Research Article GU J Sci 38(3): 1503-1516 (2025) DOI: 10.35378/gujs.1604187

Gazi University

Journal of Science

http://dergipark.org.tr/gujs

Score Test for Homogeneity of Variances in Normal Distributions

Sevgi AKSOY[1], Fikri GOKPINAR[2], Esra GOKPINAR[2, *]

[1] Ministry of National Education, Bekir Gökdağ Anatolian High School, 06500, Ankara, Türkiye
[2] Department of Statistics, Gazi University, 06500, Ankara, Türkiye

Highlights

- This paper focuses on testing the homogeneity of variances in normal distributions.
- A Score test statistic is proposed to address the homogeneity of variances in the study.
- The proposed test demonstrates superior performance compared to alternative methods.

Article Info

Received: 20 Dec 2024
Accepted: 12 May 2025

Keywords

Homogeneity of variances,
Computational approach test
Score test

Abstract

In this study, we suggest a novel test statistic based on the Score statistic for evaluating the homogeneity of variances in normal distributions. In addition to the conventional chi-square approximation of the Score statistic, we introduce a parametric bootstrap technique known as the Computational Approach Test (CAT). Through a simulation study, we evaluate the proposed test's CAT approach (referred to as CS) and assess its performance against established methods under varying group sizes and sample sizes. The results show that, regardless of the number of groups, the CAT approach of the Score test performs well when sample sizes and variances are directly proportional, even with a minimum sample size of three. Furthermore, when sample sizes and variances are inversely proportional, the proposed test significantly outperforms alternative methods. To demonstrate the application of the discussed methods, we provide two numerical examples.

1. INTRODUCTION

The analysis of variance (ANOVA) technique is widely used in many scientific applications, including health, social, and physical sciences. In the analysis of variance, one of the main assumptions is that the variances within treatments or populations are equal, known as homogeneity of variances. It is well-recognized that when this assumption is violated, the ANOVA F test performs poorly. Testing for homogeneity of variances is a critical issue not only in ANOVA but also in other fields.

Determining homogeneity is crucial in biology, medical research, agricultural production systems, manufacturing process quality control, and the creation of instructional strategies [1]. For instance, in medical studies, the variability of DNA methylation is an important biological marker associated with cancer and other complex diseases. Research has shown that, in addition to the differences in mean methylation levels between diseased and healthy individuals, variance differences among groups can provide meaningful biological insights [2]. In this context, the variance homogeneity test is a vital tool for understanding the distribution of DNA methylation markers across different groups and making accurate biological interpretations. Evaluating the homogeneity of variances is frequently a useful endpoint of analysis in quality control work [3]. For biologists, differences in population variability are significant for several reasons, such as studying adaptation mechanisms and measuring genetic diversity [1]. Furthermore, homogeneity of variance testing is often used as a precursor to discriminant analysis or dose-response modeling [4].

The literature provides a variety of tests for determining whether variances are homogeneous. One of the earliest methods was the Bartlett test, which is based on the Likelihood ratio (LR) test. This test is widely

*Corresponding author, e-mail: eyigit@gazi.edu.tr applied and commonly included in statistical software packages [5]. However, several computer simulations indicate that the Bartlett test is valid only when the number of groups is not excessively large and the sample size is moderate to large [1,4,6-8].

Over the years, numerous alternative tests for homogeneity of variance have been developed by researchers such as Cochran [9], Box [7], Levene [10], Brown and Forsythe [8], Conover et al. [3], Loh [11], Keyes and Levy [12], Bhandary and Dai [13], Liu and Xu [14], Gökpınar and Gökpınar [15], Jafari and Shaabani [16], Wang et al. [17].

In hypothesis testing problems, most statisticians initially consider using the LR test. Like the LR test, the Score (S) test is a widely used method in hypothesis testing. As noted by Bera and Bilias [18], the S test is an important technique for evaluating statistical models alongside the LR test. Based on our literature search, we found that the S test has not been addressed for this problem. Therefore, our goal is to develop an S test for addressing the variance homogeneity problem. Under the null hypothesis, the S test, like the LR test, follows an approximate chi-squared distribution with *k*-1 degrees of freedom.

It is well-established that tests relying on asymptotic distributions often perform poorly with small sample sizes, particularly concerning test size accuracy. According to Davison and Hinkley [19], likelihood-based tests such as the LR and S tests are particularly well-suited for parametric bootstrap techniques. In this study, we propose a parametric bootstrap approach for the S test, referred to as CAT (proposed by Chang et al. [20]).

The remaining sections of the manuscript are organized as follows: In section 2, we derive the S test statistic for assessing the homogeneity of variances. Additionally, we propose the CAT approach for the S test. section 3 details the simulation studies conducted to evaluate the sizes and powers of the proposed test under various scenarios. In section 4, we report the findings from the analysis of numerical examples. Finally, section 5 concludes the manuscript with some final remarks.

2. MATERIAL METHOD

Let $X_{i1}, X_{i2}, \ldots, X_{in_i}$, be a sample from normal distribution with parameters μ_i and $\sigma_i^2, i = 1, \ldots, k$. The problem of interest is to test the homogeneity of variances; that is, to test

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

$$H_1: \sigma_i^2 \neq \sigma_j^2,\ \exists i \neq j. \tag{1}$$

Note that $\hat{\mu}_i = \bar{X}_i$ and $\hat{\sigma}_i^2 = \frac{1}{n_i}\sum_{j=1}^{n_i}\left(X_{ij} - \bar{X}_i\right)^2 = S_i^2$ are the maximum likelihood estimates (MLEs) of μ_i and σ_i^2, respectively.

