ORİJİNAL ARAŞTIRMA ORIGINAL RESEARCH

DOI: 10.5336/biostatic.2017-56316

Comparison of Variance Homogenity Tests for Different Distributions

Farklı Dağılımlar İçin Varyans Homojenlik Testlerinin Karşılaştırılması

Kürşad Nuri BAYDILI,[a]
Deniz SIĞIRLI[b]

[a] Department of Biostatistics,
Uludağ University
Institute of Health Sciences,
[b]Department of Biostatistics,
Uludağ University Faculty of Medicine,
Bursa

Geliş Tarihi/*Received:* 27.04.2017
Kabul Tarihi/*Accepted:* 21.06.2017

Yazışma Adresi/*Correspondence:*
Kürşad Nuri BAYDILI
Uludağ University
Institute of Health Sciences,
Department of Biostatistics, Bursa,
TÜRKİYE/TURKEY
knuribaydili@hotmail.com

ABSTRACT **Objective:** The reliability of the hypothesis tests to be performed using variance analysis depends on the holding of the assumptions. In the analysis of variance, the sensitivity of the results obtained for the test statistic is most influenced by the heterogeneity of the variances. In this study, it is aimed to compare the variance homogeneity tests widely used in different literature for different distributions in terms of type-I error level and power. **Material and Methods:** 1000 repetitive simulations were performed for the Levene median, Levene mean, Levene trimmed mean, bootstrap Levene median, Bartlett and Cochran tests. Four cases where the sample sizes are equal were taken as $n_1 = n_2 = n_3 = 15$; $n_1 = n_2 = n_3 = 30$; $n_1 = n_2 = n_3 = 45$; $n_1 = n_2 = n_3 = 100$ and the two cases where the sample sizes are not equal were taken as $n_1 = 15, n_2 = 30, n_3 = 45$; $n_1 = 15, n_2 = 30, n_3 = 100$. For different distributions, the data were generated by taking into account the skewness and kurtosis coefficients. For power, the variance ratios of 1: 1: 2, 1: 2: 4, 1: 2: 2 and 1: 4: 4 were taken and different variances were corresponded to different sample sizes. **Results:** In terms of type-I error rates, the bootstrap Levene median test gave the best result for all distributions. For power; if the sample sizes were equal, for medium and large sample sizes bootstrap Levene median and Levene median tests gave the best results. For large sample sizes Bartlett test gave similar results to these two tests. **Conclusion:** It was seen that, while the performances of the tests were affected from the distribution, the best result against the deviations from normality was obtained for the bootstrap Levene median test and the performance of the tests was affected from the variance-sample size combination.

Keywords: Power (psychology); analysis of variance; simulation

ÖZET **Amaç:** Varyans analizi kullanılarak yapılacak olan hipotez testlerinin güvenilirliği, varsayımlarının sağlanmış olmasına bağlıdır. Varyans analizinde test istatistiği için elde edilen sonuçların duyarlılığı en çok varyansların heterojenliğinden etkilenmektedir. Bu çalışmada, farklı örneklem büyüklüklerinde, farklı dağılımlar için literatürde yaygın olarak kullanılan varyans homojenite testlerinin tip-I hata düzeyi ve güç bakımından karşılaştırılması amaçlanmıştır. **Gereç ve Yöntemler:** Levene medyan, Levene ortalama, Levene budanmış ortalama, bootstrap Levene medyan, Bartlett ve Cochran testleri için 1000 tekrarlı simülasyon çalışması gerçekleştirilmiştir. Örneklem büyüklüklerinin eşit olduğu dört durum $n_1 = n_2 = n_3 = 15$; $n_1 = n_2 = n_3 = 30$; $n_1 = n_2 = n_3 = 45$; $n_1 = n_2 = n_3 = 100$ ve örneklem büyüklüklerinin eşit olmadığı iki durum $n_1 = 15, n_2 = 30, n_3 = 45$; $n_1 = 15, n_2 = 30, n_3 = 100$ olarak alınmıştır. Farklı dağılımlar için çarpıklık ve basıklık katsayıları dikkate alınarak veriler türetilmiştir. Güç için, 1:1:2, 1:2:4, 1:2:2 ve 1:4:4 varyans oranları, farklı örneklem büyüklüklerine farklı varyans oranları denk gelecek şekilde alınmıştır. **Bulgular**: Tip-I hata oranları açısından, bootstrap Levene medyan testi tüm dağılımlar için en iyi sonucu vermiştir. Güç için; örneklem büyüklüklerinin eşit olması durumunda, orta ve yüksek örneklem büyüklüklerinde bootstrap Levene medyan ve Levene medyan testleri en iyi sonuçları vermiştir. Yüksek örneklem büyüklüklerinde Bartlett testi de bu iki teste benzer sonuçlar vermiştir. **Sonuç:** Test performanlarının dağılım yapısından etkilendiği görülmekle birlikte normallikten sapmalara karşı en iyi sonucun bootstrap Levene medyan testi için elde edildiği ve testlerin performansının varyans-örneklem büyüklüğü kombinasyonundan etkilendiği görülmüştür.

Anahtar Kelimeler: Güç (psikoloji); varyans (sapma) analizi; simülasyon

Thanks to the developments in technology and computer software, complex statistical methods have recently been developed for solving various problems encountered in practice, yet analysis of variance (ANOVA) is still one of the most widely used statistical methods in the literature.[1] One of the tests which can be used to compare more than two groups is the one-way analysis of variance test.[2]

The reliability of the hypothesis tests to be conducted by means of analysis of variance depends on whether the assumptions have been verified or not. These assumptions may be listed as normal distribution, homogeneity of variances, independence of observations and addibility of effects. When one or more of these assumptions is not verified, then both the significance level of the test and the sensitivity of the test statistics may be affected.[3,4] In cases where the data do not show normal distribution, then the data may be brought to a state of more normal distribution by means of the Box-Cox transformation methods.[5] In analysis of variance, the sensitivity of the results obtained for the test statistics is most affected by heterogeneity of the variances. If the assumption of variance homogeneity, is not met, then particularly for different sample sizes, interpretations to be made about means may be misleading and undesired results in type-I error rates may be observed.[6] Moreover, homogeneity of the variances is also tested before statistical analyses such as dose-response studies and linear-discriminant analyses are carried out.[7]

In the existing literature, various procedures have been proposed for testing homogeneity of variances, based on kurtosis adjustment, analysis of variance of mean and median absolute deviation, and resampling techniques such as bootstrapping.[8-13]

The aim of this study is to compare Levene's test, Bartlett's test, Cochran's test and the bootstrap version of Levene's test, which are mean- and median-based tests widely used in the literature, for different sample sizes, different distributions and changes in parameter values of distributions in themselves, with regard to type-I errors and powers.

