Journal of Mathematics and Statistics 6 (3): 359-366, 2010
ISSN 1549-3644

A Monte Carlo Study of Seven Homogeneity of Variance Tests

[1]Howard B. Lee, [1]Gary S. Katz and [2]Alberto F. Restori
[1]Department of Psychology,
[2]Department of Educational Psychology and Counseling,
California State University, Northridge, CA, 91330-8255, USA

Abstract: Problem statement: The decision by SPSS (now PASW) to use the unmodified Levene test to test homogeneity of variance was questioned. It was compared to six other tests. In total, seven homogeneity of variance tests used in Analysis Of Variance (ANOVA) were compared on robustness and power using Monte Carlo studies. The homogeneity of variance tests were (1) Levene, (2) modified Levene, (3) Z-variance, (4) Overall-Woodward Modified Z-variance, (5) O'Brien, (6) Samiuddin Cube Root and (7) F-Max. **Approach:** Each test was subjected to Monte Carlo analysis through different shaped distributions: (1) normal, (2) platykurtic, (3) leptokurtic, (4) moderate skewed and (5) highly skewed. The Levene Test is the one used in all of the latest versions of SPSS. **Results:** The results from these studies showed that the Levene Test is neither the best nor worst in terms of robustness and power. However, the modified Levene Test showed very good robustness when compared to the other tests but lower power than other tests. The Samiuddin test is at its best in terms of robustness and power when the distribution is normal. The results of this study showed the strengths and weaknesses of the seven tests. **Conclusion/Recommendations:** No single test outperformed the others in terms of robustness and power. The authors recommend that kurtosis and skewness indices be presented in statistical computer program packages such as SPSS to guide the data analyst in choosing which test would provide the highest robustness and power.

Key words: Robust tests, power tests, Levene test, ANOVA

INTRODUCTION

A very popular statistical package Statistical Package for the Social Sciences (SPSS now called PASW) uses the Levene Test to test for homogeneity of variance prior to conducting tests of the equality of means in the t-test and One-way ANOVA (Oladejo and Adetunde, 2009; Zeng *et al*., 2010; Mazahreh *et al*., 2009). A question arose as to whether the designers at SPSS chose the "best" test for homogeneity of variance, since there are many others available. A subsequent literature search produced some research on the Levene test (Gastwirth *et al*., 2009; Carroll and Schneider, 1985; Tomarken and Serlin, 1986). A further search found other tests of homogeneity of variance in studies by Overall and Woodward (1974; 1976); O'Brien (1981) and Levy (1975) that may have been a better choice than the Levene test.

Carroll and Schneider (1985) described the Brown and Forsyth comparison of the original Levene Test to two modifications of it. The original Levene Test used sample means. The modified Levene tests used the median and the trimmed mean. They demonstrated through Monte Carlo studies that the median and the trimmed means outperformed the original test when the homogeneity of variance assumption was violated.

The modified Z-variance test is presented by Overall and Woodward (1976). Overall and Woodward (1976) had compared the robustness and power of this modification against four other homogeneity of variance tests: (1) Z-variance unmodified, (2) Wilson-Hilferty (3) Bartlett and (4) Box. Using a series of Monte Carlo studies, Overall and Woodward (1976) demonstrated the superiority of the modified Z-variance test over the other four tests. Unfortunately, the Overall-Woodward modification of the Z-variance test is not well known. This modification appears only in a technical report that may no longer be available or easily accessible from the original source. However, a copy of this report can be obtained from the corresponding author of this article.

The O'Brien Test is mentioned by Howell (2001) but little research could be found on it. From O'Brien (1981) this test appears promising. With all of these

Corresponding Author: Howard B. Lee, Department of Psychology, California State University, Northridge, CA 91330-8255, USA Tel: 818-677-4863 Fax: 818-677-2829

more complicated formulas developed to attack the problem concerning homogeneity of variance, there is a simple one that will also be used in this study. The Fmax test developed by Hartley in 1950 (Pardo *et al*., 1997) is very simple involving no more than computing the ratio of the greatest subgroup variance and the smallest subgroup variance.

