Türk Bilim ve Mühendislik Dergisi, 6(2): 132-138, 2024

Araştırma Makalesi / Research Article

Türk Bilim ve Mühendislik Dergisi
Turkish Journal of Science and Engineering

www.dergipark.org.tr/tjse

Evaluating The Type I Error Rate Performances of Hsieh, F, and Kruskal-Wallis Tests Using Monte Carlo Simulation Technique

Malik Ergin[1*], Özgür Koşkan[1]

[1]Isparta University of Applied Sciences, Faculty of Agriculture, Department of Animal Science – Isparta - Türkiye

*****Corresponding author: malikergin@isparta.edu.tr

MAKALE BİLGİSİ

Received: 27/08/2024
Accepted: 13/12/2024

Keywords: *Analysis of variance, Hsieh test, Kruskal-Wallis test, Type I error rate*

DOI: 10.55979/tjse.1539525

ABSTRACT

In this study, the Hsieh, Classical F, and Kruskal-Wallis tests were compared in terms of Type I error probabilities using samples generated from a two-parameter exponential distribution with various parameters and a $\chi^2_{(2)}$ distribution. Among the tests examined in the study, the Hsieh test was found to have higher Type I error probabilities compared to the Classical F and Kruskal-Wallis tests, indicating that it could not maintain the Type I error at the 5.0% level. The effect of this finding is more pronounced in small sample sizes and when the assumption of homogeneity of variance is not met. In addition, even in heterogeneous variance ratios, that Levene's test neglects, the Type I error rate of the Hsieh test reached 13.3%. In conclusion, the high probability of Type I error in the Hsieh test, especially in small sample sizes and when group variances are not homogeneous, restricts its usage.

Hsieh, F ve Kruskal-Wallis Testlerinin Monte Carlo Simülasyon Tekniği Kullanılarak I. Tip Hata Olasılıkları Bakımından Değerlendirilmesi

ARTICLE INFO

Alınış tarihi: 27/08/2024
Kabul tarihi: 13/12/2024

Anahtar Kelimeler: *Varyans analizi, Hsieh testi, Kruskal-Wallis testi, I.tip hata olasılığı*

DOI: 10.55979/tjse.1539525

ÖZET

Bu çalışmada, iki parametreli üstel dağılımdan farklı parametrelerle ve $\chi^2_{(2)}$ dağılımından üretilen örnekler kullanılarak Hsieh, F ve Kruskal-Wallis testleri, I.tip hata olasılıkları açısından karşılaştırılmıştır. Çalışmada incelenen testler arasında, Hsieh testinin I.tip hata olasılıklarının, Klasik F ve Kruskal-Wallis testlerine göre daha yüksek olduğu bulunmuş ve bu durum, I.tip hatayı %5 seviyesinde tutamadığını göstermektedir. Bu durum özellikle küçük örnek genişliklerinde ve varyansların homojenliği varsayımının sağlanmadığı durumlarda daha belirgin hale gelmektedir. Ayrıca, Levene testi tarafından göz ardı edilen heterojen varyans oranlarında bile, Hsieh testinin I.tip hata oranı %13.3'e ulaşmıştır. Sonuç olarak, Hsieh testindeki yüksek I.tip hata olasılığı, özellikle küçük örnek genişliklerinde ve grup varyanslarının homojen olmadığı durumlarda kullanımını sınırlamaktadır.

1. Introduction

The actual mean of robust statistical test is that it preserves a Type I error rate at the level of 5.0% and has statistical power close to the theoretical power, even when the observations in the given dataset may not satisfy to the assumptions of the test technique (Lix et al., 1996).

In this context, parametric tests are considered more robust; however, the requirement of satisfying their assumptions is emphasized. These assumptions can be summarized as follows:

- Normality of the observations in the dataset
- Homogeneity of the variances
- The data should be continuous

Analysis of Variance or in other words in the literature, the F test, is one of the parametric tests used to test the equality of more than two group means. Blanca et al. (2017) emphasized that the Classical F-test is a powerful test when observations deviate from normality to a moderate degree, when the assumed distribution of each group's population is the same, and when working with balanced and large sample sizes. However, in this point, the researchers can be misguided by factors such as the degree of skewness in the distribution or presence of large sample sizes. In relation to that, Cessie et al. (2020) found that Student – t test is robust when the sample size of each group is equal and more than 25.

