REVSTAT – Statistical Journal
Volume 21, Number 3, July 2023, 389–404
https://doi.org/10.57805/revstat.v21i3.411

Assessing Homoscedasticity Graphically: Levene–Brown–Forsythe Approaches

Author: Elsayed A. H. Elamir ✉
– Management and Marketing Department, University of Bahrain, Kingdom of Bahrain
shabib@uob.edu.bh

Received: February 2021 Revised: January 2022 Accepted: January 2022

Abstract:

- The problem of homoscedasticity arises in several fields such as business, education, environments, and medicine, and common question in many statistical analyses. One of the most important tests in this direction is Levene test and its robust version Brown–Forsythe test. The goal of this paper is threefold. The first goal is to propose an expression that enable to develop a graphical way for Levene–Brown–Forsythe tests. The second goal is to derive the sampling distribution of the proposed expression as the generalized beta prime distribution. The third goal is to provide deep insight and understanding where the dispersion effects occur. Simulation study is carried out to study the level of significance and power of the proposed test in comparison with the original Levene–Brown–Forsythe tests. The results are of great values since the proposed method:
  (**a**) provides powerful visual tool and deep insight for testing homoscedasticity,
  (**b**) keeps the size and power of the test similar to Levene–Brown–Forsythe tests,
  (**c**) does not need to pairwise comparisons.

  Two applications are presented to show the utilities of the proposed method.

Keywords:

- *beta distribution; Bonferroni approximation; homogeneity of variance; nonnormality; test power; type I error.*

AMS Subject Classification:

- 62F03, 62J10.

✉ Corresponding author.

1. INTRODUCTION

It is known that the one-way analysis of variance (ANOVA) is one of the most frequently used tests to explore the differences among several treatment means; see, for example, Kutner *et al.* [15], Yigit and Mendes [28] and Nguyen *et al.* [19]. The homoscedasticity plays an important role in ANOVA test since the large deviations from the homoscedasticity can affects the results of F-test for equal means; see Fox and Weisberg [9] and Wang *et al.* [27]. The Levene test [17] and its robust extension Brown–Forsythe test [5] had been used to assess homogeneity of variances or homoscedasticity for several groups. These tests depend on transforming the ANOVA test of means into a test of variances based the absolute values of the differences between observations and a location measure (mean, trimmed mean and median). The assumption of homoscedasticity can be written as

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

versus

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

where k is the number of groups.

The assumption of homoscedasticity can also use on its own to compare the dispersion among several groups in a study. Kvamme *et al.* [16] used Levene test and Brown–Forsythe robust version of Levene test to compare the dispersion of the holes of the chalupa pots from the 3 different locations. The null hypothesis was that the dispersion or variation of each characteristic is the same in the three locations. Plourde and Watkins [22] utilized Levene's test to month-to-month price fluctuates to investigate whether the conduct of oil costs changed within the 1980s and got to be more like that of other goods, which head to have big cost vacillations, they utilized both the nonparametric Fligner–Killeen [8] test and the Brown–Forsythe modified of Levene test in an arrangement of post hoc pairs comparisons to evaluate the relative variations of the price fluctuates. Sant and Cowan [24] considered the effect of a privation of a profit by a company on the changeability of both the estimates of future profit and the real profit. They compared the profit and predicted of companies that excluded a profit amid the period 1963–1984 by comparing the fluctuations of the real or forecasted profit per share 2 years after the omission and 2 years before. They utilized Brown–Forsythe robust version of Levene test. Berger *et al.* [4] used a database of 6026 "echocardiograms" that perused by one of 3 similarly capable perusers to survey the contrasts in recurrence of many analyzes and related measurements. The numbers of "echocardiograms" examined by the pursuers (one, two, three) were 2702, 2101 and 1223, respectively. Levene's test was utilized to evaluate the variation in the measurements of many continuous characteristics. Nordstokke and Zumbo [20] had developed a nonparametric version of Levene test by pooling the observations from all sets, ranking the scores with taking ties in consideration, return the ranks into their original sets, and apply the Levene test on the ranks; for more details; see Nordstokke *et al.* [21] and Shear *et al.* [25]. In analytical methods Aslam and Khan [2] used Levene test to modify Chochran test to be applied for detecting outliers in the data. The goal of this paper is threefold. The first goal is to develop an expression that assist in plotting Levene–Brown–Forsythe tests. The second goal is to obtain the sampling distribution of the suggested expression as a beta prime distribution of the second type that can be used in creating a decision limit. The third goal is to provide deep insight and understanding where the dispersion effects occur. Simulation study is carried out to study the level of significance and power of the proposed test in comparison with the original Levene–Brown–Forsythe tests. The results are of great value since the proposed method provides visual and deep insight where the variation occurs and does not need to post hoc pairwise comparisons. Two applications are studied to show the usage of the proposed method.

Levene–Brown–Forsythe approach is explained in Section 2. The proposed method is introduced in Section 3. The empirical type I error and test power is presented in Section 4. The usage of the proposed method in the analysis of data from two applications is described in Section 5. Section 6 is devoted for conclusion.

2. LEVENE–BROWN–FORSYTHE APPROACH

Suppose there are k groups each follows a normal distribution with means μ_i, standard deviation σ_i, n_i the number of observations in each group, and X_{ij} the response value and n the total number of observations in all groups, $i = 1, ..., k$, $j = 1, ..., n_i$. Levene [17] proposed test to assess the equality of variances for two groups or more. The test was depending on the idea of analysis of variance (ANOVA) for the absolute deviation about mean, $|X_{ij} - X_{i\cdot}|$. Levene's test is based on the classical ANOVA method that can be written as

$$W = \frac{\sum_{i=1}^{k} n_i (Z_{i\cdot} - Z_{\cdot\cdot})^2/(k-1)}{\sum_{i=1}^{k}\sum_{j=1}^{n_i} (Z_{ij} - Z_{i\cdot})^2/(n-k)}, \tag{2.1}$$

where k is the number of groups, n_i the number of observations in group i, $i = 1, ..., k$, $n = n_1 + \cdots + n_k$ is the total number, $Z_{ij} = \left|X_{ij} - \bar{X}_{i\cdot}\right|$ is the absolute deviation about group mean, X_{ij} is the observation for j-th case from group i, $Z_{i\cdot} = \frac{1}{n_i}\sum_{j=1}^{n_i} Z_{ij}$ is the mean of Z_{ij} for group i, $Z_{\cdot\cdot} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i} Z_{ij}$ is the mean of all Z_{ij}.

