mean of each of the independent groups in the observed ordered data sets. The t statistic for each of the independent groups is calculated by using:

$$t_j = \frac{\bar{X}_j - \hat{\mu}}{S_{ej}}, \quad (6)$$

Where \bar{X}_j is the mean for each of the independent group, $\hat{\mu}$ is the grand mean for all the independent groups with population j , the t statistic with $n_j - 1$ degree of freedom. Where ν is the degree of freedom for each of the independent groups in the observed ordered data sets. The t statistic calculated for each of the groups is converted to standard normal deviates by using the [9] normalization approximation in the [3] approach. The formula is defined as:

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

$$\text{Where } c = [a \times \log_e(1 + \frac{t_j^2}{\nu_j})]^{1/2}, \quad (8)$$

$$\text{Where, } \nu_j = n_j - 1, \quad a = \nu_j - 0.5, \quad b = 48a^2 \quad (9)$$

The test statistic for the AG test is defined as:

$$A = \sum_{j=1}^J Z_j^2 \quad (10)$$

After obtaining the test statistic for the AG test, a significance level of $\alpha = 0.05$ at $(j-1)$ chi-square degree of freedom is chosen. If the p -value obtained for the AG test is > 0.05 , the test is regarded as not significant, otherwise the test is said to be significant.

3. The Winsorized Modified Alexander-Govern Test

Let the observed ordered data sets of X_1, X_2, \dots, X_n , with sample n and group sizes j . Firstly, the median of the data set is calculated by selecting the middle value from the observations. The MAD estimator is the median of the set of the absolute values of the differences between each of the score and the median. It is the median of $|X_j - M|$, ..., $|X_n - M|$. Therefore, the median absolute deviation about the median (MAD_n) estimator is calculated using the formula:

$$MAD_n = \frac{MAD}{0.6745}, \quad (11)$$

According to [30] 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:

$$\frac{|X_j - M|}{MAD_n} > K, \quad (12)$$

$$\text{Or when } \frac{|X_j - M|}{MAD_n} < -K, \quad (13)$$

Where X_j is the observed ordered random sample, 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 [30] for detecting the presence of outliers in a data distribution, because it has a very small standard error, when sampling from a normal distribution.

Equation (12) and (13) is also referred to as the MOM estimator that is used for detecting the presence of outliers in a data distribution. In this research, we modified the mean as a measure of the central tendency in Alexander-Govern test, by replacing it with the Winsorized modified one step M-estimator ($WMOM$) as the central tendency measure for the test.

The $WMOM$ estimator is applied on the data distribution, where the outlier value detected is replaced or exchanged with a preceding value closest to the position where the outlier is located. The $WMOM$ estimator is calculated by averaging the Winsorized data distribution. It is expressed as:

$$WMOM = \bar{X}_{WMOMj} = \frac{\sum_{j=1}^J X_{WMOMj}}{n}, \quad (14)$$

The $WMOM$ estimator becomes a replacement for common mean as a central tendency measure in

Alexander-Govern test, due to the following reasons:

- (i) To remove the appearance of outliers from the data distribution.
- (ii) To make the Alexander-Govern test to be robust to non-normal data

The Winsorized sample variance is calculated using:

$$S_{WMOMj}^2 = \frac{\sum_{j=1}^J (X_j - \bar{X}_{WMOMj})^2}{n-1}, \quad (15)$$

Where \bar{X}_j is the observed random ordered sample and \bar{X}_{WMOMj} , is the Winsorized *MOM* estimator for the Winsorized data distribution. The standard error of the *WMOM* is calculated using the bootstrapping technique. The bootstrapping algorithm for estimating the standard errors is obtained using the following steps.

Firstly, we select B independent bootstrap samples expressed as:

$x^{*1}, x^{*2}, \dots, x^{*B}$, where each of these random samples consists of n data values selected with replacement from x defined as:

$$x^* = (x_1, x_2, \dots, x_n), \quad (16)$$

$$F \rightarrow (x_1^*, x_2^*, \dots, x_n^*), \quad (17)$$

The indication of the symbol (*) shows that x^* is not the real data of x , but it refers to a randomized or resampled version of x . In estimating the standard error of the bootstrap samples, the number of B falls within the range of (25 – 200). According to [8] bootstrap sample of size of 50 is sufficient to give a reasonable estimate of the standard error of the *MOM* estimator. In this research, the same quantity of 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 as:

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

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 $x_i, i=1, 2, \dots, n$.

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

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

$$\text{Where } \hat{\theta}^* = \sum_{b=1}^B \hat{\theta}^*(b) / B \quad \text{and} \quad \hat{\theta}^* = s(x^*).$$

The weight w_j for the Winsorized data distribution for each of the independent groups is defined as:

$$w_j = \frac{1 / S_{eWMOMj}^2}{\sum_{j=1}^J 1 / S_{eWMOMj}^2}, \quad (20)$$

Where $\sum_{j=1}^J 1 / S_{eWMOMj}^2$ is the sum of the inverse of the square of the standard error for all the independent groups in the observed ordered random samples. Where S_{eWMOMj}^2 is the standard error of the Winsorized data distribution and is defined as:

$$S_{eWMOMj}^2 = \frac{S_j^2}{n_j}, \quad (21)$$

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

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

Where w_j is expressed as the weight for the Winsorized data distribution and \bar{X}_{WMOMj} is expressed as the mean of the Winsorized data distribution. The t statistic for each of the group is defined as:

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

Where, \bar{X}_{WMOMj} , $\hat{\mu}$, and S_e is the Winsorized *MOM*, the total mean for the Winsorized data distribution and the standard error of the Winsorized data distribution respectively. In the Alexander-Govern technique, the t_j value is transformed to standard normal by using the [9] normalization approximation and the hypothesis testing of the Winsorized sample variance of the *WMOM* estimator for μ_j is expressed as:

$$H_O : \mu_1 = \mu_2 = \dots = \mu_j$$

$$H_A : \mu_i \neq \mu_j$$

For $j = (j = 1, \dots, J)$

The normalization approximation formula for the Alexander-Govern technique, using the Winsorized Modified One Step M-estimator is defined as:

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

$$\text{Where } c = [a \times \log_e (1 + \frac{t_j^2}{\nu_j})]^{1/2},$$

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

The test statistic of the Winsorized Modified One Step M-estimator in the Alexander-Govern test for all the groups in the observed random data sample is defined as:

$$AGWMOM = \sum_{j=1}^J Z_{WMOMj}^2 \quad (24)$$

The test statistic for the *AGWMOM* test follows a chi-square distribution at $\alpha = 0.05$ level of significance

with $J - 1$ chi-square degree of freedom. The p -value is obtained from the standard chi-square distribution table. If the value of the test statistic for the *AGWMOM* is < 0.05 , the test is considered to be significant. Otherwise the test is regarded as not significant.

4. Variables Used in This Research

Five variables were used in this research, namely: balanced and unbalanced sample sizes, equal and unequal variance, group sizes, nature of pairing and types of distribution. All these variables were manipulated to show the strength and weakness of the *AG* test, the *AGMOM* test, the *AGWMOM* test, *t*-test and the *ANOVA* respectively.

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

g- (Non-negative content)	h- (Non-negative 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: [29]

Table 2. The Research Design for Two Group 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

5. The 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, 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 $ANOVA$, 5,000 data sets were simulated as a reasonable amount to give good results for the power rates of the five tests respectively. To obtain the pseudo random variates, SAS generator $RANNOR$ [25] was used with a nominal level of $\alpha = 0.05$ for the analysis of the tests in this research.

Table 3. Research Design for Four Group 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
	20:20:20:20	1:1:1:1	Balanced condition	C21
		1:1:1:36	Positive Pairing	C22
		1:4:16:36	Positive Pairing	C22
$g = 0$ and $h = 0$	15:15:15:30	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 4. Research Design for Six Group 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
$g = 0$ and $h = 0$	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		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
$g = 0.5$ $g = 0.5$ and $h = 0$	20:20:20:20:20:20	1:1:1:1:1:1	Balanced condition	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
		1:1:1:1:1:1		C80
		1:1:1:1:1:36	Positive Pairing	C81
	2:4:4:16:16:32:62	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

6. The Statistical Power of a Test

The statistical power of a test is defined as the probability that it will definitely result in significant outcomes [7]. It could also be described as the capacity of a test to recognize any effect when the effect size occurs. [7] 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 known as Type II error, which is denoted by as β . In addition, the power of a test could be explained as the

probability of not accepting the null hypothesis when it is false, and it is represented as $1 - \beta$. The power of a test is affected by three factors, namely: (i) sample size (ii) level of significance (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 size 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 of the test. The power of a test is directly proportional to the quantity of the sample sizes selected [2].

[21] 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 , and it has a value which is equal to 0.05 [7]. The level of significant chosen in this research is 0.05.

