Hacettepe Journal of Mathematics and Statistics
Volume 41 (4) (2012), 605 – 613

A TEST BASED ON THE COMPUTATIONAL APPROACH FOR EQUALITY OF MEANS UNDER THE UNEQUAL VARIANCE ASSUMPTION

Esra Yiğit Gökpınar*† and Fikri Gökpınar*

Received 05 : 01 : 2011 : Accepted 01 : 11 : 2011

Abstract

The classical F-test to compare several populations means depends on the assumption of homogeneity of variances of the population and on normality. When these assumptions - especially the equality of variance - is dropped, the classical F-test fails to reject the null hypothesis even if the data actually provide strong evidence for it. This can be considered a serious problem in some applications especially when the sample sizes are not large. To deal with this problem, a number of tests are available in the literature. Recently Pal, Lim and Ling (*A computational approach to statistical inferences,* J. Appl. Probab. Stat. **2** (1), 13–35, 2007) developed a computational technique, called the Computational Approach Test (CAT), which looks similar to a parametric bootstrap for hypothesis testing. Chang and Pal (*A revisit to the Behren-Fisher Problem: Comparison of five test methods*, Communications in Statistics - Simulation and Computation **37** (6), 1064–1085, 2008) applied CAT to test the equality of two population means when the variances are unknown and arbitrary. In this study we apply a developed CAT to test the equality of k population means when the variances are unequal. Also the Brown-Forsythe, Weerahandi's Generalized F, Parametric Bootstrap and Welch tests are recalled and a simulation study performed to compare these tests according to type one errors and powers in different combinations of parameters and various sample sizes.

Keywords: Brown-Forsythe Test, Computational Approach Test, Generalized F test, Parametric Bootstrap Test, Classic F Test, Welch Test.

2000 AMS Classification: 62 F 03, 62 F 40.

*Gazi University, Faculty of Science, Department of Statistics, Teknikokullar, Ankara, Turkey.
E-mail: (E. Y. Gökpınar) eyigit@gazi.edu.tr (F. Gökpınar) fikri@gazi.edu.tr
†Corresponding Author.

1. Introduction

In applied statistics an experimenter wants to compare two or more populations measured on independent samples. The classical F (CF) test is used under the assumption that the populations have normal distributions with the same variances. Bishop and Dudewicz [1] showed that the CF test is not robust when the population variances are unequal, especially if the sample sizes are not equal. Also Krutchkoff [8] and Lee and Ahn [9] showed that the empirical type I errors are much greater than the nominal level α, especially when sample sizes are negatively related with their population variances.

Alternative methods are developed due to this problem. For some of these test statistics the distribution is not known and the p-value can be found by simulation [13, 14]. Approximate methods are used quite often with the development of computer technology. Also these tests have been applied to solve a number of problems when conventional methods are difficult to apply or fail to provide exact solutions. In practice, some exact procedures such as the CF, Welch (W) and Brown-Forsythe (BF) tests are widely used [2, 15].

Tsui and Weerahandi [11] generalized the conventional definition of the p value so that problems such as the Behrens-Fisher problem can be resolved. Weerahandi [12] defined the notion of generalized p-value for comparing the means of k populations when the variances are not equal. Krishnamoorthy *et al.* [7] proposed a Parametric Bootstrap test (PB); Xu and Wang [16, 17] developed a generalized F-test based on the generalized value p (XW).

Recently Pal *et al.* [10] developed a computational technique, called the Computational Approach Test (CAT) which looks similar to the parametric bootstrap for hypothesis testing. Chang and Pal [3] applied CAT to test the equality of two population means when the variances are unknown and arbitrary. Also Chang *et al.* [4] have demonstrated that for homoscedastic one-way ANOVA the CAT is as powerful as the classical F test.

In this paper we apply CAT to test the equality of k population means when the variances are unknown and arbitrary. In the following section, we describe the *W, BF*, generalized F (GF) due to Weerahandi [13], PB and CAT tests. Monte Carlo comparison studies by Gamage and Weerahandi [5], Gerami and Zahiden [6] and Yiğit and Gökpınar [18] showed that, out of these and other tests, only the W, GF and PB tests emerged satisfactory provided the sample sizes are moderate or large. So we chose these tests for the simulation study. These methods are compared according to type I errors and powers in different combinations of parameters and various sample sizes.