In this study, seven homogeneity of variance tests will be compared using a Monte Carlo approach. The seven tests are (1) the original Levene Test, (2) The modified Levene Test using the median, (3) the Z-variance Test, (4) Modified Overall-Woodward Z-variance test, (5) O'Brien Test (6) Samiuddin Cube-root Test and (7) the Fmax test. A major goal of the study is to evaluate just how good the original Levene test is when compared to these other alternatives. "Goodness" of each test is determined by examining the robustness and power for each test. Should SPSS and other statistical packages consider using other tests along with the Levene?

The Levene test: In 1960, Levene proposed an alternative method to the Bartlett Test (Klotz and Johnson, 1993) for testing the assumption of homogeneity of variance for independent sample t-test and ANOVA designs. The Bartlett test works well for data that are normally or approximate normally distributed. The Bartlett test does not fare well for data that follow a leptokurtic or skewed distribution (Overall and Woodward, 1974). According to Levene (Gastwirth *et al*., 2009), the test he proposed was less sensitive to departures from normality. This says that the Levene Test had fewer Type 1 errors than the Bartlett Test for distributions that were aberrant from normality.

The Levene Test is defined as the following:

$$H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \cdots = \sigma_k^2$$

$$H_1 : \sigma_i^2 \neq \sigma_j^2 \text{ for at least one pair (i, j)}$$

If Y_{ij}'s represent the raw scores on the dependent measure:

$$W = \frac{(N-k)\sum_{i=1}^{k} n_i \left(\bar{Z}_i - \bar{Z}\right)^2}{(k-1)\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(Z_{ij} - \bar{Z}_i\right)^2}$$

Where:

N = Total sample size

n_i = Sample size for group

$Z_{ij} = \left|Y_{ij} - \bar{Y}_i\right|$

$\bar{Y}_i$ = The mean of the ith subgroup

$\bar{Z}_i$ = The group means of the Z_{ij}

$\bar{Z}$ = The overall mean of the Z_{ij}

Decision rule: Reject H_0 if W>F (α, k–1, N–k), otherwise do not reject H_0. Where: F (α, k–1, N–k) is the upper critical value of the F-distribution with k–1 and N–k degrees of freedom at a significance level of α.

The modified Levene test: The modified Levene test is nearly identical to the original Levene test. The difference is that the median is used instead of the mean in computing Z_{ij}. That is $Z_{ij} = \left|Y_{ij} - \tilde{Y}_i\right|$, where $\tilde{Y}_i$ is the median of the ith subgroup. This is the modification studied earlier by Brown and Forsyth is referenced in Carroll and Schneider (1985).

The Z-variance test: The large sample normal deviate transformation of chi-square proposed by Fisher (1995) formed the basis of the Z-variance test. The formula Fisher (1995) presents is $Z = \sqrt{2\chi^2} - \sqrt{2(df) - 1}$. It is well known that sample variances tend to have a chi-square distribution (Overall and Woodward, 1974). A normal deviate transformation is used to obtain Z-score equivalents of the sample variance. Sample variance is related to the chi-square by use of the following formula:

$$\chi_{n-1}^2 = \frac{(n-1)s^2}{\sigma^2}$$

These are then used in an F-test to determine if they are different. This F-test is presented in a number of elementary statistic textbooks (Comrey, 2009; Mendenhall and Beaver, 1991).

Overall and Woodward (1974) found this test to perform very well for data that are normally distributed. Their Monte Carlo studies discovered that this test produced too many Type 1 errors when samples are drawn from leptokurtic or skewed distributions.

