Modern Applied Science; Vol. 9, No. 10; 2015
ISSN 1913-1844 E-ISSN 1913-1852
Published by Canadian Center of Science and Education

Winsorized Modified One Step M-estimator in Alexander-Govern Test

Tobi Kingsley Ochuko [1], Suhaida Abdullah[2], Zakiyah Binti Zain[3] & Sharipah Soaad Syed Yahaya[4]

[1, 2, 3, 4]College of Arts and Sciences, Universiti Utara Malaysia

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

Received: May 14, 2015 Accepted: July 3, 2015 Online Published: September 30, 2015

doi:10.5539/mas.v9n10p51 URL: http://dx.doi.org/10.5539/mas.v9n10p51

Abstract

This research centres on independent group test of comparing two or more means by using the parametric method, namely the Alexander-Govern test. The Alexander-Govern (*AG*) test uses mean as a measure of its central tendency. It is a better alternative to the Welch test, James test and the *ANOVA*, because it has a good control of Type I error rates and produces a high power efficient for a normal data under variance heterogeneity, but not for non-normal data. As a result, trimmed mean was applied on the test under non-normal data for two group condition, but as the number of groups increased above two, the test fails to be robust. Due to this, when the *MOM* estimator was applied on the test, it was not influenced by the number of groups, but failed to give a good control of Type I error rates under skewed heavy tailed distribution. In this research, the Winsorized *MOM* estimator was applied in *AG* test as a measure of its central tendency. 5,000 data sets were simulated and analysed using Statistical Analysis Software (*SAS*). The result shows that with the pairing of unbalanced sample size with unequal variance of (1:36) and the combination of both balanced and unbalanced sample sizes with both equal and unequal variances, under six group condition, for g = 0.5 and h = 0.5, for both positive and negative pairing condition, the test gives a remarkable control of Type I error rates. In overall, the *AGWMOM* test has the best control of Type I error rates, across the distributions and across the groups, compared to the *AG* test, the *AGMOM* test and the *ANOVA*.

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

1. Introduction

This research focuses on comparing the performance of the Type I error rates and power of the *AG* test, the *AGMOM* test, the *AGWMOM* test, the *t-test* and the *ANOVA*, for two, four and six group conditions. In order to see of the five tests, which one of the tests will give good control of Type I error rates and also produce high power efficient, under skewed heavy tailed distribution. The independent group tests such as the analysis of variance (*ANOVA*) methods have been employed in many areas, for example, in medicine, economics, sociology and agriculture, as discussed by Pardo, Pardo, Vincente and Esteban (1997). Several assumptions have to be satisfied before the method can perform effectively, which are (i) homogeneity of variances, (ii) normal distribution of the data and (iii) independent observations. The *ANOVA* is a classical method of analysis that 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 distribution and is robust to small deviations from normality, especially when the sample size is large enough to guarantee normality as stated by Wilcox (1997, 2003).

Yusof, Abdullah, Yahaya and Othman (2011) observed that the two major problems confronting the analysis of variance are the presence of non-normality and variance heterogeneity in a data distribution. As a result of this, Type I error rates is increased and there is a reduction in the power efficiency of the test. When the data distribution is seen to be heavy tailed, the standard error of the mean can be badly increased (Wilcox & Keselman, 2002).This makes the standard error of the *ANOVA* to be larger than it ought to be and the power of the test would be reduced. To obtain a good test, Type I error rates should be controlled and likewise the power of the test. This implies that neither should Type I error rates be increased nor should there be a lost in the power of the test.

The *ANOVA* is very sensitive to the assumptions of homogeneity of variance such that when there is a violation, the outcome 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 genuineness of the results (Brown & Forsythe, 1974; Wilcox, Charlin & Thompson, 1986). The problem of heterogeneity of variance has been discussed by few researchers and some alternatives were proposed. Welch (1951) introduced the Welch test that is used to test for the hypothesis of two populations having equal means. It was mentioned in different literatures as an alternative to the *ANOVA* (Algina, Oshima & Lin, 1994; Keselman, 1982; Lix, Keselman, & Keselman, 1996; Wilcox et al., 1986).

James (1951) also introduced a better solution for *ANOVA* under heterogeneity of variance, namely the James test. This test is used for weighing sample means and is recommended by different researchers (Lix et al., 1996; Oshima & Algina, 1992; Wilcox, 1988). When the sample size is small, and the data distribution is non-normal, the James test fails to give a good control of Type I error rates. Both the Welch test and the James test are used for analysing a data distribution that is non-normal with unequal variance (Brunner, Dette, & Munk, 1997; Kohr & Games, 1974; Krishnamorthy, Lu, & Mathew, 2007; Wilcox & Keselman, 2003).

The Alexander-Govern test was introduced in (1994) as to deal with heterogeneity of variance under the condition of normality, but is a test that is not robust to non-normal data. The Alexander-Govern test has been compared with the Welch test and the James test and it was agreed by Schneider and Penfield (1997) and Myers (1998) that the Alexander-Govern test and the James test were better under most conditions being studied. The performance of the Alexander-Govern test was comparable to the James test, and the calculation of the Alexander-Govern test was simpler than that of the James test.