Although Levene noticed that $\left|X_{ij} - \bar{X}_{i\cdot}\right|$ are not independent within each group, he proved that the correlation is of order $1/n_i^2$ and considered that this is small dependency within each group and would not be seriously impact the distribution of W; see Gastwirth *et al.* [11]. Therefore, the test statistic W is approximated by F-distribution with $k-1$ and $n-k$ degrees of freedom, i.e., $F(\alpha; k-1, n-k)$ where F is the quantile for F-distribution and α is prechosen level of significant. In practice it may be concluded that there is heterogeneity if $W > F(\alpha; k-1, n-k)$. Brown and Forsythe [11] proposed revised version to Levene test by using median or trimmed mean rather than mean, i.e., $Z_{ij} = \left|X_{ij} - \widetilde{X}_{i\cdot}\right|$ or $Z_{ij} = \left|X_{ij} - \check{X}_{i\cdot}\right|$, $\widetilde{X}_{i\cdot}$ median and $\check{X}_{i\cdot}$ trimmed mean, with the same approximated distribution $F(\alpha; k-1, n-k)$. Brown and Forsythe carried out simulation study that indicated that median and trimmed mean performed better in heavy-tailed symmetric and skewed distributions while mean is performed best in case of normal and moderate-tailed symmetric distribution; see Brown and Forsythe [5] and Gastwirth *et al.* [11]. Although different underlying distributions give different optimal choice for location parameter, the optimal choice based on median is a recommended one as it provides a good robustness for many types of non-normal data while hold a good power in normal and symmetric distributions; see Gastwirth *et al.* [12], Wang *et al.* [27] and Nguyen *et al.* [19].

3. THE PROPOSED METHOD

The Levene–Brown–Forsythe test can be rewritten as

$$W = \frac{\sum_{i=1}^{k} n_i(Z_{i\cdot} - Z_{\cdot\cdot})^2/(k-1)}{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Z_{ij} - Z_{i\cdot})^2/(n-k)} = \sum_{i=1}^{k} \frac{n_i(Z_{i\cdot} - Z_{\cdot\cdot})^2/(k-1)}{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Z_{ij} - Z_{i\cdot})^2/(n-k)} = \sum_{i=1}^{k} U_i . \tag{3.1}$$

Hence,

$$U_i = \frac{n_i(Z_{i\cdot} - Z_{\cdot\cdot})^2/(k-1)}{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Z_{ij} - Z_{i\cdot})^2/(n-k)}, \qquad i = 1, 2, ..., k. \tag{3.2}$$

This is the ratio for each between square and all treatments squares or contribution of each between squares to mean square error. Therefore, the Levene–Brown–Forsythe tests could be plotted as

$$x_{axis} = i \;\text{ versus }\; y_{axis} = U_i \text{ with } DL, \quad \text{for } i = 1, 2, ..., k,$$

where DL is the decision limit obtained from the sampling distribution of U_i.

3.1. The sampling distribution of U_i

Under the assumptions of one-way ANOVA:

(**a**) $X_{i1}, ..., X_{kn_i}$ is a random sample of size n_i from a normal population, $i = 1, ..., k$;

(**b**) the random samples from different populations are independent;

see Johnson and Wichern [14]. Furthermore, Gastwirth *et al.* ([11], page 4) had written that "$Z_{ij} = |X_{ij} - \bar{X}_{i\cdot}|$ are treated as independent, identically distributed, normal variables, and the usual ANOVA statistic is utilized". Since $Z_{ij} = |X_{ij} - \bar{X}_{i\cdot}|$ is not normally distributed, the Levene's method takes usefulness of the reality that the ANOVA procedures for comparing means are robust to infraction of the assumption that the data follows a normal distribution; see Gastwirth *et al.* ([11], page 4) and Miller ([18], page 80). Therefore, if the null hypothesis of homogeneity of variance is true, hence, the sampling distribution of U_i can be derived as

$$n_i(Z_{i\cdot} - Z_{\cdot\cdot})^2/(k-1) \sim \sigma^2\left(\frac{n - n_i}{n(k-1)}\right)\chi^2(1) \tag{3.3}$$

and

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Z_{ij} - Z_{i\cdot})^2/(n-k) \sim \sigma^2\chi^2(n-k)/(n-k). \tag{3.4}$$

Hence,

$$U_i \sim \frac{((n-n_i)/n(k-1))\chi^2(1)}{\chi^2(n-k)/(n-k)} = \frac{gamma\left(\frac{1}{2}, \frac{n(k-1)}{2(n-n_i)}\right)}{gamma\left(\frac{n-k}{2}, \frac{n-k}{2}\right)}. \tag{3.5}$$

The sampling distribution of U_i can be obtained as

$$(3.6)\quad f_{U_i}(u) = \frac{[((n-n_i)(n-k))/n(k-1)]^{-1/2}}{B\left(\frac{1}{2}, \frac{n-k}{2}\right)}\left(1 + \frac{n(k-1)}{(n-n_i)(n-k)}u\right)^{-(n-k+1)/2} u^{-1/2},$$

where $k > 0$, $i = 1, ..., k$, and B: Beta; see Coelho and Mexia [6] and Elamir [7]. This distribution has parameters k, n_i and n and is a special type from generalized beta prime distribution with $a = 1$, $b = \frac{((n-n_i)(n-k))}{(n(k-1))}$, $p = 1/2$, $q = (n-k)/2$, $x = u$; see Coelho and Mexia [6], R Core Team [23] and GB2 package, Graf and Nedyalkova [13]. As one of the reviewers has pointed out that the distribution of U_i may also be written in terms of a scaled F-distribution. Note that U_i can be rewritten in terms of scaled F-distribution as

$$U_i \sim \frac{((n-n_i)/n(k-1))\chi^2(1)}{\chi^2(n-k)/(n-k)} = \frac{n-n_i}{n(k-1)} F(v_1 = 1, v_2 = (n-k)).$$

From Smyth ([26], page 9), the density function for scaled F-distribution ($x = (a/b)F(v_1, v_2)$) can be written as

$$f(x) = \frac{a^{v_2/2} b^{v_1/2} x^{\frac{v_1}{2}-1}}{\beta\left(\frac{v_1}{2}, \frac{v_2}{2}\right)(a+bx)^{\frac{v_1+v_2}{2}}}, \qquad x > 0.$$

The sampling distribution of U_i can be obtained from scaled F-distribution by replacing $v_1 = 1$, $v_2 = n-k$, $a = 1$, $b = (n-n_i)/(n(k-1))$.

