

Tests of equality of several variances with the likelihood ratio principle

Hyo-II Park^{1,a}

^aDepartment of Statistics, Cheongju University, Korea

Abstract

In this study, we propose tests for equality of several variances with the normality assumption. First of all, we propose the likelihood ratio test by applying the permutation principle. Then by using the p -values for the pairwise tests between variances and combination functions, we propose combination tests. We apply the permutation principle to obtain the overall p -values. Also we review the well-known test statistics for the completion of our discussion and modify a statistic with the p -values. Then we illustrate proposed tests by numerical and simulated data and compare their efficiency with the reviewed ones through a simulation study by obtaining empirical p -values. Finally, we discuss some interesting features related to the resampling methods and tests for equality among several variances.

Keywords: combination function, likelihood ratio test, normal distribution, permutation principle

1. Introduction

In statistics, the mean or location parameter has been of main interest for the statistical inference regardless of the parametric or nonparametric approach. The variance or scale parameter has always been treated as a nuisance parameter, but appears to have unnegligible roles for the results of data analysis, since sometimes one tries to check if the assumptions for the equality of variances is validated or not. For example, when one analyzes data with two-sample t -test, one assumes that the two variances are equal even though the value is not known. Then many procedures of statistical software provide outputs which tell us that the assumption between two variances are appropriate or not. Also tests for equality between two variances has been fully developed and widely applied in both parametric and nonparametric approach. For the comparison procedure among means with three or more samples, one may find some results about equality of variances in the ANOVA procedure in SAS.

As a matter of fact, the inferences about variance or variances have been scarcely reported in comparison with those about mean or means. One reason for this phenomenon may come from the fact that it would be difficult to construct any suitable form of statistic for testing equality among more than two variances or interpret the form of the likelihood ratio (LR) function and/or derive the null distribution of corresponding the LR function even under the normality assumption. There does not seem to exist either the LR procedure or an asymptotic procedure related to the LR approach in the literature except the Bartlett test (Bartlett, 1937). Only several heuristic ad hoc or modified procedures have been published and used by practitioners. The Bartlett test (Bartlett, 1937) appears first in the literature to test equality of more than two variances. The statistic resembles a function of the LR

¹ Department of Statistics, Cheongju University, 298 Dae-Sung Ro, Cheongwon-gu, Cheongju-si, Chungcheongbuk-do 28503, Korea. E-mail: hipark@cju.ac.kr

statistic to be able to obtain the null distribution even though an asymptotic result by modifying with some quantities related to the sample sizes. Hartley (1950) proposed a procedure based on the quotient of the maximum and minimum of individual sample variances. Hartley's test is easy to perform but is sensitive to departures from normality (Hand and Nagaraja, 2003). Then Levene (1960) considered a procedure with the data transformation by subtracting the sample mean and then taking absolute values for each observation. Also Brown and Forsythe (1974) modified the Levene's test by using sample medians instead of the sample means. Since then several modified tests have been proposed. All the null distributions for tests reviewed up to are asymptotic ones. Also O'Brien (1979, 1981) further modified the Levene's statistic. The main idea behind the O'Brien test is to transform the original scores so that the transformed scores reflect the variation of the original scores. Recently, Gokpinar and Gokpinar (2017), Chang *et al.* (2017) and Jayalath *et al.* (2017) considered some improvements of the tests reviewed in the above in terms of power and achieving the significance level. Especially, Gokpinar and Gokpinar (2017) and Jayalath *et al.* (2017) have tried to apply the bootstrap method.

The LR test requires exact specification or assumption of the population distributions. Especially, for the LR tests for the means and variances, one almost always assumes the normality and can use the well-known results for the null distributions of the corresponding LR statistics. Or when it would be difficult to derive the exact null distributions theoretically, it is common to obtain the limiting or asymptotic distributions based on the LR arguments. Sometimes even the derivation of an asymptotic distribution would not be possible or the serious discrepancy of an asymptotic distribution from the unknown distribution of the LR statistic might be detected from a simulation study. Then one may apply the Monte-Carlo method (Park, 2018).

With high development of computer capacity and its softwares, the distributions for test statistics have heavily been dependent on the resampling methods such as the bootstrap and permutation methods. Only the difference between both can be summarized as with replacement and without replacement when one resamples from the original data set. However it has been known that the results for both methods may be quite different (Good, 2000). If both methods can be applied for testing, the use of the permutation principle has been recommended (Good, 2000; Pesarin, 2001) since the permutation principle estimates the unknown distribution while the bootstrap method does the parameter. It is usual that the permutation principle may be applied with the Monte-Carlo approach. For the optimal number of iterations when the permutation principle can be applied, Oden (1991) and Boos and Zhang (2000) have studied some. For more discussion for the resampling methods, you may refer to Westfall and Young (1993) and Good (2000).