This finding was agreed by Myers (1998) suggesting that the Alexander-Govern test provides a good solution to the problem of variance heterogeneity. Although, the Alexander-Govern test is a good alternative to the *ANOVA* when it comes to variance heterogeneity, this test still has some drawbacks. The main weakness of the test is that it cannot handle a deviation from normality, as proven by Myers (1998). This method can acceptably put under control Type I error rates when there is heterogeneity in the variances and when the distribution of the data is normal.

Lix and Keselman (1998) proposed an alternative approach by substituting the common mean with trimmed mean in a few robust test statistics which improved the performance of the tests under non-normal data. The trimmed mean and Winsorized variance are widely used as alternative to the common mean and variance respectively, due to some good properties, such as having a remarkable control over Type I error rates and the power of the test, when there is a violation under the assumptions of homogeneity of the variance and when the distribution is normal (Wilcox, 1995).

Trimmed mean is calculated by averaging only the centred data after discarding a particular amount of the percentage of the largest and the smallest data value, while its variance is estimated by Winsorized variance. In applying trimmed mean in a data distribution, it possesses some disadvantages, which are firstly: the percentage of trimming is placed at prior, resulting in the elimination process performed regardless of the shape of the distribution. Secondly in trimming process, it should be done carefully, to minimize loss of information. Thirdly, it cannot handle large size of extreme values (Yahaya, Othman & Keselman, 2006).

One of the suggested estimator as a replacement for the trimmed mean is a better alternative referred to as the modified one-step M-estimator (*MOM*), which is able to identify the presence of outliers in a data distribution (Yusuf, Abdullah, Yahaya & Othman, 2011). It empirically trims only extreme data (Othman, Keselman, Padmanabhan, Wilcox & Fradette, 2004). The main disadvantage of using the *MOM* estimator as a measure of the central tendency, for example in Alexander-Govern test, is that it cannot control the error rates in the test when the level of skewness and kurtosis is at peak. In this research, the Winsorized *MOM* was applied in Alexander–Govern test under variance heterogeneity for non-normal data, under a skewed heavy tailed distribution, and it gave the test a remarkable control of Type I error rates and produced a high power efficient for test.

2. Method

2.1 The Alexander-Govern Test

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

2.1.1. The Alexander-Govern Test Statistic

The procedure to calculate the test statistic for the Alexander-Govern test starts by first ordering the data set, with population *j* (*j* = 1, …, *J*). Then for each of the data set, the mean is calculated using:

$$\bar{X} = \frac{\sum_j X_{ij}}{n_j} \tag{1}$$

where

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

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

where

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

$$S_{ej} = [\frac{s^2{}_j}{n_j}]^{1/2} \tag{3}$$

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

$$w_j = \frac{1/S^2{}_{e\ j}}{\sum_j 1/S^2{}_{e\ j}} \tag{4}$$

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

$$H_o : \mu_1 = \ldots \mu_j$$
$$H_A : \mu_1 \neq \ldots \mu_j \,, j = 1, \ldots, J \tag{5}$$

The alternative hypothesis contradicts the claim or statement made by the null hypothesis. The variance weighted estimate of the total mean for all the groups in the data distribution is calculated using:

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

where

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

$$t_j = \frac{\bar{X}_j - \hat{\mu}}{S_{e\ j}} \tag{7}$$

where

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

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

Where, ν is the degree of freedom for each of the groups in the order sample data. The *t* statistic calculated for each group is converted to a standard normal deviates by using the Hill's (1970) normalization approximation in the Alexander-Govern (1994) procedure. The formula is expressed as:

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

where

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

where

$$\nu_j = n_j - 1, a = \nu_j - 0.5, b = 48a^2 \tag{10}$$

The test statistic for the Alexander-Govern procedure is expressed as:

$$A = \sum_{=1j}^{J} Z^2{}_j \tag{11}$$

After obtaining the test statistic for the Alexander-Govern test, we select a significance level of $\alpha = 0.05$ with *J* – 1 chi-square degrees of freedom. The p-value is also obtained for the Alexander-Govern method. The standard chi-square distribution is used to calculate the critical value of the test. After which, we compare our chi-square value with the test statistic value of the Alexander-Govern test. If $\underline{A}$ calculated is greater than the chi-square value from the chi-square distribution table, then we reject H_O and conclude that the means of the independent groups are different from each other, otherwise, we do not.

3. The Modified Alexander-Govern Test

In this research, we modified the mean as a central tendency measure in the Alexander-Govern test by replacing it with the Winsorized modified one step M-estimator (*WMOM*) as a measure of its central tendency. The Winsorized *MOM* estimator is applied on the data distribution where the outlier detected value is replaced or exchanged with a preceding value closest to the position where the outlier is located. The Winsorized *MOM* estimator is obtained by using the formula:

$$WMOM = \bar{X}_{WMOM_j} = \frac{\sum_{j=1}^{J} X_{WMOMj}}{n} \tag{12}$$

where,

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

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

i. To remove the presence 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 expressed as:

$$S^2_{WMOMj} = \frac{\sum_{j=1}^{J}(X_j - \bar{X}_{WMOMj})^2}{n-1} \tag{13}$$

where

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

The standard error of the *WMOM* is obtained by using the bootstrapping methods in order to estimate the standard error and the bootstrapping algorithm for estimating the standard errors is defined below:

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

$$\begin{aligned} x^* &= (x_1^*, x_2^*, ..., x_n^*) \\ \hat{F} &\rightarrow (x_1^*, x_2^*, ..., x_n^*) \end{aligned} \tag{14}$$

The indication of the star shows that x^* is not the real data set 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). In this research, we made use of 50 amount of the bootstrap samples is sufficient to give reasonable estimate of the standard error of the *MOM* estimator (Efron & Tibshirani, 1998).