Effect Size: In statistics, it is observed that the probability of the null hypothesis, that is p -value, decreases as the effect size increases and the sample size increases accordingly. The effect size could also be defined as the extent a given phenomenon is observed in the population, resulting to a state whereby the null hypothesis is false in that population. The effect size shows the differences between the maximum and minimum means between two groups, divided by the standard deviation inside the population [7].

Large pattern of variability was chosen in this research for each of the tests, namely the *AG* test, the *AGMOM* test, the *AGWMOM* test, the *t*-test and the *ANOVA*, in order to obtain high power for the tests.

7. The Power Rates of the Tests

The power rates 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 [21] the power of a test must be above 0.5. It can be considered sufficient and high when its value is 0.8 and above.

The graph shows those tests that have low power, sufficient and high power with respect to the effect size indexes (d and f). In this research, the effect size index is used for the analysis of the power rates of the five different tests accordingly.

8. Graphical Representation of the Power Rates of the Tests

In Figure 1, the power of the four tests is increasing as the effect size index is increasing. In C1, the *t*-test has the highest power. The power of the four tests is above 0.5. The Power of the four tests is not up to 0.8 and is regarded as insufficient. In C2, the *t*-test has the highest power. The power of the four tests is above 0.5. The power of the four tests is below 0.8 and is said to be insufficient.

In C3, the *AG* test has the highest power. The power of the four tests is above the power value of 0.5. The power of the four tests is less than 0.8 and is considered insufficient. In C4, the *AG* test has the highest power. The power of the four tests is above 0.5. Only the power of the *AG* test is up to 0.8 and is said to be sufficient. In C5, the power of the four tests is above 0.5. The *AG* test has the highest power and the power of the test is above 0.8 and is regarded as high and sufficient.

In Figure 2, the power of the four tests is increasing as the effect size index is increasing, except in C7, that there was a decrease in the power of the *AGMOM* test, as the effect size index is increasing. In C6, the *AGWMOM* test has the highest power. The power of the four tests is below 0.5 and is regarded as very low. In C7, the *AGWMOM* test has the highest power. The power of the four tests is not above 0.5 and is considered to be very low.

In C8, the *AGWMOM* test has the highest power. The power of the four tests is below 0.5 and is considered to be very low. In C9, the *AGWMOM* test has the highest power. It can be observed that the power of the four tests is not up to 0.5. In C10, the *AGWMOM* test has the highest power. The power of the four tests is not up to the minimum power level of 0.5

Table 5. Pattern of Variability of the Effect Size Index for 4 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: [7]

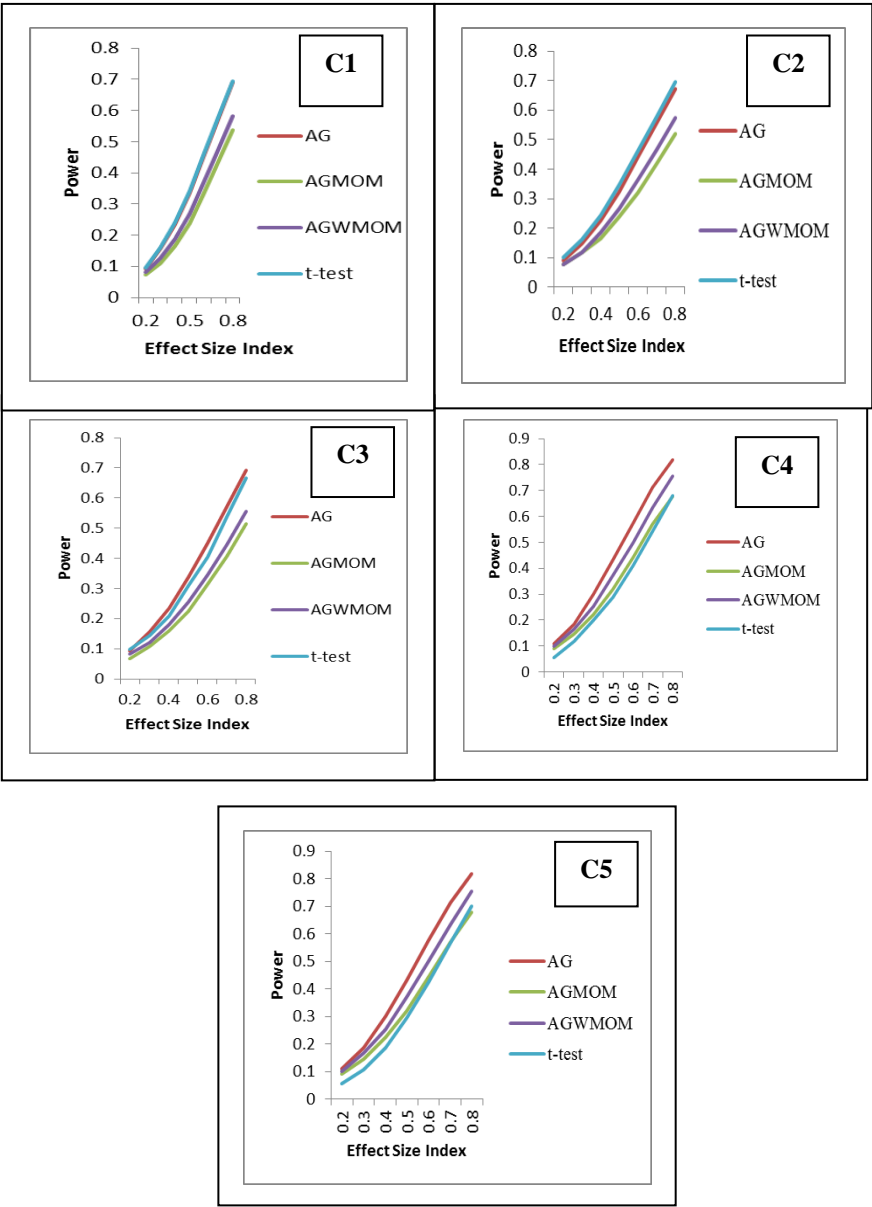


Figure 1. Power versus Effect Size Index, for Two Group Condition, Under a Normal Distribution

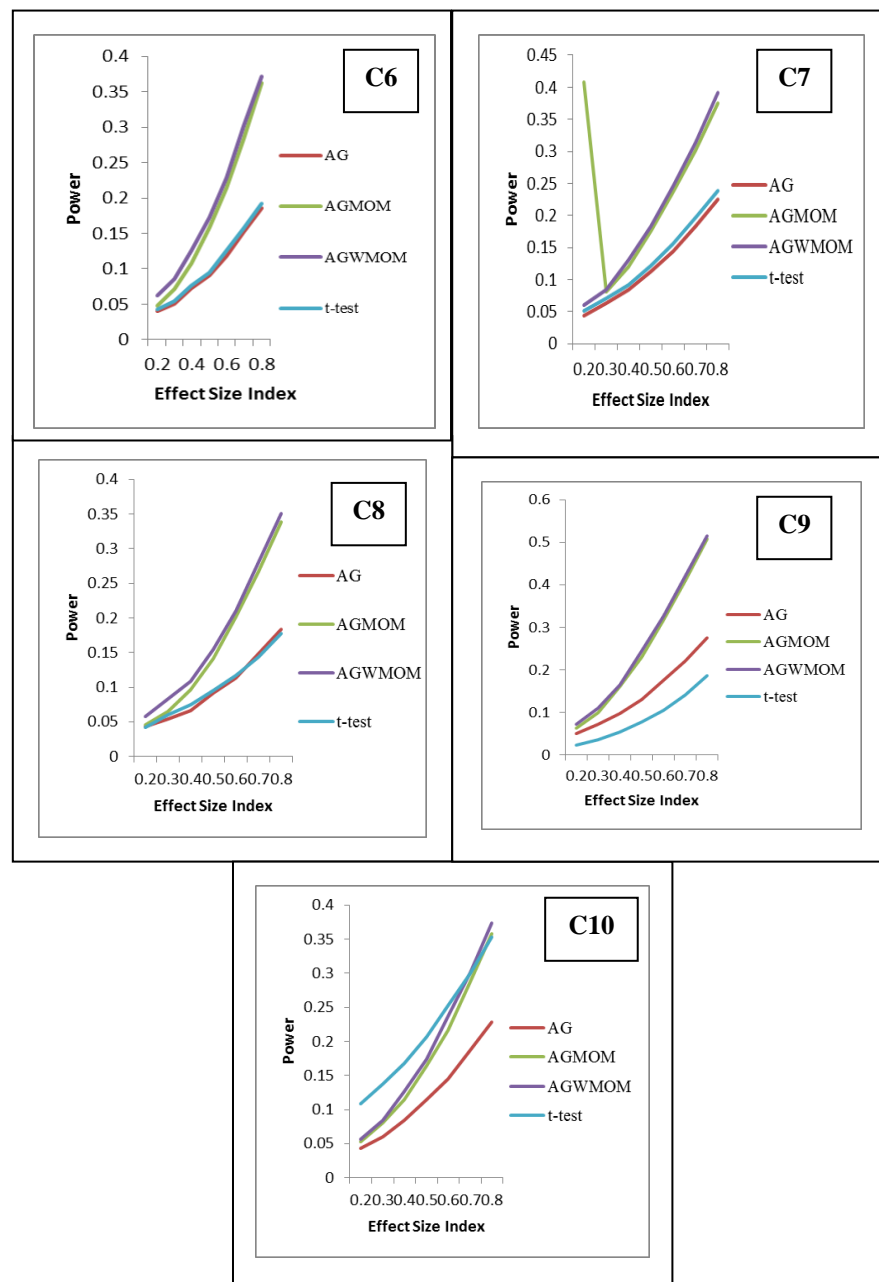


Figure 2. Power versus Effect Size Index, for a Symmetric Heavy Tailed Distribution, for Two Group Condition

