Hacettepe Journal of Mathematics and Statistics
Volume 46 (3) (2017), 493 – 510

Modified tests for comparison of group means under heteroskedasticity and non-normality caused by outlier(s)

Mustafa Cavus*† , Berna Yazici‡ , Ahmet Sezer§

Abstract

There are several approximate tests proposed such as Welch's F-test (W), the Parametric Bootstrap Test (PB) and Generalized F-test (GF) for comparing several group means under heteroskedasticity. These tests are powerful and have nominal type 1 error rates but they are not performing satisfactorily under non-normality caused by outlier(s). To handle this problem, we investigate tests that are powerful and provide nominal type 1 error rates by using robust estimators both for location and scale parameters. The performance of the modified tests are examined with Monte-Carlo simulation studies. Results of simulations clearly indicate that Generalized F-test modified with Huber's M-estimators achieves the nominal type 1 error rate and provide higher power than alternative methods.

Keywords: Outlier, Heteroskedasticity, Unbalanced, Huber's M-Estimators, Generalized F-test.

2000 AMS Classification: 60E05

Received : 23.06.2016 *Accepted* : 25.07.2016 *Doi* : Doi: 10.15672/HJMS.2017.417

*Anadolu University, Faculty of Science, Department of Statistics, Email: mustafacavus@anadolu.edu.tr

†Corresponding Author.

‡Anadolu University, Faculty of Science, Department of Statistics, Email: bbaloglu@anadolu.edu.tr

§Anadolu University, Faculty of Science, Department of Statistics, Email: a.sezer@anadolu.edu.tr

1. Introduction

Classical F-test (CF) is most commonly used procedure in testing the equality of group means when the assumptions are hold. When one or more of these assumptions are violated, CF test can give wrong results. Some modifications based on weighting are used for solving this problem especially in the case of violation of variance homogeneity [2][3][17][20][31]. Welch (1951) proposed a test procedure based on weighting for comparing the several population means with unequal variances. Welch recalculated the test statistics of Classical F-test with the ratio of sample sizes and sample variances. Equality of group means under heteroskedasticity can be tested safely by Welch procedure. With the developments in the computer simulations, many procedures are improved in the statistical analysis such as bootstrap methods. Krishnamoorthy (2007) proposed the Parametric Bootstrap test based on bootstrap sampling for comparing the group means under heteroskedasticity. Also, Weerahandi (1995) improved the Generalized F-test (GF) based on generalized p-value (GPV) method that can be calculated with Monte-Carlo simulations for comparing group means under heteroskedasticity. Generalized p-value method proposed by Tsui and Weerahandi (1989) provides test statistics independently of nuisance(unknown) parameters. Besides the Generalized F-test, many procedures are proposed based on GPV method for comparing the group means [10][23][28][30]. Gamage and Weerahandi (1998) compared the performance of the tests used for comparing the group means under heteroskedasticity with Generalized F-test and they indicated GF test is more powerful than the alternatives. GF test is not only used for normal distributed groups but also used for skewed distributions such as log-normal, inverse normal[10][12][16][21][22][23][24][25]. PB is not only used for normal distributed groups [4][5].

Some approximate tests such as W, PB and GF are considered in this study for comparing of group means. These tests are constructed on normality assumption so the violation of that assumption may cause the decrement in the power of the tests. Normality can be violated by many reasons. One of the most common reason of this violation is outlier. Outlier can be explained as extremely high or low value in the data[1]. Outlier has been a basic problem for statistical analysis for many decades. After detection of the outlier, it is ignored in the analyses in earlier studies. Later, some transformation methods are used for rasping the effect of the outlier. Robust methods are improved without using any ignoring or transformation since these methods cause missing information. First Box (1953) proposed several robust methods which can be used in the presence of outlier.