In the case of not satisfying the assumptions of the Classical F test, it is a reality that the obtained results do not reflect the truth. As well known, in the hypothesis set at the beginning of the experiment, the null (H_0) hypothesis supports the observed differences among group means are coincidental. The failure to meet the assumptions will cause an increase in the Type I error probability, which was set at 5% at the beginning of the test. In other words, when there is no difference among the group means in truth, the probability of a difference will increase, and the null hypothesis will be rejected mistakenly. If the data distribution is non-normal, applying the Classical F test without considering the assumption of normality can lead to wrong estimation of the model parameters or a decrease in the test power to detect treatment effects (Nwobi &

Akanno, 2021). In cases where data are some kinds of random sample from population which is distributed as beta, gamma, exponential, Weibull., etc., researchers often encounter difficulties in making the correct decision regarding whether to apply transformations and continue with the Classical F test or to directly apply non-parametric tests (Hammouri et al., 2020). Another option, if the assumptions are not met, is to use parametric alternative tests. In the literature, there are different parametric alternative tests such as Welch F, Brown – Forsythe, Wilcox Hm, James second-order, Alexander – Govern and Marascuilo tests (Mendeş, 2002). Author compared these parametric alternative tests in various distributions and simulation experiments in terms of type I error rate and test power by Monte Carlo simulation technique. Even though some of these tests are included in widely used package programs such as SPSS or Minitab, most of them can be conducted with libraries written in the R programming language.

Although transforming the data and then applying the classical F test is another option, some researchers who are not experts in statistics do not use appropriate transformation techniques. Wilcox (2002) reported that transformation techniques could be unable to cope with extreme values in the dataset and even some powerful transformation techniques cannot fit the observations to normal distribution.

Lix et al. (1996) reported that when non-normal distribution and heterogeneity of variances are together, using trimmed means for the location parameter and Winsorized variance for the scale parameter causes the test more robust. The use of non-parametric methods, such as the Kruskal – Wallis test, is also a frequently used method when assumptions are not met. Nonparametric methods do not have assumptions about the distribution of the data, but they are more sensitive to heterogeneity than parametric methods. However, even though non-parametric methods act independently of the assumptions of the distribution, they are very sensitive to high heterogeneity of variances. Hammouri et al. (2020) reported that using the general linear model for analysis of variance and regression, which does not require the assumption of normal distribution, may be an alternative option. On the other hand, Fan & Hancock (2012) reported that Robust Means Modelling approach is more consistent than classical F test in terms of type I error rate of skewed datasets.

In the literature, there are a lot of simulation studies in which alternative tests are evaluated by several researchers. These studies assess the type I error rate and test power in cases where the distribution shape is skewed, comparing group means (Koşkan & Gürbüz, 2009; Arıcı et al., 2011; Mendeş & Yiğit, 2013; Çavuş & Yazıcı (2020); Hammouri et al.,2020) and it is still being investigated. Mendeş & Yiğit (2013), compared classical F and ANOM tests with regards to type I error rate and test power in testing equality of the group means. The Hsieh test (also M statistic) that is fundamental point of this study, is a modified version of the likelihood ratio statistic under the H statistic. This test statistic developed by Hsieh (1986), a parametric approach that allows testing the equality of the location parameters of populations demonstrating a two-parameter exponential distribution with different scale parameters. A study has been found in the literature where the classical F-test, which could serve as an alternative to the Hsieh test in the context of a two-parameter exponential distribution, was evaluated through simulation.

Çavuş & Yazıcı (2020) compared ANOVA, logT-ANOVA (Logarithmic transformed ANOVA), and the Hsieh test for testing the equality of the location parameters of three groups in terms of type I error rate and test power. They considered the two-parameter exponential distribution as a skewed distribution. In addition, Çavuş (2021) developed a library in R programming language named as doex that consists of several heteroscedastic tests. In this library, besides the analysis of variance, and several alternatives, Hsieh test function takes part also for two parameters exponential distribution.