2. Tests for one way ANOVA

Let $X_{i1}, \ldots, X_{in_i}$ be a random sample from $N(\mu_i, \sigma_i^2)$, $i = 1, \ldots, k$. The problem of interest involves testing

(1) $\quad H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$ against H_A : Not all μ_is are equal , $i = 1, \ldots, k$.

If the σ_i^2's are unequal, then the testing procedure given in equation (1) defines the standardized between–group sum of squares in equation (2) and the standardized error sum of squares in equation (3).

$$\tilde{S}_b = \tilde{S}_b\left(\sigma_1^2, \ldots, \sigma_k^2\right) = \sum_{i=1}^{k} \frac{n_i \bar{X}_i^2}{\sigma_i^2} - \frac{\left(\sum_{i=1}^{k} \frac{n_i \bar{X}_i}{\sigma_i^2}\right)^2}{\sum_{i=1}^{k} \frac{n_i}{\sigma_i^2}}, \tag{2}$$

$$\tilde{S}_e = \sum_{i=1}^{k} \frac{(n_i - 1)\, S_i^2}{\sigma_i^2}, \tag{3}$$

where S_i^2 is the unbiased estimator of σ_i^2 and $\bar{X}_i$ is the estimator of μ_i. Most of the test statistics to test the equality of means under heteroscedasticity are based on the standardized between-group sum of squares and standardized error sum of squares. In the rest of this section some test statistics are briefly introduced.

The Welch Test

If $w_i = \frac{n_i}{S_i^2}$, Welch [15] gives the following test statistics [14]:

$$W = \frac{\tilde{S}_b\left(S_1^2, \ldots, S_k^2\right)/(k-1)}{1 + \frac{2(k-2)}{k^2-1}\sum_{i=1}^{k}\frac{1}{n_i-1}\left(1-\frac{w_i}{\sum w_j}\right)^2} = \frac{\sum_{i=1}^{k} w_i\left[\left(\bar{X}_i-\bar{X}\right)^2/(k-1)\right]}{1+\frac{2\,(k-2)}{k^2-1}\sum_{i=1}^{k}\frac{1}{n_i-1}\left(1-\frac{w_i}{\sum w_j}\right)^2}, \tag{4}$$

where $\bar{X}$ is the estimator of the overall mean. If H_0 is true, then the distribution of W is $F_{k-1,f}$, where

$$f = \frac{1}{\frac{3}{k^2-1}\sum_{i=1}^{k}\frac{1}{n_i-1}\left(1-\frac{w_i}{\sum w_j}\right)^2}.$$

For a given level α, and an observed value W_h of W, this test rejects the H_0 in equation (1) whenever the p-value is given as $P\left(F_{k-1,f} > W_h\right) < \alpha$.

The Brown-Forsythe Test

Brown and Forsythe [2] give the following test statistics.

$$B = \frac{\sum_{i=1}^{k} n_i\left(\bar{X}_i - \bar{X}\right)^2}{\sum_{i=1}^{k}\left(1-\frac{n_i}{n}\right)S_i^2}.$$

If H_0 is true, then the distribution of B is $F_{k-1,v}$, where

$$v = \frac{\left[\sum_{i=1}^{k}\left(1-\frac{n_i}{n}\right)S_i^2\right]^2}{\sum_{i=1}^{k}\frac{\left(1-\frac{n_i}{n}\right)^2 S_i^4}{n_i-1}}.$$

For a given level α and an observed value B_h of B, this test rejects the H_0 in equation (1) whenever the p-value is given as $P\left(F_{k-1,v} > B_h\right) < \alpha$.