In this study, it is aimed that to obtain powerful methods for comparing group means under heteroskedasticity and non-normality caused by outlier. Although approximate tests are powerful under heteroskedasticity, power of the tests are decreasing in the case of violation of normality and type 1 error rates of the tests inflate.[6] We modify the approximate tests because of the performance of the robust procedures in the presence of outlier. For obtaining these modifications, the maximum likelihood mean and variance estimators are replaced with robust estimators such as trimmed mean and variance, median and median absolute deviation, Huber's M-estimators. There are many studies that includes modified tests based on robust estimators [2][18][19][31][32][33][41]. Tan and Tabatabai (1985) proposed the modified Brown-Forsythe test with Huber's M-estimators and compared it with previous methods for presenting the efficiency of the proposed method. Wilcox (1995) proposed modification to F-test using with trimmed mean towards the outlier effect. Fan and Hancock (2012) proposed Robust Mean Modelling approach. Karagoz (2015) tried to compare Weibull distributed group means with modified Welch F-test using with robust estimators. Karagoz and Saracbasi (2016) proposed a modification to Brown-Forsythe test for the same purposes.

In this study, we modified the Generalized F-test, Parametric Bootstrap Test and Welch's F-test to handle the outlier(s) by replacing the MLE estimators with some robust estimators. The performances of these tests are examined in detail in terms of power of the test and type 1 error rates with Monte-Carlo simulations.

The paper is organized as follows. The approximate tests are introduced in Section 2. Some robust estimators will be replaced in test statistics are given in the Section 3. The results of the Monte-Carlo simulations are presented in Section 4. In the last section, the performance of the modified tests are explained in detail.

2. Methodology

Consider the problem of comparing the means of k populations. Assume X_{i1}, $X_{i2}, ..., X_{ni_1}, i = 1, 2, ..., k$ are observations of k independent groups from the normal distribution $N(\mu, \sigma_i^2)$. The sample means and the sample variances of the k independent groups are denoted as $\bar{\mathbf{X}} = (\bar{X}_1, \bar{X}_2, ..., \bar{X}_k)$ and $\mathbf{S}^2 = (S_1^2, S_2^2, ..., S_k^2)$ respectively, where

$$\bar{X}_i = \frac{\sum_{j=1}^{n_i} x_{ij}}{n_i} \quad and \quad S_i^2 = \frac{\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2}{n_i - 1} \tag{2.1}$$

The observed values of these random variables are $\bar{\mathbf{x}} = (\bar{x}_1, \bar{x}_2, ..., \bar{x}_k)$,$\mathbf{s}^2 = (s_1^2, s_2^2, ..., s_k^2)$ respectively. The hypotheses of this problem as follows

$$H_0 : \mu_0 = \mu_1 = ... = \mu_k$$

$$H_1 : \mu_i \neq \mu_j \text{ for some } i \neq j$$

When all σ_i^2 are equal, the F-test statistic is

$$F = \frac{\frac{\sum_{i=1}^{k} n_i \bar{x}_i^2 - n\bar{x}^2}{(k-1)}}{\frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i} x_{ij}^2 - \sum_{i=1}^{k} n_i \bar{x}_i^2}{(n-k)}} \tag{2.2}$$

where $n = \sum_{i=1}^{k} n_i$ is the total number of observations and $\bar{x} = \sum_{i=1}^{k} \bar{x}_i$ is the grand mean average of all observations. F test statistic has an F distribution with $k-1$ and $n-k$ degrees of freedom.[38]

When the population variances are equal $\sigma_1^2 = \sigma_2^2 = ... = \sigma_k^2$, the classical F-test is appropriate for testing the equality of the group means. Nevertheless, the assumption of variance homogeneity is frequently violated in practice. Some approximate tests given in below are proposed for handling this violation.

2.1. Welch test. Welch proposed a test based on weighting for comparing several group means under heteroskedasticity. The weights $w_i = \frac{n_i}{s_i^2}$ are used to define population characteristics. The sample mean is estimated by weighted mean as follows

$$SS_B = \bar{x}_w = \sum_{i=1}^{k} w_i(\bar{x}_i - \bar{x}_w)^2 \tag{2.3}$$

Now let $v_i = n_i - 1$ be the degrees of freedom in the i.th sample and a correction of degrees of freedom in the test statistics is stated as follows

$$C = \frac{\sum_{i=1}^{k}[1 - (\frac{w_i}{\sum_{i=1}^{k} w_i})]^2}{\sum_{i=1}^{k} v_i} \tag{2.4}$$

and the Welch F-test statistic is

$$W = \frac{SS_B}{(k-1)}\left[1 + \frac{2(k-1)C}{(k^2-1)}\right]^{-1} \tag{2.5}$$

has F distribution with degress of freedom $v_1 = k-1$ and $v_2 = (k^2-1)/3C$ under the null hypothesis of equality of group means. [40]

