

WINSORIZED MODIFIED ONE STEP M-ESTIMATOR AS A MEASURE OF THE CENTRAL TENDENCY IN THE ALEXANDER-GOVERN TEST

Tobi Kingsley Ochuko¹; Suhaida Abdullah²; Zakiyah Zain³;
Sharipah Syed Soaad Yahaya⁴

^{1,2,3,4} College of Arts and Sciences, School of Quantitative Sciences, Universiti Utara Malaysia,
06010 UUM Sintok, Kedah, Malaysia
^{1,2,3,4} tobikingsley@rocketmail.com

ABSTRACT

This research dealt with making comparison of the independent group tests with the use of parametric technique. This test used mean as its central tendency measure. It was a better alternative to the ANOVA, the Welch test and the James test, because it gave a good control of Type I error rates and high power with ease in its calculation, for variance heterogeneity under a normal data. But the test was found not to be robust to non-normal data. Trimmed mean was used on the test as its central tendency measure under non-normality for two group condition, but as the number of groups increased above two, the test failed to give a good control of Type I error rates. As a result of this, the MOM estimator was applied on the test as its central tendency measure and is not influenced by the number of groups. However, under extreme condition of skewness and kurtosis, the MOM estimator could no longer control the Type I error rates. In this study, the Winsorized MOM estimator was used in the AG test, as a measure of its central tendency under non-normality. 5,000 data sets were simulated and analysed for each of the test in the research design with the use of Statistical Analysis Software (SAS) package. The results of the analysis shows that the Winsorized modified one step M-estimator in the Alexander-Govern (AGWMOM) test, gave the best control of Type I error rates under non-normality compared to the AG test, the AGMOM test, and the ANOVA, with the highest number of conditions for both lenient and stringent criteria of robustness.

Keywords: Alexander-Govern (AG) test, MOM estimator, AGWMOM test, Type I error rates, Test Statistic

INTRODUCTION

In this study, five different tests were used, namely: (i) Alexander-Govern test (AG) (ii) Modified One Step *M*-estimator (*MOM*) (iii) Winsorized Modified One Step *M*-estimator in Alexander-Govern test *AGWMOM* (iv) *t*-test (v) *ANOVA*. These tests are performed under two, four and six group conditions, with each of the *g*- and *h*- distribution.

The *g*- and *h*- distribution is used for determine the level of skewness and kurtosis in a data distribution. The *ANOVA* is very useful in different areas of life, for example in agriculture, sociology, banking, economic and in medicine as stated by Pardo, Pardo, Vincente and Esteban (1997). Three basic assumptions must be satisfied before the *ANOVA* can work rightly. They are: homogeneity of the variance, normality of the data and independent observations. The *ANOVA* is very useful for comparing the differences between three or more means. It is applicable in testing the equality of the central tendency of a data set and is robust to little deviations from a normal data, mainly when the sample size is large to guarantee normality as explained by Wilcox (1997; 2003).

Researchers such as Yusof, Abdullah, Yahaya and Othman (2011) discovered that variance heterogeneity and non-normality are the problems affecting the *ANOVA*. This makes the Type I error rates to be increased and the power would be decreased. The problem of variance heterogeneity has been addressed by few researchers and some alternatives have been provided. Welch (1951) introduced the Welch test, for testing the hypothesis of two populations with equal means. It has been mentioned in different literatures as good alternative to the *ANOVA* (Algina, Oshima & Lin, 1994).

The Welch test gives a good control of Type I error rates when the variances are not equal. It is a better alternative to parametric method that uses heteroscedasticity. However, for a small sample size, the Welch test fails to give a good control of Type I error rates, as the group sizes increases (Wilcox, 1988). The James test was introduced by James (1951) as a better alternative to the *ANOVA* for variance heterogeneity. This test is used for weighing the sample means and it has been discussed in many literatures as a better alternative to the *ANOVA* (Oshima & Algina, 1992; Wilcox, 1988).

When the sample size is small under non-normal data, the James test fails to control Type I error rates. Both the Welch test and the James test are used for analysing a non-normal data with variance heterogeneity (Brunner, Dette, & Munk, 1997; Krishnamoorthy, Lu, & Matthew, 2007; Wilcox & Keselman, 2003). The Alexander-Govern test was proposed by Alexander-Govern (1994) to handle the problem of heterogeneity of variance under normal data. But the test is not robust to non-normality. Scholars such as Schneider and Penfield (1997) and Myers (1998) suggested that the Alexander-Govern test is a better alternative compared to the James test and the Welch test respectively. Myers (1998) admitted that the Alexander-Govern test gives an outstanding control of Type I error rates, for variance heterogeneity under a normal data. Lix and Keselman (1998) proposed a better alternative to the mean with the introduction of trimmed mean in few robust test statistics that increases the performance of the test under non-normality.

