Thailand Statistician
July 2021; 19(3): 617-626
http://statassoc.or.th
Contributed paper

Parametric Bootstrap Statistic based on Jackknife and Bootstrap Approaches for One-way Analysis of Variance under Unequal Variance

Noppakun Thongmual* and Nirun Nitisuk
Department of Sciences and Mathematics, Faculty of Technology and Healthy of Science,
Kalasin University, Kalasin, Thailand.
*Corresponding author; e-mail: nop_stat@hotmail.com

Received: 13 March 2020
Revised: 28 June 2020
Accepted: 22 September 2020

Abstract

The objective of this paper is to compare the test statistics for testing the equality of means of k normal populations ($k > 2$) under unequal variances. In general, F-test is a classical test statistic in one-way ANOVA. When the assumption of homogeneity of variance is violated, F-test cannot control type I error rates. In the literature, the test statistics are developed for controlling type I error rates quit well but in terms of the power of the existing tests are not satisfactory. Hence, we proposed parametric bootstrap based on Jackknife approach (PBJ) test and, parametric bootstrap based on bootstrap approach (PBB) test. PBJ test and PBB test are compared with four existing tests as Wald test, Welch test, parametric bootstrap (PB) test and generalized p-value (GP) test. Type I error rates and powers of the tests are the criteria for comparing the performance of PBJ test, PBB test, Wald test, Welch test, PB test and GP test. Under Monte Carlo simulation, the results indicate that PBJ test, PBB test and PB test control type I error rates better than Wald test, Welch test and GP test but PBB is nearly the nominal level of 0.05 more than PB test and PBJ test. By considering the results of the power of the test, the performance of PBJ test and PBB test have power of the test more than PB test. Our study suggests that PBB test is a test statistic for testing the equality of means of normal populations in one-way ANOVA under unequal variances since it is nearly the nominal level better than the other tests, and PBB test is an alternative statistic for controlling type I error, and it has height power in one-way ANOVA under unequal variances.

Keywords: Power of the test, type I error rate, unequal variance.

1. Introduction

Consider the one-way analysis of variance (ANOVA), let Y_{ij} be the observed data according to the model

$$Y_{ij} = \mu_i + \varepsilon_{ij}, \quad (i = 1,2,...,k; j = 1,2,...,n_i), \tag{1}$$

from the j^{th} subject under the i^{th} treatment, where μ_i is unknown parameter and ε_{ij} represents error terms that are independently drown from normal distribution with mean 0 and variance σ^2. Therefore, $E(Y_{ij}) = \mu_i$ and $V(Y_{ij}) = \sigma^2$, then Y_{ij} has a normal distribution with mean μ_i and variance σ^2. In general, the classical F-test is one of the test statistics in one-way ANOVA with fixed effect model that are employed to test the equality of the means of k normal populations when variances are equal. Sometimes, the assumption of homogeneity of variance is violated, which is called the Behrens-Fisher problem (unequal variance σ_i^2) as presented by Fisher (1935). Hence, the classical F-test in ANOVA is not applicable since the type I error rate could not be controlled. This is supported by Box (1954) who examined the effect of unequal variance on the classical F-test for one-way ANOVA and suggested that the test is serious if the simple sizes are in a balanced design. Moder (2007) also showed that the classical F-test had type I error rate over the nominal level 0.05 as observed for both balance and unbalance designs. The hypothesis testing is employed for testing the equality of means in one-way ANOVA are

$$H_0 : \mu_1 = \ldots = \mu_k \text{ versus } H_1 : \mu_i \neq \mu_j \text{ for some } i \neq j. \tag{2}$$

For testing the equality of means of k normal populations $(k > 2)$ in (2) with the Behrens-Fisher problem, several test statistics have been proposed by many researchers.