The moments of U_i can be obtained as

$$E({U_i}^h) = \left[\frac{(n-n_i)(n-k)}{n(k-1)}\right]^h \frac{\Gamma(0.5+h)\Gamma\left(\frac{n-k}{2} - h\right)}{\Gamma(0.5)\Gamma\left(\frac{n-k}{2}\right)}, \quad h = 1, 2, ...$$

For example,

$$E(U_i) = \left[\frac{(n-n_i)(n-k)}{n(k-1)}\right] \frac{\Gamma\left(\frac{n-k}{2} - 1\right)}{2\Gamma\left(\frac{n-k}{2}\right)} = \frac{(n-n_i)(n-k)}{n(n-k-2)}$$

and

$$V(U_i) = E({U_i}^2) - E^2(U_i) = \left[\frac{(n-n_i)(n-k)}{n(k-1)}\right]^2 \frac{3\Gamma\left(\frac{n-k}{2} - 2\right)}{4\Gamma\left(\frac{n-k}{2}\right)} - \left[\frac{(n-n_i)(n-k)}{n(n-k-2)}\right]^2.$$

When sample sizes are equal in each group $n_1 = \cdots = n_k = n_e$, the sampling distribution of U_i can be simplified to

$$f_{U_i}(u) = \frac{[1/(n_e-1)]^{-1/2}}{B\left(\frac{1}{2}, \frac{k(n_e-1)}{2}\right)}\left(1 + \frac{1}{(n_e-1)}v\right)^{-(k(n_e-1)+1)/2} u^{-1/2}.$$

This distribution has parameters k and n_e. The moments for U_i can be derived as

$$E({U_i}^h) = (n_e-1)^h \frac{\Gamma(0.5+h)\Gamma\left(\frac{k(n_e-1)}{2} - h\right)}{\Gamma(0.5)\Gamma\left(\frac{k(n_e-1)}{2}\right)};$$

see Coelho and Mexia [6].

3.2. The empirical moments of U_i

To inspect how well the beta prime distribution for U_i in different setting, a simulation study is conducted to obtain the first four empirical moments of U_i at $k = 3$ and 8, $n_i = 10$ and 25 from normal distribution, Laplace distribution (symmetric heavy-tail) and chi square distribution with 2 degrees of freedom (asymmetric heavy tail) using mean, trimmed mean (0.25) and median as a measure of location. The steps for empirical study are:

1. Select the required design for example $k = 3, n_i = 10$, normal distribution and mean as location measure;
2. Simulate data from a selected distribution with equal variance;
3. Calculate U_i, $i = 1, ..., k$, for each group;
4. Calculate the first four moments for each U_i, $i = 1, ..., k$;
5. Repeat this R times and calculate the mean for every design.

Table 1 gives the first four empirical moments for mean of U_i from normal, Laplace and chi square ($df = 2$) in addition to the theoretical value from the beta prime distribution.

Table 1: Mean of the first four empirical and theoretical (theo.) moments of mean of U_i using different setting and location measures (mean, Tri: trimmed mean (0.25) and Med: median).

k	n_i		Mean				Tri				Med			
			mean	Var.	Sk.	Ku.	mean	Var.	Sk.	Ku.	mean	Var.	Sk.	Ku.
3	10	N	0.386	0.314	3.18	19.45	0.368	0.287	3.19	20.49	0.315	0.211	3.08	21.75
		Laplace	0.443	0.396	3.24	20.02	0.377	0.302	3.44	23.51	0.34	0.237	3.22	22.35
		$\chi^2\,(df=2)$	0.708	1.043	3.24	20.53	0.539	0.675	3.73	23.61	0.364	0.276	3.81	24.62
		Theo.	0.36	0.293	3.42	23.02	0.36	0.293	3.42	23.02	0.36	0.293	3.42	23.02
	25	N	0.359	0.265	3.24	20.96	0.351	0.256	3.23	18.28	0.313	0.204	3.3	21.23
		Laplace	0.377	0.278	2.82	15.02	0.346	0.24	2.69	14.01	0.334	0.219	2.79	15.23
		$\chi^2\,(df=2)$	0.617	0.762	3.13	16.05	0.471	0.472	3.47	17.63	0.338	0.233	2.99	18.28
		Theo.	0.342	0.245	3.02	17.29	0.342	0.245	3.02	17.29	0.342	0.245	3.02	17.29
8	10	N	0.141	0.041	3.24	21.16	0.133	0.036	3.12	18.43	0.111	0.025	3.24	20.6
		Laplace	0.161	0.056	3.65	30.77	0.135	0.039	3.92	28.59	0.122	0.034	3.65	26.38
		$\chi^2\,(df=2)$	0.249	0.164	4.09	29.14	0.182	0.086	4.69	33.9	0.125	0.04	4.33	30.29
		Theo.	0.128	0.034	3.02	17.29	0.128	0.034	3.02	17.29	0.128	0.034	3.02	17.29
	25	N	0.128	0.033	2.94	15.43	0.13	0.034	2.74	16.08	0.117	0.027	2.81	15.98
		Laplace	0.137	0.039	3.11	18.29	0.126	0.033	3.11	18.49	0.121	0.03	3.1	19.12
		$\chi^2\,(df=2)$	0.221	0.114	3.56	25.74	0.164	0.065	3.72	26.1	0.123	0.033	3.55	22.76
		Theo.	0.126	0.032	2.89	15.78	0.126	0.032	2.89	15.78	0.126	0.032	2.89	15.78

This table illustrates that:

1. When the mean is the location measure, the best results (empirical is very close to theoretical) are obtained from normal distribution;
2. When the trimmed mean is the location measure, the best results (empirical is very close to theoretical) are obtained from Laplace distribution, followed by normal;
3. When the median is the location measure, the best results (empirical is very close to theoretical) are obtained from chi square distribution, followed by Laplace distribution then normal.

3.3. Decision limit

To create decision limit (DL), it must take into account k tests that required making difference between two sorts of level of significant α:

1. test-wise alpha (alpha per test $\alpha[PT]$) when working with a specific test;

2. family-wise (alpha per family or experiment alpha $\alpha[PF]$) when working with the whole experiment.