In this paper, we propose test procedures for testing equality of several variances simultaneously. For this purpose, the rest of this paper will be organized as follows. In the Section 2, first of all, we derive the LR function under the normality assumption. Then we propose the LR test which is intuitively easy to use and requires the minimal computations by applying the permutation principle to obtain the p -value. Also by examining the LR statistic in detail, we observe that the LR statistic consists of LR statistics for testing equality between two variances. Based on this fact, we consider using the combination functions to combine individual tests for equality between two variances. We also apply the permutation principle to obtain the overall p -values of the proposed combination tests. In the Section 3, we review the well-known test statistics briefly and modify a statistic with the individual p -values instead of the sample variances in the spirit of our proposed test to alleviate the discrepancy to departure of normality. Then we illustrate our procedure with numerical and simulated data and compare the efficiency of our proposed tests with other well-known ones in the Section 4. Finally we discuss some interesting features related with the test of equality of variances and re-sampling

methods and state the future research topics briefly in the Section 5.

2. Tests of equality of several variances

Suppose that we have K independent samples X_{i1}, \dots, X_{ini} from populations with distributions, $N(\mu_i, \sigma_i^2)$, $i = 1, \dots, K$. Then it is of our main interest to test

$$H_0 : \sigma_1^2 = \dots = \sigma_K^2 = \sigma^2$$

against $H_1 : \text{not } H_0$. In order to derive the LR test, first of all, we introduce notation in the following. For each i , $i = 1, \dots, K$

$$\bar{X}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij} \quad \text{and} \quad S_i^2 = \frac{1}{n_i} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2.$$

We note that all the \bar{X}_i 's and S_i^2 's are the maximum likelihood estimates of μ_i 's and σ_i^2 's, respectively under $H_0 \cup H_1$ but S_i^2 's are not unbiased. Also we have that

$$S_p^2 = \frac{1}{n} \sum_{i=1}^K n_i S_i^2,$$

where $n = \sum_{i=1}^K n_i$. We note that S_p^2 is the maximum likelihood estimate of σ^2 under H_0 but not unbiased, too. Then the LR function, $\text{LR}(\sigma_1^2, \dots, \sigma_K^2; X)$ is

$$\begin{aligned} \text{LR}(\sigma_1^2, \dots, \sigma_K^2; X) &= \frac{\sup \left\{ \prod_{i=1}^K \prod_{j=1}^{n_i} f(X_{ij}; \sigma_i^2 | H_0 \cup H_1) \right\}}{\sup \left\{ \prod_{i=1}^K \prod_{j=1}^{n_i} f(X_{ij}; \sigma^2 | H_0) \right\}} \\ &= \frac{(S_p^2)^{\frac{n}{2}}}{\prod_{i=1}^K (S_i^2)^{\frac{n_i}{2}}} \\ &= \prod_{i=1}^K \left(\frac{S_p^2}{S_i^2} \right)^{\frac{n_i}{2}}, \end{aligned} \quad (2.1)$$

where f is the probability density function for the normal. Then one may reject H_0 in favor of H_1 for some large values of $\text{LR}(\sigma_1^2, \dots, \sigma_K^2; X)$. In order to complete the LR test, we need the null distribution of $\text{LR}(\sigma_1^2, \dots, \sigma_K^2; X)$ but it would be difficult to derive the null distribution since the numerator and denominator in (2.1) are not independent. Then one may detour this difficulty via the limiting distribution approach for $2 \log \text{LR}(\sigma_1^2, \dots, \sigma_K^2; X)$, where \log means the natural logarithm and will be used with the same context in the sequel. However we have found that the discrepancy of the limiting distribution of $2 \log \text{LR}(\sigma_1^2, \dots, \sigma_K^2; X)$ from the unknown distribution was serious from a preliminary simulation study since the nominal significance level could not be achieved even for the normal distribution. For this reason, we have considered using the permutation method to obtain the null distribution G , say, of $\text{LR}(\sigma_1^2, \dots, \sigma_K^2; X)$ or $2 \log \text{LR}(\sigma_1^2, \dots, \sigma_K^2; X)$. Then we note that the LR test would be intuitively apparent and easy to use owing to the permutation distribution. Before we proceed to propose more test procedures, we briefly indicate the use of the permutation principle with

the Monte-Carlo approach to be applied to the multiple samples in the following. You may refer to Westfall and Young (1993) for more explanations and applications.

