International Journal of Statistics and Applications 2016, 6(2): 45-52
DOI: 10.5923/j.statistics.20160602.03

Using Real Life Data to Validate the Winsorized Modified Alexander-Govern Test

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

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

Abstract **Aims and Objectives:** To evaluate the efficiency and reliability of the Alexander-Govern (*AG*) test and the Winsorized Modified One Step M-estimator in the Alexander-Govern (*AGWMOM*) test, using real life data. **Methods:** Test of homogeneity of variance was done from real life data, comprising of young, middle and old groups, using the Levene's test to see if the three groups are different from each other or not as the reaction time changes. Descriptive statistics, Test of normality and Test Statistic were performed for the three independent groups, to evaluate the reliability and efficiency of the tests. **Results:** The *p*-value from the test of homogeneity of the variance is greater than 0.05, i.e 0.174 > 0.05 and it shows that we accept H_O and conclude that there is no difference between the groups as the reaction time changes. The descriptive statistics show that the *AGWMOM* test has a smaller standard error compared to the *AG* test. The result of the test statistic reveals that the *AGWMOM* test produced a *p*-value of 0.0000002869 that is considered to be significant compared to the *AG* test that produced a *p*-value of 0.0698 that is regarded as not significant, since its *p*-value is > 0.05. **Conclusions:** The *AGWMOM* test is more efficient and reliable in minimizing error as much as possible from the real life data, because the test produced a smaller standard error from the real life data in comparison to the *AG* test and is regarded as significant.

Keywords Alexander-Govern (AG) test, AGWMOM test and Test Statistic

1. Introduction

The independent group tests such as the *ANOVA* have been employed in different fields of life, such as in economics, sociology, medicine and agriculture as stated by [23]. Some assumptions have to be fulfilled before the method can perform effectively, such as: (1) homogeneity of the variances, (2) normality of the data and (3) independent observations. The *ANOVA* is 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 and is robust to small deviations from normality, mainly when the sample size is large enough to guarantee normality, as explained by [28, 29].

It is observed that the two major problems confronting the *ANOVA* is the appearance of non-normality and variance heterogeneity in a data distribution [32]. As a result, the Type I error rates are increased and the power of the test is reduced.

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 questionable, since the *p*-value becomes too conservative. Therefore, it is very important to test for the homogeneity of the variance in order to verify the equality of the variance assumptions by using the correct test, so as to increase the validity of the results [4, 30]. The problem of heterogeneity of variance has been discussed by few scholars and some alternatives have been introduced. [26] Introduced the Welch test that is used for testing the hypothesis of equality of means between two or more populations. It was discussed in different literatures as an alternative to the *ANOVA* [3, 11, 15, 30].

The Welch test gives a good control of Type I error rates for unequal variances. It is a common alternative to parametric methods which deal with unequal variances. However, for a small sample size, the Welch test fails to give a good control of Type I error rates, as the number of groups increases [27]. [8] Introduced a better alternative to the *ANOVA*, namely the James test. The James test is used for weighing sample means as discussed in different literature by different scholars [15, 21, 27]. For a small sample size and when the data distribution is non-normal, the James test fails to give a good control of Type I error rates. The Welch test and the James test are used for analysing data which are not normally distributed and have unequal variances [5, 12, 13, 29].

Alexander-Govern [2] 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. [24, 16,

* Corresponding author:
tobikingsley@rocketmail.com (Tobi Kingsley Ochuko)
Published online at http://journal.sapub.org/statistics

19] agreed that the Alexander-Govern test performs well under variance heterogeneity for a normal data, but this test fails to give a good control of Type I error rates for a non-normal data. The reason is because the test uses mean as a measure of its central tendency.

The common mean is a very good estimator for a normal distribution, but it is extremely sensitive to the presence of outliers. The common mean cannot handle any slight deviation from normality. In finding a solution to the problem of non-normality, [16] proposed the trimmed mean to handle the problem of non-normality in Alexander-Govern test. Also, [14] and [17] observed that the use of Winsorized variance and trimmed mean is capable of removing the appearance of outliers in a skewed data distribution. This shows that with the use of trimmed means, the non-normality problem can be addressed. Trimmed mean is an estimator which is used in replacing the common mean as a measure of central tendency for a non-normal data.