2.2. Parametric bootstrap test. Bootstrap methods are generally used in which obtaining the test statistics is not easy. Krishnamoorthy et al. improved the Parametric Bootstrap Test for comparing the equality of several group means under heteroskedasticity.

Let $\mathbf{x}_B = (x_1, x_2, ..., x_m)$ be a sample obtained from $\mathbf{x} = (x_1, x_2, ..., x_n)$ by taking $m\,(m < n)$ observations uniformly from $\mathbf{x}$ with replacement. Computation of the sample means and sample variances with bootstrapping of k independent groups

$$\bar{x}_{Bi} = \frac{\sum_{j=1}^{m} x_{ij}}{m} \quad \text{and} \quad s_{Bi}^2 = \frac{\sum_{j=1}^{m}(x_{ij} - \bar{x}_i)^2}{m} \tag{2.6}$$

where $\bar{x}_{Bi} \sim Z_i(0, s_i/\sqrt{n_i})$ and $Z_i \sim N(0,1)$, the Parametric Bootstrap test statistic can be calculated as follows

$$PB = \sum_{i=1}^{k} \frac{Z_i^2(n_i-1)}{\chi^2_{n_i-1}} - \frac{\sum_{i=1}^{k} \frac{\sqrt{n_i}Z_i(n_i-1)}{s_i^2\chi^2_{n_i-1}}}{\sum_{i=1}^{k} \frac{n_i(n_i-1)}{s_i^2\chi^2_{n_i-1}}} \tag{2.7}$$

where Parametric Bootstrap test statistics, PB is distributed chi-squared distribution with $(k-1)$ degrees of freedom. [20]

2.3. Generalized F-test. Generalized F-test is proposed by Weerahandi in the presence of nuisance parameter and test statistic is computed by the simulation methods. Consider standardized the sum of squares between groups

$$\tilde{s}_B = \sum_{i=1}^{k} \frac{n_i\bar{x}_i^2}{s_i^2} - \frac{(\sum_{i=1}^{k} n_i\bar{x}_i/s_i^2)^2}{\sum_{i=1}^{k} n_i/s_i^2} \tag{2.8}$$

Let the nuisance parameter s_i^2 replaced by random chi-squared variables $\chi^2_{(k-1)}$ with $(k-1)$ degrees of freedom.

$$GF = E\left[\tilde{s}_B\left(\frac{n_1 s_1^2}{Y_1}, \frac{n_2 s_2^2}{Y_2}, ..., \frac{n_k s_k^2}{Y_k}\right)\right] \tag{2.9}$$

where Generalized F-test statistic, GF is distributed chi-squared distribution with $(k-1)$ degrees of freedom and the expectation is taken with respect to the independent Y_i random variables. The computation of the test statistic can be given in the following steps:

(1) Generate a sample of large number of random numbers from each chi-squared random variable $Y_i \sim \chi^2_{n_i-1}$
(2) For each sample compute the $t = \chi^2_{(k-1)}$
(3) Compute their averages $\bar{t}$

$\bar{t}$ is the test statistic of Generalized F-test, distributed by chi-squared distribution with $(k-1)$ degrees of freedom. [37]

3. Robust estimators

Robust procedures provide methods not effected by outliers or departures from assumptions. In the case of data include outlier(s), classical estimators show poor performance so these procedures are preferred in estimating of location and shape parameters. There are some measurements for comparing the effectiveness of estimators, used in robust theory.

Let $X_1, X_2..., X_n$ be a random sample and we want to estimate location and scale parameters of this sample with robust estimators.