Under the null hypothesis H_0, let $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2 = \sigma^2$, and note that the log-likelihood function can be expressed as

$$lnL_0(\mu_1, \ldots \mu_k, \sigma^2) = -\sum_{i=1}^{k}\left(\frac{n_i}{2}\right) ln(2\pi) - \sum_{i=1}^{k}\left(\frac{n_i}{2}\right) ln(\sigma^2) - \frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\frac{\left(x_{ij}-\mu_i\right)^2}{\sigma^2}. \tag{2}$$

The restricted maximum likelihood estimates (RMLEs) of parameters μ_i and σ^2 are denoted as $\tilde{\mu}_i$ and $\tilde{\sigma}^2$, and are obtained as follows:

$$\tilde{\mu}_i = \bar{X}_i \quad \text{and} \quad \tilde{\sigma}^2 = \frac{1}{\sum_{i=1}^{k} n_i}\sum_{i=1}^{k} n_i S_i^2. \tag{3}$$

In the remainder of this section, we address likelihood-based methods for testing homogeneity of variances. To this end, we first give the LRT test, then we derive the S test for testing homogeneity of variances.

2.1. Likelihood Ratio (LR) Test

The general form of the likelihood ratio statistic is defined as follows:

$$LR = -2\left(lnL(\tilde{\mu}_1, \tilde{\mu}_2, \dots, \tilde{\mu}_k, \tilde{\sigma}^2) - lnL\left(\hat{\mu}_1, \hat{\mu}_2, \dots, \hat{\mu}_k, \hat{\sigma}_1^2, \hat{\sigma}_2^2, \dots, \hat{\sigma}_k^2\right)\right).$$

Then, the following result is obtained

$$LR = -2\left[-\sum_{i=1}^{k} \left(\frac{n_i}{2} ln\left(\frac{\tilde{\sigma}^2}{\hat{\sigma}_i^2}\right)\right)\right] = \sum_{i=1}^{k} \left[n_i \left(ln(\tilde{\sigma}^2) - ln\left(\hat{\sigma}_i^2\right)\right)\right]. \tag{4}$$

Under the null hypothesis, the LR test follows an approximate chi-square distribution with $k - 1$ degree of freedom. Furthermore, in their study, Chang et al. [20] obtained the LR test statistic as given in Equation (4).

2.2. The Proposed Score (S) Test

We propose the S test to test the homogeneity of variances of k normally distributed groups, the general form of the S test statistic is defined as follows:

$$S = U'_{\underline{\sigma}^2}\left(\underline{\tilde{\mu}}, \underline{\tilde{\sigma}}^2\right)\left(I^{\underline{\sigma}^2\underline{\sigma}^2}\left(\underline{\tilde{\mu}}, \underline{\tilde{\sigma}}^2\right)\right) U_{\underline{\sigma}^2}\left(\underline{\tilde{\mu}}, \underline{\tilde{\sigma}}^2\right). \tag{5}$$

Here, under the unrestricted model, $U'_{\underline{\sigma}^2}\left(\underline{\tilde{\mu}}, \underline{\tilde{\sigma}}^2\right)$ is the Score vector, and $I^{\underline{\sigma}^2\underline{\sigma}^2}\left(\underline{\tilde{\mu}}, \underline{\tilde{\sigma}}^2\right)$ denotes the lower right block matrix of the inverse of the information matrix (denoted as I). The RMLEs of parameters $\underline{\mu} = (\mu_1, \dots, \mu_k)$ and $\underline{\sigma}^2 = \left(\sigma_1^2, \dots, \sigma_k^2\right) = (\sigma^2, \dots, \sigma^2)$ are denoted as $\underline{\tilde{\mu}}$ and $\underline{\tilde{\sigma}}^2$.

The score vector of parameters $\underline{\sigma}^2$ are obtained as

$$U'_{\underline{\sigma}^2}\left(\underline{\tilde{\mu}}, \underline{\tilde{\sigma}}^2\right) = \left(\frac{n_1\left(-\tilde{\sigma}^2 + \hat{\sigma}_1^2\right)}{2\tilde{\sigma}^4}, \dots, \frac{n_k\left(-\tilde{\sigma}^2 + \hat{\sigma}_k^2\right)}{2\tilde{\sigma}^4}\right). \tag{6}$$

I^{-1} term is calculated using the equation shown below:

$$I^{-1} = \begin{bmatrix} I^{\underline{\mu}\underline{\mu}}\left(\underline{\mu}, \underline{\sigma}^2\right) & I^{\underline{\mu}\underline{\sigma}^2}\left(\underline{\mu}, \underline{\sigma}^2\right) \\ I^{\underline{\sigma}^2\underline{\mu}}\left(\underline{\mu}, \underline{\sigma}^2\right) & I^{\underline{\sigma}^2\underline{\sigma}^2}\left(\underline{\mu}, \underline{\sigma}^2\right) \end{bmatrix}_{2kx2k}.$$

Then, $I^{\underline{\sigma}^2\underline{\sigma}^2}\left(\underline{\tilde{\mu}}, \underline{\tilde{\sigma}}^2\right)$ term is obtained as

$$I^{\underline{\sigma}^2\underline{\sigma}^2}\left(\underline{\mu}, \underline{\sigma}^2\right) = \left(I_{\underline{\sigma}^2\underline{\sigma}^2} - I_{\underline{\sigma}^2\underline{\mu}} I_{\underline{\mu}\underline{\mu}}^{-1} I_{\underline{\mu}\underline{\sigma}^2}\right)^{-1}$$

$$I^{\underline{\sigma}^2\underline{\sigma}^2}\left(\underline{\tilde{\mu}},\underline{\tilde{\sigma}}^2\right)=\begin{bmatrix}\frac{2\tilde{\sigma}^4}{n_1} & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \frac{2\tilde{\sigma}^4}{n_k}\end{bmatrix}_{kxk} \tag{7}$$

By applying Equations (6) and (7), the S test is derived as shown:

$$S=\frac{1}{2}\sum_{i=1}^{k} n_i\left(\frac{\hat{\sigma}_i^2}{\tilde{\sigma}^2}-1\right)^2. \tag{8}$$

Under the null hypothesis, the S test follows an approximate chi-square distribution with k−1 degrees of freedom.