Beta, Gamma, Weibull, Chi-Square, and Log-normal distributions are similar with two parameters exponential distribution. Currently, Monte Carlo technique has been studied on some distributions (Babacan & Kaya, 2020). As well known, the exponential distribution is a special form of Gama distribution. Due to this similar relationship, the solution to similar problems can also be achieved. There is another relationship for Gama distribution that forming Chi-Square distribution by ensuring that integers are positive (Kim et al., 2011). In studies which related agricultural data or theoretical sampling, two parameters exponential distribution as important as normal distribution (Maurya et al., 2011). As an illustration, in the livestock datasets, the milk yields at the lactation shows the exponential distribution.

There are also some approaches proposed in the literature for testing the equality of group means in populations having a two-parameter exponential distribution. Malekzadeh & Jafari (2020) suggested that confidence interval – based estimation approaches, parametric bootstrap, and methods based on p-values may be used for parameter estimation in the case of a two-parameter exponential distribution.

Ghosh & Razmpour (1984) compared several approaches to estimate the common location parameters in populations fit to the exponential distributions with unequal scale parameters, for small sample sizes. Zhuang & Bapat (2022) proposed approaches to hypothesis testing of location parameters for two populations that demonstrate a two-parameter exponential distribution, controlling for Type I and Type II errors. Pawlitschko (2024) compared several robust methods for estimating the location parameter of a two-parameter exponential distribution. Krishnamoorthy et al. (2020) examined various likelihood ratio tests to compare the scale and location parameters in populations with a two-parameter exponential distribution and to test the homogeneity of distributions in those populations. Li et al. (2015) studied on a statistical method, using confidence intervals they derived, to test the equality of means in populations obtained from a two-parameter exponential distribution. Krishna & Goel (2018), used Bayesian approaches to estimate location and scale parameters on randomly censored data following a two-parameter exponential distribution. They reported that Bayesian approaches gave better results than maximum likelihood approaches via Monte Carlo simulation technique. Our literature research revealed that while the Hsieh test and its alternatives have been compared for testing the equality of group means in a two-parameter exponential distribution, their performance has not been evaluated in another distributions. Therefore, it is thought that this study may contribute to literature for evaluating various statistical tests comparing group means in skewed distributions.

In the present study, random numbers were generated from various two-parameter exponential distributions and a chi-square distribution with two degrees of freedom. Hsieh, Classical F, and Kruskal-Wallis tests were compared in terms of Type I error rate for testing the equality of means - medians of more than two groups in various simulation scenarios.

2. Material and Method

2.1. Simulation of random numbers

In the present study, random numbers generated from various two parameters exponential distribution and a chi-square distribution with two degrees of freedom were used, as summarized in Table 1. For manipulating parameters of the distribution in question, different combinations of variances and sample sizes were created. In the determination of group variances, the variance ratio of the last group has been manipulated with small values (such as 0.5 and 1.1). The reason for doing this is that in our preliminary simulation studies for the Hsieh test, as the variance ratio of the last group increases, the Type I error value becomes very high, and its comparison can be complex. The functions named r2exp and rchisq were used to generate random numbers from two parameters exponential distribution and a chi-square distribution, respectively. These functions are elements of the tolerance package (2010) and R Core Team (2023) which are developed in R programming language. Furthermore, for the Hsieh, Classical F test, and Kruskal-Wallis tests we used in the simulation program, we respectively employed the *HS* function derived from the doex package, the *oneway.test* function, and the *kruskal.test* function.