The Weerahandi's Generalized F test

The sample variances (MLEs) of the k populations are denoted by S_i^2, where $S_i^2 = \frac{1}{n_i}\sum_{i=1}^{n_i}\left(X_{ij} - \bar{X}_i\right)^2$. Define

$$B_j = \frac{\left(\sum_{i=1}^{j}\frac{n_i S_i^2}{\sigma_i^2}\right)}{\left(\sum_{i=1}^{j+1}\frac{n_i S_i^2}{\sigma_i^2}\right)} \sim Beta\left(\sum_{i=1}^{j}\frac{(n_i-1)}{2},\ \frac{(n_{j+1}-1)}{2}\right),\ j = 1, \ldots, k-1$$

where $\tilde{S}_e$ and B_j are all independent random variables. Note that the random variables $\frac{n_i S_i^2}{\sigma_i^2}$ can be expressed as

$$\begin{aligned}
\frac{n_1 S_1^2}{\sigma_1^2} &= \tilde{S}_e B_1 B_2 \ldots B_{k-1}, \\
\frac{n_i S_i^2}{\sigma_i^2} &= \tilde{S}_e\left(1-B_{i-1}\right) B_i \ldots B_{k-1} \text{ for } i = 2, \ldots, k-1, \\
\frac{n_k S_k^2}{\sigma_k^2} &= \tilde{S}_e\left(1-B_{k-1}\right).
\end{aligned}$$

Therefore, the generalized p value can be expressed as

$$p = 1 - E\Bigg(H_{k-1,n-k}\Bigg\{ \frac{n-k}{k-1}\tilde{s}_b\Bigg[\frac{n_1 s_1^2}{B_1 B_2 \dots B_{k-1}}, \frac{n_2 s_2^2}{(1-B_1)B_2 \dots B_{k-1}}, \dots, \frac{n_k s_k^2}{(1-B_{k-1})}\Bigg]\Bigg\}\Bigg), \tag{5}$$

where $H_{k-1,n-k}$ is the cumulative distribution function of the F-distribution with $k-1$ and $N-k$ degrees of freedom. This test rejects the H_0 in equation (1) whenever $p < \alpha$ [12].

The parametric boostrap test

The PB approach is defined as follows. In the case where the population variances σ_i^2's are unknown; a test statistic can be obtained by replacing σ_i^2 in equation (2) by S_i^2 and is given by

$$\tilde{S}_b\left(S_1^2, \dots, S_k^2\right) = \sum_{i=1}^{k} \frac{n_i \bar{X}_i^2}{S_i^2} - \frac{\left(\sum_{i=1}^{k} \frac{n_i \bar{X}_i}{S_i^2}\right)^2}{\sum_{i=1}^{k} \frac{n_i}{S_i^2}}. \tag{6}$$

As the test statistic in equation (6) is location invariant, without loss of generality, we can take the common mean to be zero [7].

Let $\bar{X}_{Bi} \sim Z_i\left(\frac{S_i}{\sqrt{n_i}}\right)$, where Z_i is a standard normal random variable and $S_{Bi}^2 \sim S_i^2 \chi_{n_i-1}^2 / (n_i - 1)$. Then the PB pivot variable can be obtained by replacing $\bar{X}$, S_i^2 in equation (6) by $\bar{X}_{Bi}$, S_{Bi}^2, and is given by

$$S_{bB} = \sum_{i=1}^{k} \frac{n_i}{S_{Bi}^2} \bar{X}_{Bi}^2 - \frac{\left[\sum_{i=1}^{k} \frac{n_i \bar{X}_{Bi}}{S_{Bi}^2}\right]^2}{\sum_{i=1}^{k} \frac{n_i}{S_{Bi}^2}} = \sum_{i=1}^{k} \frac{Z_i^2 (n_i - 1)}{\chi_{n_i-1}^2} - \frac{\left[\sum_{i=1}^{k} \frac{\sqrt{n_i} Z_i (n_i-1)}{S_i \chi_{n_i-1}^2}\right]^2}{\sum_{i=1}^{k} \frac{n_i (n_i-1)}{S_i^2 \chi_{n_i-1}^2}}.$$

H_0 is rejected if $P\{\tilde{S}_{bB}\left(Z_i, \chi_{n_i-1}^2; s_i^2\right) > \tilde{s}_b\} < \alpha$.

The test based on CAT

In this section, we give a test procedure based on CAT for one-way ANOVA when the population variances are unequal. The main point in using the CAT approach in this problem is to obtain the Restricted Maximum Likelihood (RML) estimators of the parameters. The RML estimators can be obtained as follows.