The probability of committing first error for k tests can be defined from Abdi [1] as

$$\alpha(PF) = 1 - (1 - \alpha(PT))^k. \tag{3.7}$$

Hence,

$$\alpha(PT) = 1 - (1 - \alpha(PF))^{1/k}. \tag{3.8}$$

Simpler form can be obtained using Bonferroni approximation as

$$\alpha(PT) \approx \frac{\alpha(PF)}{k}. \tag{3.9}$$

As an example, to perform $k = 8$, and the α per family $(PF) = 0.05$, based on Bonferroni approximation, the null hypothesis will be rejected its related probability is less than $\alpha(PT) \approx 0.05/8 = 0.00625$. Although the Sidak and Bonferroni corrections are closely similar, the Bonferroni correction is more conservative than Sidak and control of the expected number of type I error (Per-family error rate (PFER)) which Sidak does not. Frane [10] stated that "However, it is important to note that the Bonferroni procedure controls not only the FWER (family-wise error rate) but also the PFER (Per-family error rate (PFER))".

In addition to Bonferroni approximation, there is a good method called Benjamini–Hochberg that controls the false discovery rate (the likelihood of an incorrect rejection of a hypothesis occurs) using sequential modified Bonferroni correction for several testing rather than the family wise error rate. Benjamini and Hochberg [3] defined the false discovery rate (FDR) as the number of false discoveries in an experiment divided by the total number of discoveries in that experiment where the discovery is a test that passes one acceptance threshold. In other words, it represents one believe the result is true, but when they are accepted it is never known how many of discoveries are right or wrong. According to Benjamini and Hochberg [3], if q-value is an estimate of FDR from p-values, it may be written as $q_i = Np_i/i$, N: total p-values, p_i: i-th smallest p-value (likelihood of accepting a false result by chance), Np_i: expected value of false results if one accepts all results which have p-values of p_i or smaller, and i the number of results one accepts at i-th p-value threshold. The steps are:

(**a**) rank the p-values from all multiple hypothesis tests in an experiment;

(**b**) compute q_i;

(**c**) to ensure monotonically decreasing q-values, replace q_i with the lowest value among all lower-rank q-values that computed.

In R-software under the function "`p.adjust(p; method=""; n=length(p))`" one of the methods is BH (Benjamini–Hochberg); see R Core Team [23]. Therefore, the decision line could be proposed by using the quantile function of beta prime distribution and the Bonferroni approximation as

$$DL = \text{qgb2}\left(1-\frac{\alpha}{k},\ a=1,\ b=\frac{(n-n_i)(n-k)}{n(k-1)},\ p=0.5,\ q=\frac{(n-k)}{2}\right).$$

Moreover, the Bonferroni approximation could be replaced by BH using R-function as follows: `p.adjust(p=1−`α/k`; method="BH"; n=length(p))`; see GB2 package Graf and Nedyalkova [13]. Hence,

$$\text{if any } U_i > DL, \quad \text{for } i = 1, ..., k, \qquad H_0 \text{ is rejected.}$$

The U-plot can be plotted as

$$x_{axis} = 1:k \quad \text{versus} \quad y_{axis} = U_i, \quad \text{with decision limit } DL.$$

H_0 is rejected if any point outside DL and this will identify where the differences occur.

4. SIMULATION STUDY

The proposed method using Bonferroni (Bonf.) approximation and Benjamini–Hochberg (BH) method is compared with Levene–Brown–Forsythe methods in terms of type I error $p(\text{reject } H_0 \mid H_0 \text{ is true})$ and power of the test $p(\text{reject } H_0 \mid H_0 \text{ is false}) = 1 - p(\text{accept } H_0 \mid H_0 \text{ is false}) = 1 - \text{type II error}$.

With respect to type I error, the following steps are used in simulation:

1. Construct the desired design $k = 3, 8$, $n_i = 10, 20, 50$ and nominal $\alpha = 0.05$.
2. Simulate data from a required distribution with equal variances. The normal distribution as original distribution, Laplace distribution as symmetric heavy-tailed distribution and χ^2 $(df=2)$ as asymmetric heavy-tailed distributions are used.
3. Calculate U_i-Bonf., U_i-BH, Levene–Brown–Forsythe for each design.
4. Compute the decision limit for U_i-Bonf., U_i-BH and p-values for Levene–Brown–Forsythe.
5. Create a dummy variable by giving 1 for reject and 0 else.
6. Repeat R times and compute the mean for each design.

The results for these procedures are given in Table 2. It can be concluded about type I error that:

1. Levene test and U_i-Bonferroni using mean as location are giving a good empirical type I error in the case of normal distribution;
2. Brown–Forsythe and U_i-BH using median as location are giving a good empirical type I error in the case of chi square distribution;
3. Brown–Forsythe and U_i-Bonferroni using trimmed mean as location are giving a good empirical type I error in the case of Laplace distribution.

In general, Brown–Forsythe and U_i-BH using median as location tend to have adequate type I error control across all used distribution shapes and this is consistent with results of Wang *et al.* [27] and Nguyen *et al.* [19].

Table 2: Empirical type I error using U_i-Bonferroni (Bonf.), U_i-BH, Levene–Brown–Forsythe (LBF) methods, nominal $\alpha = 0.05$ from normal, χ^2 and Laplace distributions based on 10000 replications.

k	n_i	Bonf.	BH	LBF	Bonf.	BH	LBF	Bonf.	BH	LBF
		Mean, Normal (100,5)			Mean, Chisq ($df = 2$)			Mean, Laplace (0,4)		
3	10	0.056	0.06	0.064	0.176	0.185	0.195	0.065	0.067	0.074
	20	0.05	0.053	0.056	0.166	0.172	0.181	0.056	0.058	0.065
	50	0.048	0.051	0.053	0.16	0.168	0.178	0.047	0.05	0.054
8	10	0.071	0.073	0.074	0.314	0.322	0.37	0.103	0.094	0.101
	20	0.055	0.059	0.059	0.271	0.28	0.34	0.081	0.082	0.08
	50	0.053	0.058	0.058	0.255	0.263	0.31	0.061	0.063	0.06
		Median, Normal (100,5)			Median, Chisq ($df = 2$)			Median, Laplace (0,4)		
3	10	0.03	0.032	0.032	0.042	0.047	0.048	0.03	0.03	0.031
	20	0.034	0.037	0.036	0.041	0.044	0.044	0.037	0.04	0.043
	50	0.039	0.043	0.044	0.041	0.046	0.048	0.04	0.042	0.043
8	10	0.034	0.035	0.032	0.064	0.065	0.045	0.056	0.056	0.036
	20	0.036	0.037	0.034	0.056	0.056	0.044	0.051	0.051	0.042
	50	0.044	0.046	0.044	0.051	0.052	0.046	0.048	0.049	0.046
		Trimmed, Normal (100,5)			Trimmed, Chisq ($df = 2$)			Trimmed, Laplace (0,4)		
3	10	0.04	0.045	0.048	0.075	0.078	0.082	0.041	0.044	0.045
	20	0.044	0.046	0.049	0.059	0.062	0.066	0.038	0.042	0.043
	50	0.041	0.043	0.045	0.055	0.058	0.063	0.042	0.046	0.047
8	10	0.053	0.055	0.054	0.115	0.116	0.104	0.075	0.073	0.048
	20	0.046	0.048	0.047	0.08	0.082	0.072	0.064	0.061	0.046
	50	0.047	0.048	0.048	0.064	0.065	0.068	0.056	0.056	0.045

