***Southeast-Asian J. of Sciences*** Vol. 3, No. 2 (2014) pp. 181-194

# A COMPARISON OF TYPE I ERROR AND POWER OF BARTLETT'S TEST, LEVENE'S TEST AND O'BRIEN'S TEST FOR HOMOGENEITY OF VARIANCE TESTS

**Doungporn Hatchavanich**

*Department of Statistic, Faculty of Science and Technology*
*University of the Thai Chamber of Commerce Bangkok, Thailand*
*e-mail:doungporn_hat@utcc.ac.th*

**Abstract**

There are a good number of tests that are available for testing a hypothesis that samples come from populations with the same variance. Many studies reported that there is no test which is uniformly best for all distributions and sample size configurations.It can be seen that Bartlett's test, Levene's test and O'Brien's test offer different methods for researchers to test data. However, each test has some unigue weak points. To date, there are no studies about these tests when assumptions are violated under different situations. The aim of this paper is to compare the empirical probability of the Type I error and the power of the three statistical tests under the different types of distributions: normal, uniform,student's t, chi-square distribution and nine configurations of group size (n1, n2, n3, n4), the group variances were set as follows the ratio of 1:1:2:2, 1:2:3:4, 1:1:1:4.

It was found that no test outperformed the others in terms of robustness and power. The findings showed that the Levene's test was not the best option. Bartlett's test is a good choice to test homogeneity of variances since it is not affected by sample sizes when the data is normally or uniform distributed. Moreover, For low skew distribution, Bartlett's test is a good choice for small equal sample sizes and equal variances. O'Brien's test is the best for chi-square distributed. When the variance ratio of 1:1:1:4, low skew distribution, Bartlett's test and O'Brien's test

**Key words:** statistics for homogeneity of variance test , type I error , power of the tests, Bartletts test, Levenes test, OBriens test.

would be commended for small unequal sample sizes, since they still afford high power.

# 1. Introduction

There are a good number of tests that are available for testing a hypothesis that samples come from populations with the same variance (e.g., [2,4,14]). It is well known that classical tests for comparing variances are very sensitive to departures from normality. A large number of tests have been examined and stimulated in order to determine their robustness which is the capability to control the Type I error and their power. Many studies reported that there is no test which is uniformly best for all distributions and sample size configurations. One test which using the sample median as an estimate of the location parameter, usually stands out in terms of power and robustness against non-normality is a Levene's test [8]. After conducting extensive searches, it seems that other tests of homogeneity of variance may have been a better choice than the Levene test. Bartlett's test is extremely non-robust against non-normality [4, 7].O'Brien's test, which does fairly well for behavioral science data, is robust to data that departs from normality. It is competitive with other tests interms of power and it can be easily applied in different ANOVA designs with equal or unequal sample sizes[5].

From literature reviews, it is seen that the three statistic tests have different methods to test data and they have some different weak points. Especially, Bartlett's test is sensitive to violation of normality assumption. O'Brien's and Levene's test seem to be a good choice if robustness against non-normality is needed. Yet, there are no studies about these tests when assumptions are violated under different situations. The aim of this paper is to compare the robustness and the power of Bartlett's, Levene's and O'Brien's test. Section 2 provides a detailed description of all the tests. Section 3 reports the results of a simulation experiment on the small, moderate, and the large sample sizes performance of the tests, and the final Section gives some concluding remarks.

# 2. Description of the tests

## 2.1 Levenes test

Levene's test was defined as the one-way analysis of variance on $z_{ij} = |y_{ij} - \bar{y}_i|$, the absolute residuals and where k is the number of groups and the sample size of the ith group. The test statistic is given by:

$$L = \frac{(N-k)\sum_{i=1}^{k} n_i(\bar{z}_i - \bar{z})^2}{(k-1)\sum_{i=1}^{k}\sum_{j=1}^{n_1}(Z_{ij} - \bar{z}_i)^2}$$

where $N = \sum i = 1^k n_i,\ \bar{z}_i = \sum_{j=1}^{n_i} z_{ij}/n_i,\ \bar{z} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} z_{ij}/n.$

The $L$ is approximately distributed as an $F$ variable with $k-1$ and $N-k$ degree of freedom.[13]

## Bartlett's test

The null hypothesis is rejected when B1 is greater than the $100(1 = \alpha)_{th}$ percentile of the chi-squared distribution with $(I-1)$ degrees of freedom.