The RML method gives the ML estimator of μ and σ_i^2 based on the assumption that H_0 ($\mu_1 = \mu_2 = \dots = \mu_k = \mu$). The likelihood function of the sample $(X_{11}, \dots, X_{kn_k})$ can be given as below.

$$L = \frac{1}{(2\pi)^{\sum_{i=1}^{k} n_i/2.} \prod_{i=1}^{k} \sigma_i^{n_i}} \exp\left\{ -\frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{n_i} \frac{(X_{ij} - \mu)^2}{\sigma_i^2} \right\},$$

$$\begin{aligned} L^* = \ln(L) &= -\sum_{i=1}^{k} \frac{n_i}{2} \ln(2\pi) - \sum_{i=1}^{k} n_i \ln(\sigma_i) - \frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{n_i} \frac{(X_{ij} - \mu)^2}{\sigma_i^2} \\ &= -\sum \frac{n_i}{2} \ln(2\pi) - \sum_{i=1}^{k} \frac{n_i}{2} \ln\left(\sigma_i^2\right) - \frac{1}{2} \frac{\sum_{i=1}^{k} \left(n_i S_i^2 + n_i \left(\bar{X}_i - \mu\right)^2\right)}{\sigma_i^2}. \end{aligned} \tag{7}$$

By differentiating Equation (7) with respect to μ and σ_i^2, the following equations can be obtained.

$$\hat{\mu} = \frac{\sum_{i=1}^{k} \frac{n_i}{\hat{\sigma}_i^2} \bar{X}_i}{\sum_{i=1}^{k} \frac{n_i}{\hat{\sigma}_i^2}},$$

$$\hat{\sigma}_i^2 = S_i^2 + \left(\frac{\sum_{i=1}^{k} \frac{n_i}{\hat{\sigma}_i^2} D_{ij}}{\sum_{i=1}^{k} \frac{n_i}{\hat{\sigma}_i^2}} \right)^2,$$

where $D_{ij} = \bar{X}_i - \bar{X}_j$ and S_i^2 is the sample variance (MLEs) of the i^{th} population.

To apply the developed CAT we first express H_0 in terms of a suitable scalar η. Define η as

$$\eta = \eta(\mu_1, \ldots, \mu_k) = \sum_{i=1}^{k} n_i \left(\mu_i - \bar{\mu}\right)^2, \ \bar{\mu} = \sum_{i=1}^{k} n_i \mu_i / n, \ n = \sum_{i=1}^{k} n_i.$$

It is seen that testing H_0 against H_A in equation (1) is equivalent to testing $H_0^* : \eta = 0$ against $H_A^* = \eta > 0$. To test the hypothesis in equation (1) we use η as a test statistic. If H_0 is true then $\eta = 0$ otherwise η becomes larger than 0.

The test procedure is as follows:

1) The ML estimators of the parameter are $\hat{\mu}_{i(ML)} = \bar{X}_{i\cdot}$, $\hat{\sigma}^2_{i(ML)} = S_i^2$. Also the test statistic is $\hat{\eta}_{ML} = \sum_{i=1}^{k} n_i \left(\bar{X}_{i\cdot} - \bar{X}\right)^2$

2) If H_0 is true, $X_{ij} \sim N\left(\mu, \sigma_i^2\right) \quad (1 \le j \le n_i, \ 1 \le i \le k)$,

$$\hat{\mu}_{RML} = \frac{\sum_{i=1}^{k} \left(n_i \bar{X}_i / \hat{\sigma}^2_{i(RML)}\right)}{\sum_{i=1}^{k} \left(n_i / \hat{\sigma}^2_{i(RML)}\right)},$$

$$\hat{\sigma}^2_{i(RML)} = S_i^2 + \left(\frac{\sum_{i=1}^{k} \frac{n_i}{\hat{\sigma}^2_{i(RML)}} D_{ij}}{\sum_{i=1}^{k} \frac{n_i}{\hat{\sigma}^2_{i(RML)}}} \right)^2, \ i = 1, \ldots, k.$$

As seen from these equations, the RML estimates of the μ and σ_i^2 parameters have no closed forms. Therefore, the RML estimates of these parameters can be obtained using iteration.