MATERIAL AND METHODS

GENERAL DESCRIPTION

Where k is the number of samples, the hypotheses for homogeneity of variance tests are as follows:

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

H_1: At least one σ_i^2 value is different from the others $(i: 1, 2, \dots, \mathrm{k})$

HOMOGENEITY OF VARIANCE TESTS AND THEIR CALCULATIONS USED IN RESEARCH

Levene's Test

It is first proposed by Levene, by carrying out variance homogeneity test, to the values obtained by taking the absolute values of the differences of each measurement value from its own group mean.[14] During the following years, various modifications to the test were carried out in order to improve its robustness and power vis-a-vis normality. Brown Forsythe proposed the use of median instead of mean in the Levene method,[4] while Fligner and Killeen suggested using the trimmed mean method.[15]

Where

N: Total number of sample

k: Number of groups

n_i: Number of samples for the i-th group, the Levene test statistics for the measurement value Y_{ij} of the j-th sample in the i-th group are as follows:[16]

$$W=\frac{(N-k)(\sum_{i=1}^{k}n_i(\overline{Z_i}-\overline{Z})^2}{(k-1)\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Z_{ij}-\overline{Z_i})^2} \qquad (1)$$

The Z_{ij} value used in the test calculations is calculated in three different ways, namely:

1. $Z_{ij}=|Y_{ij}-\overline{Y_i}|$, $\overline{Y_i}$: mean of the i-th group (2)
2. $Z_{ij}=|Y_{ij}-\widetilde{Y_i}|$, $\widetilde{Y_i}$: median of the i-th group (3)
3. $Z_{ij}=|Y_{ij}-\overline{Y_i}'|$, $\overline{Y_i}'$: d% trimmed mean of the i-th group (4)

In Equation-1:

$\bar{Z}$: the general mean of the Z_{ij} values

$\overline{Z_i}$: the mean of the Z_{ij} values in the ith group

The test statistic value found is compared with the value in the F-table with the "k-1" and "N-k" degrees of freedom in α significance level. The trimmed mean value is calculated from the arithmetic mean remaining after the d-percent part of the lowest and highest values in the series is removed. Removing the end values ensures that better estimates are made, especially with populations that do not conform to normal distribution.[17]

Bartlett's Test

Bartlett's test allows you to compare the variance of two or more samples to determine whether they are drawn from populations with equal variance. It is suitable for normally distributed data. The test has the null hypothesis that the variances are equal and the alterntive hypothesis that they are not equal.[18] Bartlett's test is sensitive to departures from normality. The test statistic is calculated as follows:[19]

$$\chi_0^2=2.3026\frac{q}{c} \qquad (5)$$

In Equation 5:

$$q=(N-k)\log_{10}S_p^2-\sum_{i=1}^{k}(n_i-1)\log_{10}S_i^2 \qquad (6)$$

$$c=1+\frac{1}{3(k-1)}\left(\sum_{i=1}^{k}(n_i-1)^{-1}-(N-k)^{-1}\right) \qquad (7)$$

$$S_p^2=\frac{\sum_{i=1}^{k}(n_i-1)S_i^2}{N-k} \qquad (8)$$

In Equation 8:

2.3026 : Fixed value

n_i: Sample size of i-th group

S_i^2 : Variance of i-th group

N: Total number of sample $\left(N=\sum_{i=1}^{k}n_i\right)$

k: Number of groups

S_p^2: Common variance

The test statistic value found is compared with the χ^2 table with "k-1" degrees of freedom.

Cochran's C Test

Another test used to test homogeneity of variance is Cochran's C test. This test whether the largest variance in the group is larger than the others. The test statistic is calculated as follows:[20]

$$C_i = \frac{the\ largest\ S_i^2}{\sum_{i=1}^{k} S_i^2} \tag{9}$$

S_i^2: i-th group variance (i=1,…,k)

k : Number of groups

Critical calculations of value for Cochran's test are made as follows:

$$C(\alpha, n, N)_i = \left[1 + \frac{N-1}{F_C(\frac{\alpha}{N}, (n-1), (N-1)(n-1))}\right]^{-1} \text{ (i: 1,2,...k)} \tag{10}$$

Where:

α: Significance level

n: Number of sample having the largest variance

F_c: F table value

and $C(\alpha, n, N)$ value is calculated separately for each group and it is examined whether the groups are different from the group having the largest variance.

The Bootstrap Version of Levene's Median Test

The basis of the bootstrap method relies on the resampling of the original data. By this method, the data are retrieved by sampling with the desired amount and the operations to be applied are reapplied for each new sample. With the bootstrap method, standard error estimates can be made in a more robust manner and confidence intervals of population parameters can be estimated.[21]

The steps for the bootstrap version of Levene's median test are as follows:

i. The Levene test statistic is calculated for the observed data set.

ii. It is taken that R=0.

iii. The $Z_{ij} = |Y_{ij} - \widetilde{Y}_i|$ values are calculated ($\widetilde{Y}_i$:median of the i-th group (j=1,…, n_i; i=1,…,k).

iv. B bootstrap samples are created by drawing $e_{ij}*$ values from the sample of e_{ij} values with replacement.

v. B number of new Levene test statistics (T^*) are calculated by using the $Z_{ij}{}^*$ values for each of the B bootstrap samples. If $T^* > T$, R is increased by one.

vi. Steps i-v are repeated A times (A is the repeat number in the simulation).

vii. The bootstrap p value is calculated as R/A. If the R/A value is below the significance level α, the H_0 hypothesis is rejected.

SIMULATION STUDY

The simulation study was carried for three groups, and 1000 repetitions were made. The simulations were performed by using the "car 2.1-4", outliers 0.14" and "lawstat 3.1" packages on the R-3.2.5 software.[22-24]

For determining type-I error rates in the case of constant variance, data were created for different distributions and different sample sizes. The four situations in the groups where the sample sizes were equal were taken as $n_1 = n_2 = n_3 = 15$; $n_1 = n_2 = n_3 = 30$; $n_1 = n_2 = n_3 = 45$; $n_1 = n_2 = n_3 = 100$, and the two situations in the groups where the sample sizes were not equal were taken as $n_1 = 15, n_2 = 30, n_3 = 45$; $n_1 = 15, n_2 = 30, n_3 = 100$. When determining the distributions, skewness and kurtosis coefficients were taken into consideration. For symmetric distributions, standard normal distribution which has a skewness and kurtosis coefficients equal to 0; uniform distribution $(U \sim a = 0, b = 1)$ which is symmetric but platykurtic; t-distribution with 6 degree of freedom which is symmetric but leptokurtic were considered. For asymmetric distributions; gamma distribution with shape parameter 4 (k=4) and scale parameter 1 (b=1) and chi-squared distribution 1 degree of freedom, which has larger skewness and kurtosis coefficient than the gamma, were considered. α was taken as 0.05. The number of repetition was taken 1000. The type-I error is calculated as in equation 11.[11,25]

$$Type\ I\ Error = \frac{the\ number\ of\ \ H_0\ is\ rejection\ when\ H_0\ is\ true}{the\ number\ of\ repetition} \tag{11}$$

The simulation scenario for the calculation of type-I error rates is given in Table 1.