3.1. Trimmed mean and trimmed variance. Let $X_1, X_.., X_n$ be a random sample and let $X_{(1)} \leq X_{(2)} \leq ... \leq X_{(n)}$ be the observations in ascending order. The value $X_{(i)}$ is called the ith order statistic. The trimmed mean is computed by removing the r largest and the r smallest observations and averaging the remain values as classical mean computation. The trimmed mean is $\bar{X}_t = X_{(r+1)} + X_{(r+2)} + ... + X_{(n-r)}/(n-2r)$. Trimmed variance is calculated with the based on winsorized variance. Although of the trimmed, the idea of winsorized calculates after replacing the observations at he high and low point with the most extreme remaining values. Consider the random sample in above, it is converted $X_{(r)} \leq X_{(r)} \leq X_{(r)} \leq ... \leq X_{(n-r)} \leq X_{(n-r)} \leq X_{(n-r)}$ with winsorization. Winsorized variance can be used instead of trimmed variance as follows

$$(3.1) \qquad s_t^2 = \frac{\sum_{i=1}^{n}(W_i - \bar{W})^2}{n^2(1-2r)^2}$$

where $\overline{W}$ is the winsorized mean and W_i's are winsorized observations. [41]

3.2. Median and median absolute deviation. Median is the most popular location estimator in robust statistics. It is easily obtained as the middle observation on the data. Therefore, it is not effectted from outliers as much as the mean. Median absolute deviation is computed based on median as follows

$$(3.2) \qquad MAD = b|med_{x_i} - med(\mathbf{x})|$$

The constant b in above is needed to make estimator consistent for the parameter of interest. In the case of Normal distribution is present, the value of constant is 1.4826. Median Absolute Deviation is used instead of standard deviation so we can use its squared instead of variance.

3.3. Huber's M-estimators. Let $X_1, X_2, ..., X_n$ be a random sample from a distribution of the type $(1/\sigma)f((x-\mu)/\sigma)$. Huber proposed a method to estimate location parameter μ as follows

$$(3.3) \qquad lnL = -nln\sigma + \sum_{i=1}^{n} lnf(z_i),\ z_i = (x-\mu)/\sigma$$

If the functional form of f is known, the Maximum Likelihood estimator of μ is the solution of the equation

$$(3.4) \qquad \frac{\partial lnL}{\partial \mu} = \frac{1}{\sigma}\sum_{i=1}^{n}\xi(z_i) = 0,\ \ \xi(z) = -f'(z)/f(z)$$

Writing $w_i = w_i(z) = \xi(z_i)/z_i$ so the equation is above can be written as $\sum_{i=1}^{n} w_i(x_i - \mu) = 0$ so that $\mu = \sum_{i=1}^{n} w_i x_i / \sum_{i=1}^{n} w_i$. Huber proposed a function $\xi(z)$ as

$$(3.5) \qquad \xi(z) = \begin{cases} z, & \text{if } |z| \leq c \\ csign(z), & \text{if } |z| > c \end{cases}$$

The m-estimation equation can be solve by iteration with the known function $\xi(z)$. The choice of c is 1.345 for normal distribution in the tails[29]. Also, Huber suggested changing MAD to $MAD/0.6745$ estimator of σ.[34]

Wu (1985) examined the estimator in his study according to their robustness properties in detail. It is very extensive study for comparing the estimators with variety aspects. We emphasize the performance of the estimators on the data include outlier(s). Median and its derivatives and Huber's M-estimators show similar performance according to breakdown point and influence function. This study revealed the Huber's M-estimator is better than others according to their relative efficiencies in case of outlier(s) is present.

4. Modified tests for comparing of group means under non-normality

There many procedures are proposed for comparing of group means under heteroskedasticity such as Welch F-test, Parametric Bootstrap test and Generalized F-test. In the case of violation of normality caused by outlier, instead of these procedures, some modified tests are developed with robust estimators. Modified tests are considered in this secton respectively.