In Figure 3, the power of the four tests is increasing as the effect size is increasing, under a skewed normal tailed distribution, for two group condition. In C11, the t -test has the highest power. Both the power of the t -test and the AG test is above the power value of 0.5. The power of the four tests is regarded as insufficient. In C12, the AGMOM test has the highest power. The power of the four tests is below 0.5 and is considered to be very low. In C13, the t -test has the highest power. Only the power of the t -test is above 0.5. In C14, the AGMOM test has the highest power. Only the power of the AGMOM test is above 0.5. In C15, the AG test has the highest power. Only the power of the t -test is below 0.5. The power of the AG test is above 0.8 and is considered to be

sufficient and high.

In Figure 4, the power of the four tests is increasing as the effect size index is increasing. In C16, the AGWMOM test has the highest power. The power of the four tests is below 0.5. In C17, the AGMOM test has the highest power. The power of the four tests is not up to the minimum power value of 0.5. In C18, the AGMOM test has the highest power. The power of the four tests is below 0.5 and is considered very low. In C19, the AGMOM has the highest power. The power of the four tests is not up to the minimum power level of 0.5. In C20, the t -test has the highest power. The power of the four tests is less than the minimum power value of 0.5.

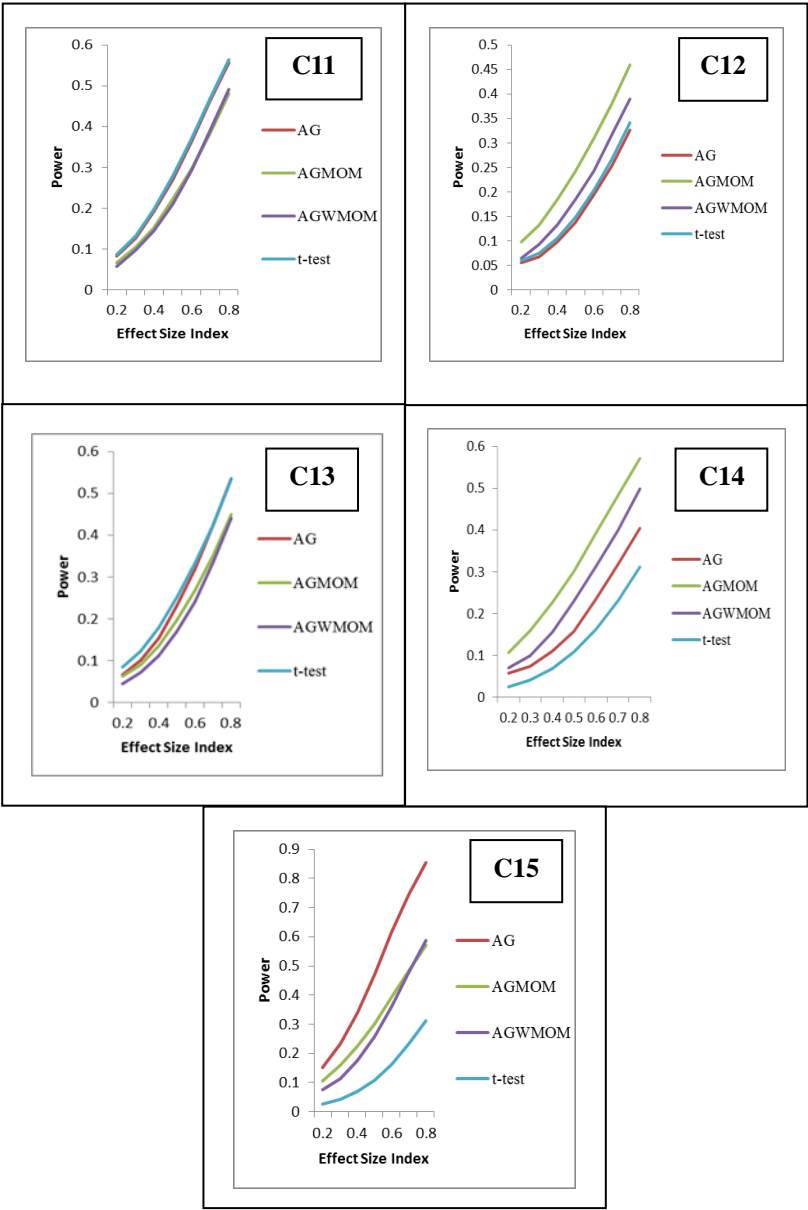


Figure 3. Power against the Effect Size Index, for Two Group Condition, Under a Skewed Normal Tailed Distribution

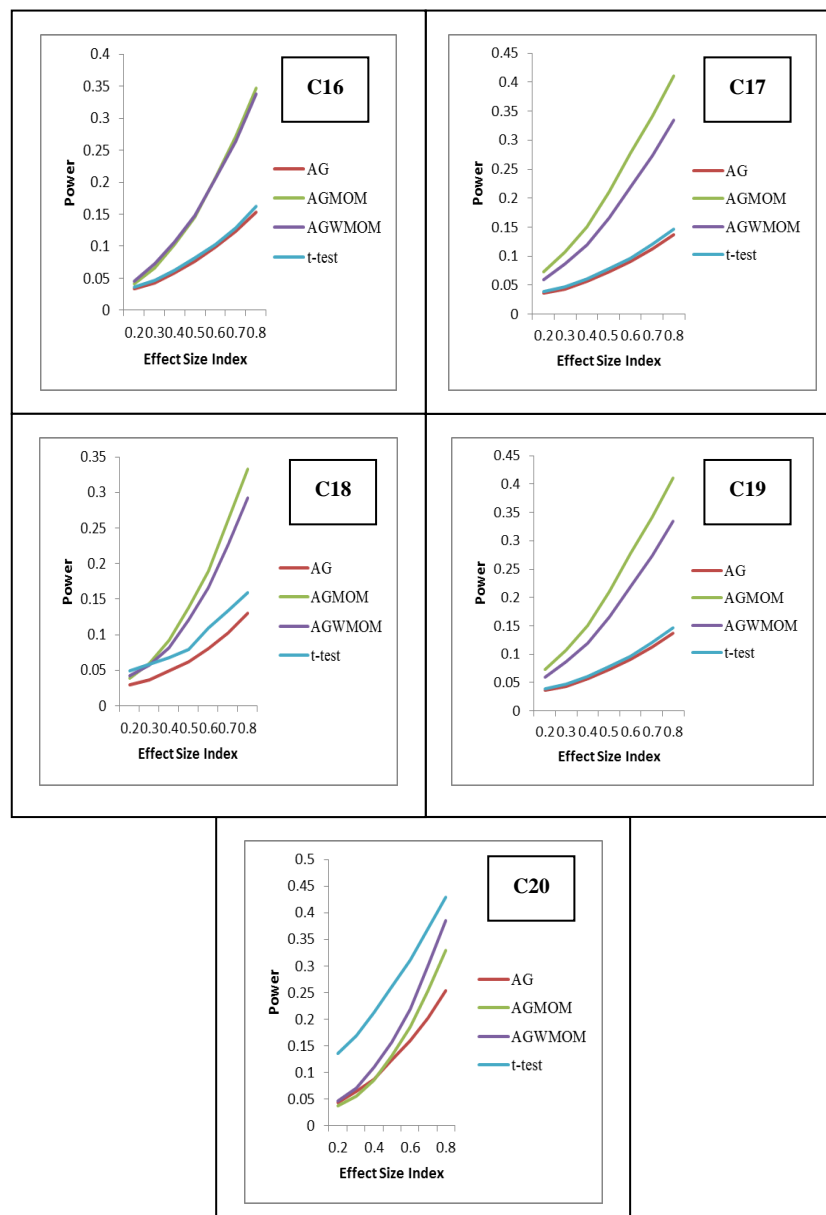


Figure 4. Power versus Effect Size Index, for Two Groups Condition, Under a Skewed Heavy Tailed Distribution

In Figure 5, In C21, the ANOVA has the highest power. The power of the four test is above 0.5. The power of the four tests is above 0.8 and is considered to be high and sufficient. In C22, the AGWMOM test has the highest power. Only the power of ANOVA is below 0.5. The power of the other three tests is above 0.8 and is said to be sufficient and high. In C23, the AG test has the highest power. The power of the four tests is below 0.5 and is regarded as very low. In C24, the power of the four tests is above 0.5 and also above 0.8. The power of the four tests is sufficient and high.