TABLE 1. Simulation scenario for calculation of Type-I error rates.

Distributions	Sample Size (n_i)	Variance Rates $\sigma_1^2:\sigma_2^2:\sigma_3^2$	Tests	Number of Repetition	Eventuation
Normal Distribution	15, 15, 15	1:1:1	Levene's Median Test	1000	T: Number of Repetition
Gamma Distribution	30, 30, 30		Levene's Mean Test		
T Distribution	45, 45, 45		Levene' Trimmed Mean Test		R: The number of test that p<0.05 (R/T)
Uniform Distribution	100, 100,100		Bartlett Test		
Chi-Square Distribution	15, 30, 45		Cochran Test		
Beta Distribution	45, 30, 15		Bootstrap Levene's Median Test		
	15, 30,100				
	100, 30,15				

For power calculations in the case of non-constant variance, data were created for different distributions and different sample sizes. The four situations in the groups where the sample numbers were equal were taken as $n_1 = n_2 = n_3 = 15$; $n_1 = n_2 = n_3 = 30$; $n_1 = n_2 = n_3 = 45$; $n_1 = n_2 = n_3 = 100$, and the two situations in the groups where the sample numbers were not equal were taken as $n_1 = 15, n_2 = 30, n_3 = 45$; $n_1 = 15, n_2 = 30, n_3 = 100$. Three same distributions were used as in the type-I error scenario. However different parameter values were determined to change the variance value. Variance rates were taken as 1:1:2, 1:2:4, 1:2:2 and 1:4:4. Different sample sizes were corresponded to different variance rates (small sample size-small variance rate and large sample size-small variance rate). The power calculated as in equation 12.[25,26]

$$power = \frac{the\ number\ of\ H_0 failed\ to\ reject\ when\ H_0\ is\ true}{the\ number\ of\ repetition} \tag{12}$$

The Simulation Scenario for the power calculations is given in Table 2.

TABLE 2. Simulation scenario for power calculations.

Distributions	Sample Size (n_i)	Variance Rates $\sigma_1^2:\sigma_2^2:\sigma_3^2$	Tests	Number of Repetition	Eventuation
Normal Distribution	15, 15, 15	1:1:2	Levene's Median Test	1000	T: Number of Test
Gamma Distribution	30, 30, 30	1:2:4	Levene's Mean Test		R: The number of result of test that p<0.05
T Distribution	45, 45, 45	1:2:2	Levene' Trimmed Mean Test		
Uniform Distribution	100,100,100	1:4:4	Bartlett Test		
Chi-Square Distribution	15, 30, 45 15,30,100	1:1:2	Cochran Test		R/T
Beta Distribution		1:2:1	Bootstrap Levene's Median Test		
		2:1:1			
		1:2:4			
		2:1:4			
		1:4:2			
		2:4:1			
		4:1:2			
		4:2:1			
		2:2:1			
		2:1:2			
		1:2:2			
		4:4:1			
		4:1:4			
		1:4:4			

RESULTS

EXAMINATION OF TYPE-I ERROR RATES IN THE TESTS

The type-I error rates obtained for data generated from different distributions are given in Table 3.

Table 3. Calculation of Type-I Error Rates.

Sample Size (n_i)	Variance Rates (σ^2)	Levene's Median Test	Levene's Mean Test	Levene Trimmed Mean Test	Bartlett Test	Cochran Test	Bootstrap Levene's Median Test
Normal Distribution ($\mu = 0, \sigma^2 = 1$)							
15, 15, 15	1:1:1	0.028	0.055	0.020	0.057	0.062	0.057
30, 30, 30	1:1:1	0.023	0.039	0.022	0.049	0.042	0.059
45, 45, 45	1:1:1	0.031	0.044	0.028	0.042	0.036	0.061
100,100,100	1:1:1	0.043	0.046	0.049	0.062	0.063	0.050
15, 30, 45	1:1:1	0.027	0.035	0.027	0.044	0.078	0.043
15,30,100	1:1:1	0.051	0.056	0.047	0.044	0.153	0.050
Uniform Distribution (a=0, b=1)							
Sample Size (n_i)	Variance Rates (σ^2)	Levene's Median Test	Levene's Mean Test	Levene Trimmed Mean Test	Bartlett Test	Cochran Test	Bootstrap Levene's Median Test
15, 15, 15	1:1:1	0.012	0.049	0.013	0.000	0.000	0.041
30, 30, 30	1:1:1	0.030	0.054	0.030	0.002	0.001	0.044
45, 45, 45	1:1:1	0.031	0.049	0.030	0.000	0.002	0.051
100,100,100	1:1:1	0.045	0.044	0.036	0.000	0.002	0.044
15, 30, 45	1:1:1	0.028	0.035	0.045	0.000	0.004	0.048
15,30,100	1:1:1	0.046	0.038	0.040	0.002	0.017	0.056
T Distribution ($\nu = 6$)							
Sample Size (n_i)	Variance Rates (σ^2)	Levene's Median Test	Levene's Mean Test	Levene Trimmed Mean Test	Bartlett Test	Cochran Test	Bootstrap Levene's Median Test
15, 15, 15	1:1:1	0.035	0.067	0.034	0.176	0.165	0.034
30, 30, 30	1:1:1	0.031	0.047	0.032	0.178	0.163	0.040
45, 45, 45	1:1:1	0.038	0.048	0.038	0.184	0.180	0.061
100,100,100	1:1:1	0.058	0.065	0.051	0.239	0.213	0.056
45, 30, 15	1:1:1	0.042	0.058	0.032	0.169	0.222	0.048
15,30,100	1:1:1	0.044	0.047	0.043	0.184	0.306	0.050
Gamma Distribution (k=4, b=1)							
Sample Size (n_i)	Variance Rates (σ^2)	Levene's Median Test	Levene's Mean Test	Levene Trimmed Mean Test	Bartlett Test	Cochran Test	Bootstrap Levene's Median Test
15, 15, 15	1:1:1	0.035	0.210	0.068	0.358	0.315	0.075
30, 30, 30	1:1:1	0.033	0.205	0.096	0.430	0.366	0.074
45, 45, 45	1:1:1	0.050	0.181	0.106	0.393	0.343	0.046
100,100,100	1:1:1	0.054	0.183	0.141	0.445	0.402	0.058
15, 30, 45	1:1:1	0.053	0.174	0.091	0.384	0.398	0.051
15,30,100	1:1:1	0.040	0.171	0.089	0.394	0.521	0.045
Chi-Square Distribution ($\nu = 1$)							
Sample Size (n_i)	Variance Rates (σ^2)	Levene's Median Test	Levene's Mean Test	Levene Trimmed Mean Test	Bartlett Test	Cochran Test	Bootstrap Levene's Median Test
15, 15, 15	1:1:1	0.035	0.287	0.100	0.527	0.445	0.034
30, 30, 30	1:1:1	0.031	0.252	0.192	0.563	0.493	0.039
45, 45, 45	1:1:1	0.053	0.290	0.239	0.596	0.535	0.053
100,100,100	1:1:1	0.052	0.263	0.290	0.634	0.568	0.054
15, 30, 45	1:1:1	0.040	0.240	0.177	0.536	0.526	0.059
15,30,100	1:1:1	0.052	0.244	0.193	0.589	0.652	0.060

α has been set to 0.05.