Table 1. Simulation design

Distributions: *r2exp(n,scale,location)* *rchisq(n, df)*	r2exp(n,2,1) r2exp(n,4,1) r2exp(n,6,1) r2exp(n,2,1.5) r2exp(n,4,1.5) r2exp(n,6,1.5)	rchisq(n, 2)
Sample Size (n)	10,35,50	4,12,20,25
Variance Ratios		1.0, 1.05, 1.1
Number of Groups (k)	3, 4, and 5	
Number of simulations	20 000	

2.2. Hsieh test

The Hsieh test (also M statistic) that is fundamental point of this study, is a modified version of the likelihood ratio statistic under the H statistic. It is a parametric approach that allows testing the equality of the location parameters of populations demonstrating a two-parameter exponential distribution with different scale parameters. When the location parameter a is zero, a single-parameter exponential distribution is formed with a mean of b. This distribution, which consists of positive integers, is a right skewed distribution. The Hsieh test statistics with a $\chi^2_{(2k-2)}$ degree of freedom was presented in Equation (1).

$$T_H = -2 \sum_{j=1}^{k} (r_j - 1) \, ln \left[\frac{S_j}{S_j + W_j}\right] \tag{1}$$

H_0: $\mu_1 = \mu_2 ... = \mu_k$

H_1: the difference between the means of at least two groups is significant

If $T_H > \chi^2(2k-2)$, H_0 (null hypothesis) will be rejected.

The probability density function is given in Equation (2) for two parameters of exponential distribution in question.

$$f(a,b) = \frac{1}{b} \, exp\left(-\frac{x-a}{b}\right), x > b, a > 0 \tag{2}$$

Where:

a: location parameter (threshold),

b: scale parameter.

2.3. F test

The mathematical model for one-way analysis of variance (ANOVA), that is commonly used for comparing means of more than two groups with respect to a single factor, can be expressed as Equation 3. For reliable results, one must be met basic assumptions of the F test which can be described as normality of the residuals, homogeneity of the variances, additive effects of the factor levels and independence of the observations.

$$Y_{ij} = \mu + \alpha_i + e_{ij} \tag{3}$$

Where:

Y_{ij}: the observed value for the j^{th} experimental unit in the i^{th} treatment group,

μ: overall population mean,

α_i: the effect of the i^{th} treatment group,

e_{ij}: random error term.

i and j: running from 1 to n

The hypothesis set that enables testing F distribution with (k – 1) and (N – k) degrees of freedom can be described as follows:

H_0: $\mu_1 = \mu_2 ... = \mu_k$

H_1: the difference between the means of at least two groups is significant

(k: treatment groups in the experiment, N: number of total observations, μ_k: population mean).

If the calculated $F = \frac{\text{Mean Square Between Treatments}}{\text{Mean Square Error}}$ value, obtained after ANOVA, is greater than the critical F-table value with (k – 1) and (N – k) degrees of freedom, the null hypothesis (H_0) will be rejected.

2.4. Kruskal-Wallis

The Kruskal-Wallis (KW) test, which is one of the distribution-free tests, is frequently used to determine whether the differences among the medians of two or more groups are due to chance or not. Hypothesis set in the KW test where H statistic is used:

H_0: $M_1 = M_2 ... = M_k$

H_1: the difference between the medians of at least two groups is significant

Where:

M_k is the median of the population

The H statistic can be calculated as Equation (4):

$$H = \left(\frac{12}{N(N+1)}\sum_{i=1}^{k}\frac{R_i}{n_i}\right) - 3(N+1) \quad (4)$$

Where:

N: sum of sample size for all samples,

k: number of samples,

R_i: sum of ranks in the ith sample,

n_i: size of the ith sample

The probability of asserting that there is a difference among group means when the true means are equal is called the type I error rate. It means that it is a probability of rejecting the true null hypothesis (H_0). In this study, the predetermined type I error rate was fixed at %5.0 level. The simulation combinations which are tabulated in Table 1, were iterated 20 000 times. The rejected H_0 hypotheses were stored after 20 000 simulations, and the type I error rate (α) was calculated by dividing the number of rejected H_0 hypotheses by the total number of simulations. The conservative criterion proposed by Bradley (1978) with the range of $4.5 < \alpha < 5.5$ was considered when comparing tests in terms of type I error rate at a significance level of 5%.

