Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2025.00061

Research Article

Homnormal: A comprehensive R package for testing the homogeneity of variances

Fikri GÖKPINAR[1], Esra GÖKPINAR[1,*]

[1]*Department of Statistics, Gazi University, Ankara, 06560, Türkiye*

ARTICLE INFO

Article history
Received: 16 September 2024
Revised: 28 November 2024
Accepted: 06 April 2025

Keywords:
Homogeneity of Variances; Type I Error; Power; R Package

ABSTRACT

Testing the homogeneity of variances among normal populations is of interest in several research areas. It is widely applied to assess uniformity in quality control, biology, agricultural production systems, human performance studies, and even in the development of educational methods. For this reason, in this paper, we introduce an R package named homnormal, which includes some of the most powerful tests for homogeneity of variances, such as Bartlett's test, Levene's test, Brown-Forsthye's test, Bandairy Dai's test, generalized p-approach, computational approach test, likelihood ratio test based on the computational approach test, and standardized likelihood ratio test. The homnormal package is designed to be user-friendly, making it accessible for researchers across multidisciplinary fields. Additionally, we compare these tests based on their empirical power and type I error rates to determine the best-performing test in specific situations. Simulated type I error rates and powers of these tests are provided and evaluated through an extensive Monte Carlo simulation study. The simulation results indicate that, regardless of the number of groups or whether sample sizes are equal or unequal, small or large, the standardized likelihood ratio test and the likelihood ratio test based on the computational approach test consistently outperform other methods in terms of performance.

Cite this article as: Gökpınar F, Gökpınar E. Homnormal: A comprehensive R package for testing the homogeneity of variances. Sigma J Eng Nat Sci 2026;44(3):1619−1634.

INTRODUCTION

In many scientific applications, testing the homogeneity of variances is widely employed. One of its applications, in experimental design, one of the primary assumptions of ANOVA is the homogeneity of variances of several normal populations. It is well known that violating of the assumption of homogeneity of variances impacts the type I error rate of the F-test, particularly in cases where sample sizes are unequal. Therefore, before ANOVA, it is common practice to test the homogeneity of variances. The ANOVA F-test is also widely used in various areas [1,2]. Furthermore, assessing uniformity is crucial in various fields such as quality control in manufacturing processes, biology, agricultural production systems, and the development of educational methods. For instance, variations in population variability are of interest to biologists for several reasons, such as the investigation of adaptation mechanisms and as a measure of general diversity. [3]. A study examining the batch-to-batch

***Corresponding author.**
*E-mail address: eyigit@gazi.edu.tr

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey

variability in laboratory animals' resistance to parasitic infestation evaluated the homogeneity of variances. In this study, rats injected with 500 larvae were analyzed to determine whether the variability in resistance differed across batches. This demonstrates that homogeneity of variance is a significant issue in biological experiments and medical research [4]. The homogeneity of variance test is used for various purposes in real estate valuation processes and market analysis studies. In particular, it is used to compare different types of real estate and sub-markets. By examining the variability levels of different types in the markets, it helps to identify which types of properties are more homogeneous and, therefore, can be evaluated more reliably [5]. The homogeneity of variance test also plays an important role in education. For example, in a study, the success scores of groups with homogeneous and heterogeneous levels of prior knowledge among students were examined to evaluate how students' learning processes were affected [6].

A review of the literature reveals that numerous methods have been developed to test the homogeneity of variances across several normal populations. Among these, the Bartlett's test [7] is the most widely used. This test is based on the likelihood ratio test (LRT), divided by an approximation of its mean, and it approximately follows a Chi-squared distribution. However, Bishop and Nair [8] and Hartley [9] noted that this test is only valid when the sample sizes are moderate to large and when the number of groups is not excessively high. Many studies have been conducted to identify a testing procedure that maintains the type I error rate while ensuring high statistical power. Hartley [9] proposed the F-max statistic, which represents the ratio of the maximum variance to the minimum variance among the groups. Subsequently, using the absolute differences of observations from their means, Levene [10] introduced an ANOVA-based testing strategy. Additionally, by substituting sample medians for sample means, Brown and Forsythe [11] modified Levene's test. To control the family-wise type I error rate, Bhandary and Dai [12] suggested a test based on a Bonferroni-type adjustment process applied to the ordered p-values.

While some of these tests have an exact distribution, most of them have an asymptotic distribution. As known, tests with asymptotic distribution do not give good results in small sample sizes, especially in terms of type I error. Especially in recent years, it is seen that methods based on resampling techniques have become popular because of the increase in the performance of computers. For this reason, many tests have been developed using the resampling technique. These methods based on the resampling methods, such as generalized p-value, parametric bootstrap, and computational approach are often used for solving the problem containing the interested parameters as well nuisance parameters. Using the generalized p-value approach, Liu and Xu [13] presented a test and evaluated it with Bartlett's test. Gökpınar and Gökpınar [14] suggested a test statistic based on the computational approach test (CAT), a specific case of parametric bootstrap, for homogeneity of variances. Chang et al. [15] suggested the CAT approach of the likelihood ratio test (LRT) and Bartlett's test for homogeneity of variances. They compared these tests with existing tests such as Bartlette's test, Levene's test, Brown- Forsythe's test, Keyes and Levy's adjusted test [16], and Loh's test [17]. According to the results of their study, the CAT approach of the LRT performs better than some existing tests under the normal model. Gökpınar [18] a proposed standardized likelihood ratio test (SLRT) which is not based on any resampling methods and intensive computer methods for homogeneity of variances under normality. She compared this test with Bartlett's test and the likelihood ratio test's Monte Carlo approach. She noted that the SLRT performs better than others.

The aim of most work on methods for testing the homogeneity of variances is to develop a test with a well-controlled type I error rate and high power whenever possible. However, tests either suffer from severe inflation of the type I error rate or lose statistical power to detect heterogeneity of variances. Therefore, much work has been and is being done to improve type I error rates and the power of tests. There are some R packages such as car [19], stats [20], lawstat [21], rstatix [22], onewaytests [23] contain very few of the tests mentioned above (i.e., Bartlett's test and Levene's test) for homogeneity of variances. However, there is no package containing recent methods, especially resampling-based tests. As mentioned above, these tests are powerful and controlling type I error rates better than others. Because of this, our main purpose in this study is to construct a way to choose the best possible testing procedure for the homogeneity of variances. To do this, initially, we create a package that includes most powerful tests obtained in recent years and some other powerful classical tests in the literature. After that, we compare these tests according to their empirical power of tests and type I error rates to detect the best-performing test in specific situations. Thus, the researcher can easily decide which test should be used in specific situations. For this purpose, we create an R package called homnormal by taking into consideration classical tests such as Bartlett's test (BT), Levene's test (LT), Brown-Forsthye's test (BFT), Bandairy Dai's test (BDT) and resample-based tests such as generalized p-approach (GPA), computational approach test (CAT), likelihood ratio test based on CAT (LRCAT), standardized LRT (SLRT). The homnormal package is available on the CRAN. We also compare these tests according to their power of test and type I error rates with comprehensive simulation study under different situations. Thus, the researcher can easily decide which test to use in which situation.

As mentioned above, the homogeneity of variances has many applications. For the sake of simplicity, in this study, we consider a one-way ANOVA model and have discussed the tests used in this field. In future studies, this package can be extended by addressing key assumptions of classical linear regression analysis, such as the homogeneity of variances. In classical linear regression analysis, there have been several tests developed to test the presence of heteroscedasticity. Among them, the tests developed by White [24], Breush and Pagan [25], Goldfeld and Quandt [26] are the most well-known tests, but there are also tests recently developed by Çelik [27,28].

The rest of this study is organized as follows. Firstly, the tests for homogeneity of variances are given. Secondly, we introduce the homnormal package and demonstrate the applicability of the package using a dataset. Thirdly, a comprehensive Monte Carlo simulation study is carried out to compare the performance of the tests regarding their type I error rates and power under various conditions. This paper concludes with several final remarks.

METHODS: TESTS FOR HOMOGENEITY OF VARIANCES

Let $X_{i1},\ldots, X_{in\_i}$, $i = 1,\ldots,k$ be a random sample of size n_i from a normal population with $E(X_{ij}) = \mu_i$ and Var $(X_{ij}) = \sigma_i^2$, i.e. $X_{ij}\sim N(\mu_i, \sigma_i^2)$.