We consider a modification of Bartlett's test investigated by Boos and Brownie (1989). The modified test statistic is $B_2 = dB_1$, where $d = 2/(\hat{\beta}_2 - 1)$ and $\hat{\beta}_2 = \frac{N \sum_{i=1}^{I} \sum_{j=1}^{n_i} (x_{ij} - x_i)^4}{\left\{ \sum_{i=1}^{I} \sum_{j=1}^{n_i} (x_{ij} - \bar{x_i})^2) \right\}^2}$

The modification is motivated by the fact that under weak regularity conditions, $B_1 \rightarrow \frac{1}{2}(\beta_2 - 1)\chi^2_{I-1}$ in distribution under $H_0$ where $\beta_2 = E(X-\mu)^4/\sigma^4$ is the kurtosis of distribution (BOX, 1953) The critical point for $B_2$ is the same as that for $B_1$ [7.]

This 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 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 this study is transformed using the following formula:

$$v_{ij} = \frac{n_i - 1.5)n_i(y_{ij} - \bar{y}_i)^2 - 0.5s_i^2(n_i - 1)}{(n_i - 1)(n_i - 2)}$$

where $\bar{y_i} = \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i}$, the mean for each subgroup $i$ and $s_i^2 = \frac{\sum_{j=1}^{n_i} y_{ij}}{n_{i-1}}$, 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 $v_{ij}$ [6]. There are two criteria to detect appropriate statistics under violation of assumptions, robustness and power [7, 6, 13]. Robustness is the capability to control the Type I error. In other words, it is the ability of the test of not falsely detecting non-homogeneous groups when the underlying data are not normally distributed and the groups are in fact homogeneous. A statistical test is designated robust, if the departure of the empirical Type I error ($\hat{\tau}$) from the nominal level

of significance ($\alpha$) does not exceed the predetermined value. In this study, robustness evaluation is established on the Cochran limit as follows:

At the 0.01 significance level, $\tau$ value is between 0.007 and 0.015;

At the 0.05 significance level, $\tau$ value is between 0.040 and 0.060

$\tau$ = the true probability of a Type I error = Probability ($H_0$ is rejected when $H_0$ is true) and $\hat{\tau}$ = the empirical probability of a Type I error = $\frac{\text{the number ofH}_0 \text{ rejection when H}_0 \text{ is true}}{\text{the number of replications 10,000 times}}$,

$\alpha$= the nominal level of significance or the theoretical alpha The statistical test is called robust when its empirical alpha values lie within Cochran limit[3]. If any actual probability of the Type I error is over the limit, it shows that the test cannot control the error rate.

The power of the test is the probability of rejecting a null hypothesis when it is false and therefore should be rejected. In this study the power of the test is calculated by subtracting the empirical probability of a Type II error from 1.0:

$$\text{power } = \frac{\text{the number of H}_0 \text{ failed to reject when H}_0 \text{ is true}}{\text{the number of replications } 10,000 \text{ times}}$$

The maximum total power of the test is 1.0; the minimum is zero.

Heterogeneity of variance may affect both the type I and type II error rates. Box (1954) showed that the effect of heterogeneous variances on the type I error rate of the ANOVA F test would be approximately proportional to a coefficient of variation of the variances (Box's coefficient:

$$c = \frac{S_{c^2}}{\sigma^2} = \frac{\sqrt{\sum_{i=1}^{k}(\sigma_i^2 - \text{mean}(\sigma^2))^2/k}}{\text{mean}(\sigma^2)}$$

where $S_{c^2}$ is the standard deviation of the variances $\sigma_i^2$. So, $c$ may be used as a measure of the degree of heteroscedasticity. For a given range of variability, Box's coefficient is largest when one variance is large and the rest are small. When sample sizes are unequal, there is an additional important effect on the error rate, which depends on the proportion of the un-weighted (ignoring sample sizes) to the weighted (by the degrees of exemption of each sample) mean of the variances (Box 1954).This "bias ratio" reflects the grade to which small sample sizes are coupled with large variations.[1]

## Monte Carlo Simulation and Results

To compare the Levene's, Bartlett's, and O'Brien's tests a series of Monte Carlo studies were done.Each statistic is measured in terms of robustness and power.For robustness, the fewer Type I error a test make (falsely claiming

unequal variances, when in fact the variances are equal), the greater the robustness. With power, the higher the number of correctly detected unequal variances, when in fact they are unequal, the greater the power of the test.

The empirical Type I errorand power of the tests are investigated in the simulation study using normal, student's t, chi-square, and low skew distributions and many combinations of the sample sizes for 4 populations. The sample variances in each group of four populations were in the ratios 1:1:1:1 (under $H_0$) and 1:1:2:2, 1:2:3:4, 1:1:1:4 (under $H_1$) which the Box's coefficients were 0.33, 0.45 and 0.74. For estimating the empirical Type I error and power estimates, nominal 5% level is used throughout the study with 10,000 Monte Carlo Simulations. The data were generated in one situation for computing Levene's test, Bartlett's test and O'Brien's test. Then these values were compared with their critical region, the values that rejected null hypothesis were counted. In case of Type II error, the values that failed to reject the null hypothesis were counted and the power of the test was calculated bysubtracting the probability of the Type II error from 1.0. The process of computation was repeated for all situations.