- (I) Combine the K samples into a one sample, X_1, \dots, X_n , say.
- (II) With a suitable random configuration scheme, rearrange X_1, \dots, X_n as X_1^p, \dots, X_n^p , say, with $X_i^p \neq X_j^p$ whenever $1 \leq i \neq j \leq n$.
- (III) Allocate $X_1^p, \dots, X_{n_1}^p$ as the first sample and do $X_{n_1+1}^p, \dots, X_{n_1+n_2}^p$ as the second sample and keep on up to $X_{n_{K-1}+1}^p, \dots, X_n^p$ as the K^{th} sample.
- (IV) Calculate the value of $\text{LR}(\sigma_1^2, \dots, \sigma_K^2; X^p)$ or $2 \log \text{LR}(\sigma_1^2, \dots, \sigma_K^2; X^p)$.
- (V) Iterate M times from (II) to (IV) and obtain empirical distribution function \hat{G}_M of $\text{LR}(\sigma_1^2, \dots, \sigma_K^2; X)$ or $2 \log \text{LR}(\sigma_1^2, \dots, \sigma_K^2; X)$ from the M number of $\text{LR}(\sigma_1^2, \dots, \sigma_K^2; X^p)$ s or $2 \log \text{LR}(\sigma_1^2, \dots, \sigma_K^2; X^p)$ s.

Then we may complete the LR test by obtaining p -value using the steps for the permutation principle stated above. On the other hand, we note that the LR function (2.1) has the form of the product of the following terms: for each $i, i = 1, \dots, K$

$$\left(\frac{S_p^2}{S_i^2} \right)^{\frac{n_i}{2}} = \left(\frac{n_i}{n} \right)^{\frac{n_i}{2}} \left(\frac{n_1 S_1^2}{n_i S_i^2} + \dots + \frac{n_{i-1} S_{i-1}^2}{n_i S_i^2} + 1 + \frac{n_{i+1} S_{i+1}^2}{n_i S_i^2} + \dots + \frac{n_K S_K^2}{n_i S_i^2} \right)^{\frac{n_i}{2}}.$$

In other words, the LR function (2.1) consists of all the components such as $(n_i S_i^2)/(n_j S_j^2)$ and $(n_j S_j^2)/(n_i S_i^2)$ for testing $H_{0ij} : \sigma_i^2 = \sigma_j^2$ against $H_{1ij} : \sigma_i^2 \neq \sigma_j^2$ for $1 \leq i \neq j \leq K$. This fact makes us to consider another type of test procedure for testing $H_0 : \sigma_1^2 = \dots = \sigma_K^2 = \sigma^2$ via applying the combination functions. Then there are $K(K-1)/2$ numbers of sub null hypotheses for testing $H_{0ij} : \sigma_i^2 = \sigma_j^2$ using $(n_i S_i^2)/(n_j S_j^2)$ or $(n_j S_j^2)/(n_i S_i^2)$. This implies that one may reject $H_{0ij} : \sigma_i^2 = \sigma_j^2$ in favor of $H_{1ij} : \sigma_i^2 \neq \sigma_j^2$ for some large or small values of $(n_i S_i^2)/(n_j S_j^2)$ or $(n_j S_j^2)/(n_i S_i^2)$. However we note that $(n_i S_i^2)/(n_j S_j^2)$ and $(n_j S_j^2)/(n_i S_i^2)$ are reciprocal forms for each other. Also we note that all the $(n_i S_i^2)/(n_j S_j^2)$ s do not have the same distribution since their distributions have different dfs even though they all are members of the F -distribution family. Therefore one cannot combine them directly. One approach to avoid this difficulty is to use the individual p -values of the individual test, $H_{0ij} : \sigma_i^2 = \sigma_j^2$. For doing this, first of all, we need following well-known result. Without loss of generality, from now, we will assume that $1 \leq i < j \leq K$.

Lemma 1. Under $H_0 : \sigma_1^2 = \dots = \sigma_K^2 = \sigma^2$, for each i and $j, 1 \leq i < j \leq K$,

$$F_{ij} = \frac{n_i(n_j - 1) S_i^2}{(n_i - 1)n_j S_j^2}$$

distributed as F with $n_i - 1$ and $n_j - 1$ dfs.