In C25, the AG test has the highest power. Only the power of the ANOVA is below 0.5. The power of the other three tests is above 0.8 and is said to be high and sufficient. In C26, the AG test has the highest power. Only the power of the ANOVA is less than 0.5. The power of the other three tests is above 0.8 and is regarded as sufficient and high. In C27, the AG test has the highest power. The power of the four tests is not up to the minimum power value of 0.5. In C28, the ANOVA has the highest power. The power of the four tests is below 0.5 and is considered to be very low.

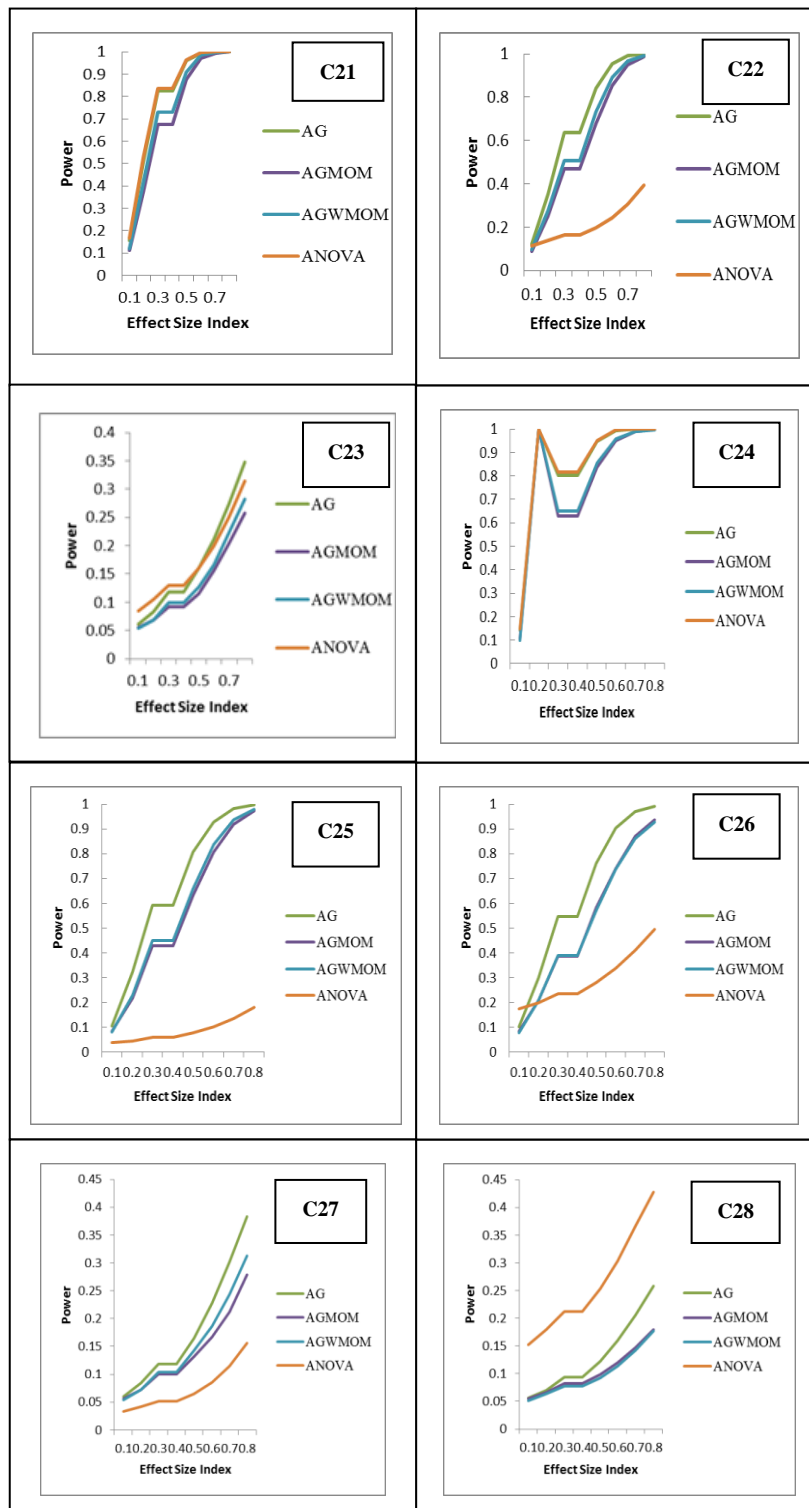


Figure 5. Power versus Effect Size Index, for Four Groups Condition, Under a Normal Distribution

In Figure 6, in C29, the *AGWMOM* test has the highest power. The power of the four tests is above 0.5. The power of the *AGWMOM* test is above 0.8 and is considered to be sufficient and high. In C30, the *AGWMOM* test has the highest power. Only the power of the *ANOVA* is below 0.5.

Both the power of the *AGMOM* test and the *AGWMOM* test is above 0.8 and are regarded as high and sufficient. In C31, the *AGWMOM* test has the highest power. The power of the four tests is above 0.5. Only the power of the *AGWMOM* test is above 0.8 and is said to be sufficient and high.

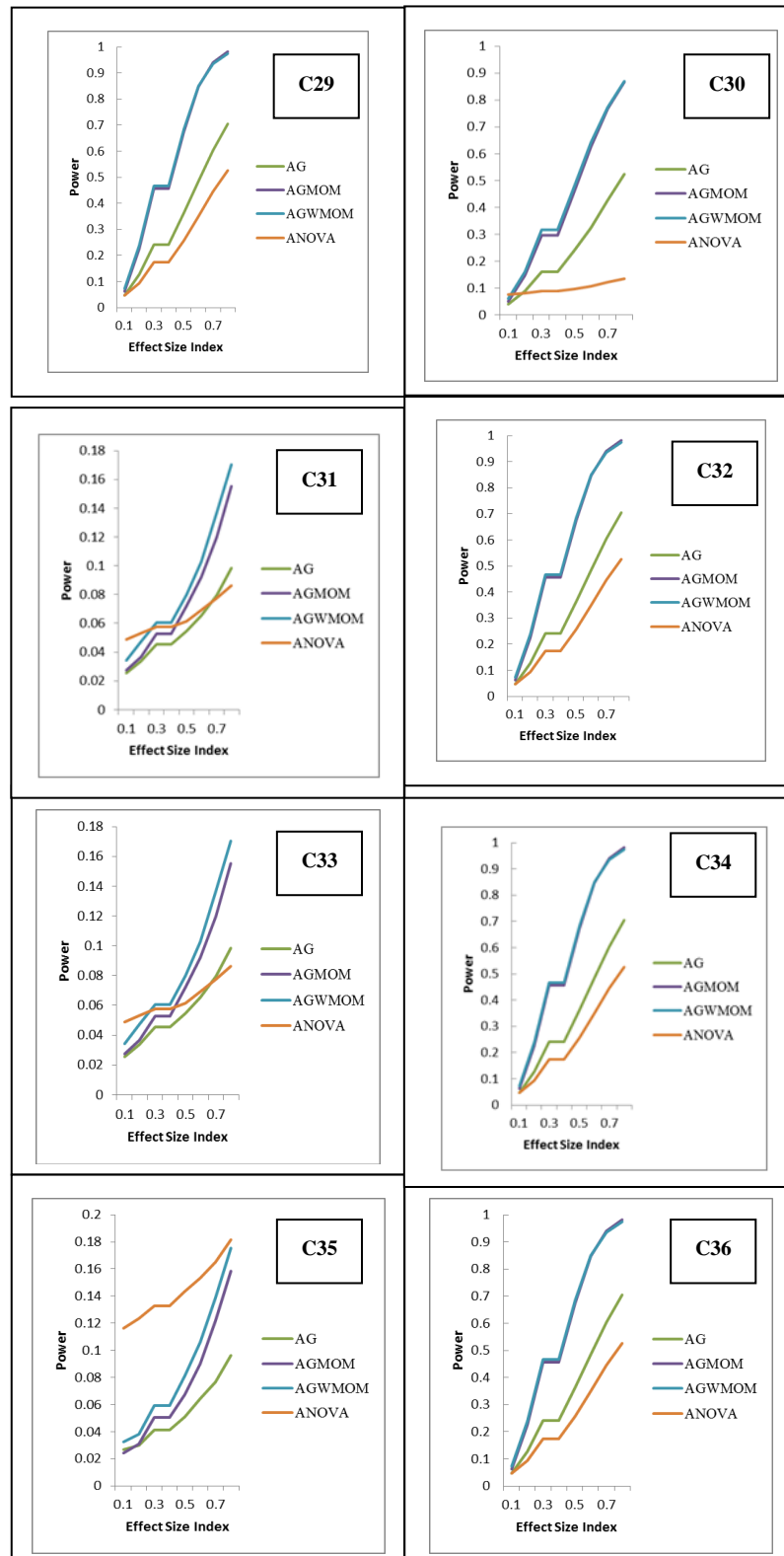


Figure 6. Power versus Effect Size Index, for Four Group Condition, Under a Symmetric Heavy Tailed Distribution

