Effect Of Sample Sizes On The Empirical Power Of Some Tests Of Homogeneity Of Variances

Ogbonna, Chukwudi J.[#1], Okenwe Idochi[2], Ifeanyichukwu Ogechukwu Sylvia[#3]

[#] *Department of Statistics, Federal University of Technology Owerri Nigeria PMB 1526, Owerri Nigeria*
[2]*Department of Statistics, School of Applied Sciences, Ken Saro Wiwa Polytechnic PMB 20, Bori, Rivers State Nigeria*

ABSTRACT

This study compares the effect of sample sizes on the empirical power of some homogeneity of variance tests that have been proposed to assess the homogeneity of within-group variances, prior to ANOVA. The Tests of Homogeneity of Variance (THV) compared are: Bartlett (1937), Levene (1960), Cochran (1941) and Hartley Fmax (1950) tests. Homogeneity of variances occurs when variances are equal across groups. Homogeneity of variance testing is a statistical method designed to provide evidence that groups are comparable by demonstrating that the variations found between groups are small enough that they are considered practically insignificant.Few recommendations exist regarding the appropriate use of these tests under varying data conditions. Monte Carlo simulation methods were used to generate data to examine and compare the power rates of the tests under conditions of equal and unequal sample sizes when the underlying distribution is normal 1,000 times through the use of R software. It was found that Hartley Fmax test performs best (highest power) when the sample sizes are equal, while, Cochran test has the worst performance. Generally, when the sample sizes are both equal and unequal, Levene test has the highest power followed by Bartlett test, hence, Bartlett and Levene tests will be recommended for both equal and unequal sample sizes since they give higher power (above 0.8).Thus it is important for researchers to conduct an initial analysis of the data in order to determine the distribution of the population and is also advised to pay attention to the amount of sample size required to obtain a powerful test.

Keywords: *Tests of Homogeneity of Variance, Empirical Power, Cochran Test, Levene Test, Hartley Fmax Test*

INTRODUCTION

Tests for homogeneity of variances are often of interest as a preliminary to other analyses such as analysis of variance or a pooling of data from different sources to yield an improved estimated variance. Many authors claim that a test of homogeneity of variancesis a prerequisite to analysis of variance. Others, like Zar (1999), confide that the testspresently available have such poor performance that they are not really useful, ANOVA beingmore robust to departures from homoscedasticity than can be detected using a test ofhomogeneity of variances, especially under conditions of non-normality. To apply the ANOVA test, several assumptions must be verified, including normal populations, homoscedasticity and independent observations. Underwood (1997) reminds us that the analysis of variance presents problems with heterogeneity in balancedsamples only when one of the variances is markedly larger than the others; it is notespecially sensitive to non-normality of the data which badly affects most of the classicaltests for homogeneity of variances.

Homogeneity of variances occurs when populations have similar variance. The classical approach to hypothesis testing usually begins with the likelihood ratio test under the assumption of normal distributions. However, the distribution of the statistic in the likelihood ratio test for equality of variances in normal populations depend on the kurtosis of the distribution (Box (1953)), which helps to explain why that test is so sensitive to departures from normality. This non robust property of the likelihood ratio test has prompted the invention of many alternative tests for variances. Some of these tests are modifications of the likelihood ratio test. Others are adaptations of the F test to test the variances rather than the sample means. Many tests are based on the non parametric methodsalthough their modification for the case in which the means are unknown often makes these tests distributionally dependent.

A good number of tests for homogeneity of variances are available for different situations. The most frequently used tests include those of Bartlett, Levene, Brown- Forsythe, Fligner Killeen, Welch tests, etc. All these tests have their various limitations. The various tests considered in this research work are Bartlett (1937), Levene (1960), Cochran (1941) and Hartley's Fmax (1950) tests.

Bartlett's test involves computing a statistic whose sampling distribution is closely approximated by the chi-square distribution with *n*-1 degrees of freedom when the random samples are from independent normal populations (Montgomery (1997)). Bartlett's statistic is designed to test for equality of variances across groups against the alternative that variances are unequal for at least two groups.

The hypothesis of Bartlett test is given by:

H_0: $\sigma_1^2 = \sigma_2^2 = ... = \sigma_k^2$; where k is the number of independent samples.
H_1: $\sigma_i^2 \neq \sigma_j^2$ for at least one pair (i, j) where $i \neq j$.

Bartlett test can be used to answer the following question: Is the assumption of equal variances valid? It is useful whenever the assumption of equalvariances is made. In particular, this assumption is made forthe frequently used one-way analysis of variance. Bartlett test is known to be powerful only if the sampled populations are normal, but badly affected by non-normality (Box, 1953). Bartlett test is the one most often presented in textbooks and taught in introductory courses because of its ease of computation.It can be easily applied in different ANOVA designs with equal or unequal sample sizes.

The Levene test uses an F test to test the null hypothesis that variances are equal across groups. The test is meant to be used with normally distributed data but can tolerate relatively low deviation from normality. Levene test is an alternative to the Bartlett test though the test is less sensitive to departures from normality than the Bartlett test. If there is strong evidence that the data do in fact come from a normal, or nearly normal, distribution, then Bartlett test has better performance. The assumptions underlying this test include: the samples from the populations under consideration are independent and approximately normal.

Cochran test is a one sided upper limit variance outlier test. The Cochrantest is used to decide if a single estimate of a variance is significantly larger than a group of variances with which the single estimate is supposed to be comparable i.e., it is used to assess the homogeneity of variances in the one factor case. The Cochrantest is essentially an outlier test. Assumptions of Cochran test include: observations in each group are normally distributed and independent among groups.