EXAMINATION OF POWER OF THE TESTS

The power obtained for data generated from different distributions are given in Table 4-8.

TABLE 4. Calculation of power for normal distribution.

Normal Distribution							
Sample Size (n_i)	**Variance Rates (σ^2)**	**Levene's Median Test**	**Levene's Mean Test**	**Levene Trimmed Mean Test**	**Bartlett Test**	**Cochran Test**	**Bootstrap Levene's Median Test**
15:15:15	**1:1:2**	0.602	0.699	0.510	0.760	0.613	0.634
	1:2:4	0.947	0.976	0.913	0.999	0.960	0.940
	1:2:2	0.425	0.562	0.403	0.697	0.301	0.491
	1:4:4	0.973	0.988	0.961	0.999	0.559	0.945
30:30:30	**1:1:2**	0.922	0.935	0.891	0.972	0.982	0.938
	1:2:4	1.000	1.000	1.000	1.000	1.000	1.000
	1:2:2	0.918	0.935	0.898	0.973	0.557	0.907
	1:4:4	1.000	1.000	1.000	1.000	1.000	1.000
45:45:45	**1:1:2**	0.992	0.994	0.985	0.998	0.998	0.993
	1:2:4	1.000	1.000	1.000	1.000	1.000	1.000
	1:2:2	0.987	0.991	0.982	0.996	0.789	0.990
	1:4:4	1.000	1.000	1.000	1.000	1.000	1.000
100:100:100	**1:1:2**	1.000	1.000	1.000	1.000	1.000	1.000
	1:2:4	1.000	1.000	1.000	1.000	1.000	1.000
	1:2:2	1.000	1.000	1.000	1.000	1.000	1.000
	1:4:4	1.000	1.000	1.000	1.000	1.000	1.000
15:30:45	**1:1:2**	0.955	0.969	0.927	0.986	0.990	0.948
	1:2:1	0.946	0.963	0.905	0.981	0.983	0.934
	2:1:1	0.780	0.831	0.687	0.868	0.949	0.812
	1:2:4	1.000	1.000	1.000	1.000	1.000	1.000
	2:1:4	1.000	1.000	1.000	1.000	1.000	1.000
	1:4:2	0.999	0.999	0.998	1.000	0.999	0.993
	2:4:1	0.900	1.000	1.000	1.000	1.000	1.000
	4:1:2	0.998	0.999	0.994	1.000	0.991	0.999
	4:2:1	0.997	0.999	0.997	1.000	0.997	0.999
	2:2:1	0.628	0.683	0.649	0.799	0.536	0.630
	2:1:2	0.829	0.841	0.799	0.868	0.627	0.809
	1:2:2	0.928	0.983	0.949	0.799	0.936	0.930
	4:4:1	0.946	1.000	1.000	1.000	0.944	1.000
	4:1:4	1.000	1.000	1.000	1.000	0.940	1.000
	1:4:4	1.000	1.000	0.999	1.000	0.922	0.996
15:30:100	**1:1:2**	0.991	0.993	0.992	1.000	1.000	0.989
	1:2:1	0.976	0.982	0.962	0.987	0.995	0.981
	2:1:1	0.831	0.869	0.744	0.872	0.970	0.862
	1:2:4	1.000	1.000	1.000	1.000	1.000	1.000
	2:1:4	1.000	1.000	1.000	1.000	1.000	1.000
	1:4:2	1.000	1.000	0.998	1.000	1.000	1.000
	2:4:1	0,998	1.000	1.000	1.000	1.000	1.000
	4:1:2	0.998	0.998	0.997	1.000	0.998	0.998
	4:2:1	0,999	0,998	0,997	0,997	0,998	0,999
	2:2:1	0.694	0.609	0.702	0.641	0.742	0.676
	2:1:2	0.965	0.963	0.950	0.990	0.880	0.951
	1:2:2	0.967	0.969	0.959	0.992	0.885	0.959
	4:4:1	0,998	0,989	0,991	0,997	0,996	0,999
	4:1:4	1.000	1.000	1.000	1.000	1.000	1.000
	1:4:4	1.000	1.000	1.000	1.000	1.000	0.999

The power of test is the probability of rejecting null hypothesis when it is false. The maximum total power of the test is 1; the minimum is 0.

TABLE 5. Calculation of power for uniform distribution.