3) Generate $\bar{X}_i$ M times such that

$$\bar{X}_{i\cdot} \sim N\left(\hat{\mu}_{RML}, \hat{\sigma}_{i(RML)}/n_i\right). \tag{8}$$

4) For each replication of $\bar{X}_i$, calculate $\hat{\eta}^*_{ML} = \sum_{i=1}^{k} n_i \left(\hat{\mu}_{i(ML)} - \hat{\bar{\mu}}_{ML}\right)^2$.

5) Calculate $p = \frac{\#\left(\hat{\eta}_{ML} > \hat{\eta}^*_{ML}\right)}{M}$; when $p < \alpha$, H_0 is rejected.

3. Simulation study

In this section we compare the CF, BF, W, GF, PB and CAT tests according to type I errors and powers in different combinations of parameters and sample sizes. We consider the balanced and unbalanced cases from smaller to larger sample sizes where $k = 3$ and $k = 5$ for comparing the tests. The values for the variances vary over a large range so that $\sigma_1^2 < \ldots < \sigma_k^2$ and $\sigma_1^2 > \ldots > \sigma_k^2$.

For each combination of n_i and σ_i^2 the rejection rate of each testing procedure is calculated and compared with the nominal level 0.05 when the means are all equal. To estimate the type I error rates of the CF, W and BF tests, we used simulation consisting

of 5000 runs for each of the sample sizes and parameter configurations. The CF, W and BF test statistics are calculated from these generated data and type I errors are estimated by the proportion of test statistics that exceed the critical values calculated from the distributions. To estimate the type I error rates of the GF, PB and CAT tests, we use a two-step simulation. For estimating the type I error rates of the CAT test we generate 5000 observed vectors in equation (8) and use 5000 runs for each observed vector to estimate the p value.

Finally the type I error rates of the CAT test are estimated by the proportion of the 5000 p-values that are less than the nominal level α. The type I error rates of the PB and GF tests are similarly estimated. In both cases of equal and unequal variances for $k = 3$ and $k = 5$ simulated type I error rates are given in Tables 1 and 2 respectively.

We observed the following from the numerical results in Tables 1 and 2.

As seen from Table 1 the new test based on CAT has empirical type I error rates close to 0.05 when $k = 3$. Also the BF, GF, PB, W tests are close to 0.05 in this situation. But the W, PB and GF tests appear to be more powerful than the CAT test when $k = 3$ and the sample sizes are small $((n_1, n_2, n_3) = (7, 9, 11))$.

When the n_i 's and σ_i^2 's are in reverse order the CF test's type I error rate is getting far away from its nominal level. When the differences between the means of the groups increase, the power of the new CAT test is superior to the other test for almost every combination of sample sizes and population variances. When the group size is 5 and the n_i 's and σ_i^2 's are in reverse order the type I error rate of CF is approximately around 0.10. This is not a acceptable level for a type I error. Also the type I error rate of GF is greater than its nominal level, especially when the sample sizes are small. The type I error of the new test based on CAT is smaller than its nominal level especially if the n_i 's and σ_i^2 's are in reverse order. When we investigate the power of the tests we can easily see that the new test is superior to the other tests.

4. Conclusion

In this simulation study for a range of choices of sample sizes and parameter configurations we compared the performance of the above tests for testing the equality of means of one-way ANOVA models under heteroscedasticity. The CF test is not an appropriate test for heteroscedasticity because its type I error rates exceed the nominal level. The type I errors of the new test based on CAT are close to the nominal level. The CAT test appears to be more powerful than the other tests under almost every situation except when $k = 3$ and the sample sizes are small $((n_1, n_2, n_3) = (7, 9, 11))$.

Table 1. Simulated type I error rates and powers when $k = 3$ for $\alpha = 0.05$