The null and alternative hypothesis for the Z-variance test is the same as the one for the Levene Test. The test statistic as written by Overall and Woodward (1974) is:

$$F = \frac{\sum_{i=1}^{k} Z_i^2}{k-1}$$

Where:

$$Z_i = \sqrt{\frac{c(n_i - 1)S_i^2}{MSE}} - \sqrt{c(n_i - 1) - \frac{c}{2}}$$

$c = 2 + 1/n_i$

S_i^2 = The unbiased estimate of variance for the ith subgroup or cell in the design

n_i = Sample size for ith subgroup

MSE = Pooled within group error variance (MSW in one-way ANOVA)

The Z_i's are assumed to be approximately unit normal with zero mean.

Decision rule: Reject H_0 if $F > F(\alpha, k-1, \infty)$, otherwise do not reject H_0. Where: $F(\alpha, k-1, \infty)$ is the upper critical value of the F-distribution with k–1 and ∞ degrees of freedom at a significance level of α.

The Overall-Woodward modified Z-variance test: To counter the distortions of the original Z-variance test, Overall and Woodward (1976) conducted a series of studies to determine a c value so that variances of the Z_i would remain stable when the sample data deviate moderately from normality. Using regression, Overall and Woodward (1976) found a c value based on sample size, skewness and kurtosis. They determined c to be a scaling coefficient that affects the variability of the Z_i values.

The new formula for c is:

$$c = 2.0\left(\frac{2.9 + .2/n_i}{K}\right)^{1.6\left(n_i - 1.8K + 14.7\right)/n_i}$$

Where:

n_i = The sample size of the ith subgroup

K = The mean of the kurtosis indices from all subgroups

The index of kurtosis used by Overall and Woodward (1976) is the 4th power of the Z-scores within each sample (subgroup) divided by n_i-2 degrees of freedom:

$$K = \frac{\sum Z_{ij}^4}{n_i - 2}$$

Where:

$$Z_{ij} = \frac{X_{ij} - \bar{X}_i}{\sqrt{\frac{n_i - 1}{n_i} S_i^2}}$$

O'Brien test: As a fifth comparison for this study, the O'Brien Test (O'Brien, 1978; 1981) was used. O'Brien (1981) has claimed that his test is a general method that does fairly well for behavioral science data. O'Brien (1981) states that the test is robust to data that departs from normality. It is also easy to program into statistical packages like SPSS, it is competitive with other tests in terms of power and it can be easily used in different ANOVA designs with equal or unequal sample sizes. O'Brien (1981) stated that not much research has been done on this statistic.

The computational operations for this test are straightforward. Every raw score, Y_{ij} in the study is transformed using the following formula:

$$V_{ij} = \frac{(n_i - 1.5)n_i\left(Y_{ij} - \bar{Y}_i\right)^2 - .5s_i^2(n_i - 1)}{(n_i - 1)(n_i - 2)}$$

Where:

$\bar{Y}_i = \frac{\sum_{j=1}^{n_i} Y_{ij}}{n_i}$, the mean for each subgroup i

$s_i^2 = \frac{\sum_{j=1}^{n_i}\left(Y_{ij} - \bar{Y}\right)_i^2}{(n_i - 1)}$, the unbiased subgroup variances

The mean of the V-values per subgroup will be equal to the variance computed for each subgroup, i.e.,

$$\bar{V}_i = \frac{\sum V_{ij}}{n_i} = s_i^2$$

The test statistic for the O'Brien Test will be the F-value computed on applying the usual ANOVA procedure on the transformed scores V_{ij}.

Samiuddin cube-root test: Samiuddin and Atiqullah (1976) developed a homogeneity of variance test which he refers to as the Bayesian test of homogeneity. Samiuddin and Atiqullah (1976) show that the "cube-root" test is superior over some other tests such as the Bartlett test when the sample distributions are not homogeneous. However, Levy (1978) has shown Samiuddin and Atiqullah's findings to be flawed or misleading. This study will re-examine the Samiuddin Test in terms of robustness and power and in comparison to some other homogeneity of variance tests not tested by Samiuddin and Atiqullah (1976) and Levy (1978).