Wald test statistic is an original test statistic for testing hypothesis in (2) under unequal variances which is (See Akritas and Papadatos 2004),

$$T(\bar{Y}_1,\ldots,\bar{Y}_k;S_1^2,\ldots,S_k^2) = \sum_{i=1}^{k}\frac{n_i}{S_i^2}\bar{Y}_i^2 - \frac{\left(\sum_{i=1}^{k} n_i\bar{Y}_i/S_i^2\right)^2}{\sum_{i=1}^{k} n_i/S_i^2}. \tag{3}$$

The null hypothesis H_0 in (2) is rejected when $T(\bar{Y}_1,\ldots,\bar{Y}_k;S_1^2,\ldots,S_k^2) > \chi^2_{k-1}$. Welch (1951) presented the test statistic for means under unequal variances based on Wald test,

$$W^* = \frac{T\left(\bar{X}_1,\ldots,\bar{X}_k;S_1^2,\ldots,S_k^2\right)/(k-1)}{1+(2(k-2)/(k^2-1))\sum_{i=1}^{k}(1/(n_i-1))(1-w_i/\sum_{i=1}^{k}w_i)^2} \sim F_{f_1,f_2} \text{ approximately} \tag{4}$$

where $f_1 = k-1$ and $f_2 = \left[\frac{3}{k^2-1}\sum_{i=1}^{k}\frac{1}{n_i-1}\left(1-w_i\Big/\sum_{i=1}^{k}w_i\right)^2\right]^{-1}$ are degree of freedom. This test rejects H_0 when $W^* > F_{f_1,f_2}$.

The concept of generalized p-value (GP) test is introduced by Weerahandi (1995a). The GP test is obtained by a generalized F-test which is described by Weerahandi (1995a). Let $V_i^2 = (n_i-1)S_i^2$ and v_i^2 is an observed value of V_i^2, $i = 1,2,\ldots,k.$ A generalized F-test is

$$GF = \frac{\sum_{i=1}^{k} \frac{n_i}{S_i^2} \bar{Y}_i^2 - \frac{\left(\sum_{i=1}^{k} n_i \bar{Y}_i / S_i^2 \right)^2}{\sum_{i=1}^{k} n_i / S_i^2}}{\sum_{i=1}^{k} \frac{n_i U_i}{v_i^2} \bar{y}_i^2 - \frac{\left(\sum_{i=1}^{k} \frac{n_i U_i}{v_i^2} \bar{y} \right)^2}{\sum_{i=1}^{k} \frac{n_i U_i}{v_i^2}}}, \tag{5}$$

where U_i, $i = 1, 2, ..., k$ are independent random variables with $U_i \sim \chi^2_{n_i - 1}$. From a generalized F-test, $\sum_{i=1}^{k} \frac{n_i}{S_i^2} \bar{Y}_i^2 - \frac{\left(\sum_{i=1}^{k} n_i \bar{Y}_i / S_i^2 \right)^2}{\sum_{i=1}^{k} n_i / S_i^2}$ has distributed as chi-square with degree of freedom $k - 1$. The generalized p-value (GP) test is given by

$$GP = P\left(\frac{\chi^2_{k-1}}{\sum_{i=1}^{k} \frac{n_i U_i}{v_i^2} \bar{y}_i^2 - \frac{\left(\sum_{i=1}^{k} \frac{n_i U_i}{v_i^2} \bar{y} \right)^2}{\sum_{i=1}^{k} \frac{n_i U_i}{v_i^2}}} > 1 \right). \tag{6}$$

The GP test rejects null hypothesis in (2) when the GP test in (6) is less than a given nominal level α. The GP test is compared in one-way ANOVA by Krishnamoorthy (2007). The result indicate that the GP test has type I error rate more than the nominal level α.

Krishnamoorthy (2007) proposed the parametric bootstrap (PB) test by improving sample mean and sample variance using bootstrap method, where the parameters are replaced by their estimates in the test statistic $T(\bar{X}_1, ..., \bar{X}_k; S_1^2, ..., S_k^2)$. The PB test can be written as

$$T(\bar{X}_{B1}, ..., \bar{X}_{Bk}; S_{B1}^2, ..., S_{Bk}^2) = \sum_{i=1}^{k} \frac{n_i}{S_{Bi}^2} \bar{X}_{Bi}^2 - \frac{\left[\sum_{i=1}^{k} n_i \bar{X}_{Bi} / S_{Bi}^2 \right]^2}{\sum_{i=1}^{k} n_i / S_{Bi}^2}, \tag{7}$$

where $\bar{X}_{Bi}$ is distributed as $Z_i(S_i / \sqrt{n_i})$ and Z_i is a standard normal random variable. Then the PB test statistic in (7) is written as

$$T_B(Z_i, \chi^2_{n_i-1}; S_i^2) = \sum_{i=1}^{k} \frac{Z_i^2 (n_i - 1)}{\chi^2_{n_i-1}} - \frac{\left[\sum_{i=1}^{k} \left(\sqrt{n_i} Z_i (n_i - 1) / S_i^2 \chi^2_{n_i-1} \right) \right]}{\sum_{i=1}^{k} \left(n_i (n_i - 1) / S_i^2 \chi^2_{n_i-1} \right)}. \tag{8}$$