n_i	σ_i	(μ_1, μ_2, μ_3)	CF	BF	GF	PB	W	CAT
10, 10, 10	1, 1.25, 1.5	(0, 0, 0)	0.061	0.058	0.045	0.048	0.049	0.047
		(0, 0.2, 0.4)	0.089	0.086	0.076	0.084	0.084	0.085
		(0, 0.4, 0.8)	0.221	0.209	0.196	0.21	0.208	0.222
		(0, 0.6, 1.2)	0.423	0.412	0.383	0.396	0.398	0.482
		(0, 0.8, 1.6)	0.655	0.642	0.621	0.632	0.633	0.757
		(0, 1, 2)	0.857	0.847	0.844	0.854	0.854	0.920
		(0, 1.2, 2.4)	0.955	0.949	0.945	0.953	0.952	0.984
30, 30, 30	1, 1.25, 1.5	(0, 0, 0)	0.051	0.050	0.047	0.050	0.048	0.047
		(0, 0.2, 0.4)	0.193	0.191	0.180	0.185	0.186	0.214
		(0, 0.4, 0.8)	0.571	0.569	0.574	0.579	0.579	0.677
		(0, 0.6, 1.2)	0.908	0.908	0.911	0.913	0.914	0.964
		(0, 0.8, 1.6)	0.995	0.995	0.995	0.995	.0995	0.999
		(0, 1, 2)	1.000	1.000	1.000	1.000	1.000	1.000
		(0, 1.2, 2.4)	1.000	1.000	1.000	1.000	1.000	1.000
7, 9, 11	1, 1.25, 1.5	(0, 0, 0)	0.046	0.055	0.039	0.048	0.047	0.039
		(0, 0.2, 0.4)	0.067	0.072	0.068	0.073	0.073	0.064
		(0, 0.4, 0.8)	0.165	0.183	0.171	0.186	0.187	0.184
		(0, 0.6, 1.2)	0.328	0.357	0.337	0.368	0.365	0.413
		(0, 0.8, 1.6)	0.561	0.593	0.565	0.596	0.594	0.665
		(0, 1, 2)	0.758	0.787	0.760	0.781	0.779	0.862
		(0, 1.2, 2.4)	0.903	0.918	0.903	0.915	0.913	0.962
	1.5, 1.25, 1	(0, 0, 0)	0.064	0.048	0.043	0.048	0.045	0.041
		(0, 0.2, 0.4)	0.103	0.077	0.077	0.083	0.083	0.072
		(0, 0.4, 0.8)	0.217	0.172	0.170	0.177	0.183	0.177
		(0, 0.6, 1.2)	0.396	0.330	0.326	0.337	0.338	0.385
		(0, 0.8, 1.6)	0.613	0.541	0.532	0.538	0.536	0.611
		(0, 1, 2)	0.820	0.758	0.769	0.771	0.771	0.823
		(0, 1.2, 2.4)	0.935	0.894	0.890	0.894	0.895	0.943
20, 25, 30	1, 1.25, 1.5	(0, 0, 0)	0.044	0.054	0.054	0.056	0.057	0.043
		(0, 0.2, 0.4)	0.128	0.148	0.151	0.155	0.158	0.177
		(0, 0.4, 0.8)	0.435	0.477	0.477	0.486	0.487	0.578
		(0, 0.6, 1.2)	0.826	0.849	0.847	0.852	0.852	0.919
		(0, 0.8, 1.6)	0.970	0.976	0.978	0.979	0.980	0.997
		(0, 1, 2)	0.998	0.999	0.998	0.998	0.998	1.000
		(0, 1.2, 2.4)	1.000	1.000	1.000	1.000	1.000	1.000
	1.5, 1.25, 1	(0, 0, 0)	0.065	0.051	0.051	0.052	0.051	0.049
		(0, 0.2, 0.4)	0.167	0.140	0.141	0.146	0.143	0.161
		(0, 0.4, 0.8)	0.496	0.446	0.459	0.461	0.461	0.554
		(0, 0.6, 1.2)	0.815	0.815	0.827	0.828	0.825	0.910
		(0, 0.8, 1.6)	0.979	0.971	0.973	0.975	0.975	0.993
		(0, 1, 2)	0.999	0.998	0.999	0.999	0.999	1.000
		(0, 1.2, 2.4)	1.000	1.000	1.000	1.000	1.000	1.000