Remark. As mentioned in Introduction section, it is well-known that LR and S tests are asymptotically equivalent. For this problem, it is also possible to show that these tests are equivalent as follows.
The LR statistic given in Equation (4) can also be expressed as $LR=-\sum_{i=1}^{k} n_i ln\left(\frac{\hat{\sigma}_i^2}{\tilde{\sigma}^2}\right)$. For any variable, say *y*, *ln y* can be expressed using the Taylor series expansion as follows:

$$\ln(y)=\ln\big(1+(y-1)\big)\cong(y-1)-\frac{1}{2}(y-1)^2.$$

If $y=\frac{\hat{\sigma}_i^2}{\tilde{\sigma}^2}$ is taken, it can be observed that the LR test statistic is equivalent to the S test statistic.

$$\ln\left(\frac{\hat{\sigma}_i^2}{\tilde{\sigma}^2}\right)\cong\left(\frac{\hat{\sigma}_i^2}{\tilde{\sigma}^2}-1\right)-\frac{1}{2}\left(\frac{\hat{\sigma}_i^2}{\tilde{\sigma}^2}-1\right)^2$$

$$-\sum_{i=1}^{k} n_i ln\left(\frac{\hat{\sigma}_i^2}{\tilde{\sigma}^2}\right)\cong-\sum_{i=1}^{k} n_i\left[\left(\frac{\hat{\sigma}_i^2}{\tilde{\sigma}^2}-1\right)+\frac{1}{2}\left(\frac{\hat{\sigma}_i^2}{\tilde{\sigma}^2}-1\right)^2\right]$$

$$LR\cong\frac{1}{2}\sum_{i=1}^{k} n_i\left(\frac{\hat{\sigma}_i^2}{\tilde{\sigma}^2}-1\right)^2$$

$$LR\cong S.$$

2.4. Computational Approach Test Approach

The S test asymptotically follows a chi-squared distribution with k-1 degrees of freedom under the H_0 hypothesis. As mentioned in the Introduction, it is known that the convergence of this test to this distribution is not sufficient for small sample sizes. To overcome this problem, we propose to use the CAT approach, which is a kind of parametric bootstrap approach, instead of the chi-square distribution for the *S* test.
The algorithm of this method is as follows:

(1) Calculate the *S* test statistics in Equation (8).
(2) Draw a pseudo random sample with size n_i from the $N(\tilde{\mu}_i,\tilde{\sigma}^2)$ for $i=1,\dots,k$.
(3) Compute the value of *S* test statistic for these generated samples.
(4) Repeat steps 2 and 3 for many times (for instance, L times). The calculated *S* test statistic is denoted by $\mathrm{S}^{(l)}$, l=1,…,L for each of these generated samples.
(5) Calculate the p-value as $\hat{p}=\sum_{l=1}^{L} I\,(\mathrm{S}^{(l)}>\mathrm{S})/L$, l=1,…,L and here I(.) express the indicator function.
(6) Reject the H_0 in Equation (1) if the $\hat{p}\leq\alpha$.

Since the LR test also asymptotically follows a chi-square distribution, Chang et al. [20] obtained the CAT approach for this test using the algorithm provided above. Therefore, we will not repeat the derivation of the CAT approach for the LR test here. In the simulation study, the CAT approximations of the S and LR tests are considered and shown as CS and CLR, respectively.

3. SIMULATION STUDY

In the simulation study, our main objective is to compare the performance of the proposed test with existing tests. The tests evaluated in this study include the CLR test, Levene (L) test [12], Brown-Forsythe (BF) test [8], generalized p-value test (GP) [14], Bhandary and Dai's test (DAI) [13], Bartlett test (B) [5], computational approach test (CAT) [15], and standardized likelihood ratio (SLR) test [21]. All tests were compared across various group sizes and sample sizes, including combinations of equal and unequal sample sizes.

For sample sizes n_i (i=1,...,k, where k is the number of groups), 10,000 random numbers were generated from a normal distribution with parameter μ=0. Additionally, Monte Carlo simulations with L=10000 iterations were used to calculate the p-values for each test. At a significance level of α=0.05, the sizes of all tests are presented in Tables 1–3 for k=3,5,7, respectively.

Table 1. *Sizes of all tests when k=3*

n	CLR	CS	GP	DAI	B	L	BF	SLR	CAT
3 3 3	0.046	0.048	0.050	0.047	0.045	0.012	0.000	0.049	0.050
5 5 5	0.051	0.050	0.047	0.046	0.047	0.048	0.001	0.048	0.047
7 7 7	0.050	0.049	0.050	0.047	0.049	0.046	0.006	0.050	0.050
9 9 9	0.050	0.049	0.050	0.046	0.048	0.050	0.011	0.048	0.050
15 15 15	0.052	0.050	0.049	0.048	0.050	0.049	0.023	0.050	0.049
30 30 30	0.051	0.052	0.048	0.045	0.049	0.049	0.037	0.049	0.048
3 5 7	0.050	0.050	0.048	0.044	0.048	0.045	0.001	0.049	0.049
3 8 13	0.053	0.056	0.048	0.044	0.045	0.044	0.012	0.048	0.049
10 15 20	0.051	0.054	0.050	0.044	0.050	0.046	0.023	0.049	0.050