This estimator has been used by different scholars in the past, because of its reliability and efficiency in controlling Type I error rates under non-normality [10, 18, 17]. The application of the trimmed mean in a data distribution has some weaknesses which are: (1) the percentage of trimming-in determining the elimination process must be set in advance. (2) It leads to loss of information, as the data is trimmed symmetrically from both tails of the data distribution. (3). It fails to handle large count of extreme values [31].

According to [1] an alternative to the use of trimmed mean in Alexander-Govern test is a highly robust estimator, known as the Modified one-step M-estimator (*MOM*). It was observed that when the distribution of the data is skewed, the *MOM* estimator gave a good control of Type I error rates. The *MOM* estimator empirically trims extreme data set depending on the nature of the distribution, be it skewed or normal. When it was applied in Alexander-Govern test, it gave a remarkable control of Type I error rates under normal or highly skewed data distribution, but this estimator fails to give a good control of Type I error rates, in an extreme condition of skewness and kurtosis [22].

According to [20] Winsorization is the process of making a replacement of an outlier value with the closest (non-outlier) value. Winsorization helps prevent loss of information in a data distribution. The sample size of the data sets is preserved unlike the trimmed mean procedure, where the data is trimmed symmetrically from both tails of the data distribution, resulting in sample size decrease.

In this research, the Winsorized Modified One Step M-estimator was applied Alexander-Govern test to overcome the weakness of the *MOM* estimator in the *AG* test, in an extreme condition of skewness and kurtosis and to make the test robust to non-normality.

The *AG* test and the *AGWMOM* test were validated using real life data from [9]. Test of Homogeneity of variances was done for the three independent groups from the real life data, comprising of young, middle and old group and the result show that the three independent groups are not different from each other as the reaction time changes. Test of normality was also performed on the three independent groups, to see which groups are normally distributed. Test statistic were calculated for the two tests, namely the *AG* test and the *AGWMOM* test and it showed that the *AGWMOM* test is more reliable and efficient in minimizing error as much as possible from the real life data, because it produced a *p*–value of 0.0000002869 compared to the *AG* test that produced a *p*-value of 0.0698.

2. The Alexander-Govern Test and Its Test Statistic

The [2] introduced the Alexander-Govern test. This test uses mean as a measure of its central tendency and it gives a good control of Type I error rates and high power, under variance heterogeneity for a normal data. This test is not robust for non-normal data. This test is used for comparing two or more means and its test statistic is obtained using the following techniques.

Firstly, to obtain the test statistic for the Alexander-Govern test, we order the data sets, comprising of J groups indexed by *j* (*j* = 1,…,*J*). Then, for each of the data sets, the mean is obtained by using the formula:

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

Where X_{ij} represent the observed ordered random observations in samples of size n_j. The mean is used as a measure of the central tendency in the [2] method. After the mean is obtained, the usual unbiased estimate of the variance is calculated, using the formula:

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

Where $\bar{X}_j$ is used to estimate μ_j for the population *j*. The standard error of the mean is obtained by using the formula:

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

The weight (w_j) for the group of the observed ordered random sample is defined, such that $\sum w_j$ equal to 1. Thus, the weight (w_j) for each of the independent groups is obtained by using the formula:

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

The null hypothesis testing for the [2] technique for the equality of the mean, under variance heterogeneity is given below:

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

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

For at least $i \neq j$

The variance weighted estimate of the total mean for all the groups in the data sets is obtained using the formula:

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

Where, w_j is the weight for each group in the data distribution and $\bar{X}_j$ is the mean of each group in the observed ordered data set. The *t* statistic for each of the group is obtained 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 groups, $\hat{\mu}$ is the grand mean for all the independent groups with population *j*. The *t* statistic, with $n_j - 1$ degrees of freedom. Denoting with ν the degree of freedom for each of the independent groups in the observed ordered data set. The *t* statistic obtained for the each of the groups and is converted to standard normal deviates by using the [7] normalization approximation in the [2] technique. The formula is expressed using:

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

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

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

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

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

After obtaining the test statistic for the *AG* test, a significance level of α = 0.05 with (*j* – 1) chi-square degree of freedom is selected. If the *p*-value of the *AG* test is greater than 0.05, it is concluded that the test is not significant otherwise, the test is significant.