Table1. The empirical Type I error under equal variance hypothesis of the Levene, Bartlett, and O'Brien tests for 0.05 significance level with normal, student's t, uniformchi-square, and low skew distribution.

| Sample Size $n_1,n_2,n_3,n_4$ | Under $H_0$ The ratio of variance 1:1:1:1 | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Normal | | | t | | | Uniform | | |
| | Levene | Bartlett | O'Brien | Levene | Bartlett | O'Brien | Levene | Bartlett | O'Brien |
| Unequal Sample sizes | | | | | | | | | |
| 10,15,20,25 | 0.0520* | 0.0479* | 0.0441* | 0.0460* | 0.0305 | 0.0382 | 0.0524* | 0.0553* | 0.0439* |
| 35,40,45,52 | 0.0508* | 0.0488* | 0.0473* | 0.0385 | 0.0316 | 0.0343 | 0.0519* | 0.0512* | 0.0475* |
| 35,50,65,80 | 0.0486* | 0.0482* | 0.0469* | 0.0390 | 0.0281 | 0.0373 | 0.0518* | 0.0506* | 0.0491* |
| 30,65,90,150 | 0.0508* | 0.0485* | 0.0513* | 0.0491* | 0.0286 | 0.0498* | 0.0496* | 0.0515* | 0.0477* |
| Equal Sample sizes | | | | | | | | | |
| 16,16,16,16 | 0.0513* | 0.0450* | 0.0410* | 0.0429* | 0.0378 | 0.0325 | 0.0497* | 0.0490* | 0.0380 |
| 20,20,20,20 | 0.0504* | 0.0478* | 0.0407* | 0.0360 | 0.0310 | 0.0285 | 0.0487* | 0.0477* | 0.0396 |
| 30,30,30,30 | 0.0523* | 0.0526* | 0.0455* | 0.0387 | 0.0316 | 0.0338 | 0.0536* | 0.0485* | 0.0468* |
| 50,50,50,50 | 0.0485* | 0.0501* | 0.0449* | 0.0377 | 0.0307 | 0.0329 | 0.0467* | 0.0467* | 0.0429* |
| 60,60,60,60 | 0.0522* | 0.0500* | 0.0478* | 0.0344 | 0.0302 | 0.0312 | 0.0483* | 0.0458* | 0.0455* |

* Type I error in control

When assumption of a normal distribution met, all three statistical tests could control the Type I error for all settings of equal and unequal sample sizes at $\alpha = 0.05$. This evidences to show when the distribution was normal, the sample size did not affect the robustness of the tests (Table 1, Figure1.).

**Table1 (continued)**

| Sample Size $n_1,n_2,n_3,n_4$ | Under $H_0$ The ratio of variance 1:1:1:1 | | | Under $H_0$ The ratio of variance 1:1:1:1 | | |
|---|---|---|---|---|---|---|
| | Chi-square | | | Low skew | | |
| | Levene | Bartlett | O'Brien | Levene | Bartlett | O'Brien |
| Unequal Sample sizes | | | | | | |
| 10,15,20,25 | 0.0656 | 0.0779 | 0.0553* | 0.0603 | 0.0531* | 0.0510* |
| 35,40,45,52 | 0.0463* | 0.0538* | 0.0428* | 0.0516* | 0.0481* | 0.0481* |
| 35,50,65,80 | 0.0515* | 0.0476* | 0.0493* | 0.0512* | 0.0489* | 0.0492* |
| 30,65,90,150 | 0.0527* | 0.0427* | 0.0528* | 0.0519* | 0.0465* | 0.0524* |
| Equal Sample sizes | | | | | | |
| 16,16,16,16 | 0.0693 | 0.0959 | 0.0548* | 0.0623 | 0.0588* | 0.0512* |
| 20,20,20,20 | 0.0559* | 0.0752 | 0.0463* | 0.0613 | 0.0555* | 0.0498* |
| 30,30,30,30 | 0.0509* | 0.0678 | 0.0452* | 0.0577* | 0.0531* | 0.0509* |
| 50,50,50,50 | 0.0489* | 0.0524* | 0.0457* | 0.0546* | 0.0492* | 0.0507* |
| 60,60,60,60 | 0.0444* | 0.0491* | 0.0402* | 0.0495* | 0.0472* | 0.0464* |

* Type I error in control

Figure 1. The empirical Type I error of Levene, Bartlett, and O'Brien tests for 0.05 significance level with normal, student's t, and uniform distribution.