The study compares four tests for homogeneity of variances based on their powers through simulation studies. The simulation is conducted by varying different sample sizes to determine which of the tests that have a higher power. In each simulation problem, the rate of rejection of the null hypothesis (at $\alpha = 5\%$) when the null hypothesis is false is the power computed after 1000 independent simulations involving the desired number of observations from the selected type of distribution which is a normal distribution.

Statement of Problem

Often, the effect of violation of assumption of one-way ANOVA result depends on the extent of the violation (such as how unequal the population variances are, or how heavy-tailed one or another population distribution is). Some small violations may have little practical effect on the analysis, while other violations may render the one-way ANOVA result incorrect or un-interpretable. In particular, small sample sizes can increase vulnerability to assumption violations. While homogeneity of variance tests have been gaining in popularity within fields such as education and psychology, as researchers are becoming more aware of the different methods, few recommendations currently exist regarding the appropriate use of the tests. The primary concern is related to whether these tests will be able to correctly detect homogeneity among groups when the groups are in fact homogenous (i.e., the test has sufficient power) and will not conclude equivalence when the groups are truly different (i.e., type I error). Typically we want the power of our test to be 0.80 or greater (i.e., we will correctly conclude equivalence 80% or more of the time). Insufficient power could lead to a conclusion of non homogeneity even if the population variances are equivalent (a type II error).

We note that the effects of the assumption of normality and constant variance are inter-related. As the violation of the assumption of homogeneity of variance is likely caused by a small sample or by the violation of normality, sample size is an important factor that conceptually influences the power of homogeneity of variance tests. For instance, small sample sizes result in larger confidence intervals than large sample sizes. As such, small sample sizes should more likely lead to conclusions of heterogeneity while large sample sizes should more likely lead to conclusions of homogeneity.

In this study, we compare the effect of sample sizes on the empirical power of the tests. Monte Carlo simulation based on normal distributions for various sample sizes both equal and unequal will reveal which of the four tests that is more powerful. Hence, we hope that our study will indicate an appropriate test of homogeneity of variances for testing the assumption of homogeneity of variancesin a given data set.

Literature Review

Diep *et al*. (2014) provided a SAS macro for testing the homogeneity of variance assumption in one way ANOVA models using fourteen different approaches. They used simulation methods to compare the fourteen tests in terms of their type I error rate and statistical power. The tests are Brown and Forsythe, Levene (absolute values), Levene(squared values), O'Brien, Bartlett, Bootstrap Brown-Forsythe, Ramsey's Conditional Test, Z-variance, MZ-variance test, Cochran C test(with arithmetic mean), Cochran C test(with harmonic mean), G test, F-max test (with arithmetic mean), F-max test(with harmonic mean). The result obtained from the simulation showed that Levene with the Squared residuals, Brown-Forsythe, O'Brien, Ramsey's conditional procedure and Bootstrap Brown-Forsythe tests were the five tests that controlled type I error adequately. O'Brien had slightly less power and Bootstrap Brown-Forsythe had slightly greater power than other tests. The O'Brien and Ramsey conditional procedure seemed to be the most robust.

Li *et al*. (2015) compared the performances of 7 homogeneity of variance tests namely: F test, Bartlett test, Levene test, trimmed –mean –based Levene test, Brown Forsythe test, Phipson and Oshlack equal variance test based on absolute difference and Phipson and Oshlack equal variance test based on squared difference. Simulation studies were used to evaluate the effects of sample size, inequality of means, non normal distributions and outliers on the performances of the 7 equal variance tests. The results showed that as the sample size increases, the performance of the 7 tests improve. That is, the estimated type I error rates are closer to the nominal value 0.05 and the estimated power is closer to the maximum value of one as the sample size increases. The inequality of means had no effect on all the7 tests. Both non normality and outliers had effects on the performance of the 7 tests. F and Bartlett's test were much more sensitive (i.e., having large type 1 error rates) to non normality and outliers than the other 5 tests.

In a simulation study carried out by Harrit*et al*. (2017) to compare empirical type I error and power of five tests namely: Anom, Bartlett's, Levenes, Brown-Forsythe and Conover tests to check homogeneity of variances, a comprehensive Monte Carlo simulation study was carried out for different number of groups (k=3,4,5 and 10), variance ratios (1,5,10,15 and 20) and sample size combinations under normality assumption. Based on the results of the simulation, carried out, it was observed that the best robust tests are the Anom and Bartlett's tests followed by Levene and Conover's test. But both Conover and Levene tests have been slightly negatively affected from increase in the number of groups when sample sizes were small. On the other hand, since the Brown-Forsythe test did not give satisfactory results for any of the experimental conditions, it was concluded that the use of the test should not be preferred for checking homogeneity of variance assumption.

Brief Description of the Tests

Let Y_{ij} be an observation on the *j*th subject of the *i*th treatment (population), then the one way ANOVA model can be written as:

$$Y_{ij} = \mu + \alpha_i + e_{ij}, \quad i = 1, 2, \ldots, t; \; j = 1, 2, \ldots, n.$$

Where μ is the overall mean for all t treatments, α_i the *i*th treatment effect with restriction $\sum_{i=1}^{t} \alpha_i = 0$ and $e_{ij's}$ represent error terms which are assumed to be mutually independent and normally distributed with $E(e_{ij}) = 0$, $V(e_{ij}) = \sigma^2$.

Our interest is on testing the null hypothesis

$H_0 : \sigma_1^2=\sigma_2^2=\ldots=\sigma_k^2$

Against the alternative hypothesis
$H_1: \sigma_i^2 \neq \sigma_j^2$ for at least one pair (i, j) for i≠j.
The test for homogeneity of variances is used to determine if the assumption of equal variances is valid.