3. Results and Discussion

The type I error rates calculated after simulation experiments conducted in various two parameters exponential distributions were shown in Table 3. In the case where the scale and location parameters were 2 and 1 respectively, and the sample sizes were 35 and 50, the type I error rates calculated after all three tests were within the Bradley's criteria. The HS test marginally deviated from Bradley criteria in the case of scale and location parameters were 4 and 1, respectively. Conversely, F and KW tests were in Bradley's criteria at all sample sizes. When the scale and location parameters were set to 6 and 1 respectively, the F and KW tests remained within the Bradley criteria under all experiments. However, the HS test could only fall within Bradley's criteria when there was a large sample sizes, especially when each group contains 50 observations. In addition, the HS test was only met Bradley's criteria when sample size was 50 in each group and distribution combinations where the location parameter was 1.5 and scale parameters were 2, 4, and 6.

In addition, the number of groups was extended to four and five across all scenarios to examine whether this had a significant effect on the Type I error rate. The simulation results, based on random numbers derived from a two-parameter exponential distribution, showed that the Type I error rates for the F and KW tests were within or close to the Bradley criteria. For the HS test, the Type I error rates tended to meet the bounds of the Bradley criteria as the sample size increased. This pattern was observed in all combinations of location and scale parameters.

For small sample sizes (n=10), having four groups controlled the Type I error rates within the Bradley criteria only for the F and KW tests. In contrast, the HS test failed to control the Type I error rates within the criteria for any combination. When the number of groups increased to five, the Type I error rates for the HS and F tests approached the Bradley criteria only as the sample size increased. However, the KW test consistently exhibited Type I error rates close to 5% across all combinations of scale, location parameters, and sample sizes.

The variance ratios that can be neglected by the Levene test were considered to compare HS, F, and KW tests under the $\chi^2_{(2)}$ distribution in the sense of Type I error rate. Therefore, the variance ratio of the last group was manipulated by multiplying random numbers with 1.05 and 1.1 constants. The Levene test statistics (F-values) and p – values were presented in Table 2. The Levene test results were not statistically significant in all combinations of sample sizes and variance ratios. This means that in cases where the variance ratios among groups can be considered small, even in the presence of heterogeneity, the Levene test has accepted the null hypothesis (H_0) and decided that the variances were homogeneous.

Table 2. F and p values of the Levene test

Parameters *rchisq(n,df)*	Variance Ratios	F - value	p-value
rchisq(**4**,2)	1:1:**1.05**	0.96	0.42
rchisq(**12**,2)	1:1:**1.05**	0.31	0.73
rchisq(**20**,2)	1:1:**1.05**	1.94	0.15
rchisq(2**5**,2)	1:1:**1.05**	1.57	0.21
rchisq(**4**,2)	1:1:**1.1**	0.92	0.43
rchisq(**12**,2)	1:1:**1.1**	0.37	0.70
rchisq(**20**,2)	1:1:**1.1**	2.12	0.13
rchisq(2**5**,2)	1:1:**1.1**	1.72	0.19

The type I error rates of the HS test compared to the F and KW tests in cases of small heterogeneity, that can even be neglected by the Levene test, were tabulated in Table 4. Even in cases where the variances were homogeneous, the F and KW tests were often close to or within the Bradley's criteria when the sample size was 12 or greater in each group. However, the HS test only approached Bradley's criteria (5.8%) when the sample size was 25 in each group. When the homogeneity of the variances was manipulated as 1:1:1.05 the type I error rate of the HS test reached 8% with sample size of 25 in each group. Contrary to this result, the KW test was found to be more conservative than other tests in the same sample size. When the variance ratios manipulated as 1:1:1.1, the type I error rates calculated after HS test simulations were bigger as the sample size increased. Conversely, the F and KW tests generally stayed within Bradley's criteria.

As the number of groups increased, the Type I error rates of the HS test also increased in scenarios with homogeneous variances. Although Type I error rates decreased as the sample size increased, the HS test failed to meet the Bradley criteria. In contrast, the F and KW tests produced similar results, with Type I error rates calculated within the acceptable bounds. When there were insignificant or small deviations in homogeneity, the Type I error rates of the HS test increased significantly, particularly as the sample size increased. Moreover, the Type I error rates for the F and KW tests remained within or very close to the Bradley criteria, as expected.