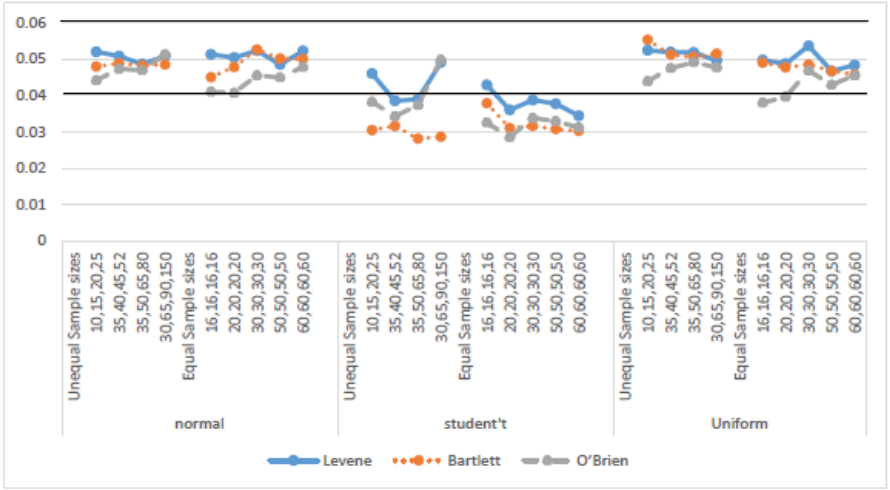


Figure 2. The empirical Type I error of Levene, Bartlett, and O'Brien tests for 0.05 significance level with chi-square, and low skew distribution.

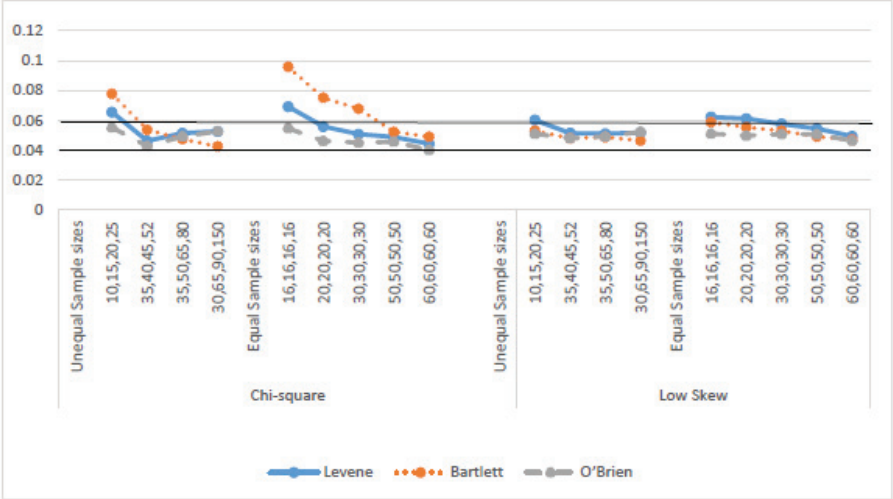
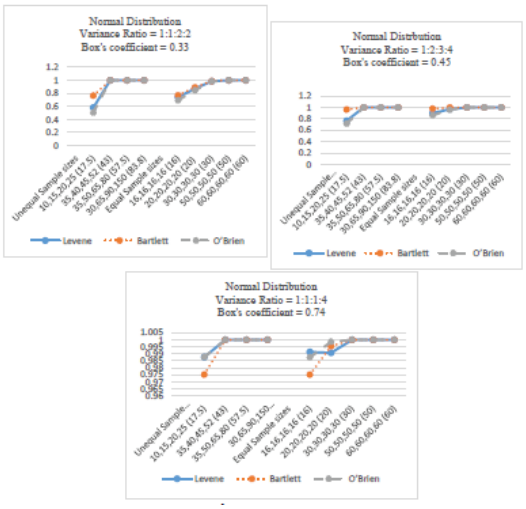



Table 2: Effect of heterogeneous variances in power of the Levene, Bartlett, and O'Brien tests of 0.05 significance level with normal, t, uniform, and chi-square distribution.