Uniform Distribution							
Sample Size (n_i)	**Variance Rates (σ^2)**	**Levene's Median Test**	**Levene's Mean Test**	**Levene Trimmed Mean Test**	**Bartlett Test**	**Cochran Test**	**Bootstrap Levene's Median Test**
15:15:15	**1:1:2**	0.186	0.329	0.190	0.109	0.205	0.277
	1:2:4	0.541	0.704	0.515	0.589	0.550	0.644
	1:2:2	0.140	0.286	0.140	0.081	0.035	0.224
	1:4:4	0.140	0.286	0.140	0.081	0.035	0.738
30:30:30	**1:1:2**	0.546	0.636	0.614	0.412	0.583	0.606
	1:2:4	0.947	0.965	0.947	0.974	0.610	0.956
	1:2:2	0.488	0.592	0.548	0.316	0.077	0.565
	1:4:4	0.988	0.892	0.848	0.916	0.577	0.987
45:45:45	**1:1:2**	0.765	0.814	0.833	0.705	0.826	0.792
	1:2:4	0.997	0.999	0.997	1.000	0.993	0.996
	1:2:2	0.741	0.804	0.811	0.622	0.142	0.747
	1:4:4	0.741	0.804	0.811	0.822	0.942	1.000
100:100:100	**1:1:2**	0.988	0.990	1.000	0.991	0.682	0.989
	1:2:4	1.000	1.000	1.000	1.000	1.000	1.000
	1:2:2	0.991	0.992	0.997	0.994	0.511	0.990
	1:4:4	1.000	1.000	1.000	1.000	1.000	1.000
15:30:45	**1:1:2**	0.314	0.561	0.645	0.765	0.720	0.650
	1:2:1	0.302	0.502	0.561	0.673	0.636	0.603
	2:1:1	0.215	0.439	0.372	0.643	0.602	0.434
	1:2:4	0.830	0.736	0.824	0.838	0.795	0.925
	2:1:4	0.815	0.877	0.929	0.907	0.825	0.986
	1:4:2	0.576	0.723	0.767	0.830	0.794	0.873
	2:4:1	0.811	0.873	0.919	0.902	0.804	0.982
	4:1:2	0.571	0.719	0.753	0.784	0.738	0.876
	4:2:1	0.660	0.753	0.812	0.790	0.739	0.921
	2:2:1	0.339	0.513	0.548	0.622	0.559	0.623
	2:1:2	0.295	0.480	0.542	0.646	0.574	0.553
	1:2:2	0.145	0.353	0.364	0.592	0.559	0.359
	4:4:1	0.562	0.675	0.757	0.716	0.614	0.892
	4:1:4	0.824	0.881	0.941	0.890	0.731	0.991
	1:4:4	0.881	0.999	0.961	0.899	0.703	0.992
15:30:100	**1:1:2**	0.629	0.633	0.777	0.617	0.759	0.709
	1:2:1	0.657	0.729	0.745	0.564	0.771	0.702
	2:1:1	0.391	0.493	0.464	0.290	0.746	0.473
	1:2:4	0.994	0.995	0.996	0.983	1.000	0.988
	2:1:4	1.000	1.000	1.000	1.000	0.997	0.999
	1:4:2	0.899	0.915	0.923	0.874	0.981	0.906
	2:4:1	1.000	1.000	0.999	1.000	0.975	0.999
	4:1:2	0.880	0.912	0.916	0.847	0.927	0.890
	4:2:1	0.976	0.984	0.967	0.989	0.939	0.984
	2:2:1	0.731	0.800	0.805	0.712	0.371	0.792
	2:1:2	0.631	0.680	0.703	0.370	0.296	0.649
	1:2:2	0.395	0.426	0.365	0.109	0.215	0.387
	4:4:1	0.948	0.968	0.958	0.896	0.736	0.899
	4:1:4	0.998	0.999	0.999	1.000	0.833	0.997
	1:4:4	0.999	0.959	0.936	1.000	0.873	0.999

The power of test is the probability of rejecting null hypothesis when it is false. The maximum total power of the test is 1; the minimum is 0.

TABLE 6. Calculation of power for T distribution.

T Distribution							
Sample Size (n_i)	**Variance Rates (σ^2)**	**Levene's Median Test**	**Levene's Mean Test**	**Levene Trimmed Mean Test**	**Bartlett Test**	**Cochran Test**	**Bootstrap Levene's Median Test**
15:15:15	**1:1:2**	0.484	0.551	0.382	0.315	0.657	0.579
	1:2:4	0.540	0.597	0.543	0.413	0.404	0.636
	1:2:2	0.545	0.687	0.641	0.360	0.323	0.634
	1:4:4	0.653	0.612	0.649	0.473	0.434	0.647
30:30:30	**1:1:2**	0.721	0.474	0.452	0.596	0.830	0.732
	1:2:4	0.764	0.517	0.557	0.606	0.529	0.775
	1:2:2	0.757	0.587	0.564	0.495	0.440	0.776
	1:4:4	0.766	0.608	0.676	0.631	0.538	0.780
45:45:45	**1:1:2**	0.849	0.960	0.471	0.682	0.949	0.778
	1:2:4	0.804	0.456	0.503	0.697	0.619	0.835
	1:2:2	0.892	0.517	0.581	0.596	0.524	0.881
	1:4:4	0.894	0.658	0.615	0.583	0.698	0.911
100:100:100	**1:1:2**	1.000	1.000	0.782	0.948	0.704	1.000
	1:2:4	0.988	0.709	0.706	0.905	0.738	0.973
	1:2:2	0.990	0.715	0.788	0.803	0.746	0.980
	1:4:4	1.000	0.830	0.828	0.949	0.895	1.000
15:30:45	**1:1:2**	0.573	0.485	0.419	0.575	0.452	0.459
	1:2:1	0.524	0.432	0.392	0.549	0.438	0.483
	2:1:1	0.584	0.285	0.267	0.453	0.406	0.415
	1:2:4	0.790	0.521	0.398	0.604	0.621	0.743
	2:1:4	0.800	0.638	0.595	0.704	0.647	0.880
	1:4:2	0.730	0.486	0.344	0.555	0.606	0.761
	2:4:1	0.812	0.609	0.630	0.738	0.695	0.634
	4:1:2	0.811	0.535	0.493	0.634	0.643	0.694
	4:2:1	0.861	0.593	0.575	0.700	0.672	0.454
	2:2:1	0.877	0.492	0.464	0.637	0.619	0.513
	2:1:2	0.717	0.484	0.435	0.609	0.579	0.569
	1:2:2	0.734	0.693	0.243	0.499	0.527	0.538
	4:4:1	0.876	0.878	0.778	0.782	0.749	0.858
	4:1:4	0.956	0.867	0.730	0.860	0.732	0.916
	1:4:4	0.997	0.832	0.840	0.825	0.947	0.927
15:30:100	**1:1:2**	0.669	0.361	0.486	0.532	0.635	0.554
	1:2:1	0.611	0.348	0.472	0.401	0.515	0.532
	2:1:1	0.609	0.352	0.465	0.342	0.494	0.535
	1:2:4	0.830	0.438	0.463	0.684	0.748	0.827
	2:1:4	0.865	0.477	0.510	0.740	0.779	0.978
	1:4:2	0.858	0.490	0.469	0.562	0.665	0.669
	2:4:1	0.849	0.337	0.526	0.656	0.688	0.676
	4:1:2	0.899	0.535	0.474	0.602	0.664	0.608
	4:2:1	0.814	0.387	0.502	0.568	0.612	0.651
	2:2:1	0.875	0.430	0.492	0.506	0.580	0.611
	2:1:2	0.755	0.390	0.459	0.544	0.610	0.778
	1:2:2	0.734	0.545	0.455	0.487	0.601	0.747
	4:4:1	0.801	0.793	0.728	0.719	0.743	0.724
	4:1:4	0.977	0.709	0.783	0.744	0.768	0.902
	1:4:4	1.000	0.853	0.865	0.871	0.755	0.987

The power of test is the probability of rejecting null hypothesis when it is false. The maximum total power of the test is 1; the minimum is 0.

TABLE 7. Calculation of power for gamma distribution.