4.1. Modified Welch's F-test. Welch proposed the Welch F-test that can be shown as *Equation 2.5.* In this test statistic, the sum of squares between groups is calculated as the follows

$$(4.1) \qquad SS_B = \bar{x}_w = \sum_{i=1}^{k} w_i(\bar{x}_i - \bar{x}_w)^2, w_i = \frac{n_i}{s_i^2}$$

where $\overline{x}_i$ and s_i^2 denoted as the maximum likelihood estimators of sample mean and variance respectively. The proposed modified Welch F-test correspond to the standard Welch F-test statistic in which the maximum likelihood estimator of sample mean and variance are replaced with robust estimators such as trimmed mean and variance, median and median absolute deviation and Huber's M-estimators.

4.2. Modified parametric bootstrap test. Krishnamoorthy et al. proposed the Parametric Bootstrap test as in *Equation 2.7* with the computation of the sample mean and variance with bootstrapping of k independent groups as in *Equation 2.6.* The maximum likelihood estimators are used for computation of the sample mean and variance. The proposed modified Parametric Bootstrap test correspond to the standard the Parametric Bootstrap test statistic in which the maximum likelihood estimator of sample mean and variance are replaced with robust estimators such as trimmed mean and variance, median and median absolute deviation and Huber's M-estimators.

4.3. Modified generalized F-test. Weerahandi proposed the Generalized F-test as in *Equation 2.9.* The generalized sum of squares between groups are used for calculating of the test statistic as follows

$$(4.2) \qquad \tilde{s}_B = \sum_{i=1}^{k} \frac{n_i \bar{x}_i^2}{s_i^2} - \frac{(\sum_{i=1}^{k} n_i \bar{x}_i / s_i^2)^2}{\sum_{i=1}^{k} n_i / s_i^2}$$

where $\overline{x}_i$ and s_i^2 denoted as the maximum likelihood estimators of sample mean and variance respectively. The proposed modified Generalized F-test correspond to the standard Generalized F-test statistic in which the maximum likelihood estimator of sample mean and variance are replaced with robust estimators such as trimmed mean and variance, median and median absolute deviation and Huber's M-estimators.

5. Simulation study

The performances of the modified tests are investigated under heteroskedasticity and non-normality caused by outlier(s). Powers and type 1 error rates of the proposed tests are calculated by Monte-Carlo simulations with 1000 replications for the nominal value $\alpha = 0.05$. In this study, three configuration factors were taken into account to evaluate the performance of type 1 error rates and powers; sample size, effect size and outlier. In the following subsections, the performance of the tests and modified tests are examined in details.

5.1. Performances of the tests. In this section, performances of the tests under heteroskedasticity and non-normality caused by outlier(s) are examined in terms of power of the test and type 1 error rates. Monte-Carlo simulation studies are configurated according to unbalanced ($n_1 = 5, n_2 = 10, n_3 = 15$)-balanced ($n_1 = 10, n_2 = 10, n_3 = 10$) designs under heteroskadastic variances ($\sigma_1^2 = 0.2, \sigma_2^2 = 0.4, \sigma_3^2 = 0.6$), various effect sizes and outlier combinations. The nominal level of $\alpha = 0.05$ is taken for computing type 1 error rates. Box-Whisker plot is used for generating outlier in simulations. The observations are defined as outliers sample variance distance from the whiskers. After 1000 replications, power of the tests and type 1 error rates are calculated and tabulated. In the tables, (o_1, o_2, o_3) display the number of outliers in the groups and (μ_1, μ_2, μ_3) display the population means.