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

$$\hat{\theta}^*(b) = s(x^{*b}), b = 1, 2, ..., B. \tag{15}$$

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

$$\hat{se}_B = \{\sum_{b=1}^{B}[\hat{\theta}^*(b) - \theta^{/*}(\cdot)]^2 / (B-1)\}^{1/2}, \tag{16}$$

where

$$\hat{\theta}^*(\cdot) = \sum_{b=1}^{B} \hat{\theta}^{\bullet}(b) / B. \tag{17}$$

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

$$w_j = \frac{1/S_e^2{}_{WMOMj}}{\sum_{j=1}^{J} 1/S_e^2{}_{WMOM\,j}} \tag{18}$$

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 groups in the ordered data distribution, from the real data distribution.

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

$$\hat{\mu}_j = \sum_{j=1}^{J} w_j \bar{X}_{WMOMj} \tag{19}$$

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

$$t_j = \frac{\bar{X}_{WMOMj} - \hat{\mu}}{S_{eWMOMj}} \tag{20}$$

where

S_{eWMOMj} is the Winsorized sample standard error from the Winsorized data distribution for each of the independent group of $\bar{X}_{WMOMj}$.

In the Alexander-Govern (1994) method, the t_j value is converted to standard normal by using the Hill's (1970) normalization approximation and the hypothesis testing of the Winsorized data distribution, where S^2_{WMOMj} is the usual estimate of the Winsorized sample variance of the *WMOM* estimator for μ_j is expressed as:

$$\begin{aligned} &H_O : \mu_1 = \ldots = \mu_J \\ &H_O : \mu_1 \neq \ldots \neq \mu_J \text{ , } for\ j = (j = 1, \ldots, J). \end{aligned} \tag{21}$$

Thus, the normalization approximation formula for the Alexander-Govern method, using the Winsorized data distribution is expressed as *AGWMOM*. The test statistic of the *AGWMOM* for all the groups in the ordered data sample is expressed as:

$$AGWMOM = \sum_{j=1}^{J} Z^2_{WMOMj} \tag{22}$$

The test statistic for the *AGWMOM* test follows a chi-square distribution at $\alpha = 0.05$ level of significance, having *J* – 1 chi-square degrees of freedom. The p- value can be determined using a chi-square distribution table.

3.1 The Variables Used in this Research

There are five different types of variables that were used in this research. They are balanced and unbalanced sample size, equal and unequal variance, group sizes, types of distribution and nature of pairing. All these variables mentioned above, were manipulated to show the strength and the weakness of the original *AG* test, the *AGMOM* test, the *AGWMOM* test, the *t-test* and the *ANOVA*.

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

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

4. The Research Design

The Alexander-Govern test is a test that is not robust for non-normal data under variance heterogeneity. For the design of this research, both balanced and unbalanced sample size 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.

For each of the test namely: the *AG* test, the *AGMOM* test, the *AGWMOM* test, the *t-test* and the *ANOVA*, 5,000 data sets were simulated in the research design. To obtain the pseudo random variates, *SAS* generator *RANNOR* (*SAS, Institute*, 1999) was used with a nominal level of $\alpha = 0.05$ for the analysis of the tests in this research. The robustness of the Winsorized modified one step M-estimator was obtained by manipulating the five listed variables as mentioned previously with regards to the Type I error rates and power of the test.

Table 2. Research Design for Two Groups

The g- and h-distribution	Balanced and Unbalanced sample size	Equal and Unequal variance	Nature of Pairing	Notations for the Conditions 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. Research Design for Four Groups

The g- and h-distribution	Balanced and Unbalanced sample size	Variance ratio	Nature of Pairing	Notations for the condition of Pairing
g = 0 and h = 0	20:20:20:20	1:1:1:1	Balanced condition	C21
		1:1:1:36	Positive Pairing	C22
		1:4:16:36	Positive Pairing	C23
	15:15:20:30	1:1:1:1		C24
		1:1:1:36	Positive Pairing	C25