Even though there are several studies in the literature that compare the performance of the HS test in two-parameter exponential distributions, the lack of simulation studies on its performance in other theoretical distribution is a limitation of our study. However, this study can be evaluated with some literature. In the case of a two-parameter exponential distribution with scale and location parameters of 2 and 1 respectively, the HS test was only within Bradley's criteria for large sample sizes. This result shows similarity to the study performed by Çavuş & Yazıcı (2020). They compared HS, LT-ANOVA and F tests in terms of type I error rate in two parameters exponential distribution that considered as skewed distribution. Çavuş and Yazıcı (2020) reported that HS test could not control type I error rate in small sample sizes. In addition, authors stated that both the F and KW tests tend to maintain type I error rate as the sample size increases. Furthermore, Lantz (2013) emphasized that the type I error rate of the Welch test were deviated from 5.0% compared with the F and KW tests in exponential distribution. The author's specific focus on unbalanced sample sizes leads to this situation. In a previous study that evaluated Type I error rates in group comparisons using Monte Carlo simulation, the authors reported that the F-test yielded acceptable Type I error rates in terms of Bradley's criteria when variances were heterogeneous (Mendeş and Yiğit, 2013). These findings are consistent with our results, as expected.

Table 3. Type I error rates (%) of the tests with different parameters of two parameters exponential distribution.

Parameters	HS			F			KW		
r2exp(n,scale,location)	k=3	k=4	k=5	k=3	k=4	k=5	k=3	k=4	k=5
r2exp(**10**,2,1)	7.52	7.64	7.91	4.19	5.44	6.09	4.54	4.365	4.37
r2exp(**35**,2,1)	5.32	5.69	6.13	5.17	5.92	6.14	4.98	5.02	5.00
r2exp(**50**,2,1)	5.42	5.65	5.60	4.94	5.47	6.00	4.84	4.695	4.90
r2exp(**10**,4,1)	7.41	7.89	8.41	4.50	5.25	6.47	4.52	4.605	4.43
r2exp(**35**,4,1)	5.55	5.84	5.93	5.19	5.64	6.45	4.80	4.93	5.13
r2exp(**50**,4,1)	5.68	5.39	5.65	5.35	5.42	6.00	5.04	4.73	4.82
r2exp(**10**,6,1)	7.82	7.45	8.54	4.57	5.17	6.39	4.69	4.36	4.49
r2exp(**35**,6,1)	5.55	5.84	5.92	4.90	5.59	5.98	4.83	4.73	4.81
r2exp(**50**,6,1)	5.48	5.60	5.96	5.13	5.66	5.85	4.92	5.15	4.94
r2exp(**10**,2,1.5)	7.72	8.04	8.35	4.86	5.59	6.33	4.59	4.52	4.40
r2exp(**35**,2,1.5)	5.63	5.715	5.94	5.49	5.76	5.93	4.94	4.84	4.89
r2exp(**50**,2,1.5)	5.52	5.27	5.43	5.06	5.45	5.87	4.78	4.88	4.82
r2exp(**10**,4,1.5)	7.28	7.75	8.32	4.55	5.38	6.28	4.38	4.54	4.44
r2exp(**35**,4,1.5)	6.03	5.93	5.92	5.31	5.69	6.34	4.88	4.75	4.83
r2exp(**50**,4,1.5)	5.48	5.69	5.71	5.35	5.65	5.81	5.02	4.91	5.00
r2exp(**10**,6,1.5)	7.12	8.03	7.96	4.33	5.41	6.38	4.17	4.44	4.47
r2exp(**35**,6,1.5)	5.56	6.01	5.38	5.35	5.80	6.22	5.10	4.81	4.88
r2exp(**50**,6,1.5)	5.45	5.53	5.45	5.32	5.67	6.01	5.10	4.88	5.00

Table 4. Type I error rates (%) of the tests by different variance ratios under $\chi^2_{(2)}$ distribution.