The PB test rejects H_0 when

$$P\left(T_B(Z_i, \chi^2_{n_i-1}; S_i^2) > T_0 \right) < \alpha \tag{9}$$

where T_0 is the observed value of $T(\overline{Y}_1,...,\overline{Y}_k;S_1^2,...,S_k^2)$. According to the results, the PB test could control type I error rates better than the Welch test. The PB test performed very satisfactorily when the sample sizes are small.

Jackknife and bootstrap approaches are technical statistics for reducing error of the test statistics where two approaches used the process of resampling technic. Jackknife approach is about computing the statistic of interest for all combinations of the data where one of the original data points are removed. Bootstrap approach attempt to estimate the sampling distribution of a population by generating new samples by drawing from the original data (Ana et al. 2011).

All the tests mentioned above are the test statistics that were used for testing the equality of the means of k normal populations under unequal variance. However, some of these tests poorly controlled type I error rates. Therefore, simple test statistic with high performance based on type I error rates as well as power of the test should be established. In this research, Jackknife samples is used for developing PB test for testing the equality of the means of k normal populations under unequal variance where Jackknife and Bootstrap approaches improve p-value of PB test for rejecting null hypothesis H_0 in (2). Hence, PB test based on Jackknife and bootstrap approaches are called parametric bootstrap based on Jackknife approach (PBJ) test and parametric bootstrap based on Bootstrap approach (PBB) test.

The objective of this paper is to compare performance of Wald test, Welch test, GP test, PB test, PBJ test and PBB test for testing the equality of the means of k normal populations when variances are unequal based on type I error rates and power of the test, which reflect the performance of the test statistic. The paper is organized as follows. PBJ test and PBB test for one-way ANOVA is described in Section 2. Section 3 is to find the critical value of the proposed test at nominal significant level and to show the performance of the tests based on type I error rates and power of the test. Finally, Section 4 contains conclusions.

2. PBJ Test and PBB Test for One-Way ANOVA

The PBJ test and PBB test is as follows:

1. For a given $(n_1,n_2...,n_k)$, $(\overline{Y}_1,\overline{Y}_2...,\overline{Y}_k)$ and $(S_1^2,S_2^2,...,S_k^2)$ are used to compute Wald test in (3) as $T(\overline{Y}_1,...,\overline{Y}_k;S_1^2,...,S_k^2)=\sum_{i=1}^{k}\frac{n_i}{S_i^2}\overline{Y}_i^2-\frac{\left(\sum_{i=1}^{k}n_i\overline{Y}_i/S_i^2\right)^2}{\sum_{i=1}^{k}n_i/S_i^2}$.

2. PB tests in (7) as $T_B(Z_i,\chi^2_{n_i-1};S_i^2)=\sum_{i=1}^{k}\frac{Z_i^2(n_i-1)}{\chi^2_{n_i-1}}-\frac{\left[\sum_{i=1}^{k}\left(\sqrt{n_i}Z_i(n_i-1)/S_i^2\chi^2_{n_i-1}\right)\right]}{\sum_{i=1}^{k}\left(n_i(n_i-1)/S_i^2\chi^2_{n_i-1}\right)}$ is calculated with m values which is data of PB test and Z_i is a standard normal random variable.