Statement of hypothesis for normality assumption of the real data collected is given by;

H_0: The data is normally distributed.

H_1: The data is not normally distributed.

A p-value less than 0.05 indicates a violation of the assumption. There are many tests of assumptions of homogeneity of variances. Commonly used tests which will be looked at in this research work are the Bartlett (1937), Levene (1960), Hartley (1950) and Cochran (1941). In what follows, we give a brief procedure of each test.

Bartlett Test
The hypothesis for Bartlett test is given as:
H_o: $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2$; where k is the number of independent samples.
H_1: $\sigma_i^2 \neq \sigma_j^2$ for at least one pair (*i*, *j*) where i≠j.

To test for equality of variances across groups against the alternative that variances are unequal for at least two groups, the Bartlett test statistic is used. It is given by:

$$B= C^{-1}[(n-k) \ln S^2- \sum_{i=1}^{k}(n_i-1)\ln S_i^2];$$

Where;

$$C=1+\frac{1}{3(k-1)}\left(\sum_{i=1}^{k}\frac{1}{n_i-1}-\frac{1}{n-k}\right);$$

S_i^2 is the variance of the ith group and is given by;

$$S_i^2 = \frac{\sum(x-\bar{x})^2}{n_i-1};$$

n is the total sample size;
n_i is the sample size of the ith group;
k is the number of groups
S^2 is the pooled variance given by;

$$S^2=\frac{\sum(n_i-1)S_i^2}{N-k};$$

The pooled variance is the weighted average of the group variances. The Bartlett test statistic follows chi-square distribution with (k-1) degrees of freedom.
The hypothesis H_0 is rejected on significance level α, when:

$$B > \chi^2_{1-\alpha,k-1};$$

Where $\chi^2_{1-\alpha,k-1}$ is the critical value of the chi-square distribution with k-1 degrees of freedom. Bartlett's test is known to be powerful if the sampled populations are normal, but badly affected by non-normality (Box, 1953).

Levene Test

Definition: The hypothesis for the Levene test is given as:
H_o: $\sigma_1^2 = \sigma_2^2 = ... = \sigma_k^2$; where k is the number of independent samples.
H_1: $\sigma_i^2 \neq \sigma_j^2$ for at least one pair (i, j) where $i \neq j$.

Given a variable Y with sample of size N divided into k subgroups, where N_i is the sample size of the ith subgroup, the Levene test statistic is defined as:

$$W = \frac{N-k}{k-1} \frac{\sum_{i=1}^{k} N_i (\bar{Z}_{i.} - \bar{Z}_{..})^2}{\sum_{i=1}^{k} \sum_{j=1}^{N_i} (Z_{ij} - \bar{Z}_{i.})^2}$$

Where;

Z_{ij}= sample observation j taken from treatment i ;

N_i =number of observations from treatment i (at least one N_i must be 3 or more);

N = overall size of combined samples;

$\bar{Z}_{i.}$ =mean of sample data from treatment i;

$\bar{Z}_{..}$ =mean of overall sample.
At least one of the treatments must have three or more observations else the Levene test statistic will be undefined (i.e., the denominator will equal zero if each treatment has 1 or 2 observations).
Z_{ij} can be computed using any of the following formulae:

1. $Z_{ij} = \left| Y_{ij} - \bar{Y}_{i.} \right|$ where $\bar{Y}_{i.}$ is the mean of the ith subgroup.
2. $Z_{ij} = \left| Y_{ij} - \tilde{Y}_{i.} \right|$ where $\tilde{Y}_{i.}$ is the median of the ith subgroup.
3. $Z_{ij} = \left| Y_{ij} - \bar{Y}_{i.}' \right|$ where $\bar{Y}_{i.}'$ is the 10% trimmed mean of the ith subgroup.

Here the 10% trimmed mean is the arithmetic mean calculated when the largest 10% and smallest 10% of the cases have been eliminated; eliminating extreme cases from the computation of the mean results in a better estimate of central tendency, especially when the data are non-normal.

The three choices for calculating Z_{ij} determine the robustness and power of Levene test. The Levene test rejects the hypothesis that the variances are equal if

$$W > F_{\alpha, k-1, N-k}$$

where $F_{\alpha, k-1, N-k}$ is the upper critical value of the F distribution with k-1 and N-k degrees of freedom at a significance level of α.

Hartley Fmax Test

The test follows the form of hypothesis testing starting with the null hypothesis.
H_0: $\sigma_1^2 = \sigma_2^2 = … = \sigma_k^2$; where k is the number of populations or treatments.
H_1: $\sigma_i^2 \neq \sigma_j^2$ for at least one pair (i, j) where $i \neq j$.

The test statistic is given by:

$$\text{Fmax} = \frac{\sigma^2 \max}{\sigma^2 \min}$$

After test statistic is calculated, a critical value is used to make determination. An *F* table is used to find a value corresponding to *a* = 0.05 and degrees of freedom= n-1 where n is the sample size drawn from each population.

Cochran Test

The hypothesis for Cochran test is given by:

$$H_0 : S^2_{max} = \sum_{i=1}^{k} S_k^{\ 2}$$

$$H_1 : S^2_{max} \neq \sum_{i=1}^{k} S_k^{\ 2}$$

The test statistic is calculated using the equation:

$$C = \frac{S^2 \max}{\sum_{i=1}^{k} S_i^{\ 2}}$$

Where;

$\sum_{i=1}^{k} S_i^{\ 2}$ is the sum of variances of the observations with equal sizes

($n_1 = n_2 = \ldots = n_k$ and $df_1 = df_2 = \ldots = df_n$ respectively);

k= number of groups

n = number of observations in each group.

df = degrees of freedom.