		36:1:1:1	Negative Pairing	C26
		1:4:16:36	Positive Pairing	C27
		36:16:4:1	Negative Pairing	C28
g = 0 and h = 0.5	20:20:20:20	1:1:1:1	Balanced condition	C29
		1:1:1:36	Positive Pairing	C30
		1:4:16:36	Positive Pairing	C31
	15:15:20:30	1:1:1:1		C32
		1:1:1:36	Positive Pairing	C33
		36:1:1:1	Negative Pairing	C34
		1:4:16:36	Positive Pairing	C35
		36:16:4:1	Negative Pairing	C36
g = 0.5 and h = 0	20:20:20:20	1:1:1:1	Balanced condition	C37
		1:1:1:36	Positive Pairing	C38
		1:4:16:36	Positive Pairing	C39
	15:15:20:30	1:1:1:1		C40
		1:1:1:36	Positive Pairing	C41
		36:1:1:1	Negative Pairing	C42
		1:4:16:36	Positive Pairing	C43
		36:16:4:1	Negative Pairing	C44
g = 0.5 and h = 0.5	20:20:20:20	1:1:1:1	Balanced condition	C45
		1:1:1:36	Positive Pairing	C46
		1:4:16:36	Positive Pairing	C47
	15:15:20:30	1:1:1:1		C48
		1:1:1:36	Positive Pairing	C49
		36:1:1:1	Negative Pairing	C50
		1:4:16:36	Positive Pairing	C51
		36:16:4:1	Negative Pairing	C52

Table 4. Research Design for Six Groups

The g- and h-distribution	Balanced and Unbalanced sample size	Variance ratio	Nature of Pairing	Notations for the Condition of Pairing

g = 0 and h = 0	20:20:20:20:20:20	1:1:1:1:1:1	Balanced condition	C53
		1:1:1:1:1:36	Positive Pairing	C54
		1:4:4:16:16:36	Positive Pairing	C55
	2:4:4:16:32:62	1:1:1:1:1:1		C56
		1:1:1:1:1:36	Positive Pairing	C57
		36:1:1:1:1:1	Negative Pairing	C58
		1:4:4:16:16:36	Positive Pairing	C59
		36:16:16:4:4:1	Negative Pairing	C60
g = 0 and h = 0.5	20:20:20:20:20:20	1:1:1:1:1:1	Balanced condition	C61
		1:1:1:1:1:36	Positive Pairing	C62
		1:4:4:16:16:36	Positive Pairing	C63
	2:4:4:16:32:62	1:1:1:1:1:1		C64
		1:1:1:1:1:36	Positive Pairing	C65
		36:1:1:1:1:1	Negative Pairing	C66
		1:4:4:16:16:36	Positive Pairing	C67
		36:16:16:4:4:1	Negative Pairing	C68
g = 0.5 and h = 0	20:20:20:20:20:20	1:1:1:1:1:1	Balanced 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
	2:4:4:16:32:62	1:1:1:1:1:1		C80
		1:1:1:1:1:36	Positive Pairing	C81
		36:1:1:1:1:1ssssssss	Negative Pairing	C82

	1:4:4:16:16:36	Positive Pairing	C83
	36:16:16:4:4:1	Negative Pairing	C84

This research design was used to determine the robustness of the modified *AG* test. By using this research design, the best procedure was obtained. As stated by Lix and Keselman (1998) the empirical rate of Type I error must fall within the interval of $0.042 \le \hat{\alpha} \le 0.058$ to judge the robustness of a given test at α level of significance. The range of the value selected for this research gave a strict condition for the robustness of the tests within this interval, with the aim of producing minimum error rate with deviation from model assumptions.

According to Abdullah, Yahaya and Othman (2007) in their research used the range of 0.042 and 0.058 to evaluate the robustness of the tests for their analysis. The interval selected by these researchers, reveal that a test is said to be robust if its Type I error rates falls within the stringent criterion of robustness, otherwise if the test falls outside this interval to Judge the robustness of the test, then the Type I error rates is out of control. According to Bradley's (1978) the lenient criteria of robustness must fall within the range of (0.025 – 0.075).

This interval of robustness is also selected in this research, to see those tests that can give remarkable control over the probability of Type I error rates.

In the Table for the Type I error rates for two, four and six group condition, the bolded values represent those values that fall within the stringent criteria of robustness. The un-bolded values represent those values that fall within the lenient criteria of robustness. The red-coloured values represent those values that are not robust at all. They neither fall within the stringent criteria of robustness nor within the lenient criteria of robustness.

Table 5. Comparison of the Type I error rates for the *AG* test, *AGMOM* test, *AGWMOM* test and *t-test* under normal distribution for two group condition.

The g- and h-distribution	**Balanced and Unbalanced Sample Size**	**Equal and Unequal Variance**	***AG***	***AGMOM***	***AGWMOM***	***t-test***
g = 0 and h = 0		1:1	**0.0508**	0.0414	0.0392	**0.0528**
	20:20	1:36	**0.0562**	**0.0528**	**0.0496**	0.0710
	16:24	1:1	**0.0484**	**0.0430**	0.0386	**0.0570**
		1:36	**0.0570**	**0.0552**	**0.0496**	0.0618
		36:1	**0.0498**	**0.0450**	**0.0438**	0.1078

In Table 5, for g = 0 and h = 0, in all the conditions of pairing, the *AG* test, the *AGMOM* test and the *AGWMOM* test have equal number of Type I error rates that are considered robust compared to the *t-test*. The *t-test* have two of its Type I error rates fall within the stringent criteria of robustness and only one of its Type I error rates fall within the lenient criteria of robustness, with the combination of balanced sample size with equal variance and the pairing of unbalanced sample size with both equal and unequal variance. The remaining two of its Type I error rates are regarded as not robust.