Let $\bar{X}_i$ and S_i^2 be the sample mean and sample variance of the ith population, respectively, where

$$\bar{X}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij}\,,\;\; S_i^2 = \frac{1}{(n_i-1)}\sum_{j=1}^{n_i}\left(X_{ij}-\bar{X}_i\right)^2,\;\;\; i = 1,\ldots,k,$$

Denote the pooled sample mean and sample variance as

$$\bar{X} = \frac{1}{N}\sum_{i=1}^{k} n_i\bar{X}_i\,,\;\;\; S_p^2 = \frac{1}{(N-k)}\sum_{i=1}^{k}(n_i-1)S_i^2,$$

respectively, where $N = \sum_{i=1}^{k} n_i$. The hypotheses of interest are

$H_0\!: \sigma_1^2 = \cdots = \sigma_k^2$ against $H_0\!: \sigma_i^2 \neq \sigma_{i'}^2$ for at least one $i \neq i'$

As mentioned in Section 1, there are many tests improved in the literature to test H_0 against H1. In the following, we describe these tests.

Bartlett's Test (BT)

Bartlett [7] proposed the following test statistic:

$$T_B = \frac{ln(S_p^2)^{n-k} - \sum_{i=1}^{k} \ln\,(S_i^2)^{n_i-1}}{1+\frac{1}{3(k-1)}(\sum_{i=1}^{k}\frac{1}{(n_i-1)}-\frac{1}{(N-k)})} \qquad (1)$$

T_B statistic has approximately χ^2_{k-1} distribution. For a given α, the H_0 is rejected if $T_B > \chi^2_{k-1,\alpha}$.

Levene's Test (LT)

Levene [10] conducted a one-way ANOVA with the dependent variable being as $Z_{ij} = |X_{ij} - \bar{X}_i|$. The Levene's test statistic is given by

$$T_L = \frac{(N-k)\sum_{i=1}^{k} n_i(\bar{Z}_i-\bar{Z})^2}{(k-1)\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(Z_{ij}-\bar{Z}_i\right)} \qquad (2)$$

where $\bar{Z}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} Z_{ij}$ and $\bar{Z} = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i} Z_{ij}$. T_L has approximately $F_{(k-1),(N-k)}$ distribution. For a given α, the H_0 is rejected if $T_L > F_{(k-1),(N-k),\alpha)}$. This test is called as LT throughout the paper.

Brown-Forsythe's test (BFT)

Brown and Forsythe [11] developed a modification of Levene's test by using the median instead of the mean as $W_{ij} = |X_{ij} - \tilde{X}_i|$ where $\tilde{X}_i$ is the median of group i. Thus, the Brown-Forsythe's test is follows as

$$T_{BF} = \frac{(N-k)\sum_{i=1}^{k} n_i(\bar{W}_i-\bar{W})^2}{(k-1)\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(W_{ij}-\bar{W}_i\right)}, \qquad (3)$$

where $\bar{W}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} W_{ij}$ and $\bar{W} = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i} W_{ij}$. BF has approximately $F_{(k-1),(N-k)}$ distribution. For a given α, the H_0 is rejected if $T_{BF} > F_{(k-1),(N-k),\alpha)}$.

Bhandary-Dai Test (BDT)

Bhandary and Dai [12] proposed a test based on Bonferroni type adjustment procedure. The F test statistics are

$$F_i = \frac{S_i^2}{S_p^2};\;\; F_i' = \frac{S_{p:i}^2}{S_i^2},\;\; i=1,\ldots,k, \qquad (4)$$

where $S_{p:i}^2 = \left((N-k)S_p^2 - (n_i-1)S_i^2\right)/\left(N-n_i-(k-1)\right)$. Let P_i and P_i' be as given below:

$$P_i = P(Y > F_i);\;\; P_i' = P(Y' > F_i'),$$

where $Y\sim F_{n_i-1,r_i}$, $Y'\sim F_{r_i,n_i-1}$ and $r_i = \sum_{i\neq j}^{k}\left(n_j-1\right)$. $P_{(1)} < P_{(2)} < \ldots < P_{(2k)}$ is obtained by sorting P_i and P_i'. For a given α, the H_0 is rejected if $P_{(i)} < \frac{i}{2k}\alpha$.

The Generalized p-value Approach (GPA)

Liu and Xu [13] proposed the generalized p-value based on the generalized test variable calculated as in (5)

$$p = P\Bigg(\sum_{i=1}^{k}\frac{(lnU_i-E(lnU_i))^2}{Var(lnU_i)} - \frac{1}{\sum_{j=1}^{k}\frac{1}{Var(lnU_j)}}\left(\sum_{i=1}^{k}\frac{lnU_i-E(lnU_i)}{Var(lnU_i)}\right)^2 \geq \sum_{i=1}^{k}\frac{(ln(n_is_i^2)-E(lnU_i))^2}{Var(lnU_i)} - \frac{1}{\sum_{j=1}^{k}\frac{1}{Var(lnU_j)}}\left(\sum_{i=1}^{k}\frac{ln(n_is_i^2)-E(lnU_i)}{Var(lnU_i)}\right)^2\Bigg) \qquad (5)$$

where $U_i = \frac{n_is_i^2}{\sigma_i^2}\sim\chi^2_{n_i-1}$, $i = 1,2,\ldots,k$. $E(lnU_i)$ and $V(lnU_i)$ terms are obtained by using Monte Carlo integration. H_0 is rejected if $p < \alpha$.

Computational Approach Test (CAT)

Gökpınar and Gökpınar [14] proposed a test based on the computational approach test (CAT), which is a kind of parametric bootstrap method, for the homogeneity of variances under normality. The test statistic, T_{CAT}, is given as

$$T_{CAT} = \sum_{i=1}^{k} n_i(\log S_i^2 - \log \bar{S}_i^2) \quad (6)$$

where $\bar{S}_i^2 = \sum_{i=1}^{k} n_i S_i^2/N$. This test is computed using Monte-Carlo simulations with algorithm:

1. Calculate T_{CAT} in (6) from the original data X_{ij}'s.
2. Under H_0, the restricted MLEs (RMLEs) of parameters (σ^2, μ_i; $i = 1,\dots, k$) are obtained in as (7).

$$\hat{\mu}_{i(RMLE)} = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij} = \bar{X}_i$$

and

$$\hat{\sigma}^2_{(RMLE)} = \frac{1}{\sum_{i=1}^{k} n_i}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(X_{ij} - \bar{X}_i\right)^2 = \overline{S^2} \quad (7)$$

3. Generate $X_{ij}^* \sim N\left(\hat{\mu}_{i(RMLE)}, \hat{\sigma}^2_{RMLE}\right), j = 1,\dots, n_i; i = 1,\dots, k$.
4. Compute the T^*_{CAT} in Eq. (7) from the generated data X_{ij}^*'s.
5. Repeat steps 3 and 4 many times, L, and recalculate the values of $T_{CAT}^{*(l)}$, $l = 1, \dots, L$.
6. Calculate the Monte Carlo estimates of the p-value as $\hat{p} = \sum_{l=1}^{L} I\left(T_{LR}^{*(l)} > T_{LR}\right)/L$, where I is the indicator function.
7. In the case of $\hat{p} < \alpha$, H_0 is rejected.

Likelihood Ratio Test based on CAT (LRCAT)

The likelihood ratio test for the homogeneity of variances under several normal populations is obtained as in Eq. (8).

$$T_{LR} = N\log\left(\sum_{i=1}^{k} n_i S_i^{2*}/N\right) - \sum_{i=1}^{k} n_i \log(S_i^{2*}) \quad (8)$$

where $S_i^{2*} = \frac{\sum_{j=1}^{n_i}(X_{ij}-\bar{X}_i)^2}{n_i}$. T_{LR} has approximately $\chi^2_{(k-1)}$ distribution. The parametric bootstrap approach of this test proposed by Chang et al. [15] is computed using Monte-Carlo simulations with algorithm:

1. Calculate T_{LR} in (8) from the original data X_{ij}'s.
2. Under H_0, the RMLEs of parameters (σ^2, μ_i; $i = 1,\dots, k$) are obtained as in Eq. (7).
3. Generate $X_{ij}^* \sim N\left(\hat{\mu}_{i(RMLE)}, \hat{\sigma}^2_{RMLE}\right), j = 1,\dots, n_i; i = 1,\dots, k$.
4. Compute the T^*_{LR} in Eq. (8) from the generated data X_{ij}^*'s.
5. Repeat steps 3 and 4 many times, L, and recalculate the values of $T_{LR}^{*(l)}$, $l = 1, \dots, L$.
6. Calculate the Monte Carlo estimates of the p-value as $\hat{p} = \sum_{l=1}^{L} I\left(T_{LR}^{*(l)} > T_{LR}\right)/L$, where I is the indicator function.
7. In the case of $\hat{p} < \alpha$, H_0 is rejected.