3. From Step 2, we obtain data of PB test with m values which is

$$\mathbf{X}=(PB_1,...,PB_{i-1},PB_i,PB_{i+1},...,PB_m).$$

Jackknife samples are selected by taking the data of PB test in Step 2 and deleting one observation from $\mathbf{X}$. The sample of Jackknife approach is

$$\mathbf{X}_{-i} = (PB_1,...,PB_{i-1},PB_{i+1},...,PB_m)$$

Bootstrap approach uses the process of resampling for PB test in $\mathbf{X}$. The sample of bootstrap approach is

$$\mathbf{X}^* = (PB_1^*,...,PB_{i-1}^*,PB_i^*,PB_{i+1}^*,...,PB_m^*)$$

4. After that we calculated p-value based on Jackknife and bootstrap samples which are $\mathbf{X}_{-i}$ and $\mathbf{X}^*$, respectively. If data in $\mathbf{X}_{-i}$ and $\mathbf{X}^*$ are more than the calculated Wald test in Step 1 set to $Q_j = 1$

5. From Step 4, we obtained p-value as

$$\text{p-value}_{PBJ} = \frac{\sum_{j=1}^{m-1} Q_j}{m-1} \text{ and } \text{p-value}_{PBB} = \frac{\sum_{j=1}^{m} Q_j}{m}, \tag{10}$$

where p-value_{PBJ} and p-value_{PBB} are called PBJ test and PBB test, respectively.

3. Type I Error and Power of the Test

Type I error rates of Wald test, Welch test, PB test, GP test, PBJ test and PBB test are calculated by Monte Carlo simulation by using the R statistical package (R Core Team 2018). We denote, without loss of generality, that $\mu_1 = \mu_2 = ... = \mu_k$, $\sigma_1^2 = 1$ and $0 < \sigma_i^2 < 1$, $i = 2,...,k$ for $k = 3$ and 6, under various values of n_i in our simulation studies. Then, the sample statistics $\bar{y}_i$ and s_i^2 are generated independently as $\bar{y}_i \sim N(0, \sigma_i^2 / n_i)$ and $s_i^2 \sim \sigma_i^2 \chi_{n_i-1}^2 \big/ (n_i - 1)$, $i = 1,...,k$. For power of the tests, we denote that $0 \le \mu_i \le 1.5$ and the sample statistics $\bar{y}_i$ and s_i^2 are generated independently as $\bar{y}_i \sim N(\mu_i, \sigma_i^2 / n_i)$ and $s_i^2 \sim \sigma_i^2 \chi_{n_i-1}^2 \big/ (n_i - 1)$, $i = 1,...,k$. To calculate the type I error rates of Wald test and Welch test are determined by the proportions of Wald test and Welch test that exceed the critical value of each test. For estimating type I error rates of GP test, PB test, PBJ test and PBB test, we used 50,000 observed vectors $(\bar{y}_1,...,\bar{y}_k, s_1^2,...,s_k^2)$ to compute the observed value in (3). We use 50,000 runs to estimate p-value of GP test, PB test, PBJ test and PBB test, respectively. Finally, the type I error rates are estimated by proportion of the 50,000 p-values that are less than the nominal level 0.05. Power of the six tests is generated from the sample statistics $\bar{y}_i \sim N(\mu_i, \sigma_i^2 / n_i)$ and $s_i^2 \sim \sigma_i^2 \chi_{n_i-1}^2 \big/ (n_i - 1)$. The values of the type I error rates are shown in Table 1, and Powers of the tests are shown in Table 2.

Table 1 The estimated type I error rates of the six tests

$k=6, \sigma_1^2=1$	$n=(5,5,5,5,5,5)$						$n=(10,10,10,10,10,10)$					
Test	Wald	Welch	PB	PBJ	PBB	GP	Wald	Welch	PB	PBJ	PBB	GP
$(\sigma_2^2,\ldots,\sigma_6^2)$												
(1, 1)	0.1228	0.0443	0.0447	0.0447	0.0447	0.0312	0.081	0.0489	0.0492	0.0492	0.0498	0.0497
(1, 0.5)	0.1250	0.0467	0.0475	0.0475	0.0476	0.0470	0.0845	0.0498	0.0515	0.0515	0.0514	0.0512
(1, 0.1)	0.1461	0.0553	0.0511	0.0511	0.0506	0.0512	0.0938	0.0511	0.0503	0.0503	0.0499	0.0621
(0.5, 0.5)	0.1243	0.0458	0.0444	0.0444	0.0444	0.0448	0.0825	0.0488	0.0473	0.0473	0.0475	0.0523
(0.5, 0.7)	0.1238	0.0469	0.0462	0.0463	0.0463	0.0464	0.0817	0.0484	0.0494	0.0494	0.0494	0.0512
(0.1, 0.1)	0.1345	0.0513	0.0474	0.0474	0.0478	0.0476	0.0861	0.0494	0.5000	0.5000	0.5000	0.0513
(0.1, 0.9)	0.1445	0.0544	0.0501	0.0500	0.0500	0.0505	0.0907	0.0516	0.5100	0.5100	0.5100	0.0612