Table 6. Comparison of the Type I error rates for the *AG* test, the *AGMOM* test, the *AGWMOM* test and the *t-test* under a symmetric heavy tailed distribution for two group condition.

The g- and h-distribution	**Balanced and Unbalanced Sample Size**	**Equal and Unequal Variance**	***AG***	***AGMOM***	***AGWMOM***	***t-test***
g = 0 and h = 0.5	20:20	1:1	0.0336	0.0262	0.0346	0.0356
		1:36	0.0340	0.0358	0.0392	0.0402
	16:24	1:1	0.0304	0.0266	0.0352	**0.0430**
		1:36	0.0394	0.0340	0.0412	0.0138
		36:1	0.0312	0.0294	0.0346	0.0814

In Table 6, for g = 0 and h = 0.5, in all the conditions of pairing, the *AG*, *AGMOM* and the *AGWMOM* test all have their Type I error rates fall within the lenient criteria of robustness and the three tests are said to be robust between the interval of 0.025 and 0.075. While the *t-test* has just one of its Type I error rate, fall within the stringent criteria of robustness, two of its Type I error rates fall within the lenient criteria of robustness and the remaining two of its Type I error rates are considered not robust.

Table 7. Comparison of the Type I error rates for the *AG* test, the *AGMOM* test, the *AGWMOM* test and the *t-test* under a skewed normal tailed distribution for two group condition.

The g- and h-distribution	Balanced and Unbalanced Sample Size	Equal and Unequal Variance	*AG*	*AGMOM*	*AGWMOM*	*t-test*
g = 0.5 and h = 0	20:20	1:1	**0.0508**	**0.0420**	0.0364	**0.0474**
		1:36	**0.0562**	**0.0534**	**0.0558**	0.0882
	16:24	1:1	**0.0480**	**0.0434**	0.0386	**0.0570**
		1:36	**0.0570**	**0.0560**	0.0588	0.0380
		36:1	**0.0498**	**0.0504**	**0.0450**	0.1538

In Table 7, for g = 0.5 and h = 0, in all the conditions of pairing, the *AG* test, the *AGMOM* test and the *AGWMOM* test are more robust compared to the *t-test*. The *AG* test and the *AGMOM* test have all of their Type I error rates fall within the stringent criteria of robustness. While the *AGWMOM* test have two of its Type I error rates fall within the stringent criteria of robustness and the remaining three of its Type I error rates fall within the lenient criteria of robustness. The *t-test* have two of its Type I error rates fall within the stringent of robustness, only one of its Type I error rates fall within the lenient criteria of robustness and the remaining two of its Type I error rates are considered not to be robust.

Table 8. Comparison of the Type I error rates for the *AG* test, the *AGMOM* test, the *AGWMOM* test and the *t-test* under a skewed heavy tailed distribution for two group condition.

The g- and h-distribution	Balanced and Unbalanced Sample Size	Equal and Unequal Variance	*AG*	*AGMOM*	*AGWMOM*	*t-test*
g = 0.5 and h = 0.5	20:20	1:1	0.0336	0.0258	0.0314	0.0288
		1:36	0.3400	0.0374	0.0370	**0.0430**
	16:24	1:1	0.0274	0.0272	0.0352	0.0370
		1:36	0.0394	0.0378	**0.0422**	0.0138
		36:1	0.0312	0.0332	0.0298	0.0878

In table 8, for g = 0.5 and h = 0.5, in all the conditions of pairing, the *AG*, the *AGMOM* and the *AGWMOM* test have equal number of Type I error rates that are regarded as robust, compared to the *t-test*. The *t-test* only has one of its Type I error rate fall within the stringent criteria of robustness and the other one of its Type I error rates fall within the lenient criteria of robustness. The remaining three of its Type I error rates are considered not robust.

Table 9. Comparison of the Type I error rates for the *AG* test, the *AGMOM* test, the *AGWMOM* test and the *ANOVA* under a normal distribution for four group condition.

The g- and h-distribution	Balanced and Unbalanced Sample Size	Equal and Unequal Variance	*AG*	*AGMOM*	*AGWMOM*	*ANOVA*
g = 0 and h = 0	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

In table 9, for g = 0 and h = 0, in all the conditions of pairing, the *AG* test, the *AGMOM* test and the *AGWMOM* have equal number of Type I error rates that are said to be robust, compared to the *ANOVA*. The *ANOVA* has two of its Type I error rates fall within the stringent criteria of robustness, the other two of its Type I error rates fall within the lenient criteria of robustness. The remaining four of its Type I error rates are said not to be robust.

Table 10. Comparison of the Type I error rates for the *AG* test, the *AGMOM* test, the *AGWMOM* test and the *ANOVA* under a symmetric heavy tailed distribution, for four group condition.