A better alternative to the use of trimmed mean is a highly robust estimator called the modified one step *M*-estimator. Othman *et al.* (2004) explained that the *MOM* estimator trims the extreme data set only, depending on the type of the data distribution. Under a skewed data distribution, the amount of trimming should not be the same at both tails of the distribution. For example, when the distribution is skewed to the right tail, more of the right tail of the distribution would be trimmed. When using any estimator that uses trimming, one thing that is significant is the process of trimming itself. Trimmed means assists to trims data symmetrically without any regard on the nature of the distribution. While the *MOM* estimator specializes in trimming only the data that is observed as outliers. When 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 is detected as outlier, it would be trimmed asymmetrically, meaning that only one tail of the data set would be trimmed. A non-normal data is a condition whereby a data is not normally distributed. In addition, Schneider and Penfield (1997) admitted that the Alexander-Govern test is a better alternative to the *ANOVA* under variance heterogeneity compared to the Welch test and the James test due its' less complexity in calculation and having a good control of Type I error rates. 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, when there is variance heterogeneity under normality it was only good for normal data, but not suitable for non-normal data, as discussed by Myer (1998).

According to scholars such as Ochuko, Abdullah, Zain, and Yahaya (2015) explained that the Winsorization process is making a substitution or an exchange for the outlier detected value with a preceding value closest to it. Winsorization has greater advantages over the trimming technique in the data distribution namely: (1) it makes a replacement or an exchange for an outlier detected value with the closest value to the position where the outlier is located (2) the sample size of the data remains the same (3) it helps to prevent loss of information.

One of the recommended estimator as a better substitute for the trimmed mean is the *MOM* estimator that is capable of detecting the appearance of outliers in a data distribution (Yusof, Abdullah, Yahaya, & Othman, 2011). The *MOM* estimator does empirically trim only extreme data sets (Othma, Keselma, Padmanabhan, Wilcox, & Fradette, 2004). However, the main constraint in using the *MOM* estimator as a central tendency measure in Alexander-Govern test is that it fails to give an excellent control of Type I error rates when $g = 0.5$ and $h = 0.5$. This study uses the Winsorized modified one step *M*-estimator in Alexander-Govern test as its central tendency measure to strengthen its weakness under non-normality, in the presence of variance heterogeneity, for $g = 0.5$ and $h = 0.5$, to give a remarkable control of Type I error rates and to produce high power for the test.

METHODS

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

The procedure in obtaining the test statistic for the Alexander-Govern test begins by first ordering the data distribution, with population sizes of j ($j = 1, \dots, J$). In each of the data sets, the mean is calculated by using the formula below:

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

Where X_{ij} represents 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 Alexander-Govern (AG) test. 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 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, where $\sum w_j$ must be equal to 1. The weight (w_j) for each of the groups is calculated using the formula below:

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

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

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

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

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 is calculated 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 the formula:

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

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 $n_j - 1$ degrees of freedom. Where ν is the degree of freedom for each of the independent groups in the observed ordered data sets. The t statistic is calculated for each of the groups and is converted to standard normal deviates using the Hill's (1970) normalization approximation formula in the Alexander-Govern (1994) approach.

The formula is defined using:

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

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

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

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, at $\alpha = 0.05$ at $(j-1)$ chi-square degree of freedom is selected. 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.

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 $|X_j - M|, \dots, |X_n - M|$. Therefore, the median absolute deviation about the median (MAD_n) estimator is calculated using the formula:

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

As stated by Wilcox and Keselman (2003) the constant value of 0.6745 is used for rescaling the MAD estimator with the purpose of estimating the σ when taking samples 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 Wilcox and Keselman (2003) in detecting the appearance of outliers in a data distribution, because it has a very small standard error, when the sample of the data is normal. Equation (12) and (13) is also referred to as the MOM estimator that is used for detecting the appearance of outliers in a data set. In this research, the mean is replaced with the modified MOM estimator as a measure of the central tendency in the Alexander-Govern 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 using:

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

The $WMOM$ estimator is a replacement for mean as a central tendency measure in the Alexander-Govern test, due to several reasons. First, to remove the appearance of outliers from the data distribution. Second, to make the Alexander-Govern test to be robust to non-normal data.