If X_{ij} (i = 1, 2,…,k; j = 1, 2,…, n_i) are normally distributed with mean μ_i and variance σ_i^2, Samiuddin and Atiqullah (1976) defines $s_i^2 = \sum_i \left(X_{ij} - M_i\right)^2$ and

$s_i^2 = S_i^2 / v_i$, where $M_i = \sum_j X_{ij} / n_i$ and $v_i = n_i - 1$. When $\sigma_1^2 = \sigma_2^2 = \ldots \sigma_k^2$, Samiuddin and Atiqullah (1976) shows $\sum (m_i - m)^2 / a_i^2$ is approximately distributed as a χ^2 with k-1 degrees of freedom where:

$$m_i = \left(\frac{v_i}{S_i^2}\right)^{1/3}\left(1 - \frac{2}{9v_i}\right)$$

$$a_i^2 = \frac{2}{9}\left(S_i^2\right)^{2/3} v_i^{1/3}$$

and:

$$m = \frac{\left(\sum \frac{m_i}{a_i^2}\right)}{\left(\sum \frac{1}{a_i^2}\right)}$$

The Fmax test: Hartley in 1950 developed the Fmax test (Pardo *et al*., 1997). It is very simple to calculate and evaluate. The null and alternative hypotheses are the same as specified for the Levene and other tests. Its test statistic is just a simple ratio between the largest subgroup variance and the smallest:

$$F_{MAX} = \frac{s_{largest}^2}{s_{Smallest}^2}$$

A table of values created by Hartley evaluates the test statistic with degrees of freedom of k and n_i-1. If the test statistic exceeds the critical value the null hypothesis is rejected. The Hartley Fmax test has appeared in many older advanced statistics textbooks such as Winer (1971) and Kirk (1994).

MATERIALS AND METHODS

Test comparisons: To compare the Levene, modified Levene, O'Brien, Fmax, Z-variance and modified Z-variance tests a series of Monte Carlo studies were performed (Agunbiade and Iyaniwura, 2010; Alabi *et al*., 2008; Rana *et al*., 2008). Each statistic is evaluated in terms of robustness and power. For robustness, the fewer Type 1 errors a test makes (falsely claiming unequal variances, when in fact the variances are equal), the greater the robustness. (Abu-Shawiesh, 2008; Vrbanek and Wang, 2007). With power, the higher the number of correctly detected unequal variances, when in fact they are unequal, the greater the power of the test.

Monte Carlo study of robustness: In this section, the first analyses were done to determine how well each test performs when there is no bias. That is, the samples are drawn from a normal distribution with equal variances. For all tests in this study, there are four groups arranged as a fixed effects completely randomized design ANOVA. With four groups, two different sample sizes were used: n = 10 and n = 30. This was the same arrangement used by Overall and Woodward (1976).

A computer program was written to carry out the analyses. In selecting a random number generator, the one described by Overall and Rhoades (1981) was used. This algorithm produced random numbers that follow a normal distribution. Three thousand simulated four-group experiments were analyzed for each sample size. The Levene, modified Levene, O'Brien, Fmax, Samiuddin, original Z-variance and modified Z-variance test statistics were computed for each simulated experiment. The computer program counted the number of times the null hypothesis was rejected at the $\alpha = 0.05$ level for 3000 experiments. The probability associated with each test statistic was computed using a subprogram developed by Jaspen (1965) and Veldman (1967). The results of these tests are given in the first column of Table 1.

The analyses were repeated for the same sample sizes with non-normal distributions that still had homogeneous variances. Following the descriptions provided by Overall and Woodward (1976) simulated experimental data were created for leptokurtic, platykurtic, chi-square (df = 6) and chi-square (df = 5) distributions. The chi-square distributions were used to approximate skewed distributions where the chi-square distribution with 6 degrees of freedom is less skewed than the one with 5 degrees of freedom. Three thousand simulated experiments were analyzed by the seven methods for each of the non-normal samples for the two sample sizes. The frequency that the null hypothesis was rejected at the $\alpha = 0.05$ level for each method (Levene, modified Levene, Fmax, Z-variance, Modified Z-variance, Samiuddin and O'Brien) for each distribution-type (normal, leptokurtic, platykurtic, moderately skewed, highly skewed) for each sample size (n = 10, n = 30) is given in Table 1.