The g- and h-distribution	**Balanced and Unbalanced Sample Size**	**Equal and Unequal Variance**	***AG***	***AGMOM***	***AGWMOM***	***ANOVA***
g = 0 and h = 0.5	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	0.0260	0.0344
		1:1:1:36	0.0238	0.0212	0.0272	0.0182
		36:1:1:1	0.0208	0.0192	0.0264	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

In Table 10, for g = 0 and h = 0.5, the *AGWMOM* test is more robust compared to the *AG* test, the *AGMOM* test and the *ANOVA*. The *AGWMOM* test have five of its Type I error rates fall within the lenient criteria of robustness, with the combination of balanced sample size with unequal variance and the pairing of unbalanced sample size with equal variance and the pairing of unbalanced sample size with unequal variance, positively and negatively. For the *AG* test, three of its Type I error rates fall within the lenient criteria of robustness and are said to be robust. The *AGMOM* test have only two of its Type I error rates fall within the lenient criteria of robustness, the remaining six of its Type I error rates are considered not robust. The *ANOVA* have only one of its Type I error rate fall within the stringent criteria of robustness, the other two of its Type I error rates fall within the lenient criteria of robustness. While the remaining five of its Type I error rates are considered not robust.

Table 11. Comparison of the Type I error rates for the *AG* test, the *AGMOM* test, the *AGWMOM* test and the *ANOVA* under a skewed normal tailed distribution for four group condition.

The g- and h-distribution	**Balanced and Unbalanced Sample Size**	**Equal and Unequal Variance**	***AG***	***AGMOM***	***AGWMOM***	***ANOVA***
g = 0.5 and h = 0	20:20:20:20	1:1:1:1	0.0620	**0.0436**	**0.0452**	**0.0550**
		1:1:1:36	0.0620	**0.0460**	**0.0472**	0.1714
		1:4:16:36	0.0756	**0.0546**	**0.0502**	0.1098
	15:15:20:30	1:1:1:1	0.0596	**0.0460**	**0.0466**	**0.0508**
		1:1:1:36	0.0872	**0.0448**	**0.0520**	0.0756
		36:1:1:1	0.0602	**0.0482**	**0.0520**	0.2330
		1:4:16:36	0.0928	**0.0502**	**0.0550**	**0.0444**
		36:16:4:1	0.0646	**0.0560**	**0.0462**	0.1954

In Table 11, both the *AGMOM* test and the *AGWMOM* test are more robust compared to the *AG* test and the *ANOVA*. For the *AGMOM* and the *AGWMOM* test, all their Type I error rates fall within the stringent criteria of robustness, in all the conditions of pairing. The *AG* test has five of its Type I error rates fall within the lenient criteria of robustness, with the combination of balanced sample size with both equal variance and unequal variance. Also, with the pairing of unbalanced sample with equal variance and unbalanced sample size with unequal variance for negative pairing condition only. For the *ANOVA*, three of its Type I error rates fall within the stringent criteria of robustness and the remaining five of its Type I error rates are regarded as not robust.

Table 12. Comparison of the Type I error for the *AG*, the *AGMOM*, the *AGWMOM* and the *ANOVA* under skewed heavy tailed distribution for four group condition.

The g- and h-distribution	Balanced and Unbalanced Sample Size	Equal and Unequal Variance	*AG*	*AGMOM*	*AGWMOM*	*ANOVA*
g = 0.5 and h = 0.5	20:20:20:20	1:1:1:1	0.0322	0.0206	0.0298	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	0.0226	0.0274	0.0240
		36:1:1:1	0.0272	0.0200	0.0266	0.1394
		1:4:16:36	0.0360	0.0266	0.0320	0.0164
		36:16:4:1	0.0266	0.0256	0.0284	0.1130

In Table 12, for g = 0.5 and h = 0.5, the *AG* test is more robust compared to the *AGMOM* test, the *AGWMOM* test and the *ANOVA*. The *AG* test have all its Type I error rates fall within the lenient criteria of robustness, in all the conditions of pairing. The *AGMOM* test have five of its Type I error rates fall within the lenient criteria of robustness and the remaining three of its Type I error rates are regarded as not robust. The *AGWMOM* test have six of its Type I error rates fall within the lenient criteria of robustness, with the combination of balanced sample size with unequal variance and the pairing of unbalanced sample size with equal variance. Also, with the pairing of unbalanced sample size with unequal variance, positively and negatively. The *ANOVA* has only one of its' Type I error rates fall within the stringent criteria of robustness, the other one of its Type I error rates fall within the lenient criteria of robustness. The remaining six of its Type I error rates are considered not robust.

Table 13. Comparison of the Type I error rates for the *AG* test, the *AGMOM* test, the *AGWMOM* test and the *ANOVA*, for six group condition, under a normal distribution

The g- and h-distribution	Balanced and Unbalanced Sample Size	Equal and Unequal Variance	*AG*	*AGMOM*	*AGWMOM*	*ANOVA*
g = 0 and h = 0	20:20:20:20:20:20	1:1:1:1:1:1	**0.0522**	**0.0440**	0.0402	**0.0530**
		1:1:1:1:1:36	**0.0522**	**0.0444**	0.0406	0.1260
		1:4:4:16:16:36	**0.0572**	**0.0488**	**0.0464**	0.0810
	2:4:4:16:32:62	1:1:1:1:1:1	0.1522	0.1864	0.1796	**0.0540**
		1:1:1:1:1:36	0.1434	0.1698	0.1724	0.0002
		36:1:1:1:1:1	0.1192	0.1432	0.1378	0.5992
		1:4:4:16:16:36	0.0920	0.0872	0.0926	0.0020
		36:16:16:4:4:1	0.1148	0.1454	0.1362	0.6878