$k=6, \sigma_1^2=1$	$n=(3, 3, 4, 5, 6, 6)$						$n=(4, 8, 12, 24, 30, 40)$					
Test	Wald	Welch	PB	PBJ	PBB	GP	Wald	Welch	PB	PBJ	PBB	GP
$(\sigma_2^2,\ldots,\sigma_6^2)$												
(1, 1)	0.1609	0.0466	0.0439	0.0439	0.0439	0.0512	0.1301	0.0576	0.0521	0.0521	0.0517	0.0522
(1, 0.5)	0.1738	0.0510	0.0462	0.0462	0.0468	0.0613	0.1417	0.0596	0.0537	0.0537	0.0537	0.0623
(1, 0.1)	0.2133	0.0679	0.0560	0.0560	0.0557	0.0702	0.1701	0.0650	0.0560	0.0560	0.0557	0.0712
(0.5, 0.5)	0.1698	0.0524	0.0465	0.0465	0.0467	0.0623	0.1412	0.0629	0.0619	0.0619	0.0619	0.0633
(0.5, 0.7)	0.1665	0.0485	0.0438	0.0438	0.0441	0.0612	0.1344	0.0609	0.0581	0.0581	0.0581	0.0622
(0.1, 0.1)	0.1974	0.0647	0.0510	0.0510	0.0510	0.0512	0.1501	0.0605	0.0506	0.0506	0.0503	0.0522
(0.1, 0.9)	0.1913	0.0622	0.0495	0.0495	0.0495	0.0623	0.1290	0.0617	0.0538	0.0538	0.0538	0.0633

Table 1 (continued)

$k=6,\sigma_1^2=1$			n = (5, 5, 5, 5, 5,5)						n = (10,10,10,10,10,10)			
Test	Wald	Welch	PB	PBJ	PBB	GP	Wald	Welch	PB	PBJ	PBB	GP
$(\sigma_2^2,\ldots,\sigma_6^2)$												
(1, 1, 1, 1, 1)	0.2159	0.0595	0.0446	0.0446	0.0446	0.0546	0.1131	0.0506	0.0448	0.0448	0.0452	0.0567
(0.1, 0.1, 0.5, 0.5, 0.5)	0.2243	0.0650	0.0465	0.0465	0.0467	0.0565	0.1194	0.0539	0.0481	0.0481	0.0481	0.0565
(0.1, 0.2, 0.3, 0.4, 0.5)	0.2230	0.0652	0.0451	0.0452	0.0454	0.0552	0.1182	0.0525	0.0502	0.0502	0.0502	0.0553
(0.1, 1, 1, 1, 1)	0.2306	0.0693	0.0491	0.0491	0.0491	0.0591	0.1233	0.0557	0.0517	0.0517	0.0516	0.0592
(0.2, 0.4, 0.4, 0.2, 0.1)	0.2198	0.0604	0.0440	0.0440	0.0440	0.0540	0.1198	0.0542	0.0497	0.0497	0.0500	0.0540
(0.5, 0.5, 0.5, 0.5, 1)	0.2162	0.0618	0.0441	0.0441	0.0441	0.0541	0.1175	0.0530	0.0495	0.0495	0.0497	0.0541
(0.3, 0.9, 0.4, 0.7, 0.1)	0.2227	0.0651	0.0459	0.0459	0.0459	0.0559	0.1235	0.0574	0.0505	0.0505	0.0500	0.0559
$k=6,\sigma_1^2=1$			n = (3, 3, 4, 5, 6, 6)						n = (4, 8, 12, 24, 30, 40)			
Test	Wald	Welch	PB	PBJ	PBB	GP	Wald	Welch	PB	PBJ	PBB	GP
$(\sigma_2^2,\ldots,\sigma_6^2)$												
(1, 1, 1, 1, 1)	0.2823	0.0775	0.0450	0.0450	0.0451	0.0550	0.1418	0.0694	0.0527	0.0527	0.0523	0.0527
(0.1, 0.1, 0.5, 0.5, 0.5)	0.2765	0.0787	0.0467	0.0467	0.0469	0.0567	0.1366	0.0692	0.0484	0.0484	0.0485	0.0584
(0.1, 0.2, 0.3, 0.4, 0.5)	0.2791	0.0796	0.0459	0.0459	0.0461	0.0559	0.1435	0.0698	0.0490	0.0490	0.0492	0.0590
(0.1, 1, 1, 1, 1)	0.2752	0.0780	0.0475	0.0475	0.0475	0.0575	0.1371	0.0695	0.0525	0.0525	0.0525	0.0565
(0.2, 0.4, 0.4, 0.2, 0.1)	0.3050	0.0881	0.0512	0.0512	0.0512	0.0552	0.1500	0.0728	0.0531	0.0531	0.0527	0.0571
(0.5, 0.5, 0.5, 0.5, 1)	0.2819	0.0766	0.0449	0.0449	0.0449	0.0549	0.1439	0.0714	0.0538	0.0538	0.0538	0.0568
(0.3, 0.9, 0.4, 0.7, 0.1)	0.3059	0.0884	0.0474	0.0474	0.0475	0.0574	0.1561	0.0742	0.0553	0.0553	0.0552	0.0573