Table 2. *Sizes of all tests when k=5*

n	CLR	CS	GP	DAI	B	L	BF	SLR	CAT
3 3 3 3 3	0.054	0.052	0.051	0.049	0.045	0.026	0.000	0.049	0.051
5 5 5 5 5	0.049	0.050	0.047	0.048	0.047	0.051	0.000	0.048	0.047
7 7 7 7 7	0.051	0.053	0.049	0.048	0.051	0.051	0.004	0.051	0.050
9 9 9 9 9	0.052	0.052	0.047	0.047	0.046	0.047	0.008	0.046	0.048
15 15 15 15	0.051	0.051	0.046	0.046	0.048	0.048	0.016	0.048	0.046
30 30 30 30 30	0.051	0.052	0.046	0.045	0.047	0.048	0.032	0.047	0.046
3 3 5 7 7	0.050	0.052	0.055	0.051	0.051	0.050	0.001	0.055	0.055
3 3 8 13 13	0.051	0.051	0.049	0.048	0.046	0.045	0.007	0.047	0.049
10 10 15 20 20	0.048	0.049	0.050	0.051	0.051	0.049	0.025	0.052	0.052

Table 3. *Sizes of all tests when k=7*

n	CLR	CS	GP	DAI	B	L	BF	SLR	CAT
3 ,…, 3	0.052	0.051	0.048	0.047	0.043	0.030	0.000	0.047	0.050
5 ,…,5	0.051	0.052	0.050	0.051	0.050	0.053	0.001	0.051	0.050
7 ,…,7	0.049	0.052	0.049	0.050	0.049	0.051	0.003	0.049	0.050
9 ,…,9	0.057	0.052	0.050	0.052	0.050	0.050	0.006	0.051	0.050
15 ,…,15	0.051	0.049	0.054	0.051	0.054	0.048	0.018	0.054	0.054
30 ,…,30	0.051	0.051	0.050	0.050	0.050	0.049	0.031	0.050	0.050
3 3 5 5 5 7 7	0.054	0.053	0.050	0.050	0.048	0.053	0.000	0.049	0.048
3 3 8 8 8 13 13	0.053	0.052	0.054	0.052	0.051	0.051	0.008	0.055	0.054
10 10 15 15 15 20 20	0.052	0.053	0.053	0.050	0.052	0.051	0.020	0.052	0.053

When Tables 1-3 are reviewed, it is seen that the sizes of all tests are very close to the nominal value. To understand which test is better in which situation, we need to look at the power values of the tests. In calculating the power values of the tests, we took different values of the population variance. For this, in addition to the case of small and large equal sample sizes, we examined the situations in which sample sizes are inversely proportional to population variances and in which population variances are directly proportional to sample sizes. The calculated powers of tests under α= 0.05 were displayed in Tables 4–6 for k = 3, 5, 7, respectively.

Table 4*. Powers of all tests when k=3*

n	CLR	CS	GP	DAI	B	L	BF	SLR	CAT
	σ^2= (0.25, 0.5, 1)								
3 3 3	0.089	0.099	0.082	0.095	0.087	0.033	0.000	0.095	0.082
5 5 5	0.162	0.175	0.150	0.165	0.167	0.130	0.010	0.169	0.152
9 9 9	0.356	0.347	0.329	0.338	0.346	0.253	0.110	0.347	0.329
15 15 15	0.605	0.585	0.586	0.582	0.598	0.475	0.356	0.598	0.588
30 30 30	0.916	0.907	0.919	0.910	0.919	0.847	0.812	0.919	0.918
3 5 7	0.167	0.085	0.131	0.108	0.130	0.087	0.009	0.163	0.160
3 8 13	0.208	0.110	0.168	0.134	0.170	0.123	0.056	0.210	0.182
10 15 20	0.572	0.458	0.538	0.493	0.542	0.409	0.323	0.575	0.602
	σ^2=(0.25, 0.75, 1.25)								
3 3 3	0.104	0.111	0.090	0.095	0.091	0.034	0.000	0.099	0.091
5 5 5	0.210	0.191	0.194	0.190	0.200	0.143	0.009	0.203	0.194
9 9 9	0.461	0.367	0.450	0.422	0.443	0.306	0.134	0.444	0.449
15 15 15	0.739	0.666	0.762	0.726	0.747	0.590	0.456	0.747	0.762
30 30 30	0.980	0.970	0.983	0.977	0.980	0.937	0.919	0.980	0.982
3 5 7	0.167	0.075	0.140	0.106	0.137	0.090	0.009	0.179	0.185
3 8 13	0.212	0.083	0.169	0.113	0.162	0.106	0.046	0.217	0.210
10 15 20	0.660	0.419	0.641	0.561	0.624	0.455	0.350	0.659	0.708
	σ^2=(1, 0.5, 0.25)								
3 5 7	0.102	0.239	0.105	0.183	0.152	0.146	0.001	0.100	0.036
3 8 13	0.133	0.343	0.156	0.271	0.222	0.236	0.041	0.128	0.028
10 15 20	0.542	0.615	0.567	0.592	0.587	0.489	0.367	0.550	0.480
	σ^2=(1.25, 0.75, 0.25)								
3 5 7	0.144	0.286	0.180	0.263	0.225	0.198	0.004	0.151	0.057
3 8 13	0.234	0.487	0.313	0.427	0.371	0.340	0.086	0.231	0.052
10 15 20	0.735	0.761	0.768	0.769	0.766	0.637	0.503	0.734	0.691