Standardized Likelihood Ratio Test (SLRT)

Gökpınar [18] proposed a standardized likelihood ratio test by using the exact mean and variance of the likelihood ratio test statistic TLR in Eq. (8). The mean and variance of T_{LR} are given as

$$E(T_{LR}) = \left\{\sum_{i=1}^{k} n_i \log n_i - N\log N\right\} - \sum_{i=1}^{k} n_i\left\{\Gamma'\left(\frac{n_{i-1}}{2}\right) - \Gamma'\left(\frac{N-k}{2}\right)\right\}$$

$$Var(T_{LR}) = \sum_{i=1}^{k} n_i^2\,\Gamma''\left(\frac{n_i - 1}{2}\right) - N^2\Gamma''\left(\frac{N-k}{2}\right)$$

where Γ'(.) and Γ''(.) are digamma and trigamma function, respectively. The standardized likelihood ratio test is obtained as follows

$$T_{SLR} = \sqrt{2(k-1)}\left(\frac{T_{LR} - E(T_{LR})}{\sqrt{Var(T_{LR})}}\right) + (k-1) \quad (9)$$

T_{SLR} has approximately $\chi^2_{(k-1)}$ distribution. For a given α, the H_0 is rejected if $T_{SLR} > \chi^2_{k-1,\alpha}$. This test is called as SLRT throughout the paper.

Some R packages such as car [19], stats [20], lawstat [21], rstatix [22], onewaytests [23] contain the tests for homogeneity of variances. Table 1 shows which tests are included in which packages. As can be seen from Table 1, there are only a few tests in the packages, although most are not new tests. Furthermore, there is no package that also contains recently proposed tests of homogeneity of variances. Onewaytests package used for one-way tests in independent groups design contains only BT, LT, and BFT. As mentioned in the introduction section, testing the homogeneity of variances is used in many fields, apart from an assumption used to apply the ANOVA-F test. For this reason, it is important to develop a package that includes many tests used in testing the homogeneity of variances. We wish to continue with the plan to build a package that contains tests not found in CRAN.

Table 1. Comparison of R packages including homogeneity of variances tests

	Tests							
R packages	**BT**	**LT**	**BFT**	**BDT**	**GPA**	**CAT**	**LRCAT**	**SLRT**
Stats	X	X	-	-	-	-	-	-
Car	-	X	-	-	-	-	-	-
Lawstat	-	X	-	-	-	-	-	-
Rstatix	-	X	-	-	-	-	-	-
Onewaytests	X	X	X	-	-	-	-	-

THE USAGE OF HOMNORMAL PACKAGE

The package including several functions provides to perform several tests for homogeneity of variances. These functions are identified by their names' initials, which are listed in the preceding sections. In this section, we use a data set to show how to use the homnormal package.

Example: As an example, FH_data is presented in the package. The data related to the survival times of patients was collected from 4 hospitals, which was a part of the data by Fleming and Harrington [29]. Our purpose is to test for the homogeneity of variances for this data.

```
# Call the homnormal package
library(homnormal)
# Call the huxtable package
library(huxtable)
# print observations of the FH_data
x1=FH_data$SurvivalTime
x1
#> [1] 105 266 227 66 24 5 155 54 58 64 15 147 42
305 92 30 82 265 237
#> [20] 208 147
x2=FH_data$HospitalNo
x2
#> [1] 1 1 1 1 2 2 2 2 2 3 3 3 4 4 4 4 4 4 4 4 4 4
```

The homnormal package includes eight tests for testing the homogeneity of variances.

Brown-Forsythe's Test (BFT): Brown_Forsythe(...)

The Brown_Forsythe function in the homnormal package is used to perform the Brown-Forsythe's test.

```
BFT = Brown_Forsythe (x1, x2, alfa = 0.05,
  table = TRUE, graph = "centerized")
```

In this code, alpha is the level of significance to assess the statistical difference. Default is set to alpha = 0.05. table is a logical value indicating whether the analysis results will be displayed in tabular form. Default is table=TRUE. Graph is a logical value for printing a boxplot of data. Default is set to graph="none". The alternative arguments of graph are "raw","centered". "raw" gives a boxplot of raw data. "centered" gives a boxplot of centered data around means of group's. When this code is run in R Programme, the following Figure 1 and Table 2 output are obtained.

Box plots provide preliminary information to understand whether the variances of groups are heterogeneous or not, but it is a visual assessment method, not a definitive test. The height of the box in a box plot is an indicator of the variability within a group. If there are significant differences in box heights between groups, this indicates that variances may be heterogeneous. Compare the "whisker" lengths of each box plot. Whiskers usually show minimum and maximum values. If the whisker lengths differ between

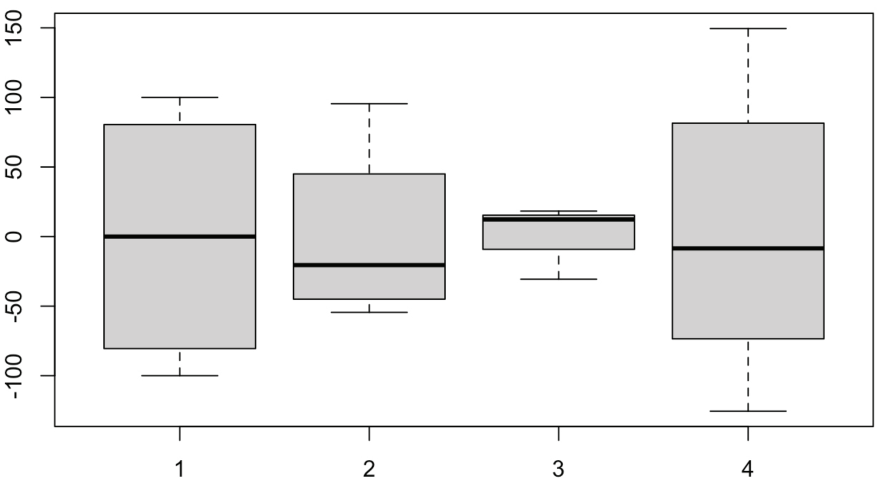

Figure 1. Boxplot based on centralized values.

Tablo 2. Brown-Forsythe's Test Result for Homogeneity of Variances

Group no	Sample size	Sample mean	Sample var	Test stat	p
1	4	166.00	9147.33		
2	4	59.50	4460.33		
3	3	45.67	714.33	2.16	0.13
4	10	155.50	9105.61		

groups, this may also indicate variance differences. In light of this information, it can be said that there is heterogeneity between the variances of the groups. But to be more precise, we need to look at the test results.

According to these results, the null hypothesis of the homogeneity of variances is not rejected (BFT=1.88, p-value=0.17) since the p-value is not smaller than 0.05. In other words, it can be said that the variances of the groups are homogenous at 5% significance level. The users who would like to use the statistics in the output of their programs can use the following codes. The BT, LT, BDT, GPA, CAT, LRCAT, and SLRT tests are also available as functions in the homnormal package. These tests are shown below.

Levene's Test (LT): levene(...)

The levene function in the homnormal package is used to perform the Levene's test.

```
LT = levene (x1, x2, alfa = 0.05,
 table = TRUE, graph = "none")
```

When this code is run in R Programme, the following Table 3 output are obtained.

Similar to BFT, according to the results in Table 3, the null hypothesis of the homogeneity of variances is not rejected (LT=2.16, p-value =0.13) since the p-value is not smaller than 0.05. In other words, according to the results of Levene's test, it can be said that the variances of the groups are homogenous at 5% significance level.

Bartlett's Test (BT): bart(...)

The bart function in the homnormal package is used to perform the Barlett's test.

```
BT = bart(x1, x2, alfa = 0.05, table=TRUE)
```

When this code is run in R Programme, the following Table 4 output are obtained.