The test value C is compared with the value of the Cochran table.

The null hypothesis (H_0) is rejected if the test value C is greater than the critical value C (probability=α; k, df).

The Bartlett, Hartley and Cochran tests test for homogeneity of variance without transformation of the data. The Levene method actually transforms the data and then tests for equality of variances. Cochran and Hartley tests assume that there are equal numbers of observations in each group. The tests of Bartlett, Cochran, Hartley and Levene may be applied for number of samples K>2. In such situation, the power of these tests turns out to be different.

Statistical Power of a Hypothesis Test.

The statistical power of a hypothesis test is the probability that a statistical test correctly rejects the null hypothesis (H_0) when the alternative hypothesis (H_1) is true. It can be equivalently thought of as the probability of accepting the alternative hypothesis (H_1) when it is true i.e., the ability of a test to detect an effect, if the effect actually exists. That is,

Power =P(reject $H_0|H_1$ is true)

The probability of not committing a Type II error is called the power of a hypothesis test. A test without sufficient statistical power not be able to provide the researcher with enough information to draw conclusions regarding the acceptance or rejection of the null hypothesis. Usually one would only consider the power of a test when one failed to reject the null hypothesis. High power is desirable from (0.7 to 1.0). High power means that there is a high probability of rejecting the null hypothesis when the null hypothesis is false.

As the power increases, there are decreasing chances of a type II error (false positive), which are also referred to as the false negative ratio (β) since the power is equal to 1-β, again, under the alternative hypothesis. A similar concept is type 1 error, also referred to as the "false positive rate" or the level of the test under the null hypothesis.

Simulation Technique

The program flow chart for the simulation study is as follows:

i. Generate random samples from the distribution specified by the alternative hypothesis.
ii. Select the sample sizes, the number of simulation times, the level of significance α.
iii. Calculate the test statistics from the simulated data under Bartlett, Levene, Cochran Hartley Fmax and determine if the null hypothesis is accepted or rejected. Tabulate the number of rejections and use this to calculate the test's significance level.
iv. Repeat steps (i), (ii) and (iii) several number of times as specified, tabulating the number of times the simulated data leads to a rejection of the null hypothesis. The power of the test is the proportion of simulated samples in step (iii) that leads to rejection of the null hypothesis.

The number of repetition of simulation has been considered to be 1,000. The alpha level (α) set at 0.05. All simulated data were generated from a normal distribution.

DATA COLLECTION AND ANALYSIS

Table 1: The number of university of Port-Harcourt students who suffered pneumonia

YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
2006	80	94	98	90	103	92	83	98	85	100	96	82
2007	90	95	87	89	94	107	92	89	103	95	102	91
2008	100	101	99	92	96	90	93	88	90	94	91	95
2009	88	100	92	93	92	108	91	100	90	95	89	93
2010	103	100	93	95	88	93	92	99	93	81	91	94
2011	107	94	93	104	109	98	111	94	90	94	94	105
2012	89	98	87	95	91	89	101	98	89	91	100	94
2013	103	93	100	107	94	91	86	95	90	89	99	85
2014	92	96	108	91	96	95	91	101	105	97	104	93
2015	98	87	96	100	95	91	97	96	88	92	91	83

Steps for the real data analysis

Checking the distribution of the data
TREATMENT
2006 2007 2008 2009 2010 2011 2012 2013 2014 2015
12 12 12 12 12 12 12 12 12 12
shapiro.test(Data4test$RESPONSE)##to check for normality of the data
Shapiro-Wilk normality test
data: Data4test$RESPONSE
W = 0.98029, p-value = 0.07579

Interpretation: since the p-value is greater than 0.05, the data is normally distributed.
Hence, the normality assumption is met.