3. The Winsorized Modified Alexander-Govern Test

Consider an observed ordered data set: $X_1, X_2, ..., X_n$, with sample size *n* and group sizes *j*. Firstly, the median of the data set is obtained by selecting the middle value from the observations. The *MAD* estimator is the median of the set of 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 obtained by using the formula below:

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

According to [29] the constant value of 0.6745 is used to rescale the MAD estimator, with the aim of making the denominator estimates σ when sampling from a normal distribution. Outliers in a data distribution can be detected by using the formula below:

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

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

Where, X_j represents 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 introduced by [29] for detecting the presence of outliers in a data set, because it has a very small standard error, when sampling from a normal distribution.

Equation (12) and (13) helps to define the *MOM* estimator 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 a central tendency measure for the test. The *WMOM* estimator is applied on the data distribution where the outlier detected value is replaced with the preceding value closest to the position the outlier is located. The *WMOM* estimator is obtained 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 the common mean as a measure of the central tendency in Alexander-Govern test, to remove the outliers from the data set and make the Alexander-Govern test robust to non-normality.

The Winsorized sample variance is obtained using:

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

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

Firstly, we chose *B* independent bootstrap samples defined as:

$x^{*1}, x^{*2}, ..., x^{*B}$, Where each of these random samples comprises of *n* data values chosen with replacement from x expressed as:

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

$$F \to (x_1^*, x_2^*, ..., x_n^*), \quad (17)$$

The indication of the symbol $(*)$ shows that x^* is not the real data set 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 [6] bootstrap sample size of 50 is sufficient enough to give a reasonable estimate of the standard error of the *MOM* estimator. In this research, the same sample size was used to estimate the standard error of the *MOM* estimator.

Secondly, we evaluate the bootstrap replication corresponding to each of the bootstrap sample define as:

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

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

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

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 for each of the independent groups is defined as:

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

$S_e{}^2_{WMOMj}$ is the squared-standard error of the Winsorized data distribution and is expressed as:

$$S_{eWMOMj} = \sqrt{\frac{s_j{}^2_{WMOMj}}{n_j}}, \quad (21)$$

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

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

Where w_j is defined as the weight for the Winsorized data distribution, and $\bar{X}_{WMOMj}$ is defined as the mean of the Winsorized data distribution.

The *t* statistic for the Winsorized data distribution for each of the group is expressed using the formula:

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

Where, $\bar{X}_{WMOMj}$, $\hat{\mu}_j$ and S_e is the Winsorized *MOM* estimator, the total mean for the Winsorized data distribution and the standard error of the Winsorized data distribution data distribution respectively. In the [2] method, the t_j value is converted to standard normal by using the [7] normalization approximation and the hypothesis testing of the Winsorized sample variance of the *WMOM* estimator for μ_j is defined as:

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

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

For $j = (j = 1, ..., J)$

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

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

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

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

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

$$AGWMOM = \sum_{j=1}^{J} Z^2_{WMOM\,j} \tag{24}$$

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 a standard chi-square distribution table. If the value of the test statistic for the *AGWMOM* test is less than 0.05, then the test is regarded as very significant, otherwise the test is referred to as not significant.

4. To Evaluate the Efficiency and Reliability of the Tests Using Real Life Data

A real life data which was obtained from [9] that comprises of three independent groups, namely: the group young, middle and old was used to evaluate the efficiency and reliability of the *AG* test and the *AGWMOM* test respectively.

Table 1. The real life data for the young, middle and old group respectively

Young (y)	Middle (m)	Old (o)
482.43	335.59	519.01
484.36	338.43	524.50
488.84	353.54	530.23
495.15	404.27	536.03
495.24	437.50	538.56
502.69	469.01	538.83
504.62	485.85	557.24
518.29	487.30	558.61
519.10	493.08	558.95
524.10	494.31	565.43
524.12	499.10	586.39
531.18	886.41	594.69
548.42	-	629.22
572.10	-	645.69
584.68	-	691.84
609.09	-	-
609.53	-	-
666.63	-	-
676.40	-	-

Source: [9]

The test of Homogeneity of variances was done for the three independent groups, using the Levene's test to determine if the three groups have different-reaction time-changes variances.