With respect to power of the test, the following steps are used in simulation:

1. Construct the desired design $k = 3, 8$, $n_i = 10, 20, 50$ and nominal $\alpha = 0.05$.

2. Simulate data from a required distribution with unequal variances. The used distributions are the normal distribution with variances 5, 5 and 10 ($k = 3$) and 5, 5, 5, 5, 10, 10, 25 and 25 ($k = 8$), Laplace distribution with $df = 2, 2, 10$ ($k = 3$) and $df = 2, 2, 2, 2, 1, 1, 5, 5$ ($k = 8$) and χ^2 with $df = 2, 2, 4$ ($k = 3$) and $df = 2, 2, 2, 2, 1, 1, 4, 4$ ($k = 8$).

3. Calculate U_i-Bonf., U_i-BH, Levene–Brown–Forsythe for each design.

4. Compute the decision limit for U_i-Bonf., U_i-BH and p-values for Levene–Brown–Forsythe.

5. Create a dummy variable by giving 1 for reject and 0 else.

6. Repeat R times and compute the mean for each design.

The results of these procedures are given in Table 3. It can be concluded that:

1. As k and n_i increase, the power becomes larger. If data is from normal and $k = 3$, n_i needs to be at least 20 to obtain good power while it will be much less if $k = 8$.
2. U_i are giving nearly power similar to Levene–Brown–Forsythe tests using the mean, trimmed mean and median.
3. U_i-BH gives slightly better results than U_i-Bonf. in terms of power.
4. With increasing the number of groups, U_i will be slightly better than Levene–Brown–Forsythe tests especially with using trimmed mean and median.

Table 3: Empirical power using U_i-Bonferroni (Bonf.), U_i-BH, Levene–Brown–Forsythe (LBF) methods, nominal $\alpha = 0.05$ from normal, χ^2 and Laplace distributions based on 10000 replications.

k	n_i	Bonf.	BH	Levene	Bonf.	BH	Levene	Bonf.	BH	Levene
		Mean, Normal var = 5, 5, 10			Mean, Chisq *df* = 2, 2, 4			Mean, Laplace scale = 5, 5, 10		
3	10	0.475	0.488	0.493	0.272	0.278	0.289	0.342	0.35	0.36
	20	0.835	0.847	0.832	0.378	0.387	0.399	0.604	0.612	0.622
	50	0.997	0.997	0.997	0.63	0.64	0.648	0.95	0.952	0.954
		var = 5, 5, 5, 5, 10, 10, 25, 25			*df* = 2, 2, 2, 2, 1, 1, 4, 4			scale = 5, 5, 5, 5, 10, 10, 20, 20		
8	10	0.997	0.997	0.999	0.575	0.598	0.695	0.907	0.924	0.968
	20	1	1	1	0.771	0.789	0.869	0.995	0.997	0.999
	50	1	1	1	0.978	0.981	0.993	1	1	1
		Median, Normal var = 5, 5, 10			Median, Chisq *df* = 2, 2, 4			Median, Laplace scale = 5, 5, 10		
3	10	0.348	0.355	0.361	0.114	0.116	0.121	0.225	0.228	0.236
	20	0.765	0.769	0.774	0.204	0.209	0.22	0.533	0.541	0.552
	50	0.998	0.998	0.998	0.518	0.526	0.538	0.943	0.945	0.947
		var = 5, 5, 5, 5, 10, 10, 25, 25			*df* = 2, 2, 2, 2, 1, 1, 4, 4			scale = 5, 5, 5, 5, 10, 10, 20, 20		
8	10	0.988	0.989	0.997	0.255	0.261	0.282	0.822	0.829	0.878
	20	1	1	1	0.492	0.513	0.639	0.991	0.993	0.999
	50	1	1	1	0.945	0.956	0.986	1	1	1
		Trimmed mean, Normal var = 5, 5, 10			Trimmed mean, Chisq *df* = 2, 2, 4			Trimmed mean, Laplace scale = 5, 5, 10		
3	10	0.42	0.43	0.435	0.162	0.168	0.177	0.272	0.277	0.285
	20	0.78	0.786	0.791	0.25	0.26	0.265	0.559	0.57	0.575
	50	0.997	0.997	0.997	0.561	0.57	0.579	0.945	0.948	0.95
		var = 5, 5, 5, 5, 10, 10, 25, 25			*df* = 2, 2, 2, 2, 1, 1, 4, 4			scale = 5, 5, 5, 5, 10, 10, 20, 20		
8	10	0.994	0.995	0.998	0.335	0.344	0.401	0.854	0.86	0.891
	20	1	1	1	0.566	0.588	0.704	0.992	0.995	0.999
	50	1	1	1	0.956	0.966	0.989	1	1	1

5. APPLICATION

Kvamme *et al.* [16] used Levene test and Brown–Forsythe robust version of Levene test to compare the dispersion of the apertures of the chalupa pots that vary in the method they arrange ceramic production from 3 locations, Dalupa (ApDg), Dangtalan (ApDg) and Paradijon (ApP). The data consists of 343 observations: ApDg that has 55 observations, ApDl that has 171 observations and ApP: that has 117 observations; see Gastwirth *et al.* [12].

Table 6 shows the mean, median and standard deviation (st. deviation) for pot data. The largest standard deviation is 12.73 (ApDg) followed by 8.13 (ApP) while the smallest standard deviation is 5.83 (ApP). Table 4 gives the results of Levene–Brown–Forsythe tests for pot data. The p-values of three tests are showing that the dispersion in every of 3 measured characteristics of the pots in different areas are statistically significant at 0.01 and 0.05.

Table 4: Levene–Brown–Forsythe tests for pot data.

	Mean	Trimmed mean	median
Test statistics	7.716	6.567	6.794
p-value	0.0005	0.0016	0.0013

On the other hand, Figure 1 illustrates the results of U-plot at both significance levels 0.01 and 0.05. Since the number of observations are not equal, the height of DL will be different.