Table 1: Powers of the tests

(n_1, n_2, n_3)		$(5, 10, 15)$				$(10, 10, 10)$			
(o_1, o_2, o_3)	(μ_1, μ_2, μ_3)	CF	GF	PB	W	CF	GF	PB	W
	(0,0,0.3)	0.152	0.210	0.181	0.150	0.167	0.172	0.158	0.146
	(0,0,0.6)	0.534	0.557	0.474	0.502	0.497	0.458	0.416	0.427
$(0, 0, 0)$	(0,0,0.9)	0.875	0.886	0.805	0.841	0.847	0.780	0.735	0.765
	(0,0,1.2)	0.989	0.992	0.971	0.985	0.986	0.952	0.916	0.952
	(0,0,1.5)	0.999	0.999	0.996	0.999	1	0.999	0.989	0.998
	(0,0,0.3)	0.085	0.144	0.125	0.107	0.133	0.121	0.095	0.095
	(0,0,0.6)	0.359	0.441	0.331	0.370	0.381	0.315	0.268	0.282
$(0, 0, 1)$	(0,0,0.9)	0.714	0.754	0.691	0.705	0.638	0.578	0.515	0.551
	(0,0,1.2)	0.920	0.932	0.920	0.905	0.831	0.774	0.752	0.756
	(0,0,1.5)	0.985	0.994	0.988	0.987	0.945	0.889	0.903	0.883
	(0,0,0.3)	0.091	0.144	0.104	0.116	0.125	0.118	0.073	0.098
	(0,0,0.6)	0.332	0.409	0.338	0.343	0.317	0.312	0.232	0.279
$(0, 1, 1)$	(0,0,0.9)	0.639	0.713	0.645	0.676	0.569	0.566	0.495	0.533
	(0,0,1.2)	0.865	0.908	0.879	0.895	0.782	0.757	0.732	0.730
	(0,0,1.5)	0.971	0.983	0.980	0.981	0.911	0.892	0.900	0.883
	(0,0,0.3)	0.104	0.140	0.117	0.120	0.124	0.092	0.085	0.091
	(0,0,0.6)	0.341	0.402	0.344	0.371	0.312	0.277	0.232	0.268
$(1, 1, 1)$	(0,0,0.9)	0.662	0.709	0.669	0.697	0.558	0.517	0.476	0.505
	(0,0,1.2)	0.881	0.905	0.891	0.886	0.772	0.726	0.710	0.711
	(0,0,1.5)	0.973	0.982	0.974	0.980	0.910	0.871	0.874	0.862
	(0,0,0.3)	0.143	0.223	0.098	0.168	0.215	0.186	0.078	0.149
	(0,0,0.6)	0.373	0.446	0.296	0.395	0.413	0.368	0.216	0.356
$(0, 0, 2)$	(0,0,0.9)	0.605	0.649	0.587	0.614	0.572	0.541	0.450	0.525
	(0,0,1.2)	0.783	0.820	0.850	0.793	0.707	0.672	0.694	0.657
	(0,0,1.5)	0.919	0.935	0.965	0.918	0.813	0.789	0.842	0.778
	(0,0,0.3)	0.136	0.202	0.106	0.164	0.188	0.186	0.064	0.143
	(0,0,0.6)	0.368	0.435	0.310	0.393	0.388	0.369	0.202	0.345
$(0, 1, 2)$	(0,0,0.9)	0.578	0.633	0.597	0.605	0.528	0.523	0.423	0.513
	(0,0,1.2)	0.749	0.803	0.829	0.763	0.657	0.650	0.689	0.632
	(0,0,1.5)	0.899	0.928	0.950	0.914	0.800	0.778	0.857	0.761
	(0,0,0.3)	0.161	0.204	0.092	0.179	0.197	0.159	0.057	0.134
	(0,0,0.6)	0.361	0.417	0.321	0.390	0.379	0.326	0.202	0.308
$(1, 1, 2)$	(0,0,0.9)	0.595	0.641	0.594	0.605	0.534	0.523	0.423	0.502
	(0,0,1.2)	0.787	0.811	0.852	0.803	0.679	0.656	0.676	0.649
	(0,0,1.5)	0.912	0.936	0.956	0.927	0.819	0.782	0.867	0.766
	(0,0,0.3)	0.185	0.266	0.089	0.213	0.310	0.282	0.096	0.246
	(0,0,0.6)	0.412	0.461	0.253	0.414	0.460	0.427	0.230	0.402
$(0, 0, 3)$	(0,0,0.9)	0.561	0.602	0.556	0.572	0.541	0.531	0.445	0.521
	(0,0,1.2)	0.701	0.747	0.817	0.703	0.612	0.603	0.647	0.600
	(0,0,1.5)	0.840	0.882	0.949	0.847	0.699	0.691	0.779	0.680
	(0,0,0.3)	0.177	0.263	0.117	0.228	0.216	0.229	0.098	0.189
	(0,0,0.6)	0.353	0.462	0.355	0.424	0.378	0.398	0.253	0.374
$(0, 2, 2)$	(0,0,0.9)	0.570	0.666	0.650	0.646	0.534	0.559	0.480	0.542
	(0,0,1.2)	0.751	0.820	0.881	0.794	0.680	0.688	0.718	0.670
	(0,0,1.5)	0.882	0.941	0.970	0.933	0.791	0.801	0.868	0.787
	(0,0,0.3)	0.205	0.218	0.162	0.242	0.263	0.217	0.125	0.237
	(0,0,0.6)	0.362	0.401	0.375	0.405	0.407	0.363	0.291	0.374
$(2, 2, 2)$	(0,0,0.9)	0.558	0.609	0.633	0.609	0.541	0.527	0.483	0.526
	(0,0,1.2)	0.750	0.808	0.808	0.798	0.679	0.669	0.685	0.666
	(0,0,1.5)	0.890	0.923	0.938	0.910	0.814	0.781	0.829	0.771