Table 1 shows the values of the estimated type I error rates of the Wald test, Welch test, PB test, GP test, PBJ test and PBB test with the number of group k = 3, 6 and sample sizes ranging from very small to moderate under equal variances and unequal variances. The results are showed as follows.

For k = 3, Wald test cannot control type I error rates at the nominal level of 0.05. In cases of balance sample size, Welch test, PB test, PBJ test and PBB test can control type I error rates quit well. We see that the type I error rates of the GP test is around the type I error rates of 0.06 for the sample size n = (10, 10, 10) under equal variance and unequal variance. For unbalance sample sizes, PB test, the PBJ test and PBB test can control the type I error rates quite well but PBB test is nearly the nominal level of 0.05 more than PB test and PBJ test. Welch test and GP test have type I error rates that exceed the nominal level of 0.05 for n = (3, 4, 5) and n = (4, 6, 20). From the reported type I error rates in Table 1, the PBB test is the best test statistics as compared to the other tests based on type I error rates.

As observed for balanced and unbalanced sample sizes with equal and unequal variances, for k = 6, PB test, PBJ test and PBB test are nearly the nominal level of 0.05 but the type I error rates of the Wald test, GP test and Welch test exceed the nominal level of 0.05. Among the six tests, PB test, PBJ test and PBB test are the best statistics in terms of type I error rates. Again, observed that PBB test is nearly the nominal level of 0.05 more than PB test and PBJ test.

In Table 2, we report the power of six tests for k = 3 and 6. We observe that Wald test, Welch test, PB test, GP test, PBJ test and PBB test appear height power of the tests in the case of $(\mu_1, \mu_2, \mu_3) = (0, 1.5, 1)$, sample sizes n = (10, 10, 10) and n = (10, 5, 15). For k = 6, PBJ test and PBB test have power of the tests comparable to the other tests in the case of n = (15, 15, 20, 20, 25, 25) and n = (19, 21, 23, 25, 27, 29) under the variances $\boldsymbol{\sigma}_a^2$ and $\boldsymbol{\sigma}_b^2$. It seems that power of PBB in table 2 have power of the test higher than PB test and PBJ test when sample sizes are equal and unequal under unequal variances.