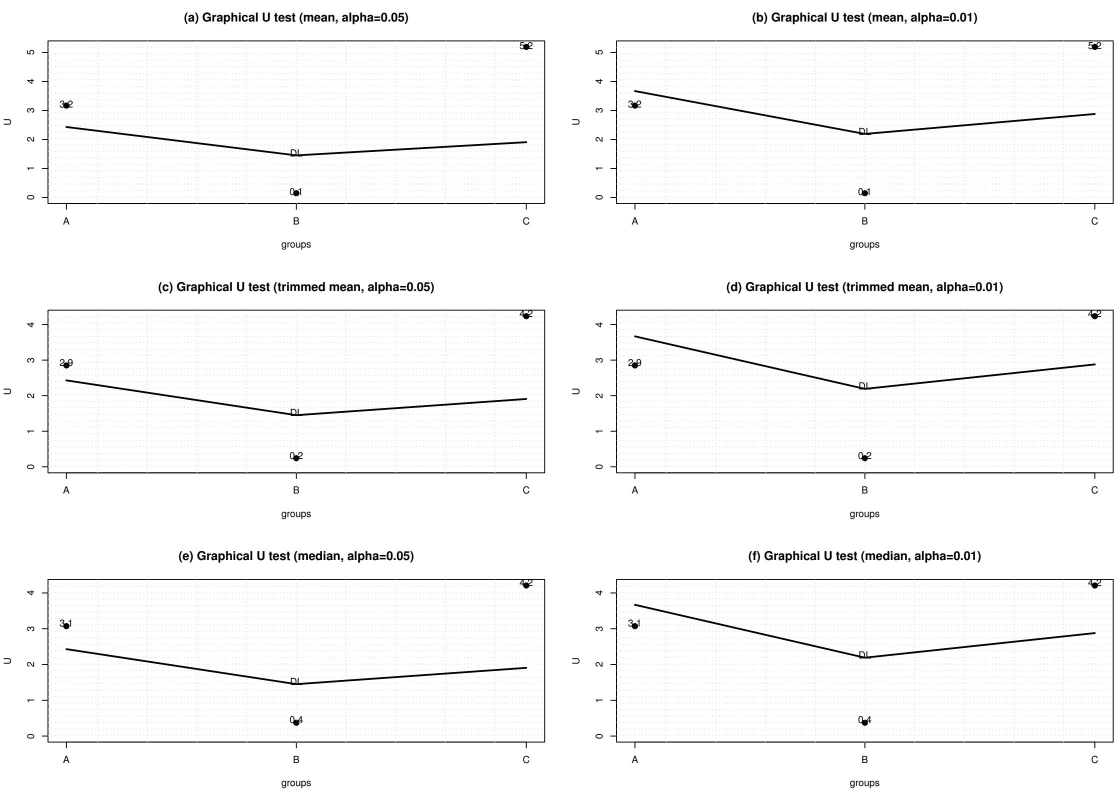

Figure 1: U plot for pot data using mean, trimmed mean and median as location measure.

For example, by using the quantile function of beta prime distribution of the second type, median as location measure and $\alpha = 0.05$, the decision limit is

$$DL = \text{qgb2}\left(1 - \frac{0.05}{3},\ p = 1,\ q = \frac{(343 - (55, 171, 117))(340)}{343(2)}, \alpha = \frac{1}{2},\ \beta = \frac{340}{2}\right).$$

This gives

$$DL = (2.43, 1.45, 1.91).$$

At 0.05, the values of U_1 and U_3 are outside the DL while the value of U_3 is outside DL for 0.01 based on mean, trimmed mean and median as location measures. Therefore, the dispersion in each of the three measured characteristics of the pots in different regions are statistically significant at 0.01 and 0.05 and the most different in dispersion comes from group 3.

The data for the second application is shown in Table 5 where these data are simulated from chi square distribution with $df = 1, 2, 2, 2, 2, 2, 2, 2$. The data consists of 8 groups and in every group, there are 20 observations.

Table 5: Simulated data from χ^2 $(df = 1, 2, 2, 2, 2, 2, 2, 2)$ distribution.

k1	k2	k3	k4	k5	k6	k7	k8
0.27	6.14	3.73	1.13	3.22	1.93	1.07	0.83
1.46	0.1	3.48	0.39	6.28	0.46	2.25	3.89
0.6	1.75	8.23	0.47	1.89	2.35	0.86	0.66
0.49	0.82	1.09	1.53	0.41	2.1	0.92	1.89
0.78	1.7	0.04	5.22	5.78	1.14	1.73	3.27
1.92	0.35	7.03	1.09	2.5	0.94	3.26	4.75
0.11	3.76	8.03	2	0.89	4.12	2.92	5.46
4.9	3.04	0.51	2.6	4.2	5.52	4.31	0.43
1.47	1.68	4.07	0.73	2.2	3.36	1.11	6.3
0.08	3.44	3.5	2.02	0.95	2.75	4.84	5.47
0.64	2.95	0.42	0.44	7.2	0.12	1.38	7.63
0.48	0.1	0.4	0.92	3.45	0.33	0.5	3.25
0.4	0.53	0.63	0.93	2.37	2.18	0.4	4.51
5.37	0.15	2.8	2.73	3.74	1.75	2.24	1.11
0.05	2.16	0.14	3.34	1.29	2.93	1.25	1.4
1.18	0.07	9.48	3.32	0.35	3.45	5.39	2.93
0.01	1.27	0.49	0.47	0.67	1.47	0.48	1.36
0.18	0.67	2.98	3.33	1.68	0.07	0.43	0.32
1.09	2.17	0.2	2.13	0.44	2.25	1.89	1.98
5.07	2.91	2.26	0.82	1.67	0.53	0.26	6.12

Table 6 shows the mean, median and standard deviation (st. deviation) for χ^2 simulated data. The largest standard deviation is 3.02 (k3) followed by 1.24 (k8) while the smallest standard deviation is 1.31 (k4) followed by second smallest 1.54 (k7).

Table 7 gives the results of Levene–Brown–Forsythe tests for simulated data from chi square distribution. The p-values of Levene–Brown–Forsythe tests are showing that the variances in each of the eight groups are statistically significant at 0.01 and 0.05.