Table 5. *Powers of all tests when k=5*

n	CLR	CS	GP	DAI	B	L	BF	SLR	CAT
	σ^2= (0.25, 0.25, 0.5,1, 1)								
3 3 3 3 3	0.116	0.144	0.083	0.114	0.104	0.068	0.000	0.113	0.083
5 5 5 5 5	0.250	0.248	0.188	0.212	0.244	0.188	0.010	0.248	0.188
9 9 9 9 9	0.535	0.514	0.488	0.429	0.539	0.394	0.166	0.539	0.488
15 15 15 15 15	0.823	0.803	0.814	0.703	0.832	0.705	0.561	0.832	0.813
30 30 30 30 30	0.993	0.992	0.992	0.975	0.993	0.975	0.964	0.993	0.992
3 3 5 7 7	0.221	0.099	0.142	0.114	0.161	0.110	0.010	0.220	0.195
3 3 8 13 13	0.283	0.098	0.192	0.130	0.215	0.148	0.055	0.284	0.226
10 10 15 20 20	0.789	0.624	0.732	0.534	0.748	0.584	0.472	0.787	0.799
	σ^2= (0.25, 0.25, 0.75,1.25, 1.25)								
3 3 3 3 3	0.142	0.154	0.097	0.136	0.133	0.086	0.000	0.145	0.097
5 5 5 5 5	0.317	0.280	0.255	0.245	0.313	0.218	0.011	0.318	0.254
9 9 9 9 9	0.676	0.592	0.645	0.518	0.676	0.486	0.222	0.676	0.646
15 15 15 15 15	0.934	0.899	0.934	0.835	0.934	0.824	0.696	0.935	0.930
30 30 30 30 30	1.000	0.999	1.000	0.997	1.000	0.997	0.995	1.000	1.000
3 3 5 7 7	0.254	0.100	0.254	0.118	0.185	0.117	0.008	0.258	0.238
3 3 8 13 13	0.305	0.087	0.305	0.123	0.219	0.138	0.046	0.315	0.278
10 10 15 20 20	0.891	0.675	0.891	0.670	0.866	0.699	0.580	0.893	0.910
	σ^2= (1, 1, 0.5, 0.25, 0.25)								
3 3 5 7 7	0.144	0.362	0.123	0.262	0.228	0.249	0.001	0.143	0.035
3 3 8 13 13	0.172	0.478	0.163	0.362	0.320	0.369	0.039	0.173	0.024
10 10 15 20 20	0.764	0.836	0.750	0.707	0.787	0.695	0.557	0.752	0.652
	σ^2= (1.25, 1.25, 0.75,0.25, 0.25)								
3 3 5 7 7	0.212	0.450	0.186	0.329	0.317	0.316	0.003	0.210	0.045
3 3 8 13 13	0.318	0.640	0.333	0.497	0.492	0.501	0.095	0.313	0.035
10 10 15 20 20	0.909	0.936	0.920	0.860	0.932	0.849	0.731	0.912	0.864

Table 6. *Powers of all tests when k=7*

n	CLR	CS	GP	DAI	B	L	BF	SLR	CAT
	σ^2= (0.25, 0.25, 0.5, 0.5, 0.5, 1, 1)								
3 3 3 3 3 3 3	0.110	0.141	0.073	0.116	0.096	0.080	0.000	0.107	0.074
5 5 5 5 5 5 5	0.217	0.241	0.144	0.181	0.207	0.168	0.005	0.209	0.145
9 9 9 9 9 9 9	0.464	0.453	0.409	0.360	0.477	0.352	0.124	0.478	0.409
15 15 15 15 15 15 15	0.769	0.746	0.734	0.631	0.763	0.637	0.462	0.763	0.733
30 30 30 30 30 30 30	0.987	0.982	0.985	0.950	0.986	0.961	0.942	0.986	0.984
3 3 5 5 5 7 7	0.214	0.136	0.126	0.125	0.164	0.122	0.007	0.214	0.170
3 3 8 8 8 13 13	0.309	0.169	0.198	0.163	0.244	0.178	0.064	0.313	0.229
10 10 15 15 15 20 20	0.753	0.632	0.679	0.527	0.716	0.571	0.448	0.754	0.750
	σ^2= (0.25, 0.25, 0.75, 0.75, 0.75, 1.25, 1.25)								
3 3 3 3 3 3 3	0.126	0.141	0.087	0.112	0.116	0.092	0.000	0.126	0.087
5 5 5 5 5 5 5	0.270	0.240	0.211	0.207	0.270	0.198	0.006	0.274	0.211
9 9 9 9 9 9 9	0.603	0.477	0.578	0.431	0.600	0.429	0.147	0.601	0.578
15 15 15 15 15 15 15	0.900	0.804	0.904	0.776	0.901	0.758	0.579	0.901	0.904
30 30 30 30 30 30 30	1.000	0.998	1.000	0.994	1.000	0.993	0.989	1.000	1.000
3 3 5 5 5 7 7	0.228	0.103	0.139	0.109	0.162	0.114	0.005	0.223	0.210
3 3 8 8 8 13 13	0.304	0.119	0.190	0.133	0.224	0.147	0.050	0.307	0.265
10 10 15 15 15 20 20	0.836	0.583	0.802	0.574	0.800	0.612	0.459	0.836	0.865
	σ^2= (1, 1, 0.5, 0.5, 0.5, 0.25, 0.25)								
3 3 5 5 5 7 7	0.134	0.292	0.112	0.216	0.199	0.222	0.001	0.134	0.036
3 3 8 8 8 13 13	0.174	0.432	0.179	0.310	0.299	0.326	0.037	0.174	0.029
10 10 15 15 15 20 20	0.696	0.758	0.699	0.646	0.744	0.631	0.459	0.705	0.593
	σ^2= (1.25, 1.25, 0.75, 0.75, 0.75, 0.25, 0.25)								
3 3 5 5 5 7 7	0.205	0.355	0.192	0.272	0.291	0.283	0.002	0.196	0.055
3 3 8 8 8 13 13	0.361	0.602	0.413	0.453	0.521	0.477	0.092	0.349	0.065
10 10 15 15 15 20 20	0.900	0.891	0.920	0.832	0.923	0.822	0.649	0.901	0.865