Monte Carlo study of power: The next major consideration is the power of each test. Will the test accurately detect real differences between the subgroups with heterogeneous variances? Overall and Woodward (1976) found the modified Z-variance test to outperform the original Z-variance test as well as several others when the true underlying distribution is

normal or platykurtic. Overall and Woodward (1976) did not make any comparison between the different homogeneity of variance tests for the other types of distribution because of the large number of Type 1 errors found during the robustness phase. However, such high levels of Type 1 errors were not found for the Levene Test or the modified Levene Test in the current study. Hence in the study reported here, the two Levene test are compared to the two Z-variance tests, the O'Brien test, the Samiuddin Cube Root test and the Fmax test across the five different distributions. Overall and Woodward (1976) found the modified Z-variance test to be slightly less powerful than the original Z-variance test.

A series of 3000 simulated 4-group experiments were created where the group means were equal, but the sample variances were different. Using the same setup as found in Overall and Woodward (1976), the group variances followed the ratio of 1:2:3:4.

RESULTS

Monte Carlo study of robustness: The results from these Monte Carlo studies demonstrate that the modified Levene performed the best in producing overall the fewest type 1 errors across all distributions. In every distribution and sample size studied, the modified Levene had values below 0.05. The next best in an overall sense is the O'Brien test. Except for small samples case (n = 10) combined with a skewed distribution, the O'Brien test also had more values below 0.05 than the other remaining methods when both samples sizes are taken into consideration. The Overall-Woodward Modified Z-variance test was the next best and matched the O'Brien test very well for the larger sample size. The unmodified Levene Test could be rated fourth with the Fmax test and the Samiuddin test tied for fifth in the comparison. The original Z-variance test fared the worst in the comparison. Overall and Woodward (1976) had previously demonstrated that the modified Z-variance test was superior to the original Z-variance test in terms of robustness. The tests of interest here are how the two versions of the Z-variance tests and the O'Brien, Samiuddin and Fmax tests fared against the highly popular Levene Test. On every comparison, the modified Z-variance test outperformed the Levene Test. In cases involving leptokurtic and skewed distributions, the Levene Test did better than the original Z-variance test. So, when there are no differences between the sample variances (null hypothesis is true), the modified Z-variance test did better than the original Z-variance and Levene tests. The O'Brien test did better than the modified Z-variance test. In almost all tests the modified Z-variance test and the O'Brien test did better for larger samples than for smaller samples. The very simple Fmax test did as well or better than the original Z-variance test. As stated by Carroll and Schneider (1985), the modified Levene test using the median outperformed the original Levene test. The Samiuddin Cube Root test was at its best when the distribution was normal. It also performed well for distributions that were platykurtic. For other distributions it was not as good as the original Levene test. The "best" values in Table 1 are in bold print.

Monte Carlo study of power: For normal and platykurtic distributions with the larger sample, n = 30, the modified Z-variance test correctly rejected the null hypothesis more often than the Levene Test. The original Z-variance test was either the best or second best when the underlying distribution was normal, leptokurtic and skewed for both n = 10 and n = 30 sample sizes.

Table 1: Observed relative frequencies of type 1 errors for analysis of samples with equal variance (Robustness)[a]

	Normal	Leptokurtic	Platykurtic	Moderate skew (df = 6)	High skew (df = 5)
n = 10					
Levene original	068	133	091	135	149
Z-VAR original	032	526	004	168	211
Z-VAR: Over-wood	057	040	055	085	087
OBRIEN	048	038	037	061	060
Hartley FMAX	058	518	004	167	196
Levene modified	032	039	022	036	038
Samiuddin cube	039	521	005	164	197
n = 30					
Levene original	064	072	093	112	120
Z-VAR original	037	546	001	181	216
Z-VAR: Over-wood	039	027	052	057	059
OBRIEN	042	032	054	054	051
Hartley FMAX	037	522	002	174	204
Levene modified	048	040	025	042	040
Samiuddin cube	035	542	002	177	217