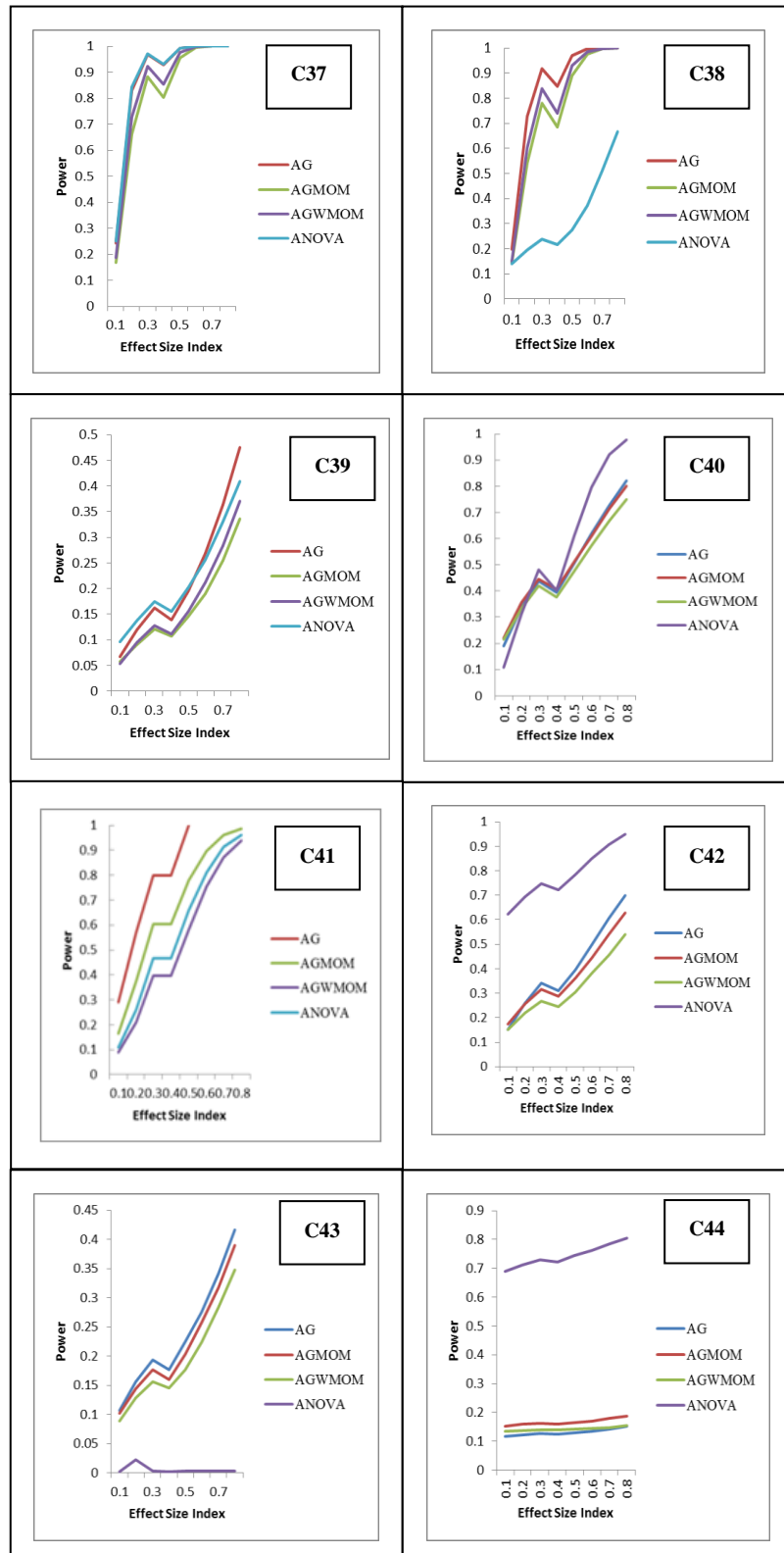


Figure 7. Power against Effect Size Index, for Four Groups Condition, Under a Skewed Normal Tailed Distribution

In C32, the *AGWMOM* test has the highest power. The power of the four tests exceeded 0.5. The power of the *AGWMOM* test is above 0.8 and is referred to as high and sufficient. In C33, the *AGWMOM* test has the highest power. The power of the four tests is below the minimum power value of 0.5 and are said to be very low. In C34, the *AGWMOM* test has the highest power. The power of the four tests reached the bench mark of 0.5. The power of the *AGWMOM* test is above 0.8 and is considered to be sufficient and high. In C35, the *ANOVA* has the highest power. The power of the four tests is less than 0.5 and are said to be very low. In C36, the *AGWMOM* test has the highest power. The power of the four tests is above the 0.5 mark. The power of the *AGWMOM* test is more than 0.8 and is referred to as sufficient and high.

In C37, the *ANOVA* has the highest power. The power of the four tests is above 0.5 and also above 0.8 and are considered to be high and sufficient. In C38, the *AGWMOM* test has the highest power. The power of the four tests is above the minimum power value of 0.5. The power of the *AG* test, the *AGMOM* test and the *AGWMOM* test is more than 0.8 and are regarded as sufficient and high. In C39, the *AGWMOM* test has the highest power. 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 said to be sufficient and high. In C40, the *AG* test has the highest power. Only the power of the *AG* test is above 0.5.

In C41, the *AGMOM* test has the highest power. The power of the four tests is above 0.8 and the tests are referred to as sufficient and high. In C42, the *AG* test has the highest power. The power of the tests is above the minimum power value of 0.5 and likewise above 0.8 and are said to be high and sufficient. In C43, the *AG* test has the highest power. Only the power of the *AG* test is above the bench mark of 0.5. In C44, the *ANOVA* has the highest power. The power of the four tests is below 0.5 and is considered to be very low.

In Figure 8, for C45, the *AGMOM* has the highest power. Only the power of the *ANOVA* is below the 0.5 mark. The power of the *AGMOM* test and the *AGWMOM* test is above 0.8 and is considered to be sufficient and high. In C46, the *AGWMOM* test has the highest power. Both the power of the *AGMOM* test and the *AGWMOM* test is more than 0.8 and is said to be high and sufficient. In C47, the *AGWMOM* test has the highest power. The power of the four tests is less than 0.5. In C48, the *AGWMOM* test has the highest power. Only the power of the *ANOVA* is below 0.5. The power of the *AGMOM* test and the *AGWMOM* test is more than 0.8 and is regarded as high and sufficient.

In C49, the *AGWMOM* test has the highest power. Only the power of the *AGWMOM* test is more than 0.8 and is considered to be sufficient and high. In C50, the *AGWMOM* test has the highest power. The power of the four tests is below 0.8 and is considered to be insufficient. In C51, the *AGWMOM* test has the highest power. The power of the four tests is below 0.5. In C52, the *AGWMOM* test has the highest power. The power of the four tests is below 0.5.

In Figure 9, considering C53, the *ANOVA* has the highest

power. The power of the four tests is above 0.8 and is considered to be high and sufficient. In C54, the *AG* test has the highest power. The power of the four tests is above 0.5. Only the power of the *ANOVA* is below 0.8 and is said to be insufficient. In C55, the *AG* test has the highest power. The power of the four tests is below 0.5 and is regarded as very low.

In C56, the *ANOVA* has the highest power. Only the power of the *AGWMOM* test is below 0.8 and is considered to be insufficient. In C57, the *AGWMOM* has the highest power. The power of the *AGMOM* test and the *AGWMOM* test is greater than 0.8 and is referred to as high and sufficient. In C58, the *ANOVA* has the highest power. Only the power of the *ANOVA* is above 0.8 and is considered to be sufficient and high. In C59, the *AG* test has the highest power. The power of the four tests is not up to 0.5 and is regarded as very low. In C60, the *ANOVA* has the highest power. Only the power of the *ANOVA* is above 0.8 and is considered to be sufficient and high.

In Figure 10, in C61, the *AGWMOM* test has the highest power. The power of the four tests is above 0.5. The power of the *AGMOM* test and the *AGWMOM* test is above 0.8 and is regarded as high and sufficient. In C62, the *AGWMOM* test has the highest power. Only the power of the *ANOVA* is below 0.5. The power of the *AGMOM* and the *AGWMOM* test is above 0.8 and is considered to be sufficient and high. In C63, the *AGWMOM* test has the highest power. The power of the four tests is below 0.5 and is said to be low.