When interpreting power tables, we first consider the situation where sample sizes are equal, and then the situations where sample sizes are proportional to variances and inversely proportional to variances. Regardless of the number of groups, when sample sizes are equal and for n =3, the CS test performs better than the other tests. As the sample sizes increase, especially for n ≥5, the powers of the CLR, SLR, and B tests become very close to each other and these tests perform slightly better than others. Considering the case where sample sizes are proportional to variances, regardless of the number of groups, it is observed that the CLR and SLR tests have significantly higher power values compared to other tests, especially for small sample sizes. With increasing sample size, the CAT method is also observed to outperform the other tests, in addition to these tests. When considering the case where sample sizes are inversely proportional to variances, the CS test has significantly higher power than the other tests. For example, for k=3, σ^2=(1, 0.5, 0.25), and n=3,8,13, the power values of the CLR, CS, GP, DAI, B, L, BF, SLR, and CAT tests are 0.133, 0.343, 0.156, 0.271, 0.222, 0.236, 0.041, 0.128, and 0.028, respectively. As the difference between the variances increases, this result has not changed. Furthermore, when we compare the CS test with other tests, it is seen that the CS test has considerably higher power values than other tests. The observed patterns do not change as the number of groups increased. It is observed that the CS test has significantly higher power values than the other tests, especially with small sample sizes. In addition, as can be seen from Table 4-6, the power values of all the tests generally converge and are very close to the value of 1 when the sample size is larger than about 30.

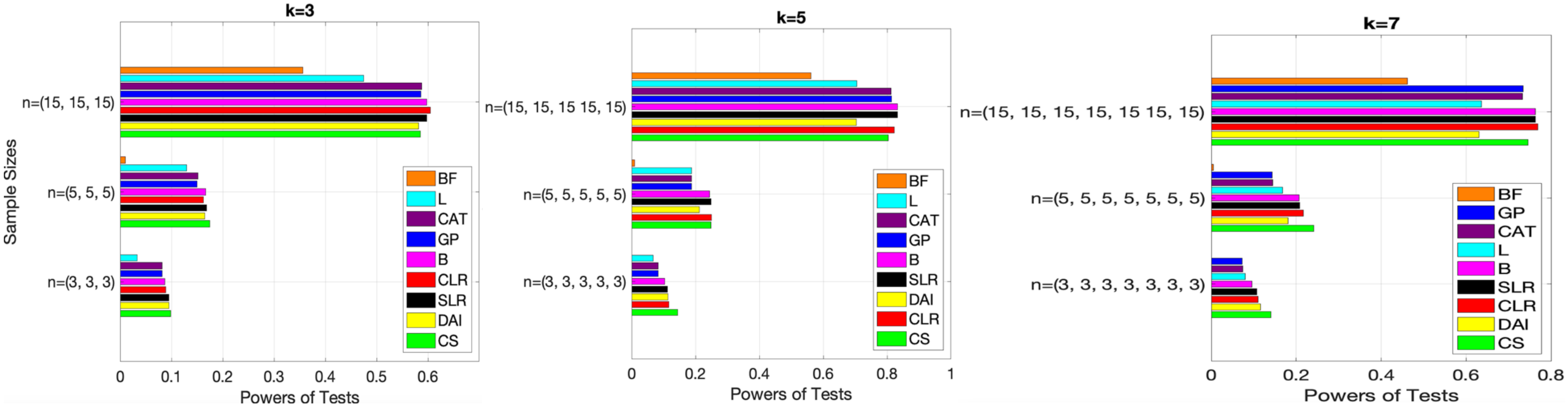

Figure 1. Powers of the tests when sample sizes are equal across different groups

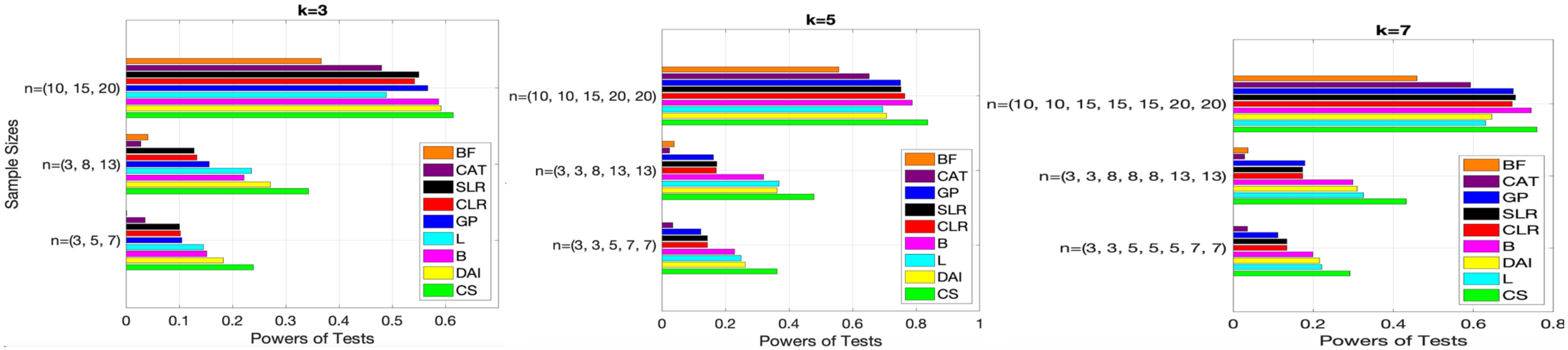

Figure 2. Powers of the tests when sample sizes are inversely proportional to variance values across different groups