The null hypothesis of the homogeneity of variances is not rejected (BT=3.06, p-value =0.38) since the p-value is not smaller than 0.05. According to the results of Bartlett's test, it can be said that the variances of the groups are homogenous at 5% significance level.

Bhandary-Dai's Test (BDT): bdai(...)

The bdai function in the homnormal package is used to perform the Bhandary-Dai's test.

```
BDT = bdai(x1, x2, alfa = 0.05,
     table = TRUE)
```

When this code is run in R Programme, the following Table 5 output are obtained.

The null hypothesis of the homogeneity of variances is not rejected (BDT=11.46, p-value =0.66) since the p-value is not smaller than 0.05. In other words, according to the results of Bhandary-Dai's test, it can be said that the variances of the groups are homogenous at 5% significance level.

Generalized p-value Approach (GPA): genp(...)

The genp function in the homnormal package is used to perform the Generalized p-value approach.

```
GPA = genp(x1, x2, alfa = 0.05,
  m = 5000, table = TRUE)
```

When this code is run in R Programme, the following Table 6 output are obtained.

The null hypothesis of the homogeneity of variances is not rejected (GPA=2.10, p-value =0.50) since the p-value is not smaller than 0.05. In other words, according to the

Table 3. Levene's test result for homogeneity of variances

Group no	Sample size	Sample mean	Sample var	Test stat	p
1	4	166.00	9147.33		
2	4	59.50	4460.33		
3	3	45.67	714.33	2.16	0.13
4	10	155.50	9105.61		

Table 4. Bartlett's Test Result for Homogeneity of Variances

Group no	Sample size	Sample mean	Sample var	Test stat	p
1	4	166.00	9147.33		
2	4	59.50	4460.33		
3	3	45.67	714.33	3.06	0.38
4	10	155.50	9105.61		

Table 5. Bhandary-Dai's test result for homogeneity of variances

Group no	Sample size	Sample mean	Sample var	Test stat	p
1	4	166.00	9147.33		
2	4	59.50	4460.33		
3	3	45.67	714.33	11.46	0.66
4	10	155.50	9105.61		

results of generalized p-value approach, it can be said that the variances of the groups are homogenous at 5% significance level.

Computational approach test (CAT): Cat_GG(...)

The Cat_GG function in the homnormal package is used to perform the computational approach test.

```
CAT = Cat_GG(x1, x2, alfa = 0.05,
    m = 2000, table = TRUE)
```

When this code is run in R Programme, the following Table 7 output are obtained.

The null hypothesis of the homogeneity of variances is not rejected (CAT=16.35, p-value =0.24) since the p-value is not smaller than 0.05. In other words, according to the results of likelihood ratio test based on CAT, it can be said that the variances of the groups are homogenous at 5% significance level.

Likelihood Ratio Test Based on CAT (LR-CAT): Cat_LR(...)

The Cat_LR function in the homnormal package is used to perform the likelihood ratio test based on CAT.

```
LR_CAT = Cat_LR(x1, x2, alfa = 0.05,
     m = 5000, table = TRUE)
```

Table 6. Generalized p-value aproach result for homogeneity of variances

Group no	Sample size	Sample mean	Sample var	Test stat	p
1	4	166.00	9147.33		
2	4	59.50	4460.33		
3	3	45.67	714.33	2.10	0.50
4	10	155.50	9105.61		

Table 7. Compuational aproach test result for homogeneity of variances

Group no	Sample size	Sample mean	Sample var	Test stat	p
1	4	166.00	9147.33		
2	4	59.50	4460.33		
3	3	45.67	714.33	16.35	0.23
4	10	155.50	9105.61		

Table 8. Likelihood ratio test result for homogeneity of variances

Group no	Sample size	Sample mean	Sample var	Test stat	p
1	4	166.00	9147.33		
2	4	59.50	4460.33		
3	3	45.67	714.33	5.85	0.27
4	10	155.50	9105.61		

Table 9. Standardized likelihood ratio test result for homogeneity of variances

Group no	Sample size	Sample mean	Sample var	Test stat	p
1	4	166.00	9147.33		
2	4	59.50	4460.33		
3	3	45.67	714.33	3.73	0.29
4	10	155.50	9105.61		

The null hypothesis of the homogeneity of variances is not rejected (LR_CAT=5.85, p-value =0.28) since the p-value is not smaller than 0.05 (Table 8).

Standardized Likelihood Ratio Test (SLRT): slrt(...)

The slrt function in the homnormal package is used to perform the standardized likelihood ratio test.

```
SLRT = slrt(x1, x2, alfa = 0.05,
         table = TRUE)
```

The null hypothesis of the homogeneity of variances is not rejected (SLRT=3.73, p-value=0.29) since the p-value is not smaller than 0.05.

The p-values of all tests indicate that the tests do not reject the null hypothesis of the homogeneity of variances at a nominal level 0.05. That is, each of the tests suggests that variation in survival times of patients is not significant among 4 hospitals. This study includes an example from survival analysis, where the sample sizes in each group are 4, 4, 4, and 10. These are quite small, and if there is genuine heterogeneity among the groups, it is crucial to use a test that performs well in detecting such differences. According to our simulation study, the LRCAT and SLCAT tests demonstrated excellent performance for small sample sizes. Therefore, it is particularly important to consider the results of these tests in this context (Table 9).

SIMULATION STUDY

In this section, we investigate the performance of the BT, LT, BFT, BDT, GPA, SLRT, CAT, and LRCAT tests by estimating the type I error rates and powers. To determine the type I error rates and powers of the tests, 5000 replications were performed. Additionally, the p-values of the CAT and LRTCAT methods were estimated using Monte Carlo simulations with m=5000. Sample size combinations include cases where the sample sizes are equal and cases where they are unequal. Random samples were generated from a normal distribution with parameters μ_i and σ_i^2 across different groups and sample sizes. Without loss of generality, μ_i; $i = 1,\ldots, k$ was set to 0.

For the nominal level of 0.05, the estimated type I error rates of all tests are provided in Table 10.

It is seen from Table 10, that while the LT tends to be liberal, the BFT test tends to be conservative in cases of equal and unequal small sample sizes. The estimated type I error rates of other tests are very close to the nominal level of 0.05 for all considered cases regardless of sample sizes and group numbers. For the specified nominal level of 0.05, the estimated powers of all tests were presented in Table 11-Table 18.

In addition, for certain cases in the tables, the power values of the tests are also presented in the following graphs. Accordingly, the first graph corresponds to k=3 with σ^2 = (1,3,9), the second graph to k=5 with σ^2 = (1,1,3,9,9), the third graph to k=7 with σ^2 = (1,1,3,3,3,9,9), and the final graph to k=9 with σ^2 = (1,1,1,3,3,3,9,9,9).

Since we take the combinations of sample sizes as equal and unequal sample sizes, tables are obtained for each group according to both equal and unequal sample sizes. This has been considered when interpreting the simulation results. While interpreting the test powers, the LT test was excluded because its estimated sizes exceeded 6%, as shown in Table 10.

As seen from Table 11 for k=3 and equal sample sizes, in cases of small sample sizes, the performances of the BDT, SLRT, and LRCAT tests are higher than the others. When the sample sizes increase, these tests have still the highest power, but the BT test approaches the powers of them. As seen from Table 12 for k =3 and unequal sample sizes, the differences between the performances of the tests become more distinctive. In cases of small sample sizes, the powers of the SLRT and LRCAT tests are remarkably higher than those of the others. When the sample sizes increase, the CAT test gives better results as well as these tests. It is expected that in equal and unequal sample sizes, as the difference between variance parameters increases, the power values of all the tests increase and close to each other. As seen from Table 13 for k=5 and equal sample sizes, the powers of the SLRT and LRCAT tests are higher than others in cases of small sample sizes. As sample sizes increase, the BT test as well as these tests have higher power values than others. As seen from Table 14 for k=5 and unequal sample sizes, the powers of the SLRT and LRCAT tests are remarkably higher than those of the others. As seen from Table 15- Table 18, the same pattern is valid for other k.