With respect to U plot, Figure 2 displays the results of U-plot at both significance levels 0.01 and 0.05 and using mean, trimmed and median as location measures. Since the number of observations are equal, the height of DL will be the same. For example, by using the quantile function of beta prime distribution of the second type and $\alpha = 0.01$, the decision limit can be computed as

$$DL = \text{qgb2}\left(1 - \frac{0.01}{8},\ p = 1,\ q = \frac{(160-20)(160-8)}{160(7)},\ \alpha = \frac{1}{2},\ \beta = \frac{160}{2}\right) = 1.35.$$

At 0.05 and 0.01, the value of U_1 is outside the DL using mean, trimmed mean and median as location measures. Therefore, the assumption of homogeneity of variances is rejected and the most different in dispersion comes from group 3.

Table 6: Summary statistics for Pot and simulation data.

	Pot data			Simulation data							
	ApDg	ApDl	ApP	k1	k2	k3	k4	k5	k6	k7	k8
#	55	171	117	20	20	20	20	20	20	20	20
Mean	170.5	163	128.6	1.33	1.79	2.98	1.78	2.56	1.99	1.87	3.18
median	170	165	130	0.62	1.69	2.53	1.33	2.04	2.02	1.31	3.09
st. devation	12.739	8.127	5.829	1.72	1.58	3.02	1.31	2.02	1.44	1.54	2.24

Table 7: Levene–Brown–Forsythe tests for simulated data.

	Mean	Trimmed mean	median
Test statistics	3.316	2.876	2.859
p-value	0.0026	0.0075	0.0078

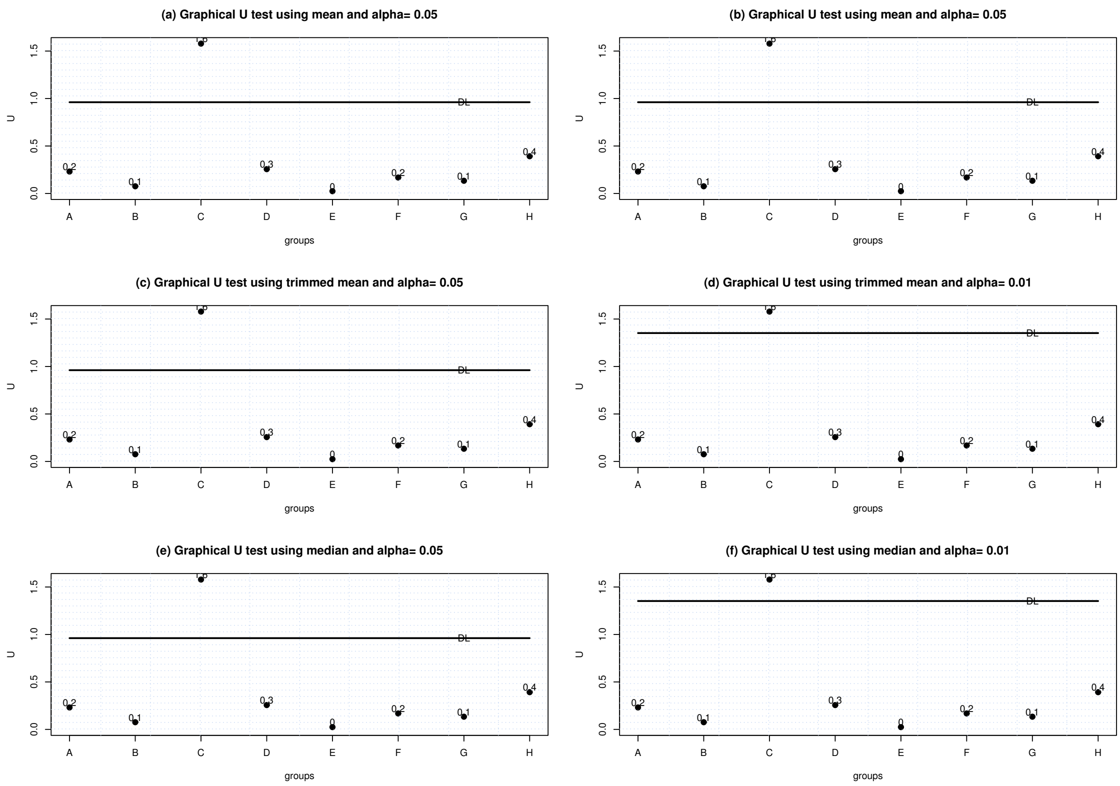

Figure 2: U plot for simulated data from chi square distribution using mean, trimmed mean and median as location measure.

6. DISCUSSION

The Levene–Brown–Forsythe test can be rewritten as

$$W = \sum_{i=1}^{k} \frac{n_i(Z_{i\cdot} - Z_{\cdot\cdot})^2/(k-1)}{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Z_{ij} - Z_{i\cdot})^2/(n-k)}.$$

This can be interpreted as an aggregate way to test whether the level factor mean absolute deviations differ from the overall mean absolute deviation. In terms of the null hypothesis, it tests for the equality of the mean absolute deviations for different factor levels. In terms of alternative hypothesis, it tests that at least two mean absolute deviations for factor levels are not equal. The U_i tests can be rewritten as

$$U_i = \frac{n_i(Z_{i\cdot} - Z_{\cdot\cdot})^2/(k-1)}{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Z_{ij} - Z_{i\cdot})^2/(n-k)}, \qquad i = 1, 2, ..., k.$$

These are simultaneous tests that show every level mean absolute deviation and the decision limit on the graph. If a value of any factor level mean absolute deviation is outside the decision limit, there is evidence that the level factor mean absolute deviation represented by that value is significantly different from the overall mean absolute deviation. In other words, these plots show whether there is statistically significant evidence of each group mean absolute deviation from centre differing from the overall mean absolute deviation from centre. In terms of alternative hypothesis, it tests at least one mean absolute deviation for factor levels is not equal the overall mean absolute deviation.

7. CONCLUSION

Assessing the homogeneity of variance is a prevalent question in many statistical analyses such as regression and analysis of variance. A graphical U_i test for homoscedasticity is proposed as the ratio for the contribution of each between squares treatment to mean square error of all treatments where the sum of the U_i is Levene–Brown–Forsythe tests. The sampling distribution of U_i is derived as beta prime distribution of the second type. By using Bonferroni approximation and Benjamini–Hochberg method, the decision line had been obtained to decide about homogeneity of variances when all values of U_i are less than decision limit or heterogeneity of variances when any value of U_i lies outside the decision line.