Parameters	Variance Ratios	HS			F			KW		
		k=3	k=4	k=5	k=3	k=4	k=5	k=3	k=4	k=5
rchisq(**4**,2)	1:1:**1**	13.2	15.1	17.1	4.1	4.2	4.4	3.9	3.4	3.1
rchisq(**12**,2)	1:1:**1**	7.0	7.5	7.6	4.4	4.3	4.5	4.7	4.6	4.6
rchisq(**20**,2)	1:1:**1**	6.3	6.3	6.5	4.6	4.7	4.6	4.7	4.8	4.8
rchisq(2**5**,2)	1:1:**1**	5.8	6.3	6.2	4.7	4.4	4.8	4.9	4.5	4.8
rchisq(**4**,2)	1:1:**1.05**	12.8	15.2	17.0	4.2	4.0	4.3	4.2	3.1	3.3
rchisq(**12**,2)	1:1:**1.05**	7.7	8.2	8.8	4.3	4.2	4.3	4.9	4.2	4.3
rchisq(**20**,2)	1:1:**1.05**	7.5	8.2	9.1	4.4	4.7	4.6	4.6	4.7	4.6
rchisq(2**5**,2)	1:1:**1.05**	8.1	8.9	10.0	4.7	4.7	4.6	5.0	4.7	4.7
rchisq(**4**,2)	1:1:**1.1**	13.3	15.7	17.8	4.0	4.2	4.6	4.4	3.4	3.3
rchisq(**12**,2)	1:1:**1.1**	8.6	10.6	11.7	4.7	4.5	4.6	5.0	4.8	3.3

4. Conclusion

At the beginning of the study, the heterogeneity of group variances was manipulated to a greater level. However, in these cases, the HS test's sensitivity to type I error rates was significantly affected, leading to a reduction in the magnitude of variance heterogeneity. One of the significant results of this study was that even in cases of small heterogeneity that the Levene test could neglect, the HS test exhibited higher type I error rates compared to the F and KW tests. This suggests that the HS test is highly sensitive to the disruption of variance homogeneity, even at low levels. In the present study, several groups were examined in all combinations and concluded that an increase in the number of groups negatively affected the HS test in terms of Type I error rates.

Conflict of Interest

Authors declare no conflict of interest.

Authors' Contribution:

The authors declare that they have contributed equally to the article.

5. References

Arıcı, K. Y., Özkan, M. M., & Kocabaş, Z. (2011). Comparison of Kruskal-Wallis test and transformed variance analysis in heterogeneous variance groups. *7th National Zootechnical Student Congress*. September 14-16, Adana, 14-16.

Babacan, E. K., & Kaya, S. (2020). Comparison of parameter estimation methods in Weibull Distribution. *Sigma Journal of Engineering and Natural Sciences*, *38*(3), 1609-1621.

Blanca Mena, M. J., Alarcón Postigo, R., Arnau Gras, J., Bono Cabré, R., & Bendayan, R. (2017). Non-normal data: Is ANOVA still a valid option? *Psicothema*, *29*(4), 552-557. doi.org/10.7334/psicothema2016.383

Bradley, J. V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology*, *31*(2), 144-152. doi.org/10.1111/j.2044-8317.1978.tb00581.x

Cavus, M. (2021). Testing the equality of normal distributed and independent groups' means under unequal variances by doex package. *The R Journal*, *12*(2), 134-154.

Cavuş, M., & Yazıcı, B. (2020). Comparison of Hsieh test and ANOVA for logtransformed on income data. *20th International Symposium on Econometrics, Operational Research and Statistics*. February 12-14, Ankara, 152-158.

Cessie, S., Goeman, J. J., & Dekkers, O.M. (2020). Who is afraid of non-normal data? Choosing between parametric and non-parametric tests. *European Journal of Endocrinology*, *182*(2), 1-3. doi.org/10.1530/EJE-19-0922

Fan, W., & Hancock, G. R. (2012). Robust means modeling: An alternative for hypothesis testing of independent means under variance heterogeneity and nonnormality. *Journal of Educational and Behavioral Statistics*, *37*(1), 137-156. doi.org/10.3102/1076998610396897

Ghosh, M., & Razmpour, A. (1984). Estimation of the common location parameter of several exponentials. *Sankhya: The Indian Journal of Statistics, Series A*, *46*(3), 383-394.