Chang et al. [15] compared LRTCAT with tests other than SLRT in their study. According to their simulation results, LRTCAT outperformed the other tests. Similarly, Jafari and Shaabani [30] stated in their study

Table 10. The estimated sizes of all the tests

k	n	BT	LT	BFT	BDT	GPA	SLRT	CAT	LRCAT
3	3,3,3	0.048	0.071	0.000	0.046	0.047	0.048	0.049	0.048
	5,5,5	0.045	0.085	0.003	0.051	0.053	0.052	0.052	0.055
	7,7,7	0.051	0.075	0.016	0.052	0.058	0.055	0.053	0.054
	10,10,10	0.048	0.075	0.041	0.048	0.056	0.053	0.059	0.050
	15,15,15	0.049	0.055	0.026	0.040	0.042	0.044	0.042	0.042
	20,20,20	0.052	0.064	0.037	0.052	0.052	0.057	0.051	0.056
	30,30,30	0.049	0.055	0.039	0.050	0.053	0.054	0.055	0.054
	3,5,7	0.043	0.069	0.002	0.043	0.046	0.049	0.051	0.049
	5,7,10	0.052	0.067	0.017	0.048	0.052	0.053	0.053	0.055
	10,15,20	0.045	0.060	0.034	0.048	0.051	0.053	0.050	0.053
	10,20,30	0.054	0.054	0.036	0.051	0.054	0.052	0.051	0.052
	20,30,40	0.049	0.050	0.037	0.043	0.051	0.050	0.049	0.050
5	3,3,3,3,3	0.048	0.124	0.000	0.044	0.048	0.050	0.046	0.052
	5,5,5,5,5	0.049	0.105	0.001	0.048	0.049	0.048	0.052	0.050
	7,7,7,7,7	0.048	0.080	0.010	0.052	0.047	0.046	0.047	0.045
	10,10,10,10,10	0.052	0.066	0.025	0.048	0.048	0.048	0.047	0.048
	15,15,15,15,15	0.048	0.071	0.032	0.050	0.053	0.051	0.053	0.052
	20,20,20,20,20	0.050	0.054	0.030	0.054	0.052	0.050	0.053	0.051
	30,30,30,30,30	0.056	0.058	0.040	0.045	0.049	0.049	0.052	0.048
	3,3,5,7,7	0.042	0.104	0.003	0.047	0.050	0.052	0.055	0.054
	5,5,7,10,10	0.048	0.084	0.018	0.057	0.056	0.053	0.052	0.054
	10,10,15,20,20	0.049	0.062	0.033	0.048	0.049	0.047	0.048	0.047
	10,10,20,30,30	0.051	0.061	0.040	0.055	0.060	0.056	0.061	0.058
	20,20,30,40,40	0.052	0.052	0.033	0.050	0.049	0.048	0.048	0.049
7	3,3,3,3,3,3,3	0.047	0.162	0.000	0.046	0.045	0.045	0.047	0.047
	5,5,5,5,5,5,5	0.049	0.097	0.002	0.051	0.048	0.045	0.046	0.046
	7,7,7,7,7,7,7	0.052	0.089	0.009	0.058	0.053	0.057	0.052	0.058
	10,10,10,10,10,10,10	0.050	0.066	0.021	0.046	0.042	0.044	0.041	0.044
	15,15,15,15,15,15,15	0.043	0.071	0.026	0.054	0.050	0.046	0.054	0.044
	20,20,20,20,20,20,20	0.046	0.062	0.030	0.049	0.052	0.053	0.057	0.054
	30,30,30,30,30,30,30	0.057	0.055	0.036	0.050	0.054	0.052	0.054	0.054
	3,3,5,5,5,7,7	0.046	0.108	0.002	0.046	0.050	0.050	0.056	0.051
	5,5,7,7,7,10,10	0.047	0.092	0.013	0.051	0.049	0.053	0.055	0.055
	10,10,15,15,15,20,20	0.054	0.069	0.031	0.046	0.041	0.044	0.044	0.046
	10,10,20,20,20,30,30	0.050	0.063	0.040	0.048	0.053	0.051	0.052	0.050
	20,20,30,30,30,40,40	0.051	0.058	0.039	0.051	0.049	0.052	0.051	0.052
9	3,3,3,3,3,3,3,3,3	0.044	0.207	0.000	0.046	0.052	0.049	0.052	0.050
	5,5,5,5,5,5,5,5,5	0.051	0.120	0.001	0.047	0.043	0.043	0.042	0.042
	7,7,7,7,7,7,7,7,7	0.051	0.081	0.008	0.049	0.050	0.048	0.050	0.047
	10,10,10,10,10,10,10,10,10	0.051	0.075	0.026	0.053	0.052	0.048	0.051	0.050
	15,15,15,15,15,15,15,15,15	0.054	0.067	0.024	0.050	0.053	0.050	0.053	0.052
	20,20,20,20,20,20,20,20,20	0.053	0.070	0.040	0.051	0.050	0.056	0.052	0.056
	30,30,30,30,30,30,30,30,30	0.050	0.059	0.036	0.050	0.048	0.052	0.049	0.052
	3,3,3,5,5,5,7,7,7	0.045	0.128	0.001	0.055	0.052	0.048	0.053	0.049
	5,5,5,7,7,7,10,10,10	0.050	0.091	0.010	0.054	0.050	0.049	0.050	0.050
	10,10,10,15,15,15,20,20,20	0.051	0.074	0.025	0.054	0.056	0.052	0.054	0.051
	10,10,10,20,20,20,30,30,30	0.049	0.062	0.029	0.046	0.049	0.046	0.046	0.047
	20,20,20,30,30,30,40,40,40	0.051	0.059	0.035	0.051	0.055	0.051	0.054	0.049

Table 11. Powers of all tests for k=3 with equal sample sizes

σ^2	n	BT	LT	BFT	BDT	GPA	SLRT	CAT	LRCAT
(1,2,4)		0.086	0.122	0.000	0.079	0.071	0.081	0.076	0.081
(1,3,9)	(3,3,3)	0.150	0.194	0.000	0.169	0.129	0.165	0.140	0.162
(1,4,16)		0.226	0.239	0.000	0.237	0.189	0.232	0.204	0.228
(1,2,4)		0.159	0.193	0.021	0.160	0.150	0.160	0.154	0.162
(1,3,9)	(5,5,5)	0.375	0.333	0.062	0.370	0.328	0.371	0.342	0.369
(1,4,16)		0.571	0.443	0.104	0.547	0.538	0.573	0.546	0.573
(1,2,4)		0.247	0.239	0.089	0.252	0.243	0.255	0.245	0.257
(1,3,9)	(7,7,7)	0.585	0.484	0.216	0.570	0.557	0.586	0.563	0.585
(1,4,16)		0.798	0.639	0.340	0.764	0.785	0.795	0.787	0.794
(1,2,4)		0.397	0.350	0.234	0.390	0.385	0.401	0.383	0.398
(1,3,9)	(10,10,10)	0.797	0.685	0.512	0.781	0.786	0.797	0.790	0.796
(1,4,16)		0.943	0.848	0.732	0.935	0.947	0.945	0.945	0.942
(1,2,4)		0.591	0.515	0.401	0.577	0.586	0.593	0.593	0.592
(1,3,9)	(15,15,15)	0.951	0.883	0.810	0.950	0.958	0.956	0.958	0.958
(1,4,16)		0.995	0.972	0.950	0.996	0.997	0.997	0.997	0.996
(1,2,4)		0.752	0.675	0.606	0.744	0.751	0.754	0.751	0.758
(1,3,9)	(20,20,20)	0.991	0.962	0.940	0.988	0.990	0.989	0.991	0.989
(1,4,16)		1.000	0.999	0.996	1.000	1.000	1.000	1.000	1.000
(1,2,4)		0.913	0.847	0.813	0.901	0.910	0.916	0.911	0.912
(1,3,9)	(30,30,30)	1.000	0.999	0.998	1.000	1.000	1.000	1.000	1.000
(1,4,16)		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 12. Powers of all tests for k=3 with unequal sample sizes