[a]Decimal points were omitted to save space

Table 2: Observed relative frequencies of type 1 errors for analysis of samples with unequal variance (power)

	Normal	Leptokurtic	Platykurtic	Moderate skew (df = 6)	High skew (df = 5)
n = 10					
Levene original	337	484	261	459	476
Z-VAR original	340	888	008	621	651
Z-VAR:Over-wood	293	233	231	291	287
OBRIEN	240	169	200	204	201
Hartley FMAX	330	893	011	601	632
Levene modified	201	236	65	146	145
Samiuddin cube	349	889	009	615	642
n = 30					
Levene original	853	913	712	826	829
Z-VAR Original	909	1000	080	952	953
Z-VAR: Over-wood	892	767	805	604	586
OBRIEN	837	561	762	432	404
Hartley FMAX	913	1000	094	945	955
Levene modified	804	839	306	533	509
Samiuddin cube	914	1000	083	950	954

The Samiuddin test did better than the original Z-variance test for both sample sizes when the distribution was normal. The Samiuddin and Fmax tests were either as good as or slightly better than the original Z-variance test when the distribution was leptokurtic. The O'Brien test did its best when the distribution was normal or platykurtic and the sample size was 30. In general, the O'Brien test was less powerful than the other tests.

In terms of power, it appears that the original Z-variance test fared the best for n = 10 distributions of varying kurtosis and skew. However, the Levene test consistently showed more power than the Overall and Woodward modified Z-variance test. However, it should be noted that nearly all of the n = 10 analyses showed minimal power, as expected with this small sample size. Notable exceptions to this were the original Z-variance test and Fmax test for leptokurtic distributions. The original Levene Test outperformed the modified Levene test on both sample sizes and for all five types of distributions. The data shows the modified Levene Test to be the worst amongst the 7 tests in terms of robustness except when the distribution was leptokurtic or platykurtic for n = 10.

For large samples, the original Z-variance test, Samiuddin test and the Fmax test exhibited the greatest power except when the distribution was platykurtic. For the platykurtic distribution for n = 30, the modified Z-variance and the O'Brien tests were the best. The "best" values are printed in bold in Table 2.

DISCUSSION

When considering robustness, the modified Z-variance test appears to be superior over the Levene and original Z-variance tests. The O'Brien test did better than the modified Z-variance test. However when looking at power, the original Z-variance test was better than the other tests for four of the distributions. Although the Levene test was not a standout in terms of robustness, power, sample size and distributional shape, it did not have the peaks and valleys as demonstrated by the two Z-variance tests. On robustness, the Levene test never attained the $\alpha = 0.05$ mark on any of the tests. It outperformed the modified Z-variance test in terms of power only for the small sample (n = 10) case for the platykurtic distribution. The O'Brien test appears to be the weakest in power. The Fmax test resembled the original Z-variance test. The simple Fmax test results were surprisingly good. This simple test seemed to do quite well in terms of robustness and power. The modified Levene Test outperformed the original Levene Test in terms of robustness. However, the reverse was true when considering power. For a normal distribution and large samples, the Samiuddin test was found to be the best of the 7 tests. The Samiuddin test did fairly well with both sample sizes except for the platykurtic distribution in terms of robustness and power.

The utility of the modified Z-variance test and the Samiuddin Cube-Root test is high, especially in those cases when the researcher with a priori evidence feels that the data approximates a normal distribution with a wider spread. Overall and Woodward (1976) recommended that a better c value be found through empirical means. The authors of this study agree that the search for a highly robust and powerful test still needs to be found.