| Distribution | Sample Size $n_1,n_2,n_3,n_4$ (The average sample size) | Under $H_1$ The ratio of variance 1:1:2:2 | | | Under $H_1$ The ratio of variance 1:2:3:4 | | | Under $H_1$ The ratio of variance 1:1:1:4 | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Levene | Bartlett | O'Brien | Levene | Bartlett | O'Brien | Levene | Bartlett | O'Brien |
| Normal | Unequal Sample sizes | | | | | | | | | |
| | 10,15,20,25 (17.5) | 0.5812 | 0.7667 | 0.5055 | 0.7636 | 0.9625 | 0.7142 | 0.9876 | 0.9752 | 0.9876 |
| | 35,40,45,52 (43) | 0.9993 | 0.9996 | 0.9991 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| | 35,50,65,80 (57.5) | 0.9999 | 1.0000 | 0.9999 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| | 30,65,90,150 (83.8) | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| | Equal Sample sizes | | | | | | | | | |
| | 16,16,16,16 | 0.7424 | 0.7703 | 0.6942 | 0.8982 | 0.9774 | 0.8633 | 0.9913 | 0.9752 | 0.9875 |
| | 20,20,20,20 | 0.8726 | 0.8927 | 0.8469 | 0.9698 | 0.9964 | 0.9586 | 0.9907 | 0.9954 | 0.9985 |
| | 30,30,30,30 | 0.9859 | 0.9878 | 0.9824 | 0.9996 | 0.9999 | 0.9991 | 1.0000 | 0.9999 | 1.0000 |
| | 50,50,50,50 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| | 60,60,60,60 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |

| Distribution | Sample Size $n_1,n_2,n_3,n_4$ (The average sample size) | Under $H_1$ The ratio of variance 1:1:2:2 | | | Under $H_1$ The ratio of variance 1:2:3:4 | | | Under $H_1$ The ratio of variance 1:1:1:4 | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Levene | Bartlett | O'Brien | Levene | Bartlett | O'Brien | Levene | Bartlett | O'Brien |
| Uniform | Unequal Sample sizes | | | | | | | | | |
| | 10,15,20,25 (17.5) | 0.4886 | 0.4936 | 0.5055 | 0.7212 | 0.7945 | 0.6427 | 0.9873 | 0.9743 | 0.9280 |
| | 35,40,45,52 (43) | 0.9754 | 0.9747 | 0.9991 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.9876 |
| | 35,50,65,80 (57.5) | 0.9971 | 0.9968 | 0.9999 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| | 30,65,90,150 (83.8) | 0.9999 | 0.9999 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| | Equal Sample sizes | | | | | | | | | |
| | 16,16,16,16 | 0.5258 | 0.4862 | 0.6942 | 0.8011 | 0.8465 | 0.7653 | 0.9540 | 0.8962 | 0.9875 |
| | 20,20,20,20 | 0.6710 | 0.6393 | 0.8469 | 0.9199 | 0.9427 | 0.9033 | 0.9875 | 0.9707 | 0.9985 |
| | 30,30,30,30 | 0.8941 | 0.8846 | 0.9824 | 0.9599 | 0.9965 | 0.9937 | 0.9995 | 0.9991 | 1.0000 |
| | 50,50,50,50 | 0.9922 | 0.9913 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| | 60,60,60,60 | 0.9999 | 0.9987 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| Low skew | Unequal Sample sizes | | | | | | | | | |
| | 10,15,20,25 (17.5) | 0.0746 | 0.2291 | 0.1205 | 0.2468 | 0.4133 | 0.2047 | 0.7033 | 0.7398 | 0.7667 |
| | 35,40,45,52 (43) | 0.1541 | 0.5706 | 0.4985 | 0.8475 | 0.9084 | 0.8331 | 0.7515 | 0.9876 | 0.7461 |
| | 35,50,65,80 (57.5) | 0.5234 | 0.6753 | 0.5784 | 0.9032 | 0.9515 | 0.8895 | 0.9316 | 0.9986 | 0.7463 |
| | 30,65,90,150 (83.8) | 0.6039 | 0.7971 | 0.6808 | 0.9352 | 0.9774 | 0.9268 | 0.9928 | 1.0000 | 0.7510 |
| | Equal Sample sizes | | | | | | | | | |
| | 16,16,16,16 | 0.2346 | 0.2406 | 0.2022 | 0.4003 | 0.4759 | 0.3524 | 0.5585 | 0.6136 | 0.7515 |
| | 20,20,20,20 | 0.2851 | 0.2923 | 0.2535 | 0.4934 | 0.5735 | 0.4533 | 0.7403 | 0.7238 | 0.7444 |
| | 30,30,30,30 | 0.4153 | 0.4289 | 0.3932 | 0.7133 | 0.7834 | 0.6898 | 0.9912 | 0.8992 | 0.7406 |
| | 50,50,50,50 | 0.6425 | 0.6535 | 0.6311 | 0.9353 | 0.9558 | 0.9309 | 0.9991 | 0.9885 | 0.7410 |
| | 60,60,60,60 | 0.7181 | 0.7338 | 0.7087 | 0.9723 | 0.9831 | 0.9700 | 1.0000 | 0.9967 | 0.7411 |
| Chi-square | Unequal Sample sizes | | | | | | | | | |
| | 10,15,20,25 (17.5) | 0.0926 | 0.2104 | 0.0759 | 0.1596 | 0.3606 | 0.1327 | 0.4693 | 0.6041 | 0.4253 |
| | 35,40,45,52 (43) | 0.2060 | 0.2961 | 0.1938 | 0.4789 | 0.6434 | 0.4577 | 0.8057 | 0.8262 | 0.7953 |
| | 35,50,65,80 (57.5) | 0.2157 | 0.3560 | 0.2034 | 0.5252 | 0.7122 | 0.5038 | 0.8965 | 0.9193 | 0.8886 |
| | 30,65,90,150 (83.8) | 0.2216 | 0.4114 | 0.2100 | 0.5464 | 0.7623 | 0.5269 | 0.9695 | 0.9843 | 0.9665 |
| | Equal Sample sizes | | | | | | | | | |
| | 16,16,16,16 | 0.1581 | 0.2119 | 0.1338 | 0.2692 | 0.3832 | 0.2362 | 0.4674 | 0.4901 | 0.4318 |
| | 20,20,20,20 | 0.1641 | 0.2242 | 0.1468 | 0.3089 | 0.4278 | 0.2796 | 0.5378 | 0.5464 | 0.5098 |
| | 30,30,30,30 | 0.1895 | 0.2496 | 0.1758 | 0.4076 | 0.528 | 0.3875 | 0.6701 | 0.6621 | 0.6552 |
| | 50,50,50,50 | 0.2763 | 0.3293 | 0.2649 | 0.5960 | 0.6914 | 0.5821 | 0.8400 | 0.8282 | 0.8343 |
| | 60,60,60,60 | 0.3149 | 0.3670 | 0.3066 | 0.6738 | 0.7522 | 0.6626 | 0.8895 | 0.8752 | 0.8858 |