Table 2 Power of the six tests

$k=3, \sigma_1^2=1$ and $\mu_1=0$		(μ_2, μ_3)						
n = (10, 10, 10)	Tests	(0, 0)	(0, 0.2)	(0, 0.5)	(0, 0.7)	(0.5, 1)	(0, 1)	(1.5, 1)
(σ_2^2, σ_3^2) = (0.3, 0.9)	Wald	0.0861	0.1175	0.2906	0.4821	0.5612	0.7698	0.9652
	Welch	0.0511	0.0736	0.2067	0.3749	0.4525	0.6748	0.9342
	GP	0.0521	0.0706	0.2027	0.3649	0.4535	0.6738	0.9242
	PB	0.0483	0.0688	0.1841	0.2890	0.3352	0.5193	0.8631
	PBJ	0.0483	0.0688	0.1841	0.2890	0.3352	0.5193	0.8631
	PBB	0.0486	0.0690	0.1843	0.2890	0.3350	0.5191	0.8636
(σ_2^2, σ_3^2) = (0.1, 0.5)	Wald	0.0910	0.1508	0.4775	0.7438	0.7011	0.9557	0.9904
	Welch	0.0525	0.0958	0.3624	0.6333	0.5876	0.9150	0.9771
	GP	0.0655	0.0988	0.3674	0.6363	0.5886	0.9160	0.9771
	PB	0.0483	0.0904	0.3205	0.5535	0.4668	0.8199	0.9777
	PBJ	0.0483	0.0904	0.3205	0.5535	0.4668	0.8199	0.9777
	PBB	0.0483	0.0906	0.3197	0.5540	0.4670	0.8200	0.9775

Table 2 (continued)

n = (10, 5, 15)	Tests	(0, 0)	(0, 0.2)	(0, 0.5)	(0, 0.7)	(0.5, 1)	(0, 1)	(1.5, 1)
(σ_2^2, σ_3^2) = (0.3, 0.9)	Wald	0.0895	0.1277	0.3312	0.5421	0.6415	0.8300	0.9206
	Welch	0.0503	0.0765	0.2260	0.4166	0.5116	0.7295	0.8575
	GP	0.0543	0.0782	0.2264	0.4263	0.5213	0.7293	0.8773
	PB	0.0483	0.0730	0.1954	0.3553	0.4297	0.6129	0.7197
	PBJ	0.0483	0.0730	0.1954	0.3553	0.4297	0.6129	0.7197
	PBB	0.0483	0.0730	0.1954	0.3556	0.4297	0.6139	0.7201
(σ_2^2, σ_3^2) = (0.1, 0.5)	Wald	0.0879	0.1575	0.5360	0.8096	0.8080	0.9815	0.9803
	Welch	0.0491	0.1575	0.4196	0.7172	0.7135	0.9616	0.9550
	GP	0.0562	0.1565	0.4193	0.7161	0.7123	0.9623	0.9451
	PB	0.0496	0.0951	0.3888	0.6600	0.6332	0.9163	0.9224
	PBJ	0.0496	0.0951	0.3888	0.6600	0.6332	0.9163	0.9224
	PBB	0.0496	0.0951	0.3888	0.6606	0.6334	0.9163	0.9224
$k = 6$ and $(\mu_1, \ldots, \mu_4) = 0$					(μ_5, μ_6)			
n = (15, 15, 20, 20, 25, 25)	Tests	(0, 0)	(0, 0.2)	(0, 0.5)	(0, 0.7)	(0.5, 1)	(0, 1)	(1.5, 1)
$\boldsymbol{\sigma}_a^2$ = (1, 0.1, 0.2, 0.3, 0.4, 0.5)	Wald	0.0813	0.1903	0.7508	0.9634	1.0000	0.9998	1.0000
	Welch	0.0501	0.1328	0.6684	0.9406	1.0000	0.9994	1.0000
	GP	0.0511	0.1363	0.6694	0.9413	1.0000	0.9995	1.0000
	PB	0.0467	0.1316	0.6312	0.9183	0.9996	0.9984	1.0000
	PBJ	0.0467	0.1316	0.6312	0.9183	0.9996	0.9984	1.0000
	PBB	0.0468	0.1310	0.6320	0.9184	0.9996	0.9984	1.0000
$\boldsymbol{\sigma}_b^2$ = (1, 0.1, 0.1, 0.3, 0.3, 0.7)	Wald	0.0821	0.1606	0.6207	0.9001	0.9999	0.9972	1.0000
	Welch	0.0517	0.1081	0.5296	0.8500	0.9998	0.9938	1.0000
	GP	0.0523	0.1092	0.5295	0.8510	0.9998	0.9936	1.0000
	PB	0.0515	0.0989	0.4896	0.8083	0.9993	0.9849	1.0000
	PBJ	0.0515	0.0989	0.4896	0.8083	0.9993	0.9849	1.0000
	PBB	0.0515	0.0989	0.4894	0.8083	0.9994	0.9848	1.0000
n = (19, 21, 23, 25, 27, 29)	Tests	(0, 0)	(0, 0.2)	(0, 0.5)	(0, 0.7)	(0.5, 1)	(0, 1)	(1.5, 1)
$\boldsymbol{\sigma}_a^2$ = (1, 0.1, 0.2, 0.3, 0.4, 0.5)	Wald	0.0713	0.0748	0.2034	0.8110	0.9842	1.0000	1.0000
	Welch	0.0511	0.0507	0.1541	0.7551	0.9742	1.0000	1.0000
	GP	0.0531	0.0508	0.1531	0.7652	0.9841	1.0000	1.0000
	PB	0.0477	0.0487	0.1482	0.7249	0.9634	1.0000	0.9998
	PBJ	0.0487	0.0487	0.1482	0.7249	0.9634	1.0000	0.9998
	PBB	0.0488	0.0490	0.1485	0.7250	0.9629	1.0000	0.9998
$\boldsymbol{\sigma}_b^2$ = (1, 0.1, 0.1, 0.3, 0.3, 0.7)	Wald	0.0913	0.0766	0.1652	0.6810	0.9351	1.0000	0.9993
	Welch	0.0511	0.0512	0.1217	0.6104	0.9096	1.0000	0.9990
	GP	0.0525	0.0532	0.1235	0.6113	0.9092	1.0000	0.9991
	PB	0.0467	0.0520	0.1177	0.5734	0.8766	0.9999	0.9961
	PBJ	0.0467	0.0520	0.1177	0.5734	0.8766	0.9999	0.9961
	PBB	0.0468	0.0520	0.1185	0.5733	0.8764	0.9999	0.9961