Table 2: Type 1 error rates of the tests

	(5, 10, 15)				(10, 10, 10)			
(o_1, o_2, o_3)	CF	GF	PB	W	CF	GF	PB	W
(0,0,0)	0.032	0.068	0.062	0.046	0.062	0.076	0.072	0.057
(0,0,1)	0.025	0.067	0.036	0.048	0.058	0.076	0.044	0.060
(0,1,1)	0.030	0.078	0.039	0.049	0.061	0.076	0.043	0.058
(1,1,1)	0.037	0.069	0.044	0.048	0.040	0.055	0.036	0.045
(0,0,2)	0.040	0.108	0.035	0.072	0.096	0.102	0.043	0.079
(0,1,2)	0.044	0.113	0.031	0.067	0.098	0.112	0.027	0.074
(1,1,2)	0.055	0.111	0.046	0.085	0.091	0.077	0.020	0.070
(0,0,3)	0.065	0.157	0.035	0.096	0.222	0.192	0.051	0.162
(0,2,2)	0.081	0.158	0.056	0.111	0.138	0.159	0.055	0.116
(2,2,2)	0.133	0.144	0.103	0.171	0.193	0.124	0.072	0.169
(0,0,0)	0.032	0.068	0.062	0.046	0.062	0.076	0.072	0.057
(0,0,1)	0.015	0.066	0.032	0.041	0.045	0.063	0.034	0.047
(0,1,1)	0.026	0.064	0.018	0.040	0.051	0.054	0.018	0.033
(1,1,1)	0.029	0.023	0.019	0.031	0.038	0.025	0.018	0.029
(0,0,2)	0.033	0.103	0.028	0.058	0.092	0.090	0.028	0.076
(0,1,2)	0.033	0.081	0.018	0.054	0.094	0.108	0.016	0.069
(1,1,2)	0.037	0.030	0.013	0.035	0.096	0.056	0.009	0.053
(0,0,3)	0.037	0.117	0.026	0.067	0.215	0.178	0.033	0.140
(0,2,2)	0.073	0.149	0.040	0.103	0.152	0.157	0.030	0.103
(2,2,2)	0.180	0.111	0.084	0.162	0.185	0.101	0.069	0.160
(0,0,0)	0.032	0.068	0.062	0.046	0.062	0.076	0.072	0.057
(0,0,1)	0.015	0.065	0.023	0.035	0.034	0.051	0.023	0.040
(0,1,1)	0.016	0.048	0.012	0.025	0.037	0.035	0.008	0.017
(1,1,1)	0.030	0.014	0.011	0.016	0.023	0.017	0.006	0.021
(0,0,2)	0.015	0.083	0.022	0.046	0.081	0.084	0.021	0.066
(0,1,2)	0.021	0.050	0.010	0.030	0.083	0.090	0.010	0.044
(1,1,2)	0.026	0.016	0.009	0.022	0.085	0.035	0.008	0.042
(0,0,3)	0.015	0.076	0.020	0.042	0.177	0.151	0.022	0.115
(0,2,2)	0.049	0.096	0.018	0.058	0.129	0.135	0.019	0.084
(2,2,2)	0.124	0.035	0.047	0.080	0.167	0.074	0.038	0.147

According the results in the **Table 1**, tests lose their powers in the presence of outlier(s). The decreasing of powers is higher in the balanced designs. Besides the decreasing power of the tests, the type 1 error rates are higher than nominal level in the presence of outlier. For handling this problem, some modifications will be proposed using the robust estimators in the following sections. Trimmed mean and variance, median and median absolute deviation, Huber's M-estimators are replaced with maximum likelihood estimators in the test statistics for obtaining modified tests. More powerful tests and nominal type 1 error rates are attempted to obtain by this modifications.