Figure 3.Power of Levene, Bartlett, and O'Brien tests as influenced by Variance Ratios and Sample Sizes with Normal Distributionat 0.05 significance level.




The Bartlett's test is most powerful in all the experimental cases when the average sample sizes are less than 20, for another case the power of the three statistic tests are approaching in 1. When the normality assumption met, but the assumptions of homogeneity of variance violated as 1:1:2:2 and 1:2:3:4, Bartlett's test was the best option as it kept up good power. However, when the variance ratio as 1:1:1:4, the power of the three statistical tests approached in 1. For all settings, the power of the three statistical tests tends to go higher as the average sample size increases and tends to go higher as the variance ratio increases (Box's coefficient increase).

Bartlett's test could control the Type I error for all of equal and unequal sample sizes, but O'Brien's test could control the Type I error for all unequal sample sizes andequal sample sizes which the average sample sizes were more than 20 (Table 1, Figure1.), for equal sample size and for all settings of unequal sample size. Nevertheless, the power of these tests tended to go higher as average sample size increased and inclined to go higher as a variance ratio increased.

When low skew distribution, Bartlett's test and O'Brien's test could control the Type I error for all settings of equal and unequal sample sizes at $\alpha = 0.05$, but Levene's test could control the Type I error for the average sample sizes was more than 20 (. As the variance ratio of 1:1:2:2 and 1:2:3:4, Bartlett's test gave the highest power for all settings of sample sizes. However, O'Brien's test gave the highest power for variance ratio of 1:1:1:4, as the average sample size less than 20, Bartlett's test gave the highest power as the average sample size more than 17.5, for equal sample size and more than 20 Levene's test gave the

Figure 4. Power of Levene, Bartlett, and O'Brien tests as influenced by Variance Ratios and Sample Sizes with Uniform Distributionat 0.05 significance level.