Table 2. Simulated type I error rates and powers when $k = 5$ for $\alpha = 0.05$

n_i	σ_i	$(\mu_1, \mu_2, \mu_3, \mu_4, \mu_5)$	CF	BF	GF	PB	W	CAT
		(0, 0, 0, 0, 0)	0.055	0.049	0.054	0.049	0.049	0.038
		(0, 0.2, 0.4, 0.6, 0.8)	0.135	0.124	0.147	0.123	0.131	0.154
		(0, 0.4, 0.8, 1.2, 1.6)	0.461	0.440	0.525	0.482	0.493	0.624
10, 10, 10,	1, 1.25, 1.5, 1.75, 2	(0, 0.6, 1.2, 1.8, 2.4)	0.847	0.832	0.847	0.853	0.859	0.959
10, 10		(0, 0.8, 1.6, 2.4, 3.2)	0.984	0.982	0.988	0.983	0.985	0.999
		(0, 1, 2, 3, 4)	0.999	0.999	1.000	0.999	0.998	1.000
		(0, 1.2, 2.4, 3.6, 4.8)	1.000	1.000	1.000	1.000	1.000	1.000
		(0, 0, 0, 0, 0)	0.055	0.054	0.048	0.044	0.045	0.045
		(0, 0.2, 0.4, 0.6, 0.8)	0.376	0.375	0.421	0.407	0.409	0.559
		(0, 0.4, 0.8, 1.2, 1.6)	0.955	0.955	0.971	0.968	0.969	0.996
30, 30, 30,	1, 1.25, 1.5, 1.75, 2	(0, 0.6, 1.2, 1.8, 2.4)	1.000	1.000	1.000	1.000	1.000	1.000
30, 30		(0, 0.8, 1.6, 2.4, 3.2)	1.000	1.000	1.000	1.000	1.000	1.000
		(0, 1, 2, 3, 4)	1.000	1.000	1.000	1.000	1.000	1.000
		(0, 1.2, 2.4, 3.6, 4.8)	1.000	1.000	1.000	1.000	1.000	1.000
		(0, 0, 0, 0, 0)	0.042	0.062	0.065	0.053	0.063	0.045
		(0, 0.2, 0.4, 0.6, 0.8)	0.099	0.142	0.156	0.141	0.145	0.177
		(0, 0.4, 0.8, 1.2, 1.6)	0.415	0.504	0.537	0.512	0.519	0.659
	1, 1.25, 1.5,1.75,2	(0, 0.6, 1.2, 1.8, 2.4)	0.815	0.870	0.908	0.886	0.889	0.968
		(0, 0.8, 1.6, 2.4, 3.2)	0.982	0.991	0.994	0.994	0.994	0.998
		(0, 1, 2, 3, 4)	1.000	1.000	1.000	0.999	0.999	0.999
7, 9, 11,		(0, 1.2, 2.4, 3.6, 4.8)	1.000	1.000	1.000	1.000	1.000	1.000
13, 15		(0, 0, 0, 0, 0)	0.102	0.053	0.063	0.054	0.057	0.036
		(0, 0.2, 0.4, 0.6, 0.8)	0.207	0.133	0.158	0.132	0.141	0.150
		(0, 0.4, 0.8, 1.2, 1.6)	0.553	0.428	0.505	0.455	0.466	0.604
	2, 1.75, 1.5, 1.25, 1	(0, 0.6, 1.2, 1.8, 2.4)	0.906	0.808	0.900	0.857	0.869	0.937
		(0, 0.8, 1.6, 2.4, 3.2)	0.994	0.978	0.989	0.986	0.986	0.995
		(0, 1, 2, 3, 4)	1.000	0.999	1.000	0.999	0.999	0.999
		(0, 1.2, 2.4, 3.6, 4.8)	1.000	1.000	1.000	1.000	1.000	1.000
		(0, 0, 0, 0, 0)	0.037	0.047	0.053	0.048	0.050	0.048
		(0, 0.2, 0.4, 0.6, 0.8)	0.273	0.325	0.362	0.352	0.353	0.460
		(0, 0.4, 0.8, 1.2, 1.6)	0.872	0.898	0.932	0.927	0.927	0.988
	1, 1.25, 1.5, 1.75, 2	(0, 0.6, 1.2, 1.8, 2.4)	0.999	0.999	1.000	1.000	1.000	1.000
		(0, 0.8, 1.6, 2.4, 3.2)	1.000	1.000	1.000	1.000	1.000	1.000
		(0, 1, 2, 3, 4)	1.000	1.000	1.000	1.000	1.000	1.000
19, 22, 25,		(0, 1.2, 2.4, 3.6, 4.8)	1.000	1.000	1.000	1.000	1.000	1.000
28, 30		(0, 0, 0, 0, 0)	0.078	0.056	0.058	0.052	0.052	0.044
		(0, 0.2, 0.4, 0.6, 0.8)	0.356	0.306	0.335	0.326	0.326	0.435
		(0, 0.4, 0.8, 1.2, 1.6)	0.923	0.891	0.932	0.924	0.925	0.985
	2, 1.75, 1.5, 1.25, 1	(0, 0.6, 1.2, 1.8, 2.4)	0.999	0.998	1.000	0.999	1.000	1.000
		(0, 0.8, 1.6, 2.4, 3.2)	1.000	1.000	1.000	1.000	1.000	1.000
		(0, 1, 2, 3, 4)	1.000	1.000	1.000	1.000	1.000	1.000
		(0, 1.2, 2.4, 3.6, 4.8)	1.000	1.000	1.000	1.000	1.000	1.000