In Table 13, for g = 0 and h = 0, the four different tests have equal number of Type I error rates that are said to be robust. With the combination of balanced sample size with both equal and unequal variance, the *AG* test and *AGMOM* test have three of their Type I error rates fall within the stringent criteria of robustness. While the remaining five of their Type I error rates are said not to be robust. The *AGWMOM* test has only one of its Type I error rate fall within the stringent criteria of robustness, the other two of its Type I error rates fall within the lenient of robustness. The remaining five of its Type I error rates are considered not robust. The *ANOVA* have two of its Type I error rates fall within the stringent criteria of robustness, the other one of its Type I error rates fall within the lenient criteria of robustness. The remaining five of its Type I error rates are regarded as not robust.

Table 14. The Comparison of the Type I error rates, for the *AG* test, *AGMOM* test, the *AGWMOM* test and the *ANOVA,* under symmetric heavy tailed distribution, for six group condition.

The g- and h-distribution	Balanced and Unbalanced Sample	Equal and Unequal	*AG*	*AGMOM*	*AGWMOM*	*ANOVA*

	Size	Variance				
g = 0 and h = 0.5	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	0.0950	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:!6:4:4:1	0.0898	0.0956	0.1008	0.4732

In Table 14, for g = 0 and h = 0.5, the *AG* test is more robust compared to the *AGMOM* test, the *AGWMOM* test and the *ANOVA*. The *AG* test have four of its Type I error rates fall within the lenient criteria of robustness, with the combination of balanced sample size with both equal and unequal variance and the pairing of unbalanced sample size with unequal variance, for positive pairing only. The *AGMOM* test has only one of its Type I error rate fall within the stringent criteria of robustness. The remaining seven of its Type I error rates are considered not robust. For the *AGWMOM* test, only one of its Type I error rates fall within the stringent criteria of robustness, the other two of its Type I error rates fall within lenient criteria of robustness. The remaining five of its Type I error rates are regarded as not robust. The *ANOVA* have three of its Type I error rates fall within the lenient criteria of robustness, the remaining five of its Type I error rates are said not to be robust.

Table 15. Comparison of the Type I error rates for the *AG* test, the *AGMOM* test, the *AGWMOM* test and the *ANOVA* under a skewed normal tailed distribution for six group condition

The g- and h-distribution	Balanced and Unbalanced Sample Size	Equal and Unequal variance	*AG*	*AGMOM*	*AGWMOM*	*ANOVA*
g = 0.5 and h = 0	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**	0.0270
		1:4:4:16:16:36	0.0860	**0.0576**	**0.0514**	0.1184
	2:4:4:16:32:62	1:1:1:1:1:1	0.2080	0.1944	0.2118	0.0670
		1:1:1:1:1:36	0.2734	0.1692	0.2188	0.0060
		36:1:1:1:1:1	0.1678	0.1600	0.1740	0.5692
		1:4:4:16:16:36	0.2514	0.0880	0.1430	0.0034
		36:16:16:4:4:1	0.1418	0.1636	0.1620	0.6722

In Table 15, for g = 0.5 and h = 0, the *AGMOM* test, the *AGWMOM* test and the *ANOVA* have equal number of Type I error rates that are said to be robust, compared to the *AG* test. For both the *AGMOM* and the *AGWMOM* test, with the combination of balanced sample size with both equal and unequal variance, their Type I error rates fall within the stringent criteria of robustness. The *AG* test have two of its Type I error fall within the lenient criteria of robustness and the remaining six of its Type I error rates are regarded as not robust.

The *ANOVA* has only one of its Type I error rate fall within stringent criteria of robustness and the other two of its Type I error rates fall within the lenient criteria of robustness. While the remaining five of the Type I error rates of the *ANOVA* are considered not robust.

Table 16. Comparison of the Type I error rates for the *AG* test, the *AGMOM* test, the *AGWMOM* test and the *ANOVA* for six group condition, under a skewed heavy tailed distribution

The g- and h-distribution	Balanced and Unbalanced Sample Size	Equal and Unequal Variance	*AG*	*AGMOM*	*AGWMOM*	*ANOVA*
g = 0.5 and h = 0.5	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	0.0386	0.0186	0.0292	0.1028
		1:4:4:16:16:36	0.0400	0.0246	0.0300	**0.0574**
	2:4:4:16:32:62	1:1:1:1:1:1	0.1212	0.1136	0.1200	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

In Table 16, for g = 0.5 and h = 0.5, the *AGWMOM* test is more robust compared to the *AG* test, the *AGMOM* test and the *ANOVA*. The *AGWMOM* test have three of its Type I error rates fall within the lenient criteria of robustness and only one of its Type I error rate fall within the stringent criteria of robustness, with the combination of balanced sample size with both equal and unequal variance and the pairing of unbalanced sample size with unequal variance, for positive pairing only. The *AG* test have three of its Type I error rates fall within the lenient criteria of robustness and the remaining five of its Type I error rates are regarded as not robust. The *AGMOM* test has only one of its Type I error rate fall within the stringent criteria of robustness and the remaining seven of its Type I error rates are considered not robust. The *ANOVA* has only one of its Type I error rate fall within the stringent criteria of robustness and the other one of its Type I error rate fall within the lenient criteria of robustness. The remaining six of its Type I error rates are regarded as not robust.