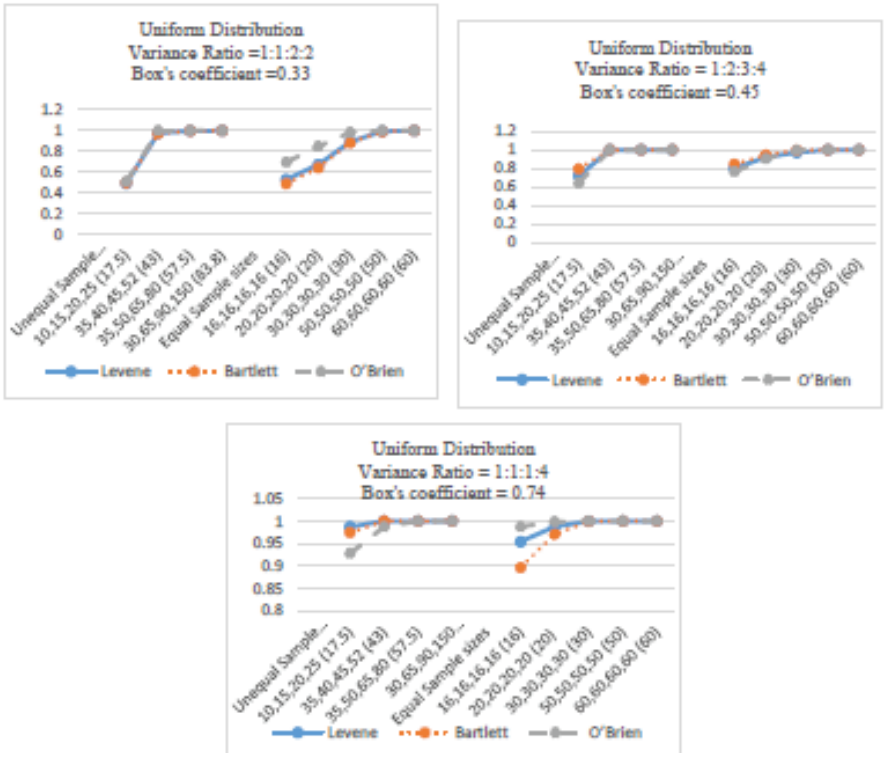
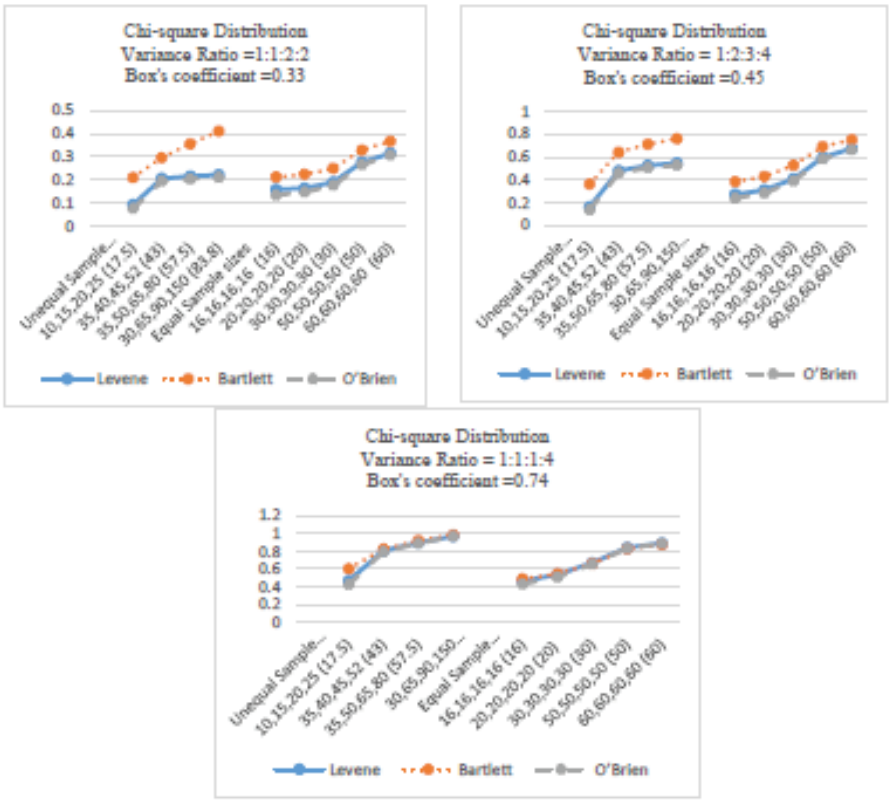


Figure 5. Power of Levene, Bartlett, and O'Brien tests as influenced by Variance Ratios and Sample Sizes with Low Skew Distribution at 0.05 significance level.

Figure 6. Power of Levene, Bartlett, and O'Brien tests as influenced by Variance Ratios and Sample Sizes with Chi-square Distribution at 0.05 significance level.




highest power. It likewise found that when normal assumption, homogeneity of variance assumption and equality of the sample sizes violated, Bartlett's test still was an honest choice that paid the highest power. It likewise found that the power of all tests tended to go higher as the Box's coefficient increases.

For chi-square distribution, O'Brien's test could control Type I error for all the settings of equal and unequal sample sizes, Levene's test for equal sample size of 20, 30, 40, 50 and 60 and Bartlett's test for sample size of 50 and 60. For unequal sample size, Levene's test and Bartlett's test could control Type I error of the for the average sample sizes more than 17.5. As the variance ratio of 1:1:2:2 and 1:2:3:4, Bartlett's test had the highest power for all settings of sample sizes, as the variance ratio of 1:1:1:4, and all settings of unequal sample size, Bartlett's test had a little higher power more than the others, for the equal sample size of 30, 50, 60 Levene's test had a little higher power more than the others.It likewise found that when the average sample sizes or the

variance ratio increased, the power of these tests also increased.