4. Conclusions

The F-test is a classical statistic for one-way ANOVA under equal variance when the assumption of variance is violated. F-test cannot control type I error rates and is examined by Box (1954) and Moder (2007). In this paper, we proposed PBJ test and PBB test, and compared with Wald test, Welch test, GP test and PB test based on type I error rates and power of the tests. For $k = 3$, the PB test, PBJ test and PBB test can control type I error rates very well where three tests are better than the Wald test and Welch test at the nominal level 0.05 under equal variance and unequal variance. Considering power of the tests for $k = 3$, the results indicated that all the tests have power of the test quit well. When $k = 6$, the PB test, PBJ test and PBB test can control type I error rates very well when the nominal level is 0.05. Based on power the tests in the case of $k = 6$, the six tests have high power of the test. From the results of the type I error rates, PB test, PBB test and PBJ test are the best statistics since these are nearly the nominal level of 0.05 more than the other tests. However, we suggest PBB test to be used as an alternative approach for testing the equality of means when the assumptions of homogeneity of variance are violated in one-way ANOVA since it controls type I error rates very well comparable to PB test and PBJ test, and it has high power of the test better than PB test and PBJ test.

Acknowledgments

The author would like to thank Kalasin University for supporting the research.

References

Akritas M, Papadatos N. Heteroscedastic one-way ANOVA and lack-of-fit tests. J Am Stat Assoc. 2004; 99: 368-382.

Ana S, Joao AC, Ashley RD, Mario R, Francisco RP. Evaluation of Jackknife and bootstrap for defining confidence intervals for pairwise agreement measures. PLoS One. 2011 [cited 2019 January 5]; 6(5). Available from: https://doi.org/10.1371/journal.pone.0019539.

Box GEP. Some theorems on quadratic forms in the study of analysis of variance problems. I. Effect of inequality of variance in the one-way classification. Ann math stat.1954; 25: 290-302.

Fisher RA. The fiducial argument in statistical inference. Ann Eugen. 1935; 8: 391-398.

Krishnamoorthy K, Lu F, Mathew T. A parametric bootstrap approach for ANOVA with unequal variances: Fixed and random models. Comput Stat Data Anal. 2007; 51: 5731-5742.

Moder K. How to keep the type I error rate in ANOVA if Variances are Heteroscedastic. Austrian J Stat. 2007; 36: 179-188.

R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing. Vienna: Austria; 2018 [cited 2019 Jan 20]. Available from: https://www.R-project.org.

Weerahandi S. ANOVA under unequal error variances. Biometrics. 1995a; 51: 589–599.

Welch BL. On the comparison of several mean values: An alternative approach. Biometrika. 1951; 38: 330-336.