Gamma Distribution							
Sample Size (n_i)	**Variance Rates (σ^2)**	**Levene's Median Test**	**Levene's Mean Test**	**Levene Trimmed Mean Test**	**Bartlett Test**	**Cochran Test**	**Bootstrap Levene's Median Test**
15:15:15	**1:1:2**	0.190	0.369	0.215	0.507	0.455	0.197
	1:2:4	0.378	0.550	0.409	0.668	0.541	0.442
	1:2:2	0.258	0.443	0.299	0.569	0.442	0.249
	1:4:4	0.444	0.618	0.493	0.348	0.444	0.441
30:30:30	**1:1:2**	0.329	0.495	0.446	0.603	0.553	0.788
	1:2:4	0.780	0.833	0.844	0.871	0.788	0.813
	1:2:2	0.572	0.682	0.680	0.750	0.598	0.794
	1:4:4	0.867	0.892	0.920	0.914	0.648	0.876
45:45:45	**1:1:2**	0.469	0.612	0.614	0.688	0.679	0.405
	1:2:4	0.916	0.930	0.961	0.949	0.903	0.849
	1:2:2	0.465	0.562	0.612	0.735	0.513	0.679
	1:4:4	0.975	0.977	0.991	0.978	0.822	0.926
100:100:100	**1:1:2**	0.844	0.875	0.948	0.917	0.893	0.722
	1:2:4	1.000	1.000	1.000	0.999	0.995	0.997
	1:2:2	0.989	0.986	0.997	0.984	0.948	0.965
	1:4:4	1.000	1.000	1.000	1.000	0.989	1.000
15:30:45	**1:1:2**	0.363	0.497	0.496	0.639	0.615	0.321
	1:2:1	0.373	0.527	0.474	0.649	0.617	0.284
	2:1:1	0.218	0.377	0.278	0.506	0.537	0.220
	1:2:4	0.702	0.568	0.575	0.622	0.610	0.604
	2:1:4	0.838	0.665	0.686	0.600	0.601	0.580
	1:4:2	0.615	0.600	0.605	0.697	0.588	0.572
	2:4:1	0.855	0.684	0.687	0.613	0.607	0.683
	4:1:2	0.582	0.600	0.639	0.662	0.517	0.562
	4:2:1	0.693	0.672	0.633	0.621	0.524	0.633
	2:2:1	0.332	0.474	0.422	0.563	0.464	0.350
	2:1:2	0.326	0.467	0.449	0.587	0.497	0.287
	1:2:2	0.162	0.320	0.257	0.497	0.449	0.151
	4:4:1	0.619	0.636	0.633	0.629	0.673	0.677
	4:1:4	0.861	0.892	0.712	0.710	0.677	0.805
	1:4:4	0.894	0.871	0.815	0.879	0.643	0.831
15:30:100	**1:1:2**	0.476	0.582	0.678	0.727	0.776	0.401
	1:2:1	0.402	0.531	0.541	0.632	0.704	0.359
	2:1:1	0.229	0.419	0.298	0.548	0.650	0.241
	1:2:4	0.832	0.857	0.905	0.903	0.936	0.745
	2:1:4	0.945	0.940	0.970	0.947	0.927	0.869
	1:4:2	0.611	0.703	0.721	0.790	0.867	0.567
	2:4:1	0.948	0.963	0.963	0.954	0.916	0.912
	4:1:2	0.621	0.714	0.735	0.788	0.824	0.568
	4:2:1	0.823	0.881	0.848	0.900	0.843	0.776
	2:2:1	0.471	0.605	0.575	0.654	0.623	0.434
	2:1:2	0.354	0.488	0.527	0.632	0.678	0.310
	1:2:2	0.194	0.339	0.357	0.515	0.610	0.188
	4:4:1	0.978	0.984	0.990	0.977	0.896	0.959
	4:1:4	0.917	0.926	0.959	0.952	0.884	0.859
	1:4:4	0.642	0.700	0.769	0.814	0.836	0.574

The power of test is the probability of rejecting null hypothesis when it is false. The maximum total power of the test is 1; the minimum is 0.

TABLE 8. Calculation of power for chi-square distribution.

Chi-Square Distribution							
Sample Size (n_i)	**Variance Rates (σ^2)**	**Levene's Median Test**	**Levene's Mean Test**	**Levene Trimmed Mean Test**	**Bartlett Test**	**Cochran Test**	**Bootstrap Levene's Median Test**
15:15:15	**1:1:2**	0.163	0.408	0.273	0.590	0.510	0.169
	1:2:4	0.433	0.597	0.531	0.699	0.572	0.358
	1:2:2	0.153	0.365	0.252	0.543	0.427	0.128
	1:4:4	0.446	0.644	0.560	0.737	0.495	0.387
30:30:30	**1:1:2**	0.832	0.545	0.579	0.718	0.628	0.847
	1:2:4	0.822	0.810	0.865	0.663	0.798	0.912
	1:2:2	0.894	0.468	0.523	0.647	0.521	0.812
	1:4:4	0.940	0.877	0.940	0.902	0.720	0.997
45:45:45	**1:1:2**	0.843	0.604	0.768	0.762	0.685	0.835
	1:2:4	0.913	0.932	0.985	0.940	0.883	0.790
	1:2:2	0.809	0.556	0.720	0.687	0.597	0.893
	1:4:4	0.957	0.956	0.995	0.937	0.817	0.857
100:100:100	**1:1:2**	0.844	0.875	0.948	0.917	0.893	0.822
	1:2:4	1.000	1.000	1.000	0.999	0.995	0.997
	1:2:2	0.989	0.986	0.997	0.984	0.948	0.965
	1:4:4	1.000	1.000	1.000	1.000	0.989	1.000
15:30:45	**1:1:2**	0.314	0.561	0.645	0.765	0.720	0.272
	1:2:1	0.302	0.502	0.561	0.673	0.636	0.255
	2:1:1	0.215	0.439	0.372	0.643	0.602	0.598
	1:2:4	0.730	0.636	0.624	0.638	0.595	0.532
	2:1:4	0.715	0.777	0.729	0.707	0.725	0.797
	1:4:2	0.576	0.523	0.667	0.630	0.594	0.795
	2:4:1	0.611	0.673	0.619	0.602	0.604	0.610
	4:1:2	0.571	0.519	0.653	0.584	0.638	0.498
	4:2:1	0.660	0.653	0.612	0.590	0.639	0.576
	2:2:1	0.339	0.513	0.548	0.622	0.559	0.269
	2:1:2	0.295	0.480	0.542	0.646	0.574	0.245
	1:2:2	0.145	0.353	0.364	0.592	0.559	0.135
	4:4:1	0.881	0.904	0.957	0.916	0.714	0.888
	4:1:4	0.824	0.881	0.941	0.890	0.731	0.931
	1:4:4	0.562	0.675	0.761	0.799	0.703	0.450
15:30:100	**1:1:2**	0.420	0.616	0.786	0.800	0.829	0.328
	1:2:1	0.382	0.572	0.703	0.712	0.760	0.310
	2:1:1	0.222	0.406	0.406	0.611	0.664	0.196
	1:2:4	0.775	0.831	0.927	0.896	0.916	0.646
	2:1:4	0.932	0.939	0.988	0.953	0.953	0.822
	1:4:2	0.620	0.724	0.809	0.823	0.867	0.533
	2:4:1	0.932	0.948	0.986	0.936	0.903	0.844
	4:1:2	0.874	0.704	0.793	0.772	0.822	0.492
	4:2:1	0.782	0.833	0.911	0.842	0.833	0.675
	2:2:1	0.467	0.624	0.684	0.677	0.683	0.665
	2:1:2	0.504	0.489	0.639	0.659	0.696	0.649
	1:2:2	0.645	0.341	0.430	0.608	0.689	0.631
	4:4:1	0.972	0.977	0.994	0.965	0.880	0.920
	4:1:4	0.890	0.899	0.960	0.910	0.878	0.768
	1:4:4	0.899	0.702	0.816	0.816	0.852	0.998

The power of test is the probability of rejecting null hypothesis when it is false. The maximum total power of the test is 1; the minimum is 0.