Let Λ_{ij} be the p -value for testing $H_{0ij} : \sigma_i^2 = \sigma_j^2$ based on F_{ij} . Then we combine $K(K-1)/2$ numbers of p -values by choosing a suitable combination function. The famous and widely used combination functions are as follows. Pesarin (2001) presented a concise review and investigated extensively the properties of the combination functions.

(1) Fisher combination (FC) function

$$FC = -2 \sum_{1 \leq i < j \leq K} \log(1 - \Lambda_{ij}).$$

(2) Liptak combination (LC) function

$$LC = \sum_{1 \leq i < j \leq K} \Psi^{-1}(\Lambda_{ij}),$$

where Ψ is a distribution function and Ψ^{-1} , its inverse.

(3) Tippett combination (TC) function

$$TC = \min \{ \Lambda_{ij}; 1 \leq i < j \leq K \}.$$

Then the testing rule would be to reject $H_0 : \sigma_1^2 = \dots = \sigma_K^2 = \sigma^2$ for some small values of each combination function. To complete the test $H_0 : \sigma_1^2 = \dots = \sigma_K^2 = \sigma^2$, we have to derive the null distributions of the combination functions. However since the components are not independent, it would be difficult to obtain the null distributions in theoretic manner. Therefore we will use the permutation principle to obtain the null distributions for the combination functions.

Then we may carry out to test $H_0 : \sigma_1^2 = \dots = \sigma_K^2 = \sigma^2$ by choosing a suitable test from the proposed ones and applying the permutation principle using the steps stated previously for estimating the null distributions of the combination functions. In relation to our proposal of test procedures for testing $H_0 : \sigma_1^2 = \dots = \sigma_K^2 = \sigma^2$, we briefly review some test statistics which are famous and widely used to compare the structure of test statistics with the proposed ones. In the next section, we will deal with this subject in some detailed fashion.

3. Other test statistics

In this section, we review the well-known test statistics and modify one of them using Λ_{ij} 's which are p -values for testing $H_{0ij} : \sigma_i^2 = \sigma_j^2$ for $i \neq j$ introduced in the previous section. First of all, we note that Bartlett (1937) proposed a test basically based on the LR function (2.1). By abusing the usage of notation, we assume that S_i^2 's and S_p^2 are unbiased estimates of σ_i^2 's and σ^2 , under $H_0 \cup H_1$ and H_0 , respectively. Then the Bartlett statistic BA can be defined as

$$BA = \frac{(n - K) \log(S_p^2) - \sum_{i=1}^K (n_i - 1) \log(S_i^2)}{1 + \frac{1}{3(K-1)} \left(\sum_{i=1}^K \frac{1}{n_i - 1} - \frac{1}{n - K} \right)} = \frac{\sum_{i=1}^K (n_i - 1) \log\left(\frac{S_p^2}{S_i^2}\right)}{1 + \frac{1}{3(K-1)} \left(\sum_{i=1}^K \frac{1}{n_i - 1} - \frac{1}{n - K} \right)}.$$

Then we note that one may obtain the numerator of BA by taking logarithm, multiplying 2 and subtracting 1 from n_i for each i for (2.1). The purpose of a modification of $2 \log \text{LR}(\sigma_1^2, \dots, \sigma_K^2; X)$ into BA was to make the approximation to the chi-square distribution better (Bartlett, 1937). This in turn implies that the limiting distribution of $2 \log \text{LR}(\sigma_1^2, \dots, \sigma_K^2; X)$ cannot be a chi-square with $K - 1$ df. In relation with the sample variances, Hartley (1950) also proposed a test using the following statistic, HA

$$HA = \frac{\max \{ S_i^2; 1 \leq i \leq K \}}{\min \{ S_i^2; 1 \leq i \leq K \}}$$

Table 1: Doughnut fat absorption data

Amount of fat absorbed			
Fat 1	Fat 2	Fat 3	Fat 4
164	178	175	155
172	191	193	166
168	197	178	149
177	182	171	164
156	185	163	170
195	177	176	168

the ratio of the largest sample variance relative to the smallest sample variance. Then the testing rule would be reject H_0 for some large values of HA. Pearson and Hartley (1970) have provided critical values of HA for some selected sample sizes and significance levels. It is also known to be very sensitive to departures from the normality assumption. However we note that HA resembles the Tippett combination function. For this reason, we may modify the statistic HA using Λ_{ij} such as

$$\text{THA} = \frac{\min \{ \Lambda_{ij}; 1 \leq i < j \leq K \}}{\max \{ \Lambda_{ij}; 1 \leq i < j \leq K \}}.$$