The Winsorized sample variance is defined as:

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

Where 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: $x^{*1}, x^{*2}, \dots, x^{*B}$, for each of these random samples that consists of n data values that are 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 symbol (*) show that x^* is not the exact value of x , but it refers to a 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 Efron and Tibshirani (1998) 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:

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

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

The weight w_j for the Winsorized data distribution is defined using:

$$w_j = \frac{1 / S_e^2{}_{WMOMj}}{\sum_{j=1}^J 1 / S_e^2{}_{WMOMj}}, \quad (20)$$

Where $\sum_{j=1}^J 1 / S_e^2{}_{WMOMj}$ 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 $S_e^2{}_{WMOMj}$ is the standard error of the Winsorized data distribution and is defined using:

$$S_e^2{}_{WMOMj} = \frac{S_j^2{}_{WMOMj}}{n_j}, \quad (21)$$

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:

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

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 Hill's (1970) normalization approximation formula and the hypothesis testing for the Winsorized sample variance of the *WMOM* estimator for μ_j is expressed using:

$$H_o : \mu_1 = \mu_2 = \dots = \mu_j \quad \text{For } j = (j = 1, \dots, J)$$

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

The normalization approximation formula for the Alexander-Govern (AG) technique, for the *AGWMOM* test is defined using:

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

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

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

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 using:

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

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 using the standard chi-square distribution table. When the value of the test statistic for the *AGWMOM* is < 0.05 , the test is referred to as significant. Otherwise the test is considered not significant.

The variables used in this research are 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 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: Wilcox (1997)

The Type I error rates of the five different tests that were used in this research must fall under three criteria of robustness. They are (i) those tests that fall within the stringent criteria of robustness (ii) those tests that fall within the lenient criteria of robustness and (iii) those tests that do not fall on neither stringent criteria of robustness nor the lenient criteria of robustness and are regarded as not

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

RESULTS AND DISCUSSIONS

Table 2, 3, 4 and 5, define type I error rates for two groups condition. Next, Table 6, 7, 8, and 9 show type I error rates for four groups condition. Table 10, 11, 12 and 13 explain type I error rates for six groups condition. Within those tables, the bolded and italicized values are those values that falls strictly within the stringent criteria of robustness. The bolded values are those values that are within the lenient criteria of robustness. The un-bolded values are referred to as not robust.

Table 2 *AG* test, *AGMOM* test, *AGWMOM* test and *t*-test for the Type I Error Rates under Two Groups Condition, for $g = 0$ and $h = 0$

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

Table 3 *AG* test, *AGMOM* test, *AGWMOM* test and *t*-test for the Type I Error Rates for $g = 0$ and $h = 0.5$, for Two Groups Condition

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

Table 4 *AG* test, *AGMOM* test, *AGWMOM* test and *t*-test the Type I Error Rates for $g = 0.5$ and $h = 0$, for Two Groups Condition

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

Table 5 *AG* test, *AGMOM* test, *AGWMOM* test and *t*-test for the Type I Error Rates for $g = 0.5$ and $h = 0.5$, for Two Groups Condition

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

Table 6 *AG* test, *AGMOM* test, *AGWMOM* test and *t*-test for the Type I Error Rates for $g = 0$ and $h = 0$, for Four Groups Condition

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

Table 7 *AG* test, *AGMOM* test, *AGWMOM* test and *t*-test for the Type I Error Rates for $g = 0$ and $h = 0.5$, for Four Groups Condition

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

Table 8 *AG* test, *AGMOM* test, *AGWMOM* test and *t*-test for the Type I Error Rates for $g = 0.5$ and $h = 0$, for Four Groups Condition

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

Table 9 *AG* test, *AGMOM* test, *AGWMOM* test and *t*-test for the Type I Error Rates for $g = 0.5$ and $h = 0.5$, for Four Groups Condition

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

Table 10 *AG* test, *AGMOM* test, *AGWMOM* test and *t*-test for the Type I Error Rates for $g = 0$ and $h = 0$, for Six Groups Condition

Sample Size	Equal and Unequal Variance	<i>AG</i>	<i>AGMOM</i>	<i>AGWMOM</i>	<i>ANOVA</i>
20:20:20:20:20:20	1:1:1:1:1:1	0.0522	0.0440	0.0402	0.0530
	1:1:1:1:1:36	0.0522	0.0444	0.0406	0.1260
	1:4:4:16:16:36	0.0572	0.0448	0.0464	0.0810
2:4:4:16:32:62	1:1:1:1:1:1	0.1522	0.1864	0.1796	0.0640
	1:1:1:1:1:36	0.1434	0.1698	0.1724	0.0002
	36:1:1:1:1:1	0.1192	0.1432	0.1378	0.5992
	1:4:4:16:16:36	0.0920	0.0872	0.0926	0.0020
	36:16:16:4:4	0.1148	0.1454	0.1362	0.6878

Table 11 *AG* test, *AGMOM* test, *AGWMOM* test and *t*-test for the Type I Error Rates for $g = 0$ and $h = 0.5$, for Six Groups Condition