Hammouri, H. M., Sabo, R. T., Alsaadawi, R., & Kheirallah, K. A. (2020). Handling skewed data: A comparison of two popular methods. *Applied Sciences*, *10*(18), 6247. doi.org/10.3390/app10186247

Hsieh, H. K. (1986). An exact test for comparing location parameters of k exponential distributions with unequal scales based on type II censored data. *Technometrics*, *28*(2), 157-164. doi.org/10.1080/00401706.1986.10488117

Kim, B. S., Park, S. G., You, Y. K., & Jung, S. I. (2011). *Probability & statistics for engineers & scientists*. New York, Pearson.

Koskan, Ö., & Gürbüz F. (2009). Comparison of F test and resampling approach for type I error rate and test power by simulation method. *Journal of Agricultural Science*, *15*(1), 105-111.

Krishna, H., & Goel, N. (2018). Classical and Bayesian inference in two parameter exponential distribution with randomly censored data. *Computational Statistics*, *33*, 249-275. doi.org/10.1007/s00180-017-0725-3

Krishnamoorthy, K., Nguyen, T., & Sang, Y. (2020). Tests for comparing several two-parameter exponential distributions based on uncensored/censored samples. *Journal of Statistical Theory and Applications*, *19*(2), 248-260. doi.org/10.2991/jsta.d.200512.001

Lantz, B. (2013). The impact of sample non-normality on ANOVA and alternative methods. *British Journal of Mathematical and Statistical Psychology*, *66*(2), 224-244. doi.org/10.1111/j.2044-8317.2012.02047.x

Li, J., Song, W., & Shi, J. (2015). Parametric bootstrap simultaneous confidence intervals for differences of means from several two-parameter exponential distributions. *Statistics & Probability Letters*, *106*, 39-45. doi.org/10.1016/j.spl.2015.07.002

Lix, L. M., Keselman, J. C., & Keselman, H. J. (1996). Consequences of assumption violations revisited: A quantitative review of alternatives to the one-way analysis of variance F test. *Review of Educational Research*, *66*(4), 579-619. doi.org/10.3102/0034654306600457

Malekzadeh, A., & Jafari, A. A. (2020). Inference on the equality means of several two-parameter exponential distributions under progressively Type II censoring. *Communications in Statistics - Simulation and Computation*, *49*(12), 3196-3211. doi.org/10.1080/03610918.2018.1538452

Maurya, V., Goyal, A., & Gill, A. N. (2011). Simultaneous testing for the successive differences of exponential location parameters under heteroscedasticity. *Statistics & Probability Letters*, *81*(10), 1507-1517. doi.org/10.1016/j.spl.2011.05.010

Mendeş, M. (2002). *The Comparison of some alternative parametric tests to one - way analysis of variance about Type I error rates and power of test under non - normality and heterogeneity of variance*. (PhD thesis, Ankara University).

Mendeş, M., & Yiğit, S. (2013). Comparison of ANOVA-F and ANOM tests with regard to type I error rate and test power. *Journal of Statistical Computation and Simulation*, *83*(11), 2093-2104. doi.org/10.1080/00949655.2012.679942

Nwobi, F. N., & Akanno, F. C. (2021). Power comparison of ANOVA and Kruskal–Wallis tests when error assumptions are violated. *Advances in Methodology and Statistics / Metodološki zvezki*, *18*(2), 53-71. doi.org/10.51936/ltgt2135

Pawlitschko, J. (2001). Robust estimation of the location parameter from a two-parameter exponential distribution (No. 2001, 36). Technical Report.

R Core Team (2023) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna. URL https://www.R-project.org/

Wilcox, R. R. (2002). Understanding the practical advantages of modern ANOVA methods. *Journal of Clinical Child & Adolescent Psychology*, *31*(3), 399-412.

Young, D. S. (2010). Tolerance: an R package for estimating tolerance intervals. *Journal of Statistical Software*, *36*, 1-39. doi.org/10.18637/jss.v036.i05

Zhuang, Y., & Bapat, S.R. (2022). On comparing locations of two-parameter exponential distributions using sequential sampling with applications in cancer research. *Communications in Statistics - Simulation and Computation*, *51*(10), 6114-6135. doi.org/10.1080/03610918.2020.1794007