Table 17. Overall Summary of the Type I error rates for the five tests, for all the distributions and group sizes:

Notation for the robustness of the tests	*AG*	*AGMOM*	*AGWMOM*	*t-test*/*ANOVA*
SR	21	32	26	17
LR	35	19	34	17
NR	28	33	24	50
TOTAL	84	84	84	84

In Table 17, SR represents the number of each test that fall within the stringent criteria of robustness, LR represents the number of each test that fall within the lenient criteria of robustness and NR represents the number of each test that are considered not robust.

Note. We have eighty-four conditions of pairing from the research design i.e C1 – C84 to determine the Type I error rates of each of the tests in this research.

5. Discussion and Conclusion

In this research, the stringent criteria of robustness to give a strict control of Type I error rates must fall within the interval of 0.042 – 0.058, as stated by Lix and Keselman (1998) and the lenient criteria of robustness must fall within the interval of 0.025 – 0.075, as discussed by Bradley's (1978). These two conditions were selected in this research to give a strict control of the Type I error rates for the test.

In Table 17, the overall results of the Type I error rates shows 56 out of 84 of the Type I error rates of the *AG* test are said to be robust, where 21 of its Type I error rates fall within the stringent criteria of robustness, 35 of its Type I error rates fall within the lenient criteria of robustness and 28 of its Type I error rates are regarded as not robust. The *AGMOM* test has 51 out 84 of its Type I error rates that are said to be robust, where 32 of its Type I error rates fall within the stringent criteria of robustness, 19 of its Type I error rates fall within the lenient criteria of robustness and 33 of its Type I error rates are considered not robust. The *AGWMOM* test has 60 out of 84 of its Type I error rates are considered to be robust, where 26 of its Type I error rates fall within the stringent criteria of robustness, 34 of its Type I error rates fall within the lenient criteria of robustness and 24 of its Type I error rates are regarded as not robust. The *ANOVA* has 34 out of 84 of its Type I error rates that are said to be robust, where 17 of its Type I error rates fall within the stringent criteria of robustness, 17 of its Type I error rates fall within the lenient criteria of robustness and the remaining 50 of its Type I error rates are regarded as not robust.

In all, the *AGWMOM* test has the highest number of Type I error rates that are regarded as robust. Hence, the *AGWMOM* test is the best test compared to the other four tests in the control of Type I error rates. Also, for g = 0.5 and h = 0.5, under six group condition, the *AGWMOM* test has more of its Type I error rates that are said to be robust compared to the *AG* test, the *AGMOM* test and the *ANOVA*.

Acknowledgement

This research is dedicated to "God Almighty", the author of wisdom, knowledge and understanding. All thanks, praises, honour, adoration, worship and glory goes to "God Almighty" for everything. Secondly, I want to acknowledge my beloved and blessed parents, in person of Mr. and Mrs. D.K.O. Tobi, for always giving me all the needed support and encouragement. I love and appreciate them very greatly.

References

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

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.

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. Retrieved from, http://www.jstor.org/stable/1165297

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

Brown, M. B., & Forsythe, A. B. (1974). The small sample behaviour of some statistics which test the equality of several means. *Technometics, 16*, 129-132.

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.

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

Hill, G. W. (1970). Algorithm 395, Student's *t*-distribution. *Communications of the ACM, 13*, 617-619.

Keselman, J. J. J. C. (1982). Parametric Alternatives to the Analysis of Variance. Author (s): Jennifer J. Clinch and H. J. Keselman Source: Journal of Educational Statistics, 7(3) (Autumn, 1982), 207 -214. Published by: American.

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

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

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

Lix, L. M, Keselman, J. C., & Keselman, H. J. (1996). Consequence 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.

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

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

Othman, A. R., Keselman, H. J., Padmanabhan, A. R., Wilcox, R.R., & Fradette, K. (2004). Comparing measures of the "typical" score across treatment groups. *The British journal of mathematical and statistical psychology, 57*(Pt 2), 215-234. Retrieved from http://www.ncbi.nih.gov/pubmed/15511305

Pardo, J. A., Pardo, M. C., Vicente, 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. http://dx.doi.org/10.1016/S0167-9473(96)00080-1

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

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

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 means: An alternative approach Biometrica, 38, 330-336.

Wilcox, R. R. (1988). A new alternative to the ANOVA F and new results on James's second-order method. *British Journal of Mathematical and Statistical Psychology, 42*, 203-213.

Wilcox, R. R. (1995). ANOVA: A paradigm for low power and misleading measures of effect size?

Review of Educational Research, 65, 51-77.

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

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

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

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

Wilcox, R. R. (1995). ANOVA: A paradigm for low power and misleading measures of effect size? *Review of Educational Research,* 65, 51-77. Retrieved from http:www.jstor.org/stable/1170478.

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

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*.