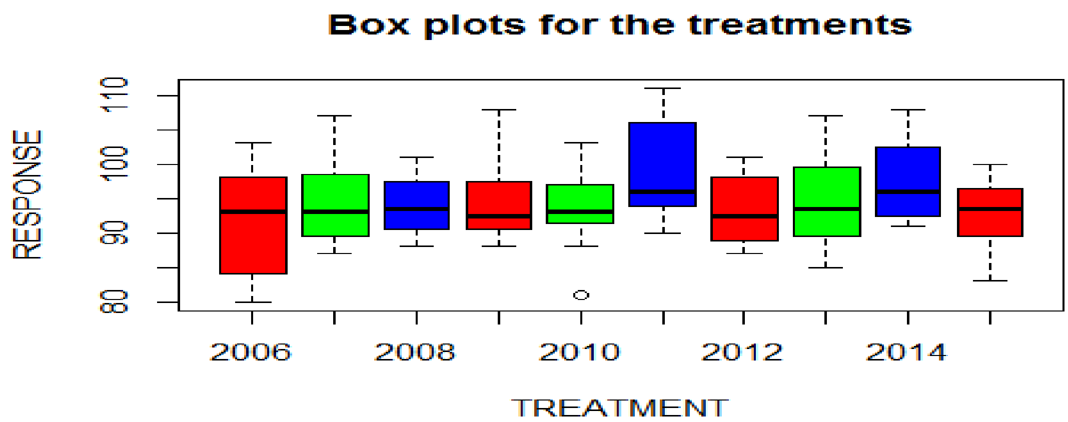

Fig.1: Box plots

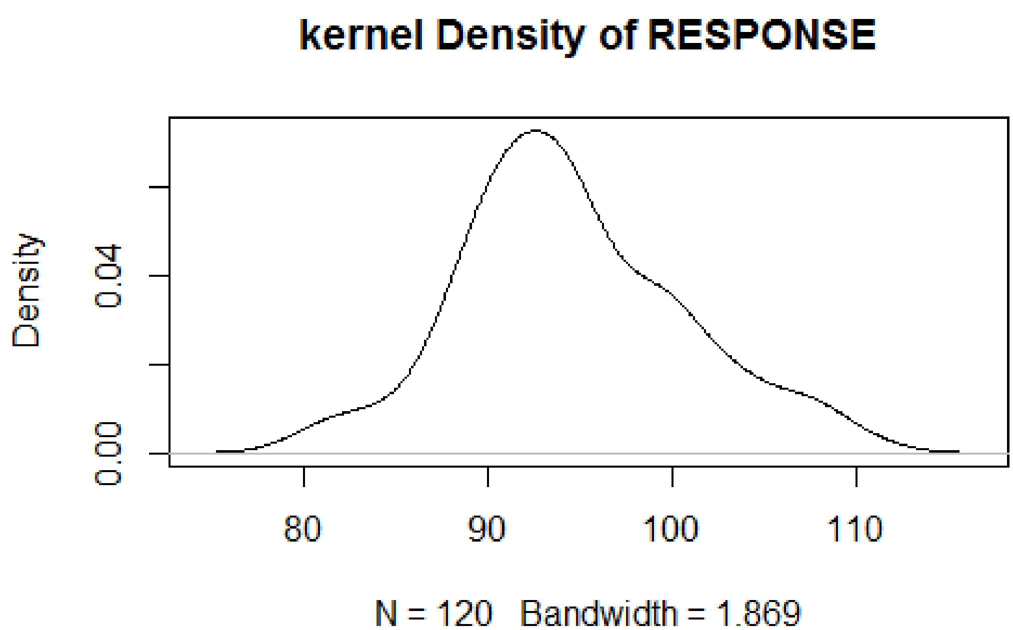

Fig. 2: Density plot

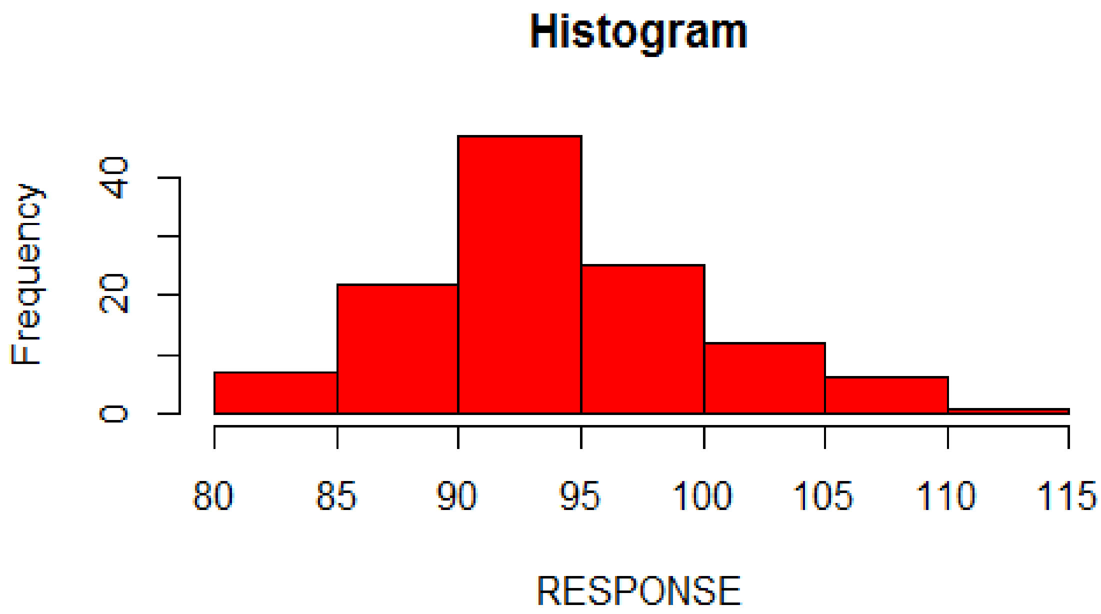

Fig. 3: Histogram plot

Interpretation: it is observed that the shape of the data implies normal distribution

Bartlett test of homogeneity of variances

data: RESPONSE by TREATMENT

Bartlett K-squared = 6.5441, df = 9, p-value =0.6845

Interpretation:

Since the p-value is greater than 0.05, there is no variation in the treatment groups(number of years).

Levene test of homogeneity.

Levene Test for Homogeneity of Variance (center = median)

	Df	F value	Pr(>F)
group	9	0.8869	0.5394

Interpretation:

Since the p-value is greater than 0.05, there is no variation in the treatment groups (number of years).

Cochran Test for Homogeneity

Cochran test of homogeneity of variances

data: lm(data = Data4test, RESPONSE ~ TREATMENT)

C = 0.16297, n = 12, k = 10, p-value = 0.7293

alternative hypothesis: Group 2006 has outlying variance

sample estimates:

2006	2007	2008	2009	2010	2011	2012
59.4773	40.0909	17.9015	33.1136	32.8182	53.1742	23.3636

2013	2014	2015
46.0606	33.3561	25.6061

Interpretation:

Since the p-value is greater than 0.05, there is no variation in the treatment groups(number of years).