Then the testing rule would be to reject H_0 for some small values of THA. The null distribution of THA can be obtained by the permutation principle. We will use THA instead of HA when we consider to show examples and carry out simulation study. In order to avoid the sensitivity to departures from normality assumption, Levene (1960) proposed a test using the following statistic,

$$W = \frac{(n - K) \sum_{i=1}^K n_i (\bar{Z}_i - \bar{Z})^2}{(K - 1) \sum_{i=1}^K \sum_{j=1}^{n_i} (Z_{ij} - \bar{Z}_i)^2},$$

where $Z_{ij} = |X_{ij} - \bar{X}_i|$. Also \bar{Z}_i and \bar{Z} are the means of Z_{i1}, \dots, Z_{in_i} and all the Z_{ij} 's respectively. Then the testing rule would be to reject H_0 for some large values of W and the limiting distribution of W is known as F -distribution with $K - 1$ and $n - K$ df. The Levene test has been popular and widely used with various modification forms. Especially Brown and Forsythe (1974) modified the Levene statistic by using median instead of the mean in computing the spread within each group. Brown and Forsythe (1974) performed several simulation studies by varying \bar{Z} such as trimmed mean and median.

Then it would be interesting to compare the efficiency among those proposed and reviewed tests for testing $H_0 : \sigma_1^2 = \dots = \sigma_K^2 = \sigma^2$ in the next section. For the Hartley's test, we will use THA rather than HA since it is known that the test based on HA is sensitive to departure from normality. Then we will begin the next section by illustrating our proposed tests with two examples which are real and simulated data.

4. Examples and simulation results

In this section, we begin with the illustration of our test procedure with a numerical example for absorption in various amounts of fat during cooking doughnuts reported from the Iowa Agricultural Experiment Station (Lowe, 1935) summarized in Table 1. For each of four fats, six batches of doughnuts were prepared. The data in Table 1 are the grams of fat absorbed per batch. Initially, Snedecor and Cochran (1989) performed to test $H_0 : \sigma_1^2 = \dots = \sigma_4^2 = \sigma^2$ against $H_1 : \text{not } H_0$, by applying the Bartlett test and obtained insignificant result. We consider 7 different tests proposed and reviewed

Table 2: *p*-values for doughnut data

Test	LR	FC	LC	TC	THA	BA	W
<i>p</i> -value	0.5709	0.5070	0.5877	0.6038	0.5971	0.6258	0.3613

LR = likelihood ratio; FC = Fisher combination; LC = Liptak combination; TC = Tippett combination; THA = Tippett and Hartley test; BA = Bartlett statistic; W = Levene test statistic.

Table 3: Simulated data

Data for class			
1	2	3	4
7.40	8.84	8.09	7.55
6.18	6.69	7.96	5.65
6.86	7.12	5.31	6.69
7.76	7.42	7.39	6.50
6.39	6.83	0.51	5.46
5.95	5.06	7.84	7.40
7.48	5.35	6.28	8.37

Table 4: *p*-values for simulated data

Test	LR	FC	LC	TC	THA	BA	W
<i>p</i> -value	0.3804	0.1643	0.2032	0.2507	0.2768	0.0106	0.1247

LR = likelihood ratio; FC = Fisher combination; LC = Liptak combination; TC = Tippett combination; THA = Tippett and Hartley test; BA = Bartlett statistic; W = Levene test statistic.

ones up to now. Especially for the Hartley test, we use the THA rather than the HA for the test statistic in order to sensitivity to departure of normality. The distributions used for the Bartlett and Levene tests are chi-square with 3 df and *F* with 3 and 20 df, respectively. For the rest, we applied the permutation principle with the Monte-Carlo approach to obtain the *p*-values with 10,000 repetitions on SAS/IML with PC-version. The respective *p*-values for this example are summarized in Table 2 and show the insignificance for equality among variances.