References

[1] Bishop, T. A. and Dudewicz, E. J. *Heteroscedastic ANOVA*, Sankhya **43** B, 40–57, 1981.

[2] Brown, M. B. and Forsythe, A. B. *The small sample behavior of some statistics which test the equality of several means*, Technometrics **16**, 129–132, 1974.

[3] Chang, C. H. and Pal, N. *A revisit to the Behren-Fisher Problem: Comparison of five test methods*, Communications in Statistics-Simulation and Computation **37** (6), 1064–1085, 2008.

[4] Chang, C. H., Pal, N. and Lim, W K. *Comparing several population means: a parametric bootstrap method, and its comparison with usual ANOVA F test as well as ANOM,* Computational Statistics **25** (1), 71–95, 2010.

[5] Gamage, J. ve Weerahandi, S. *Size performance of some tests in one-way ANOVA*, Communications in Statistics Simulations **27** (3), 625–640, 1998.

[6] Gerami, A. and Zahedian, A. *Comparing the means of normal populations with unequal variances* (Proceedings of the 53rd Session of Intenational Statistical Institute, Seoul, Korea, 2001).

[7] Krishnamoorthy, K., Lu, F. and Thomas, M. *A parametric boostrap approach for ANOVA with unequal variances: fixed and random models*, Computational Statistics and Data Analysis, **51**, 5731–5742, 2006.

[8] Krutchkoff, R. G. *One-way fixed effects analysis of variance when the error variances may be unequal*, J. Statist. Comput. Simulation **30**, 259–271, 1988.

[9] Lee, S. and Ahn, C. H. *Modified ANOVA for unequal variances*, Communications in Statistics Simulations **32**, 987–1004, 2003.

[10] Pal, N., Lim, W K. and Ling, C H. *A computational approach to statistical inferences,* J. Appl. Probab. Stat. **2** (1), 13–35, 2007.

[11] Tsui, K. and Weerahandi, S. *Generalized p-values in significance testing of hypotheses in the presence of nuisance parametres*, Journal of the American Statistical Assocation **84**, 602–607, 1989.

[12] Weerahandi, S. *ANOVA under unequal error variances*, Biometrica **38**, 330–336, 1995.

[13] Weerahandi, S. *Exact Statistical Method for Data Analysis* (Springer-Verlag, New York, 1995).

[14] Weerahandi, S. *Generalized Inference in Repeated Measures: Exact Methods in MANOVA and Mixed Models* (Wiley, New York, 2004).

[15] Welch, B. L. *On the comparison of several mean values: An alternative approach*, Biometrica **38**, 330–336, 1951.

[16] Xu, L. and Wang, S. *A new generalized p-value for ANOVA under heteroscedasticity*, Statistics and Probability Letters **78**, 963–969, 2007.

[17] Xu, L. and Wang, S. *A new generalized p-value and its upper bound for ANOVA under unequal erros variances*, Communications in Statistics Theory and Methods **37**, 1002–1010, 2007.

[18] Yiğit, E. and Gökpınar, F. *A simulation study on tests for one-way ANOVA under the unequal variance assumption*, Commun. Fac. Sci. Univ. Ank. Series A **12**, 15–34, 2010.