HartleyFmax test

[1] 9.89927

Analysis using simulated data

Bartlett test of homogeneity of variances

data: values by ind

Bartlett's K-squared = 6.3235, df = 11, p-value = 0.8509

Levene test of homogeneity of variances

Levene's Test for Homogeneity of Variance (center = median)

Df F value Pr(>F)

group 11 0.3321 0.9769

108

Hartley Fmax test of homogeneity of variances

[1] 9.89927

Cochran test of homogeneity of variances

Cochran test of homogeneity of variances

data: lm(values ~ ind, data = test)

C = 0.14436, n = 10, k = 12, p-value = 0.8106

alternative hypothesis: Group X1 has outlying variance

sample estimates:

X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
1.6404	0.6230	1.2263	0.9492	0.8048	0.4231	0.6888	0.7005	1.3340	0.8940	0.9575	1.1220

Interpretation: The results of the various tests show that there is no variation in the groups. Hence the assumption of homogeneity of variance is met.

Simulation for Power of the Various Tests

The R codes used for the simulation are given in the appendix.

Table 2: Simulation Results for power of Bartlett test

Equal sample size	Power	Unequal sample size	Power
15,15,15; N=45	0.4968	15,20,30; N=65	0.6662
20,20,20; N=60	0.6688	20,30,45; N=95	0.8614
30,30,30; N=90	0.8716	40,50,60; N=150	0.9866
50,50,50; N=150	0.9866	50,50,100; N=200	0.9994

Interpretation: It is observed that as the sample sizes increase (for both equal and unequal), the power increase. This implies that increase in sample size improves the power of a test.

Table 3 : Simulation Results for power of Levene test

Equal sample size	Power	Unequal sample size	Power
15,15,15; N=45	0.9518	15,20,30; N=65	0.9538
20,20,20; N=60	0.9463	20,30,45; N=95	0.9725
30,30,30; N=90	0.9597	40,50,60; N=150	0.9988

Interpretation:

It is observed that as the sample sizes increase (for both equal and unequal), the power of the test increase with greater powers occurring at unequal sample sizes.

Table 4: Simulation results for power of Cochran test

Equal sample size	Power	Unequal sample size	Power
15,15,15; N=45	0.3333	15,20,30; N=65	0.4615
20,20,20; N=60	0.3333	20,30,45; N=95	0.4737
30,30,30; N=90	0.3333	40,50,60; N=150	0.4801
50,50,50; N=150	0.3333	50,50,100; N=200	0.5000

Interpretation:
It is observed that when the sample sizes are equal, the power values remain the same and very low. But when the sample sizes are unequal, the power of the test increases but considerably low compared to the powers of Bartlett and Levene tests.

Table 5: Simulation results for Hartley Fmax test

Equal sample size	Power	Unequal sample size	Power
15,15,15; N=45	1.0000	15,20,30; N=65	0.4444
20,20,20; N=60	1.0000	20,30,45; N=95	0.5000
30,30,30; N=90	1.0000	40,50,60; N=150	0.6667
50,50,50; N=150	1.0000	50,50,100; N=200	0.5000

Interpretation: It is observed that when the sample sizes are equal, the power values remain the same with the perfect value of one.. But when the sample sizes are unequal, the power of the test increases but considerably low compared to the powers of Bartlett and Levene tests.
Thus, Cochran and Hartley Fmax test behave alike.

Table 6: Summary of the simulation results for the four tests.

Sample size	Power			
Equal	Bartlett	Levene	Cochran	Hartley Fmax
15,15,15; N=45	0.4968	0.9518	0.3333	1.0000
20,20,20; N=60	0.6662	0.9463	0.3333	1.0000
30,30,30; N=90	0.8614	0.9597	0.3333	1.0000
50,50,50; N=150	0.9866	0.9952	0.3333	1.0000
Sample size	Power			
Unequal	Bartlett	Levene	Cochran	Hartley's Fmax
15,20,30; N=65	0.6688	0.9538	0.4615	0.4444
20,30,45; N=95	0.8716	0.9725	0.4737	0.5000
40,50,60; N=150	0.9866	0.9988	0.4801	0.6667
50,50,100; N=200	0.9994	0.9997	0.5000	0.5000

Interpretation: The power of Levene test is similar to Bartlett test. When the sample sizes increase, the powers approach one for both equal and unequal cases. Cochran test performs better when the sample size is unequal.

Summary and Conclusion

The purpose of this research was to perform a quantitative assessment of the statistical power of four homogeneity of variance tests namely: Bartlett, Levene, Cochran and Hartley Fmax tests by varying different sample sizes. The results obtained in this study showed that Levene test have the highest power. However, Bartlett and Levene tests have similar power (above 0.8), while Cochran's testperforms poorly (power around 0.3) when the sample sizes are equal. The study showed that equal sample sizes had no effect on the powers of Cochran and Hartley's Fmax tests. Conclusively, Levene and Bartlett tests appear to be the superior over Cochran and Hartley Fmax tests. Marked improvement in power occurs at larger sample sizes.

REFERENCES

[1] Bartlett, M. S. (1937b) Properties of sufficiency and statistical tests.*Proc. Roy. Statist. Soc. Ser. A* 160: 268-282.

[2] Bishop, T. A. (1976). Heteroscedastic ANOVA, MANOVA, and multiple comparisons. Unpublished doctoral dissertation, Ohio State University, Columbus.

[3] Box, G. E. P. (1953) Non-normality and tests on variances.*Biometrika*40: 318- 335.

[4] Brown, M. B., Forsythe. A. B. (1974a). The small sample behavior of some statistics which test the equality of several means. *Technometrics*, 16, 129- 132.

[5] *Brown, Morton B.; Forsythe, Alan B. (1974). "Robust tests for the equality of variances". Journal of the American Statistical Association.*69*: 364–367.*

[6] Cochran, W. G. (1941). The distribution of the largest of a set of estimated variances as a fraction of their total. Ann. Eug., 11, 47–52.