CONCLUSION

In light of these findings, perhaps computer programs should not be limited to only one statistical test for homogeneity of variance. It would be highly useful for the researcher to be able have several of these tests available in a computer output. The user of

statistical packages such as SPSS (aka PASW) should check descriptive statistics concerning the kurtosis and skews of the data. If the index for kurtosis is negative then the distribution is platykurtic. If it is zero or near zero it is mesokurtic (possibly normal) and if the kurtosis index is positive, the distribution is leptokurtic. Likewise, the computed index of skewness can also be evaluated. A zero or near zero value indicates a symmetric distribution. If the index value is negative or positive then the distribution is skewed. By using these simple guidelines along with the robustness and power values, the data analyst can decide on which test of homogeneity of variance should be used and interpreted. For those computer programs that do not provide a measure of skewness and/or kurtosis, the following formulas can be used:

$$\text{Skewness} = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^3 \Big/ n}{S^3}$$

$$\text{Kurtosis} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^4 \Big/ n}{S^4} - 3$$

The modified Levene test was evaluated in this study and was found to be superior in terms of robustness and power to the original Levene test. In any case, researchers need to be aware that when homogeneity of variance tests fails to reject the null hypothesis of equal variance, the probability of a Type 2 error may be high depending on the test used. With SPSS (now PASW), the Levene test appears to have mediocre power that gets better with larger sample sizes.

In conclusion, the original Levene test (used by SPSS) is not the best choice. There are better tests that can be used and some are more preferable depending on the distributional shape of the data. The mere substitution of the median for the mean in the original Levene Test computations will alleviate some of the problems. It appears that a combination of skewness and kurtosis should give better information to the data analyst to select an appropriate homogeneity of variance test. None of the tests presented here were difficult to compute. Hence the computer overhead in including some or all of these tests is small. Relying upon defaults set in some statistical analytic software may lead to an increase in Type 1 errors associated with homogeneity of variance testing and/or loss of power that can be avoided or minimized.

REFERENCES

Abu-Shawiesh, M.O.A., 2008. A simple robust control chart based on MAD. J. Math. Stat., 4: 102-107. http://www.scipub.org/fulltext/jms2/jms242102-107.pdf

Agunbiade, D.A. and J.O. Iyaniwura, 2010. Estimation under multicollinearity: A comparative approach using Monte Carlo methods. J. Math. Stat., 6: 183-192. http://www.scipub.org/fulltext/jms2/jms262183-192.pdf

Alabi, O.O., K. Ayinde and T.O. Olatayo, 2008. Effect of multicollinearity on power rates of the ordinary least squares estimators. J. Math. Stat., 4: 75-80. http://www.scipub.org/fulltext/jms2/jms24275-80.pdf

Carroll, R.J. and H. Schneider, 1985. A note on Levene's tests for equality of variances. Stat. Probab. Lett., 3: 191-194. DOI: 10.1016/0167-7152(85)90016-1

Comrey, A.L., 2009. Elementary Statistics: A Problem Solving Approach. 4th Edn., Lulu.com, Dubuque, IA., pp: 188.

Fisher, R.A., 1995. Statistical Tables for Biological. Agricultural and Medical Research. 6th Edn., Longman Group United Kingdom, UK., pp: 156.

Gastwirth, J.L., Y.R. Gel and W. Miao, 2009. The impact of Levene's test of equality of variances on statistical theory and practice. Stat. Sci., 24: 343-360. DOI: 10.1214/09-STS301

Howell, D.C., 2001. Statistical Methods for Psychology. 5th Edn., Wadsworth Publishing, Pacific Grove, CA., pp: 816.

Jaspen, N., 1965. The calculation of probabilities corresponding to values of Z, t, F and chi-square. Educ. Psychol. Measur., 25: 877-880. DOI: 10.1177/001316446502500319

Kirk, R.E., 1994. Experimental Design: Procedures for the Behavioral Sciences. 3rd Edn., Wadsworth Publishing, Belmont, CA., pp: 944.

Klotz, S. and N.L. Johnson, 1993. Breakthroughs in Statistics: Volume 1: Foundations and Basic Theory. 1st Edn., Springer, New York, pp: 680.