The power performance of the tests can be seen in Figure 1 and Figure 2. Since there are many cases as seen in the tables, due to space constraints, graphs comparing the power values of the tests were obtained only for remarkable cases. For this purpose, Figure 1 shows the graphs comparing the power values of the tests when the sample sizes are equal for different number of groups. Here, σ^2=(1, 0.5, 0.25) for *k*=3, σ^2= (0.25, 0.25, 0.5,1, 1) for *k*=5, and σ^2= (0.25, 0.25, 0.5, 0.5, 0.5, 1, 1) for *k*=7. In Figure 2, graphs comparing the power values of the tests are presented when the sample sizes vary across groups and are inversely proportional to the variance values. Specifically, σ^2=(1, 0.5, 0.25) for *k* = 3, σ^2= (1, 1, 0.5, 0.25, 0.25) for *k* = 5 and σ^2= (1, 1, 0.5, 0.5, 0.5, 0.25, 0.25) for *k* = 7. It is observed that the results obtained from the tables are also clearly reflected in the graphs.

4. NUMERICAL EXAMPLES

In this section of the paper, we will present two data sets and show how the homoscedasticity tests work.

Example 1. In this section, we use the iris dataset that Anderson [22] gathered, which is still widely used in many studies. Examples of these studies include those that deal with data mining, multivariate normality assessment, and other topics [23, 24]. This dataset contains measurements of the sepals and characteristics of iris flowers, specifically their length and width in centimeters. It also includes a grouping variable indicating the type of iris flower, which can be categorized as setosa, virginica and versicolor. To demonstrate the application of the tests, we use the sepal length measurements as the dependent variable, while the iris types serve as the independent variable. The summary statistics of iris data are presented in Table 9.

Table 9. *The summary statistics of iris data*

	n	$\overline{X}_i$	S_i^2
setosa	50	5.01	0.124
versicolor	50	5.94	0.266
virginica	50	6.59	0.404

When we performed the Shapiro-Wilk test to check the normality of the data in each group, the p-values obtained for each group are 0.460, 0.465, and 0.258 respectively. As we can see, each group has a normal distribution. To calculate the p-values for all tests, we performed 10000 replications.

Table 10. *The results of tests for iris data*

Test	CLR	CS	GP	B	L	BF	SLR	CAT
p-value	0.0001	0.0003	0.0004	0.0003	0.0008	0.0022	0.0000	0.0002

Table 10 demonstrates that every test yields the same result, i.e., every test rejects the H_0 provided in Equation (2) at the nominal level of 0.05. As a result, we can conclude that there are differences among iris species.

Example 2. The data set is drawn from a study by Chang et al. [20]. The subjects in the study were 45 rape survivors, each subject randomly assigned to one of four treatment groups: (i) Treatment-1=Stress inoculation treatment (SIT), where subjects learned various coping skills. (ii) Treatment-2=Prolonged Exposure (PE), where subjects mentally replayed the rape in their minds repeatedly for seven sessions. (iii) Treatment-3=Supportive Counselling (SC), which is a standard treatment group; and (iv) Treatment-4=Waiting List (WL), control.

Summary statistics for symptom scores obtained from subjects under four different treatments are given in Table 11.

Table 11. *The summary statistics of rape victims*

	n	$\bar{X}_i$	S_i^2
Treatment-1	14	11.071	14.495
Treatment-2	10	15.400	111.240
Treatment-3	11	18.091	46.265
Treatment-4	10	19.500	45.450

When we performed the Shapiro-Wilk test to check the normality of the data in each group, the p-values obtained for each group are 0.997, 0.312, 0.355, and 0.114 respectively. As we can see, each group has a normal distribution. To calculate the p-values for all tests, we performed 10000 replications. The p-values of the tests are given in Table 12.

Table 12. *The results of tests for rape victims*

Test	**CLR**	**CS**	**GP**	**B**	**L**	**BF**	**SLR**	**CAT**
p-value	0.016	0.013	0.019	0.014	0.003	0.004	0.000	0.025

Table 12 demonstrates that every test rejects the H_0 provided in Equation (2) at the nominal level of 0.05. As a result, we can conclude that the treatments of rape victims have difference variances.

5. CONCLUSION

In this paper, we propose a CAT-based approach for the S test, referred to as the CS test, to assess the homogeneity of variances under normality. A simulation study was conducted to evaluate the performance of the proposed test and compare it with its competitors.

Using Monte Carlo simulations, all tests were compared in terms of size and power across various sample sizes and group numbers. According to the simulation results, the sizes of all tests are very close to the nominal level. When comparing the tests based on power values, the following observations were made: Regardless of the group size, the CS test outperforms the other tests for n=3. As sample sizes increase, particularly for n≥5, the power values of the CLR, SLR, and B tests slightly better than those of the other methods. Additionally, as the sample sizes increase, the power values of all the tests tend to converge. When sample sizes and variances are inversely proportional, the CS test consistently exhibits significantly higher power values than the other tests, regardless of group size. The simulation study highlights that the CS and CLR tests perform differently across all scenarios, particularly for small sample sizes. Furthermore, numerical examples have shown that the proposed method is highly accurate even for small sample sizes or large sample sizes.

Small sample sizes often arise due to time constraints, cost limitations, or insufficient data availability, making the choice of test critical. Therefore, we recommend researchers consider the CS test, especially alongside the CLR test, when sample sizes and variances are inversely proportional. Additionally, for other hypothesis testing problems, researchers may explore other likelihood-based tests beyond the LR test.