[7] Cochran, W. G. and Cox, G. M. (1957).Experimental design. New York: John Willey and Sons Inc.

[8] Diep, T. Nguyen, ThanhV. Pham, Patricia Rodriguez de Gil, Tyler Hicks, Yan Wang, Isaac Li, Aarti Bellara, Jeanine L. Romano, EunSook Kim, Harold Holmes, Yi-Hsin Chen, Jefferey D. Kromry (2014). ANOVA-HOV: A SAS MACRO for Testing Homogeneity of Variances in one-Factor ANOVA models, University of South Florida.

[9] Gartside, P.S. (1972), "A Study of Methods for Comparing Several Variances". J. Amer. Statist. Assoc., 67, 342 -346

[10] Hartley, H. O. (1950) The maximum F-ratio as a short-cut test for heterogeneity of variance. *Biometrika* 37:308-312.

[11] Levene, H. (1960). Robust tests for equality of variances. In I. Olkin (Ed.), Contributions to Probability and Statistics, pp. 278–292. Palo Alto, California: Stanford University Press.

[12] LiX, Qiu W, Morrow J DeMeo DL, Weiss ST, Fu Y.(2005). "A Comparative Study of Tests for Homogeneity of Variances with Application to DNA Methylation Data.PLOSONE10(12):eo145295,doi:0.1371/journal.pone.0145295.

[13] Mu, Zhiqiang. (2006)" Comparing the Statistical Tests for Homogeneity of Variances". Electronic Theses and Dissertations.Paper 2212. http://dc.etsu.edu/etd/2212.

[14] Neter, J, Kutner, M.H., Nachtsheim, C.J. and Wasserman, W. (1996). Applied linear regression model.3rd ed. Illinois: Irwin.

[15] Ostertagová, E. and Ostertag, O.(2013). "Methodology and Application of One Way ANOVA".American Journal of Mechanical Engineering, 1(7), 256- 261.

[16] 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.

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

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

[19] Rogan, J. C., and Keselman, H. J. (1977). Is the ANOVA F -test robust to variance heterogeneity when sample sizes are equal? An investigation via a coefficient of variation. American Educational Research Journal, 14, 493–498.

[20] Underwood, A. J. (1997). *Experiments in ecology – Their logical design and interpretation using analysis of variance.*Cambridge: Cambridge University Press.

[21] Wilcox, R. R., Charlin, V. L. and Thomson, K. L. (1986). New Monte Carlo results on the robustness of the ANOVA F, W and F ∗ statistics. J. Statist. Comput. Simul.,15, 33–43.

[22] Zar, J. H. (1999). *Biostatistical analysis* (Fourth Edition). Upper Saddle River, N.J.:Prentice Hall.

APPENDIX A

THE CODES FOR LOADING AND TRANSFORMATION OF THE REAL DATA INTO R SOFTWARE

```
data4test <- read_excel("C:/Users/user/Documents/STAT.TEST/Data4test.xlsx",col_names =TRUE,sheet = 2)
attach(data4test)## to load the data into R.
str(Data4test)##to check the structure of the data.
data4test$TREATMENT <-as.factor(data4test$TREATMENT)## converting the treatment to factors.
table(TREATMENT)##to show the different levels of the treatments.
```

R codes for graphical representation of the real data.

```
plot(RESPONSE~TREATMENT,data = Data4test,col = rainbow(3))##boxplots of         the treatments.
plotmeans(response ~ trt, xlab="Treatment", ylab="Response",
main="Mean Plot\nwith 95% CI")
qqPlot(fit, labels=row.names(TREATMENT), id.method="identify",
simulate=TRUE, main="Q-Q Plot")
 h <- hist(Data4test$RESPONSE,breaks=12,col="red",xlab="RESPONSE",main="Colored histogram with 12
bins")## Checking the distribution of the data using histogram plot.
d <- density(Data4test$RESPONSE)##visualising the distibution of the data using density and polygon.
plot(d,main = "kernel Density of RESPONSE")
polygon(d,col="red",border = "blue",main="kernel Density of RESPONSE")
```

APPENDIX B

CODES FOR HOMOGENEITY OF VARIANCE TESTS

Bartlett test of homogeneity:

```
bartlett.test(RESPONSE~TREATMENT,data=Data4test)## testing homogeneity of variances using bartlett test.
```

Levene test of homogeneity:

```
library(car) ## package required for the test
leveneTest(Data4test$RESPONSE,Data4test$TREATMENT) ## levene test for homogeneity of variance.
```

Cochran test of homogeneity:

```
library(GAD) ##package required for the test.
C.test(lm(data = Data4test,RESPONSE~TREATMENT))## cochran's test of homogeneity of variance.
```

Hartley's test of homogeneity

```
library(SuppDists) ## package required for the test.
qmaxFratio(p = 0.05,df = 9,k = 10,lower.tail = FALSE)## Hartley's fmax test using the package "SuppDist"
```

R codes for the different tests using simulated data.

```
set.seed(1000)
sims = 1000
pvals<-matrix(NA,sims)
## simulate data under the null hypothesis.
for(var in 1:1000){
  i=1
output<- matrix(ncol=12,nrow=10)
while(i <= 1000){
output[,i] <- rnorm(10,mean=0,sd=1)
  i=i+1
}
## Fit the models
data=(data.frame(output))
test<- stack(data)
## Bartlett test of homogeneity
f <- bartlett.test(values~ind,data=test)
f
```

```
##Levene test of homogeneity
library(car)
g <- leveneTest(values~ind,data=test)
g
## Hartley's Fmax test
library(SuppDists)
h <- qmaxFratio(values~ind,p = 0.05,df = 9,k = 10,lower.tail = FALSE)
h
## Cochran's test of homogeneity
library(GAD)
m <- C.test(lm(values~ind,data=test))
m
```

APPENDIX C

CODES FOR POWER SIMULATION OF HOMOGENEITY OF VARIANCE TESTS.