DISCUSSION

When type-I error rate was examined, it was observed that for normal distribution, Bartlett's test, the bootstrap Levene's median test and Levene's mean test gave the best results for all sample sizes. It was observed that Levene's median test and Levene's trimmed mean tests gave results that were considerably below nominal level for small and medium sample sizes. Cochran's test, however, gave results that were considerably different from nominal level for all sample sizes.

The best results for uniform distribution, which was symmetric and flatter than normal, were obtained from the bootstrap Levene's median test. However, it was seen that the Levene's mean test gave results close to nominal level except in cases where there were different sample sizes. It was observed that the poorest results were obtained from Bartlett's and Cochran's tests and that the type-I error levels of these tests were considerably below nominal level. However, it was seen that Levene's median test gave a Type-I error rate close to nominal level for large sample sizes when sample sizes were equal and not equal, while for small and medium sample sizes, it gave more conservative values.

For t-distribution which is symmetric but leptokurtic, it was seen that except for small samples, the best result was obtained with the bootstrap Levene's median test. It was observed that while results close to nominal level were obtained with Levene's median test, Levene's mean test and Levene's trimmed mean test for medium and large sample sizes; poorer results were obtained in cases where sample sizes were different.

For asymmetric distributions, gamma distribution which shows lower deviation from normality, the type-I error rates in Levene's median test were quite close to nominal level and that this test gave the best result. Besides, it was observed that the bootstrap Levene's median test also gave results close to nominal level for medium and large sample sizes. It was seen that the poorest results were obtained for Bartlett's and Cochran's tests independently of sample sizes. For gamma distribution the simulation results showed no variation according to whether the sample sizes were equal or not.

For chi-squared, another asymmetric distribution which shows greater deviation from normality, the best result was obtained for Levene's median test for type-I error rate. It was seen that for medium and large sample sizes Levene's median test gave results quite close to nominal level. On the other hand, it was seen that the bootstrap Levene's median test gave results close to nominal level for medium and large samples where sample sizes were equal, whereas it gave poorer results for small samples where sample sizes were different. While the other tests gave results that were considerably different from nominal level, it was seen that the worst results were obtained for Bartlett's and Cochran's tests.

When the power of the tests is considered, it was found that for normal distribution, apart from small sample sizes, for all other situations the power values of the other tests was high except for Cochran's test. When sample sizes related to the groups were equal, it was observed that there was an increase in power values as sample size increased for all tests. When sample sizes related to the groups were different, for small and medium sample sizes, Cochran's test gave quite low power values. For small and medium sample sizes, in cases where the difference between variance rates were small, a decrease in power values of the tests occurred. Particularly for different sample sizes, when low variance rate corresponded to small samples and high variance rate corresponded to medium and large samples, low power values were obtained for all tests.

For uniform distribution, quite low power values were obtained for small samples for all tests. Particularly for Cochran's test, it was seen that the worst results were obtained. For medium sample sizes, in cases where the difference between variance rates were large, high power values were obtained for all tests other than Cochran's test. For cases where sample sizes were equal, however, it was observed that for large sample sizes, when the difference between variance rates was small, Cochran's test gave lower power values, while in all cases other than this, higher power values were obtained for all tests. However, in cases where sample sizes were different, it was observed that as the difference between variance rates increased, power values increased for all tests.

For gamma distribution which shows small deviation from symetry, it was observed that for small sample sizes, low power values were obtained in all situations for all the tests. For medium sample sizes, when the difference between sample sizes was high, it was seen that higher power values were obtained in all tests except for the bootstrap Levene's median test; however, when the difference was low, lower power values were obtained. Yet for the bootstrap Levene's median test, in cases where the difference between variance rates was both small and large, high power values were obtained. For large sample sizes, it was observed that as the difference between variance rates increased, high power values were obtained for all tests in all situations. When sample sizes of the groups were different, however, it was observed that when high variance corresponded with small sample size and low variance corresponded with large sample size, although poorer results were obtained; at the same time power values increased as the difference between variance rates increased.

For chi-squared distribution showing a high degree of deviation from symmetry, all tests gave rather low power values for small sample sizes. For other sample sizes, the best results were obtained for Levene's median test and the bootstrap Levene's median test. For medium sample sizes, however, when the difference between variance rates was large, higher power values were obtained. For large sample sizes, it was seen that there was no difference in terms of power between the tests. For unequal sample sizes, it was seen that when the difference between variance rates were small, rather low power values were obtained, while as the difference between variance rates increased, power values increased for all tests.

In the present paper, it was seen that the bootstrap Levene's median test gave the best results for all distributions in terms of type-I error rates. Yet in terms of power, it was observed that the bootstrap Levene's median test gave better results for symmetric distributions and that for asymmetric distributions it gave higher power values as the difference between variance rates increased. Parra-Frutos et al. (2009), also reported that the bootstrap Levene's median test gave good power values for all distributions only when the difference between variance rates was high.[27] In our study it was seen that while Levene's median test gave type-I error rates close to nominal level for all distributions, particularly with large sample sizes. For small sample sizes, type-I error values were considerably below nominal level. Shoemaker (2003) also stated in his study, which was conducted with various symmetric and asymmetric distributions, that type-I error rates for Levene's median test were at the nominal level with large sample sizes, but that conservative values were obtained with small sample sizes.[9] Similarly, Algina et al. (1995) stated that for all distributions, Levene's median test gave good results for all sample sizes other than small sample sizes.[28] In terms of power, in our study for all distributions other than normal distribution, the test gave low power values for small sample sizes and when the difference between variance rates was small; for different sample sizes, especially when large samples had high variance, it gave higher power values. These findings are similar to those of Algina et al. (1989).[29]

In our study, Levene's mean test gave results close to nominal level for normal distribution and symmetric platykurtic distributions; whereas for both leptokurtic and platykurtic asymmetric distributions, it gave quite liberal results. Similarly, Veitch et al. (1974) stated that Levene's mean test gave more liberal results for leptokurtic distributions. For platykurtic distributions, however, they stated that the test gave good power and type-I error rates.[30]

In our study while Bartlett's test gave type-I error rates close to nominal level for normal distribution, it gave rather conservative rates for normal distribution that was symmetric and platykurtic, and extremely liberal rates for other distributions. Veitch et al. (1974) and Shoemaker (2003), similarly, stated that type-I error rates for Bartlett's test were quite different from nominal level for distributions other than normal distribution.[9,30]

Veitch et al. (1974) stated that Cochran's test gave rather liberal results compared to the other tests[30]. Similarly, we found that, Cochran's test gave rather conservative results for uniform distribution, while it gave rather liberal results for t-distribution, gamma distribution and chi-squared distribution.