Overall, the simulation results showed that the performance of U_i plot is similar to Levene–Brown–Forsythe tests using different designs of number of groups and the number of observations in terms of type I error and test power. Therefore, it can be concluded that U_i plot using mean and trimmed means as a location is suited to symmetric distributions and U_i plot using median as a location was suited to asymmetric distribution. Moreover, if there are no ideas about the shape of the data, the U_i based on median should be used as a general test where it gives a good control for type I error and reasonable power in case of asymmetric distributions while hold a reasonable type I error control and test power in symmetric distributions.

There are many advantages of using U_i plot:

(**a**) provides a powerful visual tool for testing homogeneity of variances;

(**b**) keeps the size and power of the test like Levene–Brown–Forsythe tests;

(**c**) does not need to pairwise comparisons where it could be considered as a complement method to original test.

ACKNOWLEDGMENTS

The author acknowledge the valuable remarks and suggestions made by the referees and editor which improved the presentation of the paper.

REFERENCES

[1] Abdi, H. (2007). *The Bonferroni and Sidak corrections for multiple comparisons*. In "Encyclopedia of Measurement and Statistics" (Neil Salkind), Thousand Oaks (CA), Sage.

[2] Aslam, M. and Khan, N. (2021). Normality test of temperature in Jeddah city using Cochran's test under indeterminacy, *MAPAN: Journal of Metrology Society of India*, **36**(3), 589–598.

[3] Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple hypothesis testing, *Journal of Royal Statistical Society B*, **57**(1), 289–300.

[4] Berger, A.K.; Gottdiener, J.S.; Yohe, M.A. and Guerro, J.L. (1999). Epidemiologic approach to quality assessment in echocardiographic diagnosis, *Journal of the American College of Cardiology*, **34**(6), 1831–1836.

[5] Brown, M.B. and Forsythe, A.B. (1974). Robust tests for equality of variances, *Journal of the American Statistical Association*, **69**(346), 364–367.

[6] Coelho, C. and Mexia, J. (2007). On the distribution of the product and ratio of independent generalized gamma-ratio random variables, *Sankhya: The Indian Journal of Statistics*, **69**(2), 221–255.

[7] Elamir, E.A.H. (2021). Simultaneous test for means: An Unblind way to the F-test in One-way analysis of variance, *Statistics, Optimization and Information Computing*, **X**, Month 2021, 0–14.

[8] Fligner, M.A. and Killeen, T.J. (1976). Distribution-free two-sample tests for scale, *Journal of the American Statistical Association*, **71**(353), 210–213.

[9] Fox, J. and Weisberg, S. (2019). *An R Companion to Applied Regression*, Sage Publication Inc.

[10] Frane, A.V. (2015). Are per-family type I error rates relevant in social and behavioral science, *Journal of Modern Applied Statistical Methods*, **14**(1), 12–23.

[11] Gastwirth, J.; Yulia, R. and Miao, W. (2009). The impact of Levene's test of equality of variances on statistical theory and practice, *Statistical Science*, **24**(2), 343–360.

[12] Gastwirth, J.; Yulia, R.; Wallace, H.; Lyubchich, V.; Miao, W. and Noguchi, K. (2009). *Lawstat: tools for biostatistics, public policy and law*, Version 3.4. `https://cran.r-project.org/web/packages/lawstat/index.html`

[13] Graf, M. and Nedyalkova, D. (2015). *Generalized beta distribution of the second kind: properties, likelihood, estimation.* `https://cran.r-project.org/web/packages/GB2/index.html`

[14] Johnson, A.R. and Wichern, D.W. (2007). *Applied Multivariate Statistical Analysis*, 6th Ed., Pearson, Prentice Hall.

[15] Kutner, M.; Nachtsheim, C.; Neter, J. and William, L. (2004). *Applied Linear Statistical Models*, 5th Ed., McGraw-Hill/Irwin.

[16] Kvamme, K.L.; Stark, M.T. and Longacre, M.A. (1996). Alternative procedures for assessing standardization in ceramic assemblages, *American Antiquity*, **61**(1), 116–126.

[17] Levene, H. (1960). *Robust testes for equality of variances.* In "Contributions to Probability and Statistics" (I. Olkin), Stanford Univ. Press, Palo Alto, CA. MR0120709, 278–292.

[18] Miller, R.G. (1986). *Beyond ANOVA: Basics of Applied Statistics*, Wiley, New York.

[19] Nguyen, D.; Kim, E.; Wang, Y.; Pham, T.V.; Chen, Y.-H. and Kromrey, J.D. (2019). Empirical comparison of tests for one-factor ANOVA under heterogeneity and non-normality: a Monte Carlo study, *Journal of Modern Applied Statistical Methods*, **18**(2), 2–30.

[20] Nordstokke, D.W. and Zumbo, B.D. (2010). A new nonparametric Levene test for equal variances, *Psicologica*, **31**(2), 401–430.

[21] Nordstokke, D.W.; Zumbo, B.D.; Cairns, S.L. and Saklofske, D.H. (2011). The operating characteristics of the nonparametric Levene test for equal variances with assessment and evaluation data, *Practical Assessment, Research and Evaluation*, **16**(5), 1–8.

[22] Plourdes, A. and Watkins, G.C. (1998). Crude oil prices between 1985 and 1994: How volatile in relation to other commodities?, *Resource and Energy Economics*, **20**(2), 245–262.

[23] R Core Team (2021). *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria. `https://www.R-project.org/`

[24] Sant, R. and Cowan, A.R. (1994). Do dividends signal earnings?the case of omitted dividends, *Journal of Banking and Finance*, **18**(6), 1113–1133.

[25] Shear, B.R.; Nordstokke, D.W. and Zumbo, B.D. (2018). A note on using the nonparametric Levene test when population means are unequal, *Practical Assessment, Research, and Evaluation*, **23**(2), 1–11.

[26] Smyth, G.K. (2004). Linear models and empirical Bayes methods for assessing differential expression in microarray experiments, *Statistical Applications in Genetics and Molecular Biology*, **3**(1), 1–11.

[27] Wang, Y.; Rodriguez de Gil, P.; Chen, Y.; Kromrey, J.; Kim, E.; Pham, T.; Nguyen, D. and Romano, J. (2017). Comparing the performance of approaches for testing the homogeneity of variance assumption in One-Factor ANOVA models, *Educational and Psychological Measurement*, **77**(3), 305–329.

[28] Yigit, S. and Mendes, M. (2018). Which effect size measure is appropriate for one-way and two-way ANOVA models? A Monte Carlo simulation study, *REVSTAT – Statistical Journal*, **16**(3), 295–313.