R code for power simulation of Bartlett test.

```
## simulation for equal and unequal sample sizes.
asim<-1000  #simulate with 1000 replications.
alpha<- 0.05
pv<-rep(NA,asim)
for(i in 1:asim)
{
set.seed(i)
 n1<-50
 n2<-50
 n3<-100
mu<-0
 sd1<-sqrt(25)
 sd2<-sqrt(50)
 sd3<-sqrt(100)
 g1<-rnorm(n1,mu,sd1)
 g2<-rnorm(n2,mu,sd2)
 g3<-rnorm(n3,mu,sd3)
 x=c(g1,g2,g3)
 group=c(rep(1,n1),rep(2,n2),rep(3,n3))
 N=200
 k=3
 v1=var(g1)
 v2=var(g2)
 v3=var(g3)
 #pooled variance
 A=((n1-1)*v1+(n2-1)*v2+(n3-1)*v3)/(N-k)
 #calculate B
 B=((N-k)*(log(A)))-((n1-1)*log(v1)+(n2-1)*log(v2)+(n3-1)*log(v3))
 #calculate C
 C=1+(1/(3*(k-1))*(((1/(n1-1))+(1/(n2-1))+(1/(n3-1)))-(1/(N-k))))
 #calculate layard estimator
 xbar1=mean(g1)
 xbar2=mean(g2)
 xbar3=mean(g3)
 sum1=sum((g1-xbar1)^4)
 sum2=sum((g2-xbar2)^4)
 sum3=sum((g3-xbar3)^4)
 sum4=sum((g1-xbar1)^2)
 sum5=sum((g2-xbar2)^2)
 sum6=sum((g3-xbar3)^2)
 y= (N*(sum1+sum2+sum3))/((sum4+sum5+sum6)^2)
```

```
  #calculate bartlett modified statistic
  bar2=B/(C*(1/2)*(y-1))
  bar2
pv[i]<-pchisq(bar2,2,lower=FALSE)
}
power<- mean(pv<0.05)
power
```

R code for power simulation of Levenetest

```
library(car)
## mean version of Levene
asim<- 1000 #simulate with 1000 replications.
pval<- rep(NA,asim)
alpha<- 0.05
power<- rep(NA,length(t)) #allocate a vector to store the calculated power in.
for(i in 1:asim)
{
set.seed(i)
t <-3 #number of group
n1=15#sample size for group1
n2=20 #sample size for group2
n3=30 #sample size for group3
N <- n1+n2+n3 #overall sample size,group1+group2+group3
#generated data
g1 <- rnorm(n1,0,1) #n1=15; mean=0;sd=1
g2 <- rnorm(n2,0,1) #n2=15;mean=0;sd=1
g3 <- rnorm(n3,0,1) #n3=15;mean=0;sd=1
xbar1 <- mean(g1)
xbar2 <- mean(g2)
xbar3 <- mean(g3)
z1 <- abs(g1-xbar1)
z2 <- abs(g2-xbar2)
z3 <- abs(g3-xbar3)
zbar1 <- mean(z1)
zbar2 <- mean(z2)
zbar3 <- mean(z3)
zbar<- (sum(z1)+sum(z2)+sum(z3))/(n1+n2+n3)
numerator1 <- n1*(zbar1-zbar)^2
numerator2 <- n2*(zbar2-zbar)^2
numerator3 <- n3*(zbar3-zbar)^2
numerator<- (numerator1+numerator2+numerator3)
denominator1 <- sum((z1-zbar1)^2)
denominator2 <- sum((z2-zbar2)^2)
denominator3 <- sum((z3-zbar3)^2)
denominator<-(denominator1+denominator2+denominator3)
Flevene<- numerator/denominator
Fvalue<- (1-pf(Flevene,2,(n1+n2+n3-3)))
pv[i] <- Fvalue
}
Power<- mean pv[i]
power
```

R code for power simulation of Cochran test

```
library(SuppDists)
asim<- 1000
pv<- rep(NA,asim)
alpha<- 0.5
```

```
for(i in 1:asim)
{
set.seed(i)
 t <- 3
#generated data
   g1 <- rnorm(n1,0,1)
   g2 <- rnorm(n2,0,1)
   g3 <- rnorm(n3,0,1)
 n1=15
 n2=15
 n3=15
 N <- n1+n2+n3
 sd1 <- 15
 sd2 <- 20
 sd3 <- 30
sd<- c(sd1,sd2,sd3)
numerator<- max(sd)
denominator<- sum(sd1+sd2+sd3)
}
Fcochran<- numerator/denominator
pv[i] <- Fcochran
power<- pv[i]
power
```

R code for power simulation of HartleyFmax test

```
library(SuppDists)
asim<- 1000
pv<- rep(NA,asim)
alpha<- 0.5
for(i in 1:asim)
{
set.seed(i)
t <- 3
#generated data
 g1 <- rnorm(n1,0,1)
 g2 <- rnorm(n2,0,1)
 g3 <- rnorm(n3,0,1)
n1=15
n2=15
n3=15
N <- n1+n2+n3
sd1 <- 15
sd2 <- 20
sd3 <- 30
sd<- c(sd1,sd2,sd3)
numerator<- min(sd)
denominator<- max(sd)
}
Fmax<- numerator/denominator
pv[i] <- Fmax
power<- pv[i]
power
```