In C64, the *AGWMOM* test has the highest power. The power of the four tests is not up to the minimum power value of 0.5 and is considered to be very low. In C65, the *AGWMOM* test has the highest power. Both the power of the *AGMOM* test and the *AGWMOM* test is above 0.5. In C66, the *AGWMOM* test has the highest power. The power of the *AGMOM* test and the *AGWMOM* test reached the bench mark of 0.5. In C67, the *AGMOM* test has the highest power. The power of the four tests is below 0.5. In C68, the *ANOVA* has the highest power. Only the power of the *ANOVA* is above the bench mark of 0.5.

In Figure 11, for C69, the *ANOVA* has the highest power. The power of the four tests is above 0.8 and is considered to be high and sufficient. In C70, the *AG* test has the highest power. The power of the *AG* test, the *AGMOM* test and the *AGWMOM* test is more than 0.8 and is regarded as sufficient and high. In C71, the *AG* test has the highest power. The power of the *AG* test and the *ANOVA* is above 0.5. In C72, the *ANOVA* has the highest power. The power of the four tests is more than 0.5. Only the power of the *ANOVA* is above 0.8 and is regarded as high and sufficient.

In C73, the *AGMOM* test has the highest power. Both the power of the *AG* test and the *AGMOM* test is above 0.5. The power of the *AGMOM* test is more than 0.8 and is said to be sufficient and high. In C74, the *ANOVA* has the highest power. Both the power of the *AG* test and the *ANOVA* is above 0.5. The power of the *ANOVA* is above 0.8 and is regarded as high and sufficient. In 75, the *AG* test has the highest power. The power of the *AG* test and the *AGWMOM*

test is above the minimum power value of 0.5. In C76, the *ANOVA* has the highest power. Only the power of the *ANOVA* is above the bench mark of 0.5.

In Figure 12, for C77, the *AG* test has the highest power. Only the *ANOVA* did not produced power value of 0.5. The power value of the *AGMOM* test and the *AGWMOM* test is above 0.8 and is regarded as high and sufficient. In C78, the *AGMOM* test has the highest power. Only the power of the *ANOVA* is below 0.5. Both the power of the *AGMOM* test and the *AGWMOM* test is above 0.8 and is considered to be high and sufficient. In C79, the *AGWMOM* test has the highest power. The power of the four tests is below 0.5. In

C80, the *AG* test has the highest power. Only the power of the *ANOVA* is less than 0.5 and is said to be very low.

In C81, the *AG* test produced the highest power. The power of the *AG* test, the *AGMOM* test and the *AGWMOM* test is above 0.5. In C82, the *AG* test produced the highest power. Only the power of the *AG* test is above 0.5. In C83, the *AG* test produced the highest power. The power of the four is below 0.5 and considered to be very low. In C84, the *ANOVA* has the highest power. The power of the four tests is less than the minimum power value of 0.5 and is said to be very low.

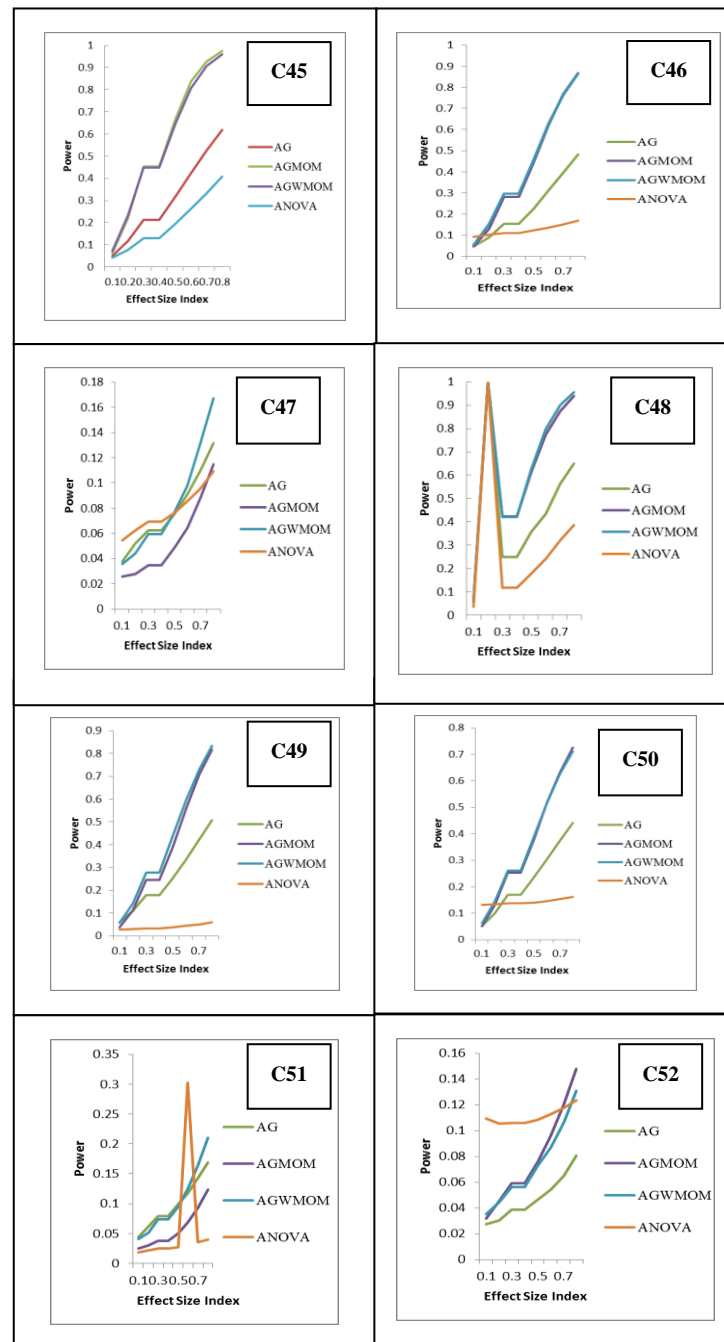


Figure 8. Power versus Effect Size Index, for Six Groups Condition, Under a Skewed Heavy Tailed Distribution, for Four Groups Condition

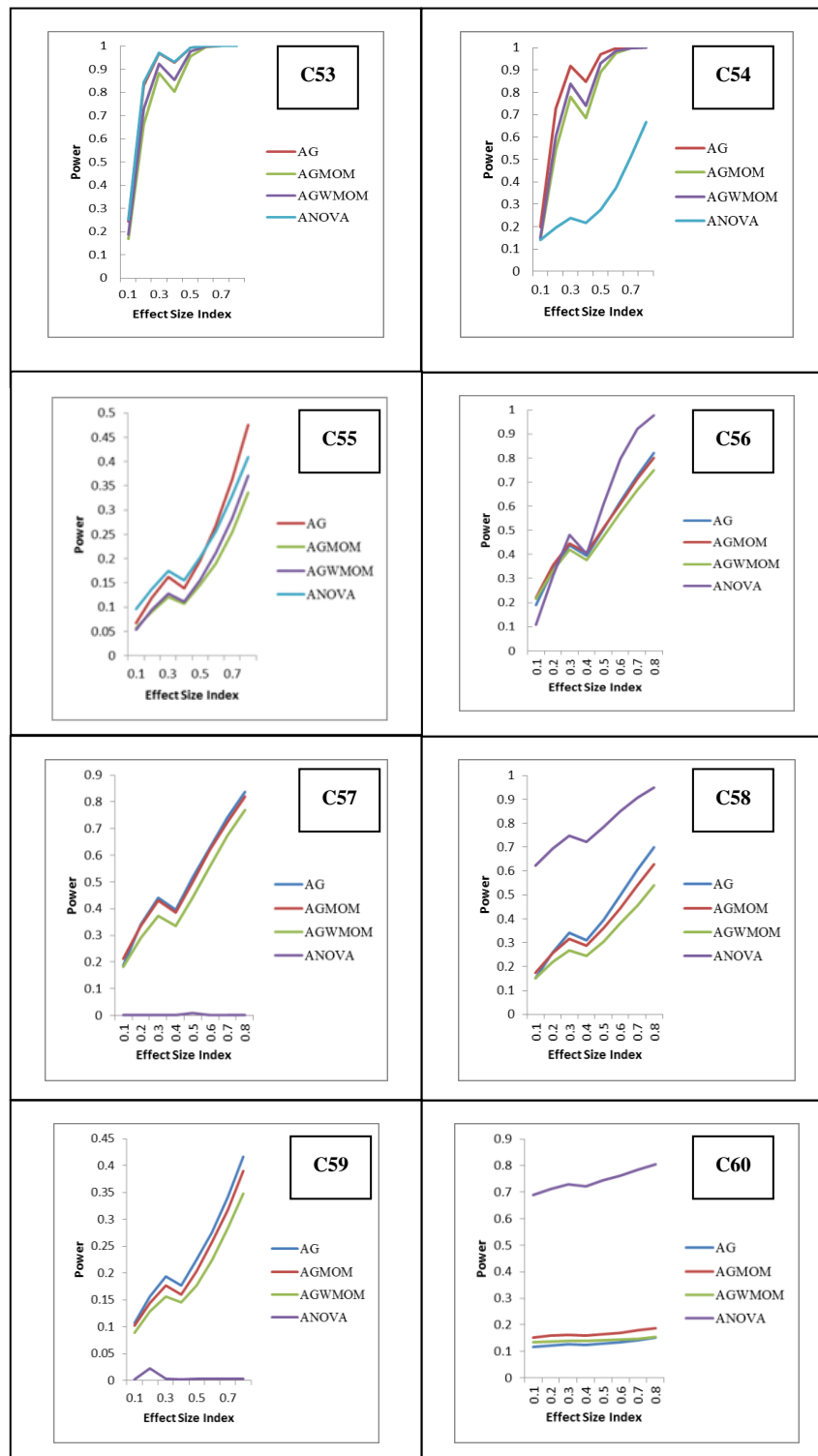


Figure 9. Power versus Effect Size Index, for Six Groups Condition, Under a Normal Distribution