Snedecor and Cochran (1989) also considered simulated data tabulated in Table 3 for testing $H_0 : \sigma_1^2 = \dots = \sigma_4^2 = \sigma^2$ against $H_1 : \text{not } H_0$ in order to observe initially the behaviors of the Bartlett and Levene tests when the assumption of normality is violated. In Table 3, four independent samples with $n_i = 7$ for all *i* were drawn from the *T* distribution with 3 df (a symmetrical long-tailed distribution) with the number 7 added. Snedecor and Cochran (1989) applied both Bartlett and Levene tests and observed that the Bartlett test yielded significance result which is absurd since all the data were drawn from the same distribution. Also we considered 7 tests used in the previous example and obtained *p*-values which are summarized in Table 4. We note that except the Bartlett test, all the other ones show insignificance in terms of *p*-values. Also we will note once again this absurdity in the Bartlett test from the sequel simulation study

We compare efficiency among 7 tests considered up to through simulation study by obtaining empirical *p*-values. For this, we considered the three-sample case and four different distributions such as normal, Laplace, logistic and uniform, which are all symmetric. We have drawn random numbers with unit variance for all cases and varied values of the standard deviation from 1 to 2 with increment 0.2 only for the third sample. We considered two cases for the sample sizes such as $n_1 = n_2 = n_3 = 15$ and $n_1 = 15, n_2 = 20, \text{ and } n_3 = 25$. We have carried out 10,000 simulations and applied the permutation principle to obtain the distribution with 2,000 repetitions within a simulation for each case. The nominal significance level has been chosen as 0.05 for all case. All the computations were carried out with SAS/IML with PC-version. The obtained empirical *p*-values are summarized in

Table 5: Empirical p -values for normal distribution

Test	(n_1, n_2, n_3)	$(\sigma_1, \sigma_2, \sigma_3)$					
		(1, 1, 1)	(1, 1, 1.2)	(1, 1, 1.4)	(1, 1, 1.6)	(1, 1, 1.8)	(1, 1, 2)
LR	(15, 15, 15)	0.0464	0.1145	0.2801	0.4837	0.6667	0.7908
FC		0.0470	0.0993	0.2153	0.3313	0.4206	0.4798
LC		0.0491	0.1056	0.2521	0.4302	0.6010	0.7294
TC		0.0470	0.1084	0.2583	0.4499	0.6270	0.7575
THA		0.0480	0.1042	0.2523	0.4485	0.6297	0.7629
BA		0.0498	0.0980	0.2388	0.4326	0.6240	0.7718
W		0.0586	0.1039	0.2256	0.3961	0.5572	0.6927
LR		(15, 20, 25)	0.0474	0.1354	0.3810	0.6342	0.8066
FC	0.0481		0.0991	0.2355	0.3714	0.4628	0.5116
LC	0.0481		0.1091	0.2873	0.5179	0.7036	0.8321
TC	0.0482		0.1129	0.3138	0.5571	0.7475	0.8660
THA	0.0482		0.1081	0.3135	0.5808	0.7483	0.8698
BA	0.0476		0.1111	0.3155	0.5808	0.7865	0.9035
W	0.0575		0.1092	0.2815	0.5019	0.7021	0.8349

LR = likelihood ratio; FC = Fisher combination; LC = Liptak combination; TC = Tippett combination; THA = Tippett and Hartley test; BA = Bartlett statistic; W = Levene test statistic.

Table 6: Empirical p -values for Laplace distribution

Test	(n_1, n_2, n_3)	$(\sigma_1, \sigma_2, \sigma_3)$					
		(1, 1, 1)	(1, 1, 1.2)	(1, 1, 1.4)	(1, 1, 1.6)	(1, 1, 1.8)	(1, 1, 2)
LR	(15, 15, 15)	0.0483	0.0916	0.1918	0.3180	0.4435	0.5490
FC		0.0493	0.0822	0.1402	0.2115	0.2805	0.3404
LC		0.0481	0.0868	0.1711	0.2819	0.3991	0.5058
TC		0.0478	0.0874	0.1764	0.2910	0.4140	0.5251
THA		0.0488	0.0857	0.1766	0.2907	0.4126	0.5192
BA		0.2424	0.2894	0.3953	0.5151	0.6283	0.7309
W		0.0651	0.0968	0.1702	0.2698	0.3851	0.4996
LR		(15, 20, 25)	0.0503	0.0990	0.2423	0.4099	0.5594
FC	0.0479		0.0785	0.1480	0.2327	0.3077	0.3671
LC	0.0496		0.0824	0.1838	0.3242	0.4654	0.5885
TC	0.0502		0.0841	0.1927	0.3453	0.4931	0.6192
THA	0.0524		0.0838	0.1927	0.3435	0.4933	0.6169
BA	0.2447		0.3181	0.4678	0.6263	0.7559	0.8516
W	0.0645		0.0901	0.1870	0.3272	0.4779	0.6201

LR = likelihood ratio; FC = Fisher combination; LC = Liptak combination; TC = Tippett combination; THA = Tippett and Hartley test; BA = Bartlett statistic; W = Levene test statistic.