Levy, K.J., 1975. An empirical comparison of the z-variance and Box-Scheffé tests for homogeneity of variance. Psychometrika, 40: 519-524. DOI: 10.1007/BF02291553

Levy, K.J., 1978. An empirical study of the cube-root test for homogeneity of variance with respect to the effects of non-normality and power. J. Stat. Comput. Simul., 7: 71-78. DOI: 10.1080/00949657808810209

Mazahreh, A.S., H. Hammad and H. Abu-Jaber, 2009. The attitudes of instructors and faculty members about the quality of technical education programs in community colleges in Jordan. J. Soc. Sci., 5: 401-407. http://www.scipub.org/fulltext/jss/jss54401-407.pdf

Mendenhall, W. and R. Beaver, 1991. Introduction to Probability and Statistics. 8th Edn. PWS Kent, Boston, ISBN: 0-534-92409-3, pp. 365-366.

O'Brien, R.G., 1978. Robust techniques for testing heterogeneity of variance effects in factorial designs. Psychometrika, 43: 327-342. DOI: 10.1007/BF02293643

O'Brien, R.G., 1981. A simple test for variance effects in experimental designs. Psychol. Bull., 89: 570-574. DOI: 10.1037/0033-2909.89.3.570

Oladejo, N.K. and I.A. Adetunde, 2009. Assessing Computer knowledge in senior high school: A case study of the Upper East region in Ghana. J. Math. Stat., 5: 287-297. http://www.scipub.org/fulltext/jms2/jms254287-297.pdf

Overall, J.E. and J.A. Woodward, 1974. A simple test for homogeneity of variance in complex factorial design. Psychometrika, 39: 311-218. DOI: 10.1007/BF02291705

Overall, J.E. and J.A. Woodward, 1976. A robust and powerful test for heterogeneity of variance. University of Texas Medical Branch Psychometric Laboratory.

Overall, J.E. and H.M. Rhoades, 1981. The significance of treatment effects across several independent studies. J. Psychiatric Res., 16: 241-257. DOI: 10.1016/0022-3956(81)90018-2

Pardo, J.A., M.C. Pardo, M.L. Vicente and M.D. Esteban, 1997. A statistical information theory approach to compare the homogeneity of several variances. Comput. Stat. Data Anal., 24: 411-416. DOI: 10.1016/S0167-9473(96)00080-1

Rana, M.S., H. Midi and A.H.M. Rahmatullah Imon, 2008. A robust modification of the goldfeld-quandt test for the detection of heteroscedasticity in the presence of outliers. J. Math. Stat., 4: 277-283. http://www.scipub.org/fulltext/jms2/jms244277-283.pdf

Samiuddin, M. and M. Atiqullah 1976. A test of equality of variance. Biometrika, 63: 206-208. DOI: 10.1093/biomet/63.1.206

Tomarken, A.J. and R.C. Serlin, 1986. Comparison of ANOVA alternatives under variance heterogeneity and specific noncentrality structures. Psychol. Bull., 99: 90-99. DOI: 10.1037/0033-2909.99.1.90

Veldman, D.J., 1967. Fortran Programming for the Behavioral Sciences. 1st Edn., Holt, Rinehart and Winston, New York, pp: 406.

Vrbanek Jr., J. and W. Wang, 2007. Data-driven forecasting schemes: Evaluation and applications. J. Comput. Sci., 3: 747-753. http://www.scipub.org/fulltext/jcs/jcs39747-753.pdf

Winer, B.J., 1971. Statistical Principles in Experimental Designs. 2nd Edn., McGraw-Hill Companies, New York, pp: 907.

Zeng, H.Z., R.W. Leung and M. Hipscher, 2010. An examination of teaching behaviors and learning activities in physical education class settings taught by three different levels of teachers. J. Soc. Sci., 6: 18-28. http://www.scipub.org/fulltext/jss/jss6118-28.pdf