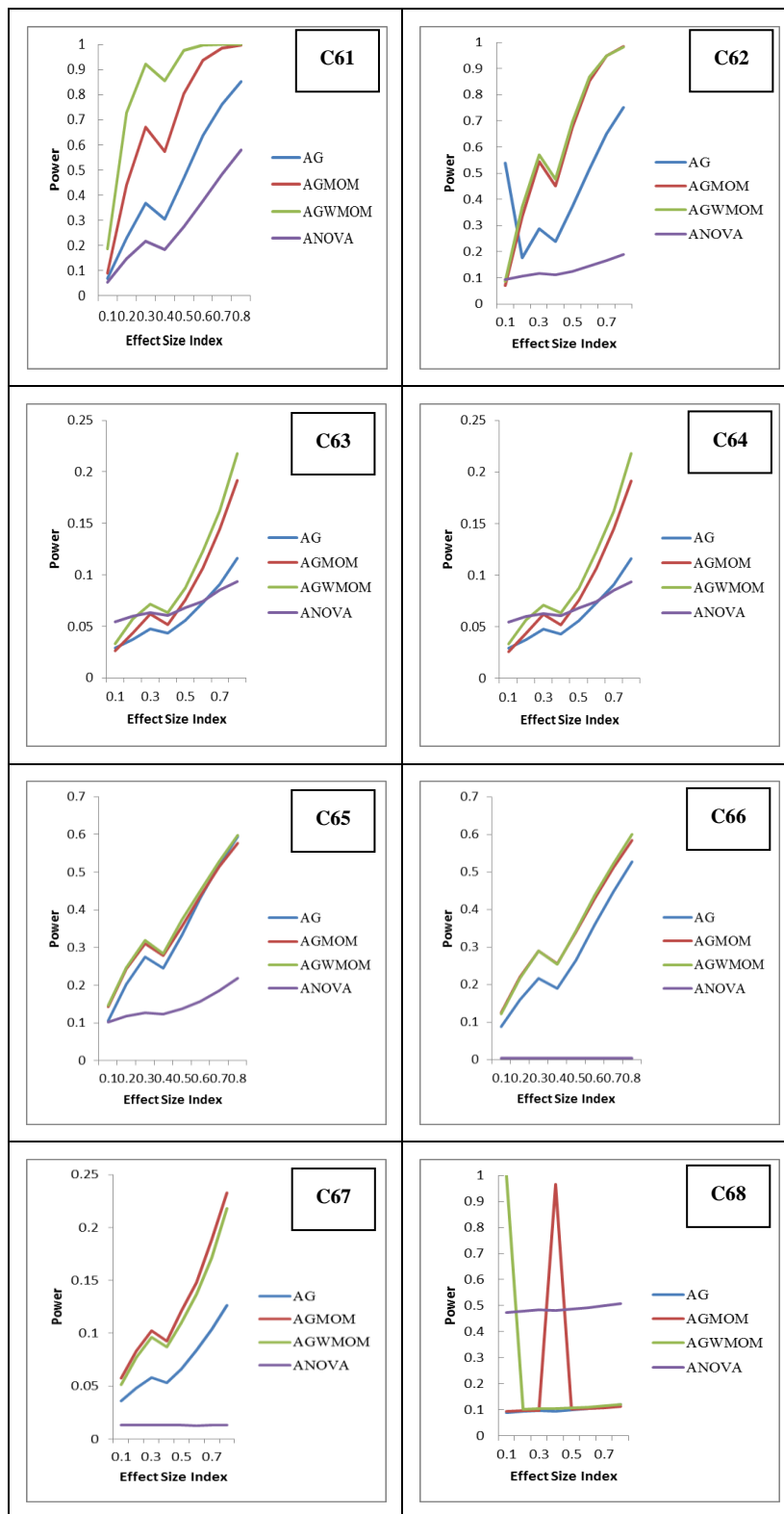


Figure 10. Power against the Effect Size Index, for Six Groups Condition, Under a Symmetric Heavy Tailed Distribution

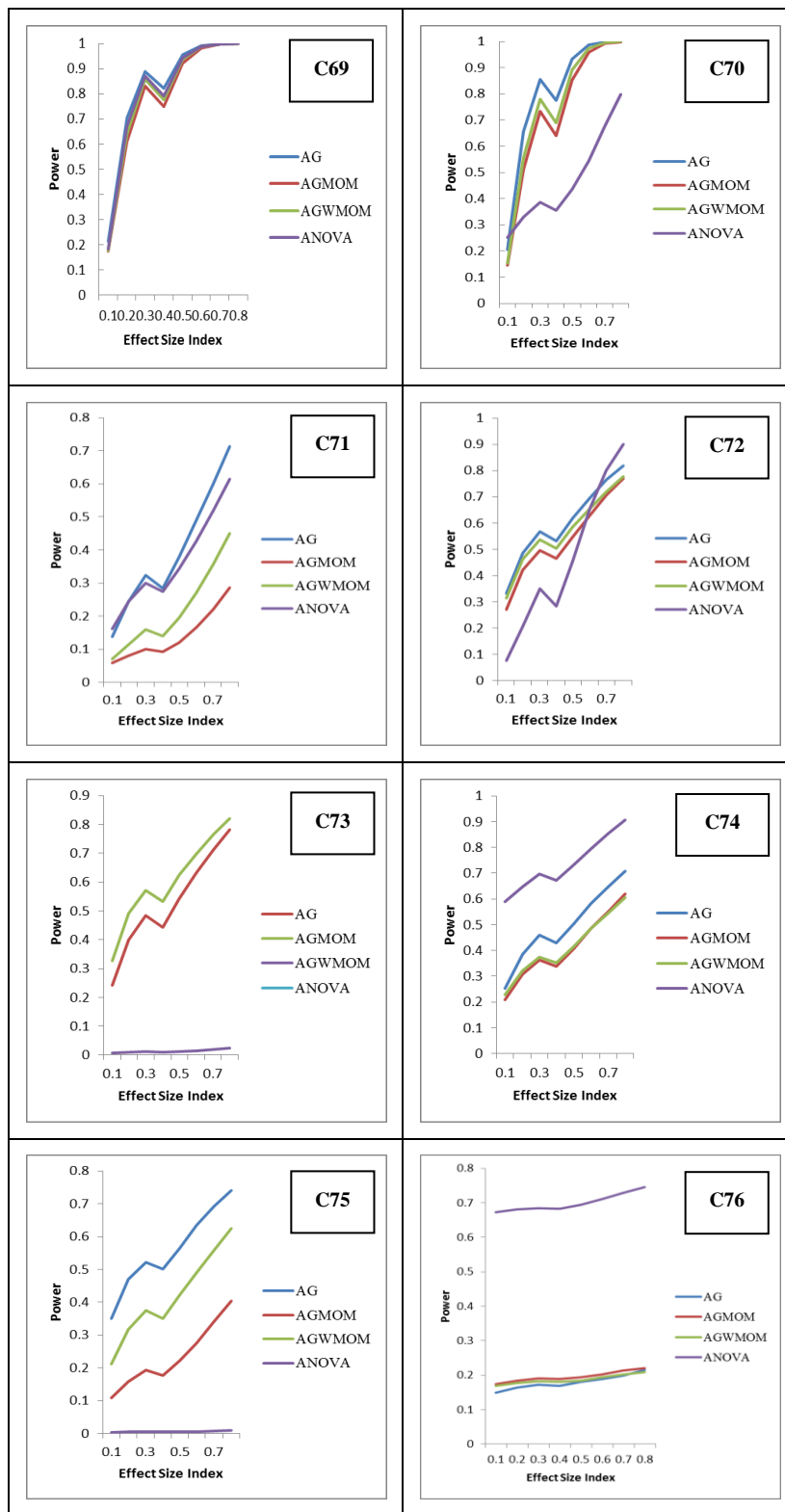


Figure 11. Power versus Effect Size Index, for Six Groups Condition, Under a Skewed Normal Tailed Distribution