Sample Size	Equal and Unequal Variance	<i>AG</i>	<i>AGMOM</i>	<i>AGWMOM</i>	<i>ANOVA</i>
20:20:20:20:20:20	1:1:1:1:1:1	0.0260	0.1092	0.0266	0.0350
	1:1:1:1:1:36	0.0258	0.0186	0.0256	0.0922
	1:4:4:16:16:36	0.0248	0.0216	0.0288	0.0520
2:4:4:16:32:62	1:1:1:1:1:1	0.0794	0.1092	0.1092	0.0988
	1:1:1:1:1:36	0.0656	0.0450	0.0896	0.0040
	36:1:1:1:1:1	0.0796	0.0896	0.0982	0.3890
	1:4:4:16:16:36	0.0348	0.0486	0.0442	0.0130
	36:16:16:4:4:1	0.0898	0.0456	0.1008	0.0473

Table 12 *AG* test, *AGMOM* test, *AGWMOM* test and *t*-test for the Type I Error Rates for $g = 0.5$ and $h = 0$, for Six Groups Condition

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

Table 13 *AG* test, *AGMOM* test, *AGWMOM* test and *t*-test for the Type I Error Rates for $g = 0.5$ and $h = 0.5$, for Six Groups Condition

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

Across the distribution and across the whole group for both stringent and lenient criteria of robustness, the *AGWMOM* test produced 60 out of the total 84 conditions of pairing that is within the stringent and lenient criteria of robustness. The *AG* test has 56 out of 84 conditions of pairing that falls within the lenient and stringent criteria of robustness. The *AGMOM* test has 51 out of 84 conditions of pairing that are within the interval of stringent and lenient criteria of robustness. The *ANOVA* has a total of 34 out of 84 conditions of pairing that falls within the lenient and stringent criteria of robustness.

CONCLUSIONS

The *AGWMOM* test gave the best control of Type I error rates under non-normality, compared to the *AG* test, the *AGMOM* test and the *ANOVA* because the test always gives the highest number of conditions for both stringent and lenient criteria of robust.

REFERENCES

- 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.
- 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. When There Are Two Groups. *Journal of Educational and Behavioural Statistics*, 19(3), 275-291.
- Bradley, J. V. (1978). Robustness?. *British Journal of Mathematical and Statistical Psychology*, (31), 144-152.
- Brunner, e., Dette, H., & Munk, A. (1997). Box-Type Approximations in Nonparametric Factorial Designs. *Journal of the American Statistical Association*, 92(440), 1494-1502.
- Efron, B., & Tibshirani. (1998). *An introduction to the bootstrap*. New York: Chapman & Hall.
- Hill, G. W. (1970). Algorithm 395, Student's *t*-distribution. *Communication of the ACM*, 13, 67-619.
- James, G. S. (1951). The comparison of several groups of observations when the ratios of the population variances are unknown. *Biometrika*, 38, 324-329.
- 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.
- Kulinskaya, E., Staudte, R. G., & Gao, H. (2003). Power Approximations in Testing for Unequal Means in a One-Way ANOVA Weighted for Unequal Variances. *Communication in Statistics – Theory and Methods*, 32(12), 2353-2371. Doi: 10.1081/STA-12002538.
- Krishnamoorthy, K., Lu, 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.

- Lix, L. M, Keselman, J. C., & Keselman, H. J. (1998). To trim or not to trim. *Educational and Psychological Measurement*, 58(3), 409-429.
- 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 Computation, Computational*, 60, 207-222.
- Ochuko, T. K., Abdullah, S., Zain, Z., & Yahaya, S. S. S. (2015). Winsorized Modified One Step M-estimator in Alexander-Govern test. *Modern Applied Science*, 9(11), 51-67.
- Oshima, T. C., & Algina, J. (1992). Type I error rates for James's second-order test and Wilcoxon's Hm test under heteroscedaticity and non-normality. *British Journal of Mathematical and Statistical Psychology*, 45, 255-263.
- Othman, A. R., Keselman, H. J., Padmanabhan, A. R., Wilcox, R. R., & Fradette, K. (2004). Comparing measures of the "typical" score across treatments groups. *The British Journal of Mathematical and Statistical psychology*, 57(Pt 2), 215-234.
- 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.
- 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. (1988). A new alternative to the ANOVA F and new results on James's second-order method. *British Journal of Mathematical and Statistical Psychology*, 42, 203-213.
- Wilcox, R. R. (1997). *Introduction to robust estimation and hypothesis testing*. San Diego, CA: Academic Press.
- Wilcox, R. R., & Keselman, H. J. (2003). Modern Robust Data Analysis Methods: Measures of Central Tendency. *Psychological Methods*, 8(3), 254-274.
- Wilcox, R. R. (2003). Multiple comparisons based on a modified one-step M-estimator. *Journal of Applied Statistics*, 30, 1231-1241.
- Wilcox, R. R., & Keselman, H. J. (2000). 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 Strategy for TWO Groups case. *Modern Applied Science*, 5(4), 236-242.