σ^2	n	BT	LT	BFT	BDT	GPA	SLRT	CAT	LRCAT
(1,2,4)		0.130	0.127	0.021	0.107	0.104	0.153	0.150	0.155
(1,3,9)	(3,5,7)	0.274	0.235	0.047	0.226	0.217	0.331	0.305	0.334
(1,4,16)		0.410	0.293	0.076	0.352	0.341	0.500	0.466	0.502
(1,2,4)		0.243	0.202	0.110	0.195	0.195	0.272	0.262	0.271
(1,3,9)	(5,7,10)	0.539	0.404	0.248	0.475	0.469	0.579	0.572	0.576
(1,4,16)		0.766	0.561	0.391	0.699	0.707	0.807	0.796	0.810
(1,2,4)		0.550	0.444	0.358	0.489	0.507	0.572	0.572	0.571
(1,3,9)	(10,15,20)	0.927	0.827	0.764	0.900	0.914	0.939	0.942	0.937
(1,4,16)		0.992	0.954	0.929	0.988	0.993	0.995	0.996	0.995
(1,2,4)		0.626	0.519	0.466	0.562	0.578	0.667	0.660	0.665
(1,3,9)	(10,20,30)	0.972	0.904	0.883	0.944	0.956	0.977	0.978	0.977
(1,4,16)		0.998	0.983	0.975	0.997	0.997	0.998	0.998	0.998
(1,2,4)	(20,30,40)	0.885	0.811	0.782	0.862	0.873	0.899	0.899	0.901
(1,3,9)		0.999	0.997	0.996	0.998	0.999	0.999	0.999	0.999
(1,4,16)		1.000	0.999	0.999	1.000	1.000	1.000	1.000	1.000

Table 13. Powers of all tests for k=5 with equal sample sizes

σ^2	n	BT	LT	BFT	BDT	GPA	SLRT	CAT	LRCAT
(1,1,2,4,4)		0.104	0.231	0.000	0.116	0.086	0.118	0.095	0.117
(1,1,3,9,9)	(3,3,3,3,3)	0.239	0.363	0.000	0.217	0.146	0.246	0.168	0.246
(1,1,4,16,16)		0.392	0.487	0.000	0.305	0.232	0.401	0.284	0.400
(1,1,2,4,4)		0.244	0.264	0.025	0.195	0.174	0.234	0.188	0.236
(1,1,3,9,9)	(5,5,5,5,5)	0.609	0.514	0.096	0.461	0.498	0.606	0.526	0.602
(1,1,4,16,16)		0.819	0.655	0.171	0.646	0.741	0.814	0.764	0.812
(1,1,2,4,4)		0.395	0.392	0.112	0.319	0.340	0.411	0.352	0.409
(1,1,3,9,9)	(7,7,7,7,7)	0.831	0.706	0.348	0.667	0.782	0.822	0.791	0.824
(1,1,4,16,16)		0.964	0.867	0.567	0.891	0.959	0.971	0.961	0.971
(1,1,2,4,4)		0.593	0.517	0.351	0.486	0.558	0.604	0.568	0.603
(1,1,3,9,9)	(10,10,10,10,10)	0.960	0.888	0.761	0.889	0.953	0.964	0.955	0.963
(1,1,4,16,16)		0.997	0.975	0.924	0.990	0.998	0.999	0.998	0.999
(1,1,2,4,4)		0.830	0.734	0.595	0.687	0.797	0.822	0.799	0.819
(1,1,3,9,9)	(15,15,15,15,15)	0.997	0.985	0.957	0.987	0.997	0.998	0.998	0.998
(1,1,4,16,16)		1.000	0.999	0.997	1.000	1.000	1.000	1.000	1.000
(1,1,2,4,4)		0.940	0.878	0.821	0.855	0.929	0.939	0.932	0.937
(1,1,3,9,9)	(20,20,20,20,20)	1.000	0.998	0.996	1.000	1.000	1.000	1.000	1.000
(1,1,4,16,16)		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
(1,1,2,4,4)		0.995	0.984	0.974	0.974	0.994	0.994	0.994	0.994
(1,1,3,9,9)	(30,30,30,30,30)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
(1,1,4,16,16)		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 14. Powers of all tests for k=5 with unequal sample sizes

σ^2	n	BT	LT	BFT	BDT	GPA	SLRT	CAT	LRCAT
(1,1,2,4,4)		0.168	0.161	0.022	0.112	0.110	0.215	0.179	0.218
(1,1,3,9,9)	(3,3,5,7,7)	0.391	0.300	0.051	0.220	0.254	0.484	0.398	0.488
(1,1,4,16,16)		0.613	0.419	0.099	0.314	0.427	0.706	0.611	0.709
(1,1,2,4,4)		0.343	0.274	0.136	0.208	0.258	0.392	0.360	0.394
(1,1,3,9,9)	(5,5,7,10,10)	0.759	0.580	0.381	0.485	0.659	0.815	0.782	0.816
(1,1,4,16,16)		0.941	0.760	0.551	0.725	0.907	0.972	0.959	0.971
(1,1,2,4,4)		0.762	0.615	0.512	0.533	0.691	0.783	0.780	0.784
(1,1,3,9,9)	(10,10,15,20,20)	0.995	0.967	0.939	0.953	0.991	0.995	0.995	0.995
(1,1,4,16,16)		1.000	0.997	0.992	0.998	1.000	1.000	1.000	1.000
(1,1,2,4,4)		0.822	0.708	0.645	0.582	0.761	0.861	0.857	0.862
(1,1,3,9,9)	(10,10,20,30,30)	0.999	0.988	0.979	0.971	0.999	1.000	1.000	1.000
(1,1,4,16,16)		1.000	0.999	0.999	0.999	1.000	1.000	1.000	1.000
(1,1,2,4,4)		0.987	0.961	0.954	0.931	0.985	0.990	0.990	0.991
(1,1,3,9,9)	(20,20,30,40,40)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
(1,1,4,16,16)		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 15. Powers of all tests for k=7 with equal sample sizes

σ^2	n	BT	LT	BFT	BDT	GPA	SLRT	CAT	LRCAT
a_1		0.105	0.276	0.000	0.100	0.063	0.096	0.067	0.094
a_2	(3,3,3,3,3,3,3)	0.216	0.446	0.000	0.204	0.132	0.231	0.149	0.233
a_3		0.338	0.572	0.001	0.290	0.188	0.348	0.232	0.351
a_1		0.213	0.274	0.020	0.186	0.156	0.214	0.167	0.212
a_2	(5,5,5,5,5,5,5)	0.520	0.518	0.087	0.400	0.399	0.530	0.432	0.530
a_3		0.768	0.682	0.169	0.602	0.672	0.769	0.698	0.773
a_1		0.330	0.347	0.093	0.265	0.264	0.332	0.273	0.334
a_2	(7,7,7,7,7,7,7)	0.768	0.679	0.313	0.609	0.702	0.761	0.710	0.762
a_3		0.938	0.859	0.535	0.846	0.919	0.949	0.925	0.948
a_1		0.525	0.478	0.296	0.412	0.471	0.529	0.476	0.527
a_2	(10,10,10,10,10,10,10)	0.937	0.865	0.724	0.839	0.928	0.939	0.934	0.938
a_3		0.995	0.973	0.914	0.976	0.995	0.996	0.994	0.996
a_1		0.764	0.675	0.523	0.618	0.723	0.761	0.729	0.755
a_2	(15,15,15,15,15,15,15)	0.997	0.975	0.943	0.973	0.993	0.995	0.994	0.995
a_3		1.000	0.999	0.997	1.000	1.000	1.000	1.000	1.000
a_1		0.903	0.844	0.768	0.795	0.895	0.910	0.897	0.909
a_2	(20,20,20,20,20,20,20)	1.000	0.997	0.994	0.999	1.000	1.000	1.000	1.000
a_3		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
a_1		0.990	0.961	0.942	0.948	0.984	0.986	0.983	0.987
a_2	(30,30,30,30,30,30,30)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
a_3		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

a_1 = (1,1,2,2,2,4,4); a_2 = (1,1,3,3,3,9,9); a_3 = (1,1,4,4,4,16,16)

Table 16. Powers of all tests for k=7 with unequal sample sizes

σ^2	n	BT	LT	BFT	BDT	GPA	SLRT	CAT	LRCAT
a_1		0.161	0.193	0.016	0.112	0.091	0.208	0.160	0.212
a_2	(3,3,5,5,5,7,7)	0.413	0.366	0.074	0.262	0.239	0.492	0.381	0.499
a_3		0.635	0.506	0.136	0.398	0.412	0.737	0.594	0.737
a_1		0.317	0.294	0.129	0.219	0.231	0.377	0.314	0.378
a_2	(5,5,7,7,7,10,10)	0.735	0.597	0.363	0.474	0.585	0.776	0.716	0.778
a_3		0.928	0.795	0.581	0.715	0.862	0.950	0.925	0.952
a_1		0.716	0.630	0.492	0.515	0.631	0.741	0.716	0.745
a_2	(10,10,15,15,15,20,20)	0.992	0.964	0.926	0.947	0.990	0.996	0.996	0.996
a_3		1.000	0.998	0.995	0.998	1.000	1.000	1.000	1.000
a_1		0.814	0.730	0.651	0.601	0.747	0.848	0.828	0.848
a_2	(10,10,20,20,20,30,30)	0.999	0.992	0.986	0.976	0.996	0.999	0.998	0.999
a_3		1.000	0.999	0.999	1.000	1.000	1.000	1.000	1.000
a_1		0.980	0.948	0.927	0.907	0.973	0.984	0.983	0.983
a_2	(20,20,30,30,30,40,40)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
a_3		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

a_1 = (1,1,2,2,2,4,4); a_2 = (1,1,3,3,3,9,9); a_3 = (1,1,4,4,4,16,16)