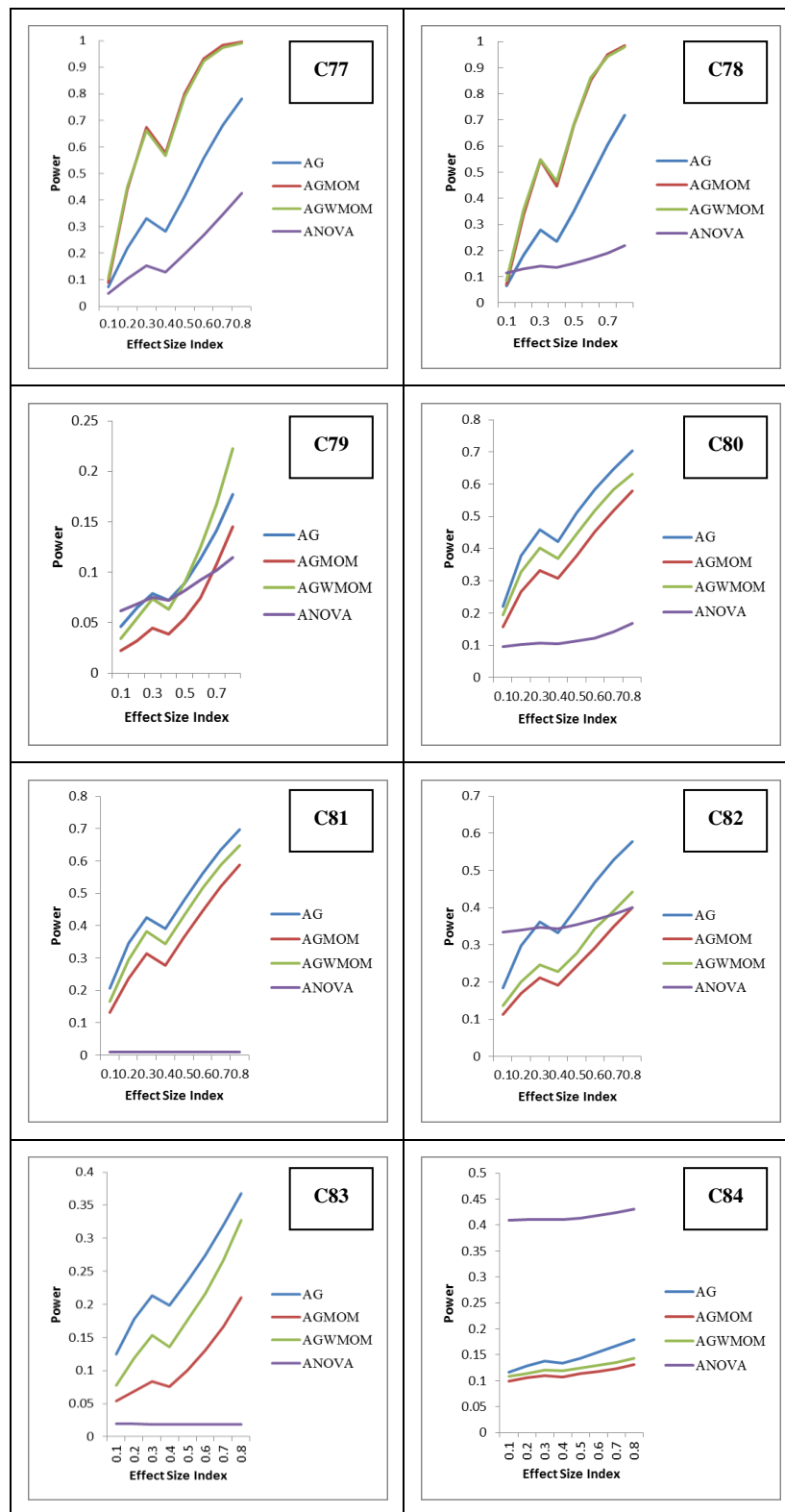


Figure 12. Power versus Effect Size Index, for Six Groups Condition, Under a Skewed Heavy Tailed Distribution

9. Conclusions

According to [21] the power of a test must be 0.5 and above and is referred to as sufficient and high when its power value is 0.8 and above. In this research, the AGWMOM test produced a high power value of 0.9562 with the pairing of unbalanced sample size of (15:15:20:30) with equal variance of (1:1:1:1) and high power value of 0.8336, with the pairing of unbalanced sample size of (15:15:20:30) with unequal variance of (1:1:1:36) under skewed heavy tailed distribution, for four group conditions only, compared to the AG test, the AGMOM test, the *t*-test and the ANOVA and the power of test is referred to as high and sufficient.

ACKNOWLEDGEMENTS

I give all the thanks, praises, glory, honor, adoration, power and worship to God Almighty for everything. He is the author of wisdom, knowledge and understanding. The Everlasting Father, "The beginning and ending of everything".

I also want to thank and acknowledge my wonderful, blessed, ever-loving, caring and ever-dynamic parents, in person of Mr. and Mrs. D.K.O. Tobi, for their constant encouragement, love, support, backings, sacrifice and goodwill in the course of carrying out this research. I love and appreciate them very greatly.

REFERENCES

- [1] Abdullah, S., Yahaya, S.S.S., & Othman, A.R. (2007). Proceedings of The 9th Islamic Countries Conference on Statistical Sciences. In Modified One Step M-Estimator as a Central Tendency Measure for Alexander-Govern Test, 834-842.
- [2] Abdullah, S., Syed Yahaya & 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.
- [3] Alexander, R.A., & Govern, D.M. (1994). A New and Simpler Approximation for ANOVA Under Variance Heterogeneity. *Journal of Education Statistics*, 19(2), 91-101.
- [4] 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.
- [5] Brown, M.B., & Forsythe, A.B. (1974). The small sample behavior of some statistics which test the equality of several means. *Technometrics*, 16, 129-132.
- [6] Brunner, E., Dette, H., & Munk, A. (1997). Box-Type Approximations in Nonparametric Factorial Designs. *Journal of the American Statistical Association*, 92(440), 1494-1502.
- [7] Cohen, J. (1988). Statistical power analysis for the behavioral sciences. New York: Chapman & Hall.
- [8] Efron, B., & Tibshirani (1998). *An introduction to the bootstrap*. New York: Chapman & Hall.
- [9] Hill, G. W (1970). Algorithm 395. Student's *t*-distribution. *Communications of the ACM*, 13, 617-619.
- [10] James, G. S. (1951). Variances are Unknown when the ratios of the population variances, 38(3/4), 324-329.
- [11] 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.
- [12] Keselman, J. J. C. and H. J. (1982). Parametric Alternative to the Analysis of Variance Author (s): Jennifer J. Clinch and H. J. Keselman Source: *Journal of Educational Statistics*, 7(3), 207-214.
- [13] 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.
- [14] Krishnamoorthy, K., F., & Matthew, T. (2007). A parametric bootstrap approach for ANOVA with unequal variances: Fixed and random models. *Computational Statistics & Data Analysis*, 51(12), 5731-5742.
- [15] Lix, Lisa, M., & Keselman, J.C., & Keselman, H. J (1995). Approximate degrees of freedom tests. A unified perspective on testing for mean equality. *Psychological Bulletin*, 117(3), 547-560.
- [16] 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.
- [17] Lix, L. M, & Keselman, H. J. (1998). To trim or not to trim. *Educational and Psychological Measurement*, 58(3), 409-429.
- [18] 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.
- [19] 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.
- [20] 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 Computation*, 60, 207-222.
- [21] Murphy, K.R., & Myers, B. (1998). Statistical power analysis: A simple and general model for traditional and modern hypothesis tests. Mahwah, NJ: Lawrence Erlbaum.
- [22] 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.

- [23] Othman, A. R., Keselman, H. J., Padmanabban, A. R., Wilcox, R. 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(2), 215-234.
- [24] Pardo, J. A., Pardo, M. C., Vincente, M. L., & Esteban, M. D. (1997). A statistical information theory approach to compare the homogeneity of several variances. *Computational Statistics & Data Analysis*, 24(4), 411-416.
- [25] SAS Institute Inc. (1999). *SAS/IML User's Guide Version 8*. Cary, NC: SAS Institute Inc.
- [26] Schneider, P. J., & Penfield, D. A. (1997). Alexander-Govern's Approximation: Providing an alternative to ANOVA Under Variance Heterogeneity. *Journal of Experimental Education*, 65(3), 271-287.
- [27] Welch, B. L. (1951). On the comparison of several means: An alternative approach. *Biometrika*, 38, 330-336.
- [28] 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.
- [29] Wilcox, R. R. (1997). *Introduction to robust estimation and hypothesis testing*. San Diego, CA: Academic Press.
- [30] Wilcox, R. R., & Keselman, H. J. (2003). Modern Robust Data Analysis Methods: Measures of Central Tendency. *Psychological Methods*, 8(3), 254-274.
- [31] 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.
- [32] 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.
- [33] Yusof, 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*, 5(4), 1-7.