CONFLICTS OF INTEREST

No conflict of interest was declared by the authors.

ACKNOWLEDGEMENTS

This article is derived from the doctoral dissertation titled “Alternative tests for the equality of K group means and variances” completed at Gazi University and numbered 724628.

REFERENCES

[1] Boos, D.D., Brownie, C., “Comparing variances and other measures of dispersion”, Statistical Science, 19: 571-578, (2004). DOI: https://doi.org/10.1214/088342304000000500.

[2] Li, X., Qiu, W., Morrow, J., DeMeo, D.L., Weiss, S.T., Fu, Y., Wang, X., “A comparative study of tests for homogeneity of variances with application to DNA methylation data”, PloS one, 10(12): (2015). DOI: https://doi.org/10.1371/journal.pone.0145295

[3] Conover, W.J., Johnson, M.E., Johnson, M.M., “A comparative study of tests for homogeneity of variances, with applications to the outer continental shelf bidding data”, Technometrics, 23: 351-361, (1981). DOI: https://doi.org/10.1080/00401706.1981.10487693

[4] Cahoy, D.O., “A bootstrap test for equality of variances”, Computational Statistics and Data Analysis, 54: 2306-2316, (2010). DOI: https://doi.org/10.1016/j.csda.2010.01.015

[5] Bartlett, M.S., “Properties of Sufficiency and Statistical Test”, Proceedings of the Royal Society: A, 160: 268-282, (1937). DOI: https://doi.org/10.1098/rspa.1937.0109

[6] Bishop, D.J., Nair, U.S., “A note on certain methods of testing for the homogeneity of a set of estimated variances”, Journal of the Royal Statistical Society, 6: 89-99, (1939).

[7] Box, G.E., “Non-Normality and Tests on Variances”, Biometrika, 40: 318-335, (1953). DOI: https://doi.org/10.1093/biomet/40.3-4.318

[8] Brown, M.B., Forsythe, A.B., “Robust tests for the equality of variances”, Journal of the American Statistical Association, 69: 364-367, (1974). DOI: https://doi.org/10.2307/2285659

[9] Cochran, W.G., “Testing a linear relation among variances”, Biometrics, 7: 17-32, (1951). DOI: https://doi.org/10.2307/3001666

[10] Levene, H., “Robust Tests for Equality of Variances”, In Contributions to Probability and Statistics: Essays in honor of Harold Hotelling, 2: 278-292, (1960).

[11] Loh, W.Y., “Some modifications of Levene's test of variance homogeneity”, Journal of Statistical Computation and Simulation, 28: 213-226, (1987). DOI: https://doi.org/10.1080/00949658708811047

[12] Keyes, T.K., Levy, M.S., “Analysis of Levene’s Test under Design Inbalance”, Journal of Educational and Behavioral Statistics, 22: 227-236, (1997). DOI: https://doi.org/10.3102/10769986022003227

[13] Bhandary, M., Dai, H., “An alternative test for the equality of variances for several populations when the underlying distributions are normal”, Communications in Statistics-Simulation and Computation, 38: 109-117, (2008). DOI: https://doi.org/10.1080/03610910802387378

[14] Liu, X., Xu, X., “A new generalized p-value approach for testing the homogeneity of variances”, Statistics and Probability Letters, 80: 1486-1491, (2010). DOI: https://doi.org/10.1016/j.spl.2010.04.018

[15] Gökpınar, E., Gökpınar, F., “Testing equality of variances for several normal populations”, Communications in Statistics-Simulation and Computation, 46(1): 38–52, (2017). DOI: https://doi.org/10.1080/03610918.2015.1093934

[16] Jafari, A.A., Shaabani, J., "Comparing scale parameters in several gamma distributions with known shapes", Computational Statistics, 35(4): 1927-1950, (2020). DOI: https://doi.org/10.1007/s00180-019-00926-0

[17] Wang, J., Li, X., Liang, H., "A new exact p-value approach for testing variance homogeneity", Statistics Theory Related Fields, 6(1): 81-86, (2022). DOI: https://doi.org/10.1080/24754269.2021.2021010

[18] Bera, A.K., Bilias, Y., "Rao's score, Neyman's C (α) and Silvey's LM tests: an essay on historical developments and some new results", Journal of Statistical Planning and Inference, 97: 9-44, (2001**).** DOI: https://doi.org/10.1016/S0378-3758(00)00236-2

[19] Davison, A.C., Hinkley, D.V., "Bootstrap methods and their application", Cambridge: Cambridge University Press, (1997). DOI: https://doi.org/10.1017/CBO9780511802843

[20] Chang, C.H., Pal, N., Lin, J.J., "A revisit to test the equality of variances of several populations", Communications in Statistics-Simulation and Computation, 46: 6360-6384, (2017). DOI: https://doi.org/10.1080/03610918.2016.1148143

[21] Gökpınar, E., "Standardized likelihood ratio test for homogeneity of variance of several normal populations", Communications in Statistics- Simulation and Computation, 51: 6309-6319, (2022). DOI: https://doi.org/10.1080/03610918.2021.1943494

[22] Anderson, E., "The Irises of the Gaspe Peninsula", Bulletin of the American Iris Society, 59: 2-5, (1935).

[23] Hu, Y.C., "A new fuzzy-data mining method for pattern classification by principal component analysis", Cybernetics and Systems, 36: 527-547, (2005). DOI: https://doi.org/ 10.1080/01969720590913116

[24] Korkmaz, S., Goksuluk, D., Zararsiz, G., "MVN: An R Package for Assessing Multivariate Normality", The R Journal, 6: 151-162, (2014). DOI: https://doi.org/10.32614/RJ-2014-031