Table 17. Powers of all tests for k=9 with equal sample sizes

σ^2	n	BT	LT	BFT	BDT	GPA	SLRT	CAT	LRCAT
a_4		0.115	0.396	0.000	0.126	0.071	0.126	0.081	0.130
a_5	(3,3,3,3,3, 3,3,3,3)	0.300	0.622	0.000	0.244	0.151	0.323	0.187	0.322
a_6		0.477	0.751	0.000	0.324	0.249	0.500	0.325	0.502
a_4		0.282	0.380	0.029	0.235	0.204	0.307	0.222	0.304
a_5	(5,5,5,5,5, 5,5,5,5)	0.697	0.653	0.097	0.469	0.528	0.685	0.560	0.686
a_6		0.907	0.824	0.227	0.669	0.805	0.910	0.831	0.909
a_4		0.447	0.462	0.108	0.325	0.361	0.471	0.376	0.467
a_5	(7,7,7,7,7, 7,7,7,7)	0.904	0.824	0.418	0.695	0.847	0.903	0.856	0.902
a_6		0.991	0.951	0.675	0.901	0.984	0.991	0.985	0.992
a_4		0.687	0.625	0.404	0.482	0.624	0.692	0.633	0.692
a_5	(10,10,10,10,10,10,10,10,10)	0.987	0.964	0.865	0.913	0.983	0.988	0.985	0.987
a_6		0.999	0.998	0.979	0.993	1.000	1.000	1.000	1.000
a_4		0.900	0.815	0.673	0.718	0.876	0.894	0.877	0.895
a_5	(15,15,15,15,15,15,15,15,15)	1.000	0.998	0.993	0.994	1.000	1.000	1.000	1.000
a_6		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
a_4		0.971	0.942	0.895	0.868	0.973	0.977	0.972	0.977
a_5	(20,20,20,20,20,20,20,20,20)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
a_6		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
a_4		0.999	0.997	0.995	0.984	1.000	1.000	0.999	0.999
a_5	(30,30,30,30,30,30,30,30,30)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
a_6		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

a_4 = (1,1,1,2,2,2,4,4,4); a_5 = (1,1,1,3,3,3,9,9,9); a_6 = (1,1,1,4,4,4,16,16,16)

Table 18. Powers of all tests for k=9 with unequal sample sizes

σ^2	n	BT	LT	BFT	BDT	GPA	SLRT	CAT	LRCAT
a_4		0.212	0.240	0.018	0.112	0.109	0.247	0.188	0.253
a_5	(3,3,3,5,5,5,7,7,7)	0.534	0.471	0.082	0.264	0.304	0.648	0.492	0.652
a_6		0.772	0.621	0.161	0.372	0.517	0.858	0.729	0.860
a_4		0.411	0.393	0.172	0.224	0.287	0.487	0.410	0.485
a_5	(5,5,5,7,7,7,10,10,10)	0.886	0.743	0.488	0.504	0.732	0.904	0.852	0.905
a_6		0.983	0.908	0.740	0.753	0.959	0.993	0.987	0.993
a_4		0.859	0.773	0.647	0.568	0.798	0.885	0.868	0.886
a_5	(10,10,10,10,15,15,15,20,20,20)	1.000	0.994	0.982	0.971	0.999	1.000	0.999	1.000
a_6		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
a_4		0.922	0.867	0.804	0.634	0.878	0.948	0.938	0.949
a_5	(10,10,10,20,20,20,30,30,30)	1.000	0.999	0.998	0.988	1.000	1.000	1.000	1.000
a_6		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
a_4		0.998	0.991	0.987	0.955	0.996	0.999	0.998	0.999
a_5	(20,20,20,30,30,30,40,40,40)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
a_6		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

a_4 = (1,1,1,2,2,2,4,4,4); a_5 = (1,1,1,3,3,3,9,9,9); a_6 = (1,1,1,4,4,4,16,16,16)

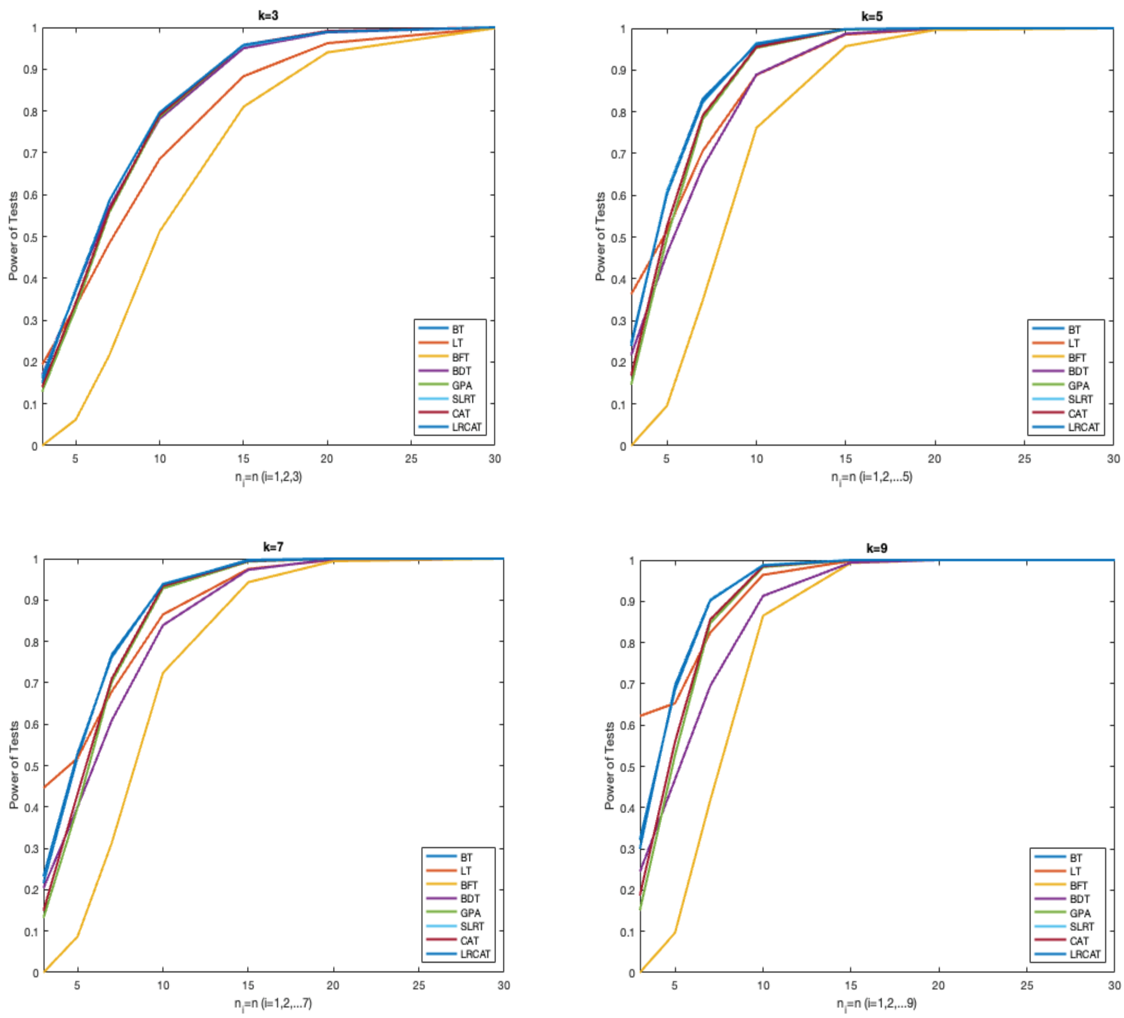

Figure 2. Power values of the tests for k=3,5,7,9 with equal sample sizes.

that SLRT is a good alternative for testing the homogeneity of variances compared to the tests they examined, excluding LRTCAT. Our literature review indicates that no study has compared SLRT and LRTCAT simultaneously. However, previous research shows that LRTCAT performed better than the methods it was compared against, and similarly, SLRT outperformed the methods it was tested against. In our simulation study, both SLRT and LRTCAT were considered alongside other tests. The results demonstrate that both tests performed better than the alternatives, confirming that our findings align with those of previous studies.