In terms of power, it was seen that Levene's trimmed mean test gave type-I error values close to nominal level for medium sample sizes in symmetric distributions, while the type-I error values were considerably above nominal level for skewed distributions. While Parra-Frutos (2009), stated differently to our study that Levene's trimmed mean test also gave type-I error rates close to nominal level for small sample sizes in symmetric distributions, whereas, similarly to our study findings, it is reported that for skewed distributions, rather liberal type-I error rates were obtained.[27]

CONCULUSION

Generally in the present study, the best results were obtained for all tests with normal distribution, followed by uniform distribution and t-distribution from symmetric distributions. In terms of type-I error rates, it was observed that the bootstrap Levene's median test gave the best results, and that the power of the tests was affected by the difference between variance rates and by the combination of sample size-variance magnitude.

Conflict of Interest

Authors declared no conflict of interest or financial support.

Authorship Contributions

Idea/Concept: *Deniz Sığırlı;* ***Design:*** *Deniz Sığırlı;* ***Control/Supervision:*** *Deniz Sığırlı;* ***Analysis and/or Interpretation:*** *Kürşad Nuri Baydılı;* ***Literature Review:*** *Kürşad Nuri Baydılı;* ***Writing the Article:*** *Kürşad Nuri Baydılı;* ***Critical Review:*** *Kürşad Nuri Baydılı;* ***References and Fundings:*** *Kürşad Nuri Baydılı.*

REFERENCES

1. Jan SL, Shieh G. Sample size determinations for Welch's test in one-way heteroscedastic ANOVA. Br J Math Stat Psychol 2014;67(1):72-93.
2. Kaps M, Lamberson W. Biostatistics for Animal Science. London: 2014. p.445.
3. Cochran WG. Some consequences when the assumptions for the analysis of variance are not satisfied. Biometrics 1947;3(1):22-38.
4. Bishop TA, Dudewicz EJ. Exact analysis of variance with unequal variances: test procedures and tables. Technometrics 1978;20(4):419-30.
5. Özdamar K. SPSS İle Biyoistatistik. Eskişehir: Kaan Kitabevi; 1999. p.317-8.
6. Montgomery DG. Design and Analysis of Experiments. 3rd ed. New York: John Wiley & Sons Inc; 1991. p.565.
7. Conover ME, Johnson ME, Johnson MM. A comparative study of test for homogeneity of variances, with applications to the outer continental shelf bidding data. Technometrics 1981;23(4):351-61.
8. Box GEP, Andersen SL. Permutation theory in the derivation of robust criteria and the study of departure from assumption. J R Stat Soc Series B 1955;17(1):1-34.
9. Shoemaker LH. Fixing the F test for equal variances. Am Stat 2003;57(2):105-14.
10. Levene H, Olkin I. Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling. 1st ed. Stanford, Calif.; Stanford University Press; 1960. p.278-92.
11. Brown B, Forsythe A. Robust tests for equality of variances. J Am Stat Assoc 1974;69(346):364-7.
12. Boos DD, Brownie C. Bootstrap methods for testing homogeneity of variances. Technometrics 1989;31(1):69-82.
13. Boos DD, Brownie C. Comparing variances and other measures of dispersion. Stat Sci 2004;19(4):571-8.
14. Levene H. Robust tests for equality of variances. Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling. 1st ed. Stanford, Calif.; Stanford University Press; 1960. p.278-92.
15. Fligner MA, Killeen TJ. Distribution-free two-sample tests for scale. J Am Stat Assoc 1976;71(353):210-3.
16. Filliben, JJ, Heckert A. Levene's test for equality of variances, 2000. Erişim: http://www.nist.gov/itl/div898/handbook/eda/section3/eda358.htm (8 Şubat 2017, 14.47).
17. Neter J, Kutner MH, Nachtsheim CJ, Wasserman W. Applied Linear Regression Model. 3rd ed. Illinois: Irwin Peechawanich; 1996. p.720.
18. Montgomery DC, Runger GC. Applied Statistics and Probability for Engineers. 4th ed. John Wiley & Sons; 2007. p.768.
19. Bartlett MS. Properties of sufficiency and statistical tests. Proceedings of the Royal Society of London Series A 1937;160:268-82.
20. Snedecor GW, Cochran WG. Statistical Methods. 7th ed. Ames, Iowa; Iowa State University Press; 1980. p.507.
21. Manly BFJR. Randomization, Bootstrap and Monte Carlo Methods in Biology. 2nd ed. Boca Raton, Florida; Chapman & Hall/CRC Press; 1997. p.424.
22. Fox J, Weisberg S. An R Companion to Applied Regression. 2nd ed. Thousand Oaks CA: Sage; 2011. p.449.
23. Gastwirth JL, Gel YR, Hui WL, Lyubchich V, Miao W, Noguchi K. Tools for Biostatistics, Public Policy, and Law. Version 3.2. 2017.
24. Komsta L. Tests for outliers. 2015. p.2-5. https://cran.r-project.org/web/packages/outliers/outliers.pdf.
25. Hatchavanich D. 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. Southeast-Asian J of Sciences 2014;3(2):181-94.
26. Vorapongsathorn T, Taejaroenkul S, Viwatwongkasem C. A comparison of type I error and power of Bartlett's test, Levene's test and Cochran's test under violation of assumptions. Songklanakarin J Sci Technol 2004;26(4):537-47.

27. Parra-Frutos I. The behaviour of the modified Levene's test when data are not normally distributed. Computational Statistics 2009;24(4):671-93.

28. Algina J, Blair RC, Coombs WT. A maximum test for scale: type I error rates and power. Journal of Educational and Behavioral Statistics 1995;20(1):27-39.

29. Algina J, Olejnik S, Romer O. Type I error rates and power estimates for selected two-sample tests of scale. Journal of Educational Statistics 1989;14(4):373-84.

30. Veitch WR, Roscoe JT. Homogeneity of variance: an empirical comparison of 4 statistical tests. The Journal of Experimental Education 1974;43(2):73-8.