Tables 5–8. First of all, we note that except normal distribution, the Bartlett test hardly have achieved its nominal significance level whose phenomenon have already been pointed out by many statisticians. Among other tests, the LR test performs the best in terms of the empirical powers and is convenient to use since it only requires the computation of the LR function (2.1) or $2 \log \text{LR}(\sigma_1^2, \dots, \sigma_K^2; X)$ without any further consideration. Only the Tippett and Hartley tests may compete with the LR test in terms of powers but require the three individual tests for $H_{0ij} : \sigma_i^2 = \sigma_j^2, 1 \leq i < j \leq 3$. As one may expect, the Tippett and Hartley with THA yielded very similar results for all distributions and sample sizes. Also the LR test protects the departure of normality well. In general, we note that the Fisher test showed relatively low performance.

5. Some concluding remarks

In general, the research in variance or scale parameters have been retarded compared with that of

Table 7: Empirical p -values for logistic distribution

Test	(n_1, n_2, n_3)	$(\sigma_1, \sigma_2, \sigma_3)$					
		(1, 1, 1)	(1, 1, 1.2)	(1, 1, 1.4)	(1, 1, 1.6)	(1, 1, 1.8)	(1, 1, 2)
LR	(15, 15, 15)	0.0486	0.1042	0.2392	0.4036	0.5532	0.6813
FC		0.0480	0.0890	0.1735	0.2689	0.3529	0.4164
LC		0.0483	0.0958	0.2110	0.3593	0.5091	0.6298
TC		0.0487	0.0974	0.2202	0.3755	0.5274	0.6521
THA		0.0488	0.0982	0.2212	0.3737	0.5255	0.6491
BA		0.1218	0.1756	0.3061	0.4633	0.6146	0.7419
W		0.0585	0.0953	0.1951	0.3269	0.4726	0.6124
LR	(15, 20, 25)	0.0499	0.1254	0.3153	0.5285	0.6991	0.8143
FC		0.0476	0.0926	0.1883	0.2979	0.3864	0.4433
LC		0.0509	0.0982	0.2390	0.4247	0.5917	0.7282
TC		0.0492	0.0992	0.2604	0.4578	0.6296	0.7620
THA		0.0516	0.0995	0.2581	0.4600	0.6285	0.7612
BA		0.1221	0.1991	0.3874	0.5971	0.7660	0.8774
W		0.0559	0.1026	0.2406	0.4278	0.6156	0.7669

LR = likelihood ratio; FC = Fisher combination; LC = Liptak combination; TC = Tippett combination; THA = Tippett and Hartley test; BA = Bartlett statistic; W = Levene test statistic.

Table 8: Empirical p -values for uniform distribution

Test	(n_1, n_2, n_3)	$(\sigma_1, \sigma_2, \sigma_3)$					
		(1, 1, 1)	(1, 1, 1.2)	(1, 1, 1.4)	(1, 1, 1.6)	(1, 1, 1.8)	(1, 1, 2)
LR	(15, 15, 15)	0.0473	0.1648	0.4735	0.7542	0.9050	0.9637
FC		0.0462	0.1491	0.3952	0.5191	0.5965	0.6336
LC		0.0467	0.1519	0.4229	0.6892	0.8474	0.9319
TC		0.0468	0.1544	0.4393	0.7213	0.8814	0.9528
THA		0.0481	0.1475	0.4374	0.7238	0.8891	0.9564
BA		0.0037	0.0149	0.1075	0.3395	0.6235	0.8283
W		0.0594	0.1297	0.3240	0.5513	0.7300	0.8489
LR	(15, 20, 25)	0.0493	0.2481	0.6639	0.9147	0.9822	0.9969
FC		0.0480	0.1723	0.4250	0.5793	0.6365	0.6561
LC		0.0483	0.1826	0.5300	0.8157	0.9414	0.9826
TC		0.0488	0.1995	0.5765	0.8660	0.9665	0.9929
THA		0.0511	0.1951	0.5776	0.8739	0.9714	0.9955
BA		0.0030	0.0242	0.1969	0.5608	0.8539	0.9639
W		0.0556	0.1717	0.4587	0.7290	0.8913	0.9595

LR = likelihood ratio; FC = Fisher combination; LC = Liptak combination; TC = Tippett combination; THA = Tippett and Hartley test; BA = Bartlett statistic; W = Levene test statistic.