In Table 4, the mean of the three independent groups, namely: the young, middle and old are displayed above. The standard errors for the group young, middle and old are regarded as very high, with values 59.7266, 144.6221 and 49.5377 respectively, for the three independent groups. This is as a result of the presence of outliers in the real life data for the *AG* test.

Table 2. The Winsorized Data Distribution from the Real Life Data

Winsorized Young	Winsorized Middle	Winsorized Old
482.43	404.27	519.01
484.36	404.27	528.50
488.84	404.27	530.23
495.15	437.50	536.03
495.24	469.01	538.56
502.69	485.85	538.83
504.62	487.30	557.24
518.29	493.08	558.61
519.10	494.31	558.95
524.10	499.10	565.43
524.12	499.10	586.39
531.18	-	594.69
548.42	-	629.22
572.10	-	645.69
584.68	-	645.69
609.09	-	-
609.53	-	-
609.53	-	-
609.53	-	-

Source: [20]

Table 3. Test of Homogeneity of Variances

Reaction Time			
Levene Statistic	df1	df2	Sig.
1.821	2	43	.174

$\alpha = 0.05$
H_O = There is no difference between the groups
H_1 = There is difference between the groups
If the *p*-value is < 0.05, we reject H_O and accept H_1. When the *p*-value is > 0.05, we accept H_O and reject H_1. The *p*-value from the test of homogeneity of the variance is > 0.05, i.e $0.174 > 0.05$, implies that we accept H_O and conclude that there is no difference between the groups as the reaction time changes variances.

Table 4. Descriptive Statistics for the Young, Middle and Old Groups using the *AG* test with 50 bootstrap samples

Descriptive Statistics	Young (Y)	Middle (M)	Old (O)
Usual Mean	544.0511	473.6992	571.6813
Standard Error of Mean	59.7266	144.6221	49.537

Table 5. Descriptive Statistics for the Winsorized Young, Middle and Old Groups Using the *AGWMOM* test with 50 bootstrap samples

Descriptive Statistics	Winsorized Young	Winsorized Middle	Winsorized Old
Usual Mean	505.8433	456.8608	551.0392
Standard Error of Mean	4.9059	12.1963	6.7518

BOX PLOT

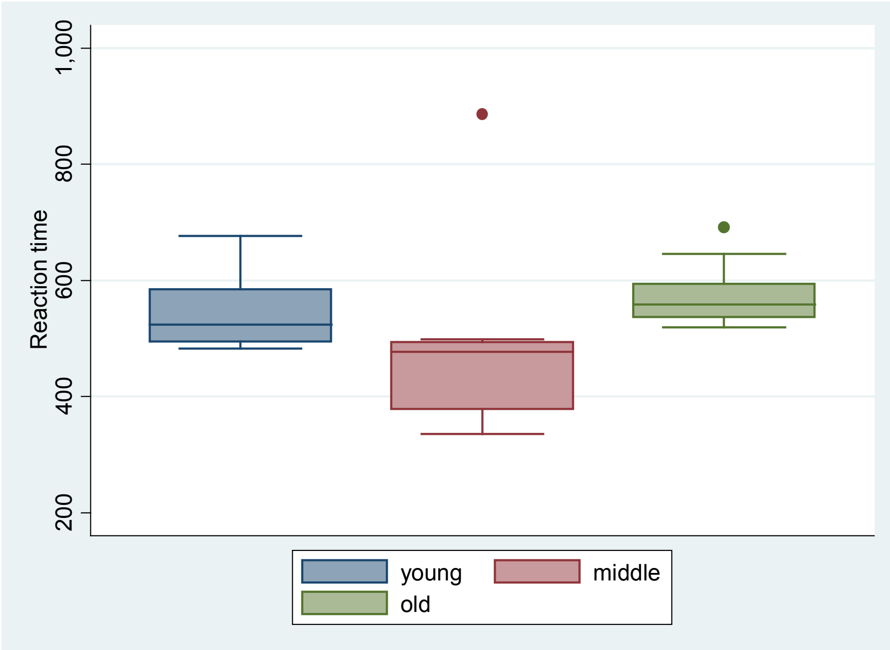

Figure 1. Boxplots on reaction time against the young, middle and old groups

In Table 5, the Winsorized mean for the three independent groups, namely: the young, middle and old are: 505.8433, 456.8608 and 551.0392 are observed to be smaller in comparison to the mean of the young, middle and old groups respectively of the *AG* test. The standard errors for the Winsorized young, middle and old groups are: 4.9059, 12.1963 and 6.7518 are considered to be far smaller compared to the standard errors for the young, middle and old groups of the *AG* test in Table 4. This is as a result of the elimination of the outliers from the real life data that have been replaced with the preceding values closest to the outlier values from the real life data.