## Conclusion & Discussion

Under the normal assumption, homogeneity of variances, equal and unequal sample sizes, the empirical Type I error of Levene's, Bartlett's and O'Brien's test fall within the Cochran limits at =0.05. In particular, the empirical probability of Type I error of O'Brien's test was the most modest. This was the same as the studies by [9, 10, 11, 12] that may have been a safer option than the Levene's test. This current study showed that unequal sample sizes had not touched on the empirical probability of a Type I error of the three statistical tests. When the normality were met, but the homogeneity of variance was violated, variance ratio of 1:1:2:2 (Box's coefficient = 0.33) and 1:2:3:4 (Box's coefficient = 0.45), for equal and unequal sample sizes, Bartlett's test had the highest power. However, when the variance ratio was increased to 1:1:1:4 (Box's coefficient = 0.74), and equal sample sizes, power of the three tests were similar.

In terms of robustness, for normal distribution, Levene's test and Bartlett's test were as well as O'Brien's test. For uniform distribution, Levene's test and Bartlett's test appeared the superior over O'Brien's test. For chi-square distribution, O'Brien's test did its best and Levene's test can control the Type I error when the average sample sizes is more than 20. In terms of power, it appeared that Levene's test and Bartlett's test were as well as O'Brien's test. For the normal distribution and the average sample sizes were smaller than 20, Bartlett's test appeared the superior over Levene's test and O'Brien's test when 1:1:2:2 and 1:2:3:4 ratio of variance. Nonetheless, for low skew and Chi-square distribution, Bartlett's test appeared the superior over O'Brien's test and Levene's test for 1:1:2:2 and 1:2:3:4 ratio of variance in all equal and unequal sample sizes. For Chi-square distribution and variance ratio of 1:1:1:4, Bartlett's test were as well as the others.

Bartlett's test also seems to be robust to nominal significant level according to Cochran's criterion. However, only the empirical Type I error rate of Bartlett's test falls within the narrow interval. Bartlett's test shows a good performance even in asymmetric distributions like the low skew distribution, but it does not control the Type I error rate in highly asymmetric distributions, like the Chi-square distribution. Nevertheless, it is slightly more powerful than Levene's test and O'Brien's test.

In closing, no test outperformed the others in terms of robustness and power. The Levene's test was not the best option. There were better testing could be utilized, and some were more preferable depending on the distributional form of the data. For a normal distribution, entirely of the ratios of variance, the average sample sizes are more than 30, the power approach to 1. For uniform

distribution, the power of the Levene's test is similar to Bartlett's test, when the average sample sizes increase, the power approach to 1. For low skew distribution, the power of Bartlett's test is better than O'Brien's test. For student's distribution, no statistical test could control the Type I error for all of sample sizes.

Because of these determinations are founded on one set of simulation experiment, generalizations require caution. For instance, the operating characteristics of the tests may vary when there are more than four groups. Also, the conclusions may not be applicable if the true variances are very different than those investigated here.

# References

[1] Box, G.E.P.,Non-normality and Tests on Variances. Biometrika, 40, 318-335 (1953)

[2] Brown, M.B. and A.B. Forsythe, Robust tests for equality of variances, J. Amer. Statist. Assoc., 69, 364-367 (1974)

[3] Cochran, W.G.,The combination of estimates from different experiments. Biometrics,10:101-129(1954)

[4] Conover, W.J., M.E. Johnson and M.M. Johnson, A comparative study of tests for homogeneity of variances, with applications to the outer continental shelf bidding data, Technometrics, 23 351-361 (1981)

[5] Howell, D.C., Statistical Methods for Psychology, fourth ed. Wadsworth, Belmont (2001)

[6] Katz Gary S., Restori Alberto F., Lee Howard B. [on-line]. Available at http://www.questia.com/library/journal/1G1-213084806/a-monte-carlo-study-comparing-the-levene-test-to-other/ [Accessed 20 December 2013]

[7] Lim, T.S. and Loh, W.Y.,A Comparison of Tests of Equality of Variances. Computational Statistics and Data Analysis 22 287-301 (1996)

[8] Levene, H.,Robust Testes for Equality of Variances, in Contributions to Probability and Statistics. California: Stanford University Press.(1960)

[9] Levy, K.J.,An empirical comparison of the z-variance and Box-Scheff'e tests for homogeneity of variance. Psychometrika, 40: 519-524. DOI:10.1007/BF02291553 (1975)

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

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

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

[13] Parra-Frutos, I., The behaviour of the modified Levene's test when data are not normally distributed. Compute Stat, Springer, 671-693. (2009)

[14] Sharma, S.C., A new jackknife test for homogeneity of variances, Comm. Statist.: Simulation Comput.,20, 479-495 (1991)