mean or location parameters. The reason for this may come from the deficiency of demands in the application aspect. However it cannot be denied either that the distributions of the LR functions and related statistics have not been fully developed until now. This phenomenon already has been confirmed by Park (2018) for the study of testing procedure for the covariance matrix even for the one-sample problem. If we confess the difficulty for the derivation of the LR statistic, we have found during the preliminary simulation study that the limiting distribution of $2 \log \text{LR}(\sigma_1^2, \dots, \sigma_K^2; X)$ did not work properly but just fail simply since the size of test could hardly be achieved for any given nominal significance level. This should have been found by Bartlett (1937) since he proposed the modified form, BA of $2 \log \text{LR}(\sigma_1^2, \dots, \sigma_K^2; X)$ to be adjusted for the chi-square distribution with $K - 1$ df, which was successful for normal distribution but not for the others. Even though our approach cannot be called as a ground-breaking method, one should consider seriously adopting an alternative approach. Therefore our results may be far from completion but can be a bridge to cross the unfinished

battle ground for the tests of variance and/or scale parameters.

Nowadays the simultaneous test procedures for the mean and variance or location and scale parameters have appeared in the media of statistical journals frequently (Park, 2015, 2017). However the lack of the results for the variance and scale parameters could be obstacles for this research. Therefore one may take a privilege from this result to further ones research for simultaneous tests or the tests among variances.

References

- Bartlett MS (1937). Properties of sufficiency and statistical tests. In *Proceedings of the Royal Statistical Society, Series A*, **160**, 268–282.
- Boos DD and Zhang J. (2000). Monte Carlo evaluation of resampling-based hypothesis tests, *Journal of American Statistical Association*, **95**, 486–492.
- Brown MB and Forsythe AB (1974). Robust tests for the equality of variances, *Journal of the American Statistical Association*, **69**, 364–367.
- Chang CH, Pal N, and Lin JJ (2017). A revisit to test the equality of variances of several populations, *Communications in Statistics-Simulation and Computation*, **46**, 6360–6384.
- Gokpinar E and Gokpinar F (2017). Testing equality of variances for several normal populations, *Communications in Statistics-Simulation and Computation*, **46**, 38–52.
- Good P (2000). *Permutation Tests, A Practical Guide to Resampling Methods for Testing Hypotheses*, Springer, New York.
- Hand HA and Nagaraja HN (2003). *Order Statistics* (3rd ed), Wiley, New York.
- Hartley HO (1950). The Use of Range in Analysis of Variance, *Biometrika*, **37**, 271–280.
- Jayalath KP, Ng HKT, Manage AB, and Riggs KE (2017). Improved tests for homogeneity of variances, *Communications in Statistics-Theory and Methods*, **46**, 7423–7446.
- Levene H (1960). Robust tests for equality of variances, In Ingram Olkin; Harold Hotelling; *et al. Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, Stanford University Press. pp. 278–292.
- Lowe B (1935). Data from the Iowa Agricultural Experiment Station.
- O'Brien RG (1979). A general ANOVA method for robust tests of additive models for variance, *Journal of American Statistical Association*, **74**, 877–880.
- O'Brien RG (1981). A simple test for variance effects in experimental designs, *Psychological Bulletin*, **89**, 570-574.
- Oden NL (1991). Allocation of effort in Monte Carlo simulation for power of permutation tests, *Journal of American Statistical Association*, **86**, 1074–1076.
- Park HI (2015). Simultaneous test for the mean and variance with an application to the statistical process control, *Journal of Statistical Theory and Practice*, **9**, 868–881.
- Park HI (2017). On tests detecting difference in means and variances simultaneously under normality, *Communications in Statistics-Theory and Methods*, **46**, 10025–10035.
- Park HI (2018). A note on the test for the covariance matrix under normality, *Communications for Statistical Applications and Methods*, **25**, 71–78.
- Pearson ES and Hartley HO (1970). *Biometrika Tables for Statisticians*, Vol 1.
- Pesarin F (2001). *Multivariate Permutation Tests with Applications in Biostatistics*, Wiley, West Sussex, England.
- Piepho HP (1996). A Monte Carlo test for variance homogeneity in linear models, *Biometrical Journal*, **38**, 461–473

Snedecor GW and Cochran WG (1989). *Statistical Methods*, Eighth Edition, Iowa State University Press.

Westfall PH and Young SS (1993). *Resampling-Based Multiple Testing, Examples and Methods for p -Value Adjustment*, Wiley, New York.

Received January 10, 2018; Revised March 21, 2018; Accepted May 8, 2018