Table 6. Test of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	Df	Sig.	Statistic	Df	Sig.
Young	0.185	18	0.200	0.924	18	0.319
Middle	0.347	11	0.000	0.721	11	0.001
Old	0.199	14	0.200	0.935	14	0.431

Shapiro-Wilk Test is a test that is most suitable for sample sizes that is not up to 50 samples. This test can also handle sample sizes that is as large as 2000 [25]. As a result, the Shapiro-Wilk Test is used to test for the normality of the three independent groups, namely the group young, middle and old. At the significance level of α = 0.05, if the significant value of any of the three groups is greater than 0.05, then the data is considered to be normally distributed. Otherwise, if the significant value is less than 0.05, then the data distribution is non-normal.

The results from Table 6 show that the *p*-value for the group young and old are greater than 0.05, hence both groups are normally distributed i.e young with *p*-value of 0.319 and old with *p*-value of 0.431. The middle group has a *p*-value of 0.001which is < 0.05 and is regarded as non-normally distributed.

In Figure 1 above, shows the boxplots of the reaction time against the young, middle and old groups. It can be seen very clearly from the plots that there is no extreme value present in the group young and old, hence the data distribution is regarded as normally distributed. It can be observed that there is an extreme value in the group middle and this shows that the data distribution for the group middle is non-normally distributed.

Table 7. The Test Statistic for the *AG* test and the *AGWMOM* test

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

In Table 7, the test statistic for the *AG* test has a value of 5.3237, with a *p*-value of 0.06982 at α = 0.05 level of significant. This shows that the *AG* test is not significant, since its *p*-value of 0.06982 > 0.05. While the test statistic value of the *AGWMOM* test produced a value of 30.1280, which is almost six times that of the *AG* test.

The *AGWMOM* test has a *p*-value of 0.0000002869 at α = 0.05 level of significant. The *AGWMOM* test is regarded as significant, since its *p*-value of 0.0000002869 is < 0.05 compared to the *AG* test. The standard error of the Winsorized *AGMOM* from the real life data for the young, middle and old group is far smaller compared to the standard error of the *AG* test from the original real life data.

5. Conclusions

The *AGWMOM* test is more efficient and reliable in minimizing error as much as possible from the real life data, by making a replacement for the presence of outliers in the real life data with a smaller standard error in comparison to the *AG* test.

ACKNOWLEDGEMENTS

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

I also want to acknowledge and thank my blessed, wonderful, special, very caring and ever-dynamic parents in person of Mr. and Mrs. D.K.O. Tobi, for their constant encouragement, love, sacrifice, support and goodwill. 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] 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.

[3] 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.

[4] 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.

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

[6] Efron, B., & Tibshirani (1998). *An introduction to the bootstrap.* New York: Chapman & Hall.

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

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

[9] Keselman, H., Wilcox, R., Lix, L.M., Algina, J., & Fradette, K. (2007). Adaptive robust estimator and testing *British Journal of Mathematical Psychology,* 60, 267-293.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] Lix, Lisa, M., & Keselman, J.C., & Keselman, H. J (1995). Approximate degrees of freedom tests. A unified perspective on testing for mean equality. *Pschological Bulletin,* 117(3), 547-560.

[15] 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.

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

[17] 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.

[18] 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.

[19] 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.

[20] Ochuko, T. K., Abdullah, S., Zain, Z., & Yahaya, S. S. S. (2015). Modifying and Evaluating the Alexander-Govern Test Using Real Data. *Modern Applied Science,* 9(12), 1-11.

[21] 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.

[22] 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 Psycholofy*, 57(2), 215-234.

[23] 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.

[24] 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.

[25] Shapiro, S. S. and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrka,* 52, 591-611.

[26] Welch, B. L. (1951). On the comparison of several means: An alternative approach. Biometrica, 38, 330-336.

[27] 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.

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

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

[30] 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.

[31] 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.

[32] 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.