The widely used BT test, despite its popularity, exhibited notable limitations under conditions of small and unequal sample sizes, confirming its sensitivity to such scenarios as highlighted in [13,15,18]. This simulation study highlights the importance of alternative approaches such as SLRT and LRCAT in this context. These tests have been observed to give much better results than other tests, especially at small sample sizes. Since data collection is challenging due to various constraints, especially in real-life applications, the practical implications of these findings are getting more significant. For instance, in survival analysis, sample sizes are often limited by the availability of specimens, patient recruitment difficulties, or high costs associated with longitudinal studies. Similarly, in environmental studies, obtaining measurements from remote or hazardous locations can restrict sample sizes. In clinical trials, ethical considerations and the scarcity of rare disease patients often result in small sample sizes. In genetics, experiments on rare genotypes or high-throughput sequencing can also face limitations due to the cost and complexity of data collection. These scenarios underscore the importance of using tests like the SLRT and LRCAT, which perform robustly even under small and unequal sample sizes, as supported by our findings and consistent with prior research.

CONCLUSION

In this study, we introduced the homnormal package, which provides a comprehensive set of tests for assessing the homogeneity of variances under normality assumptions. Unlike existing R packages, homnormal offers a broader selection of homogeneity of variances tests, addressing an essential gap in statistical software.

Through an extensive simulation study, we evaluated these tests based on their Type I error rates and statistical power across different scenarios involving varying group numbers and sample sizes. Results indicate that the standardized likelihood ratio test (SLRT) and likelihood ratio test based on computational approach (LRCAT) outperform other tests, particularly in cases of unequal sample sizes. These findings are consistent with previous research and suggest that SLRT and LRCAT are good alternative tests for testing the homogeneity of variances. Despite its widespread use, Bartlett's test (BT) has shown significant limitations, especially under small and unequal sample sizes, confirming its sensitivity in such conditions. This highlights the need for tests that also perform well under these conditions. The SLRT and LRCAT tests also show to be good alternative methods under these conditions. Given the difficulties in data collection, especially in clinical trials, environmental studies, genetic and survival analysis, it is important to use tests that perform well in small sample sizes.

While this study focuses on the homogeneity of variances within a one-way ANOVA framework, the concept has broader applications. For this reason, this package is a limited study addressing the topic of homogeneity of variances. Future work could extend the homnormal package to address the homogeneity of variances in other statistical models, such as classical linear regression, where tests for homogeneity of variances (e.g., White, Breusch-Pagan, Goldfeld-Quandt) are widely used. Expanding the package to incorporate such tests would enhance its applicability across various fields.

AUTHOR'S CONTRIBUTIONS

G.F: Conceptualization, Methodology, Data Curation, Writing - Original Draft, Writing - Review & Editing. G.E: Conceptualization, Methodology, Data Curation, Writing - Original Draft, Writing - Review & Editing. All authors discussed the results and contributed to the final manuscript.

ACKNOWLEDGEMENTS

The authors thank the editor, the associate editor, and the referees for their constructive suggestions that helped us improve the early manuscript.

REFERENCES

[1] Buyukada M, Evrendilek F. Color and cod removals by photocatalytic degradation: an experimental design approach and cost analysis. Sigma J Eng Nat Sci 2017;8:217–226.
[2] Erdogdu S, Nayır S, Kandil U, Kurbetci SK, Nas M. Evaluation of the dependency of the compressive strength of concrete on the core drilling direction through Anova test. Sigma J Eng Nat Sci 2020;38:1879–1895.
[3] Boos DD, Brownie C. Comparing variances and other measures of dispersion. Stat Sci 2004;19:571–578. [CrossRef]
[4] Riffenburgh RH, Gillen DL. Statistics in medicine. Cambridge:Academic Press; 2020.
[5] Cupal M. Sales comparison approach indicating heterogeneity of particular type of real estate and corresponding valuation accuracy. Acta Univ Agric Silvic Mendel Brun 2017;65:977–985. [CrossRef]
[6] Puntambekar S, Gnesdilow D, Yavuz S. Understanding the effect of differences in prior knowledge on middle school students' collaborative interactions and learning. Int J Comput Support Collab Learn 2023;18:531–573. [CrossRef]
[7] Bartlett MS. Properties of sufficiency and statistical tests. Proc R Soc Lond A Math Phys Sci 1937;160:268-282. [CrossRef]
[8] Bishop DJ, Nair US. A note on certain methods of testing for the homogeneity of a set of estimated variances. J R Stat Soc 1939;6:89–99. [CrossRef]
[9] Hartley HO. The maximum F-ratio as a short cut test for heterogeneity of variances. Biometrika 1950;37:308–312. [CrossRef]
[10] Levene H. Contributions to probability and statistics: essay in honor of Harold Hotellings. In: Olkin L (editor). California, USA: Stanford University Press; 1960. p. 178–292.
[11] Brown MB, Forsythe AB. Robust tests for the equality of variances. J Am Stat Assoc 1974;69:364–367. [CrossRef]
[12] Bhandary M, Dai H. An alternative test for the equality of variances for several populations when the underlying distributions are normal. Commun Stat Simul Comput 2009;38:109–117. [CrossRef]
[13] Liu X, Xu X. A new generalized p-value approach for testing the homogeneity of variances. Stat Probab Lett 2010;80:1486–1491. [CrossRef]
[14] Gokpinar E, Gokpinar F. Testing equality of variances for several normal populations. Commun Stat Simul Comput 2017;46:38–52. [CrossRef]
[15] Chang CH, Pal N, Lin JJ. A revisit to test the equality of variances of several populations. Commun Stat Simul Comput 2017;46:6360–6384. [CrossRef]
[16] Keyes TK, Levy MS. Analysis of Levene's test under design imbalance. J Educ Behav Stat 1997;22:227–236. [CrossRef]

[17] Loh WY. Some modifications of Levene's test of variance homogeneity. J Stat Comput Simul 1987;28:213–226. [CrossRef]
[18] Gokpinar E. Standardized likelihood ratio test for homogeneity of variance of several normal populations. Commun Stat Simul Comput 2020;51:1–11. [CrossRef]
[19] Fox J, Weisberg S. An R Companion to Applied Regression. 2nd ed. Thousand Oaks: Sage; 2011
[20] R Core Team. R: A Language and Environment for Statistical Computing. R Project 2017.
[21] Gastwirth JL, Gel YR, Hui WW, Lyubchich V, Miao W, Noguchi K. lawstat: Tools for Biostatistics, Public Policy and Law R package version 3.1. R Project 2017.
[22] Kassambara A. rstatix: Pipe-Friendly Framework for Basic Statistical Tests. R Project 2021.
[23] Dag O, Dolgun A, Konar NM. onewaytests: an R package for one-way tests in independent groups designs. R J 2018;10:175–199. [CrossRef]
[24] White H. A heteroscedasticity -consistent covariance matrix and a direct test for heteroscedasticity. Econometrica 1980;48:817–838. [CrossRef]
[25] Breusch TS, Pagan AR. A simple test for heteroscedasticity and random coefficient variation. Econometrica 1979;48:1287–1294. [CrossRef]
[26] Goldfeld SM, Quandt RE. Some tests for homoscedasticity. J Am Stat Assoc 1965;60:539–547. [CrossRef]
[27] Celik R. A new test to detect monotonic and non-monotonic types of heteroscedasticity. J Appl Stat 2017;44:342–361. [CrossRef]
[28] Celik R. RCEV heteroscedasticity test based on the studentized residuals. Comm Stat Theory Methods 2019;48:3258–3268. [CrossRef]
[29] Fleming TR, Harrington DP. Counting processes and survival analysis. New York: John Wiley and Sons; 2011.
[30] Jafari AA, Shaabani J. Comparing scale parameters in several gamma distributions with known shapes. Comput Stat 2020;35:1927–1950. [CrossRef]