5.2. Performances of the modified tests by the trimmed mean and variance. In this section, trimmed mean and variance are used instead of maximum likelihood estimators to estimate location and scale parameters in the test statistics for comparing group means under heteroskedasticity and nonnormality. The performance of the modified tests are examined with Monte-Carlo simulations in terms of power of the test and type 1 error rates.

Table 3: Powers of the modified tests by trimmed mean and variance

(n_1, n_2, n_3)		(5, 10, 15)				(10, 10, 10)			
(o_1, o_2, o_3)	(μ_1, μ_2, μ_3)	CF	GF	PB	W	CF	GF	PB	W
	(0,0,0.3)	0.196	0.602	0.544	0.550	0.191	0.531	0.505	0.512
	(0,0,0.6)	0.542	0.874	0.845	0.856	0.517	0.808	0.785	0.796
(0, 0, 0)	(0,0,0.9)	0.868	0.984	0.978	0.979	0.818	0.956	0.947	0.953
	(0,0,1.2)	0.983	0.999	0.999	0.999	0.968	0.997	0.996	0.997
	(0,0,1.5)	0.999	1	1	1	0.998	1	1	1
	(0,0,0.3)	0.091	0.541	0.497	0.500	0.086	0.464	0.423	0.440
	(0,0,0.6)	0.345	0.856	0.815	0.829	0.327	0.734	0.701	0.713
(0, 0, 1)	(0,0,0.9)	0.715	0.974	0.966	0.968	0.653	0.917	0.904	0.907
	(0,0,1.2)	0.943	0.998	0.997	0.998	0.891	0.992	0.986	0.989
	(0,0,1.5)	0.994	1	1	1	0.986	0.999	0.998	0.999
	(0,0,0.3)	0.051	0.498	0.448	0.468	0.041	0.401	0.353	0.369
	(0,0,0.6)	0.265	0.802	0.768	0.778	0.231	0.670	0.633	0.642
(0, 1, 1)	(0,0,0.9)	0.622	0.965	0.951	0.954	0.527	0.899	0.881	0.886
	(0,0,1.2)	0.888	0.995	0.995	0.995	0.818	0.985	0.978	0.981
	(0,0,1.5)	0.980	1	1	1	0.958	1	0.998	0.999
	(0,0,0.3)	0.047	0.366	0.328	0.336	0.033	0.366	0.329	0.342
	(0,0,0.6)	0.233	0.757	0.702	0.716	0.206	0.659	0.628	0.642
(1, 1, 1)	(0,0,0.9)	0.624	0.965	0.947	0.950	0.501	0.886	0.871	0.880
	(0,0,1.2)	0.908	0.996	0.995	0.996	0.810	0.979	0.975	0.977
	(0,0,1.5)	0.990	1	1	1	0.958	0.999	0.999	0.998
	(0,0,0.3)	0.052	0.518	0.471	0.489	0.064	0.426	0.389	0.403
	(0,0,0.6)	0.246	0.802	0.767	0.773	0.263	0.669	0.631	0.650
(0, 0, 2)	(0,0,0.9)	0.596	0.958	0.941	0.944	0.555	0.848	0.821	0.830
	(0,0,1.2)	0.866	1	0.997	0.996	0.796	0.965	0.956	0.961
	(0,0,1.5)	0.982	1	1	1	0.938	0.994	0.993	0.993
	(0,0,0.3)	0.040	0.518	0.471	0.485	0.036	0.370	0.329	0.348
	(0,0,0.6)	0.216	0.775	0.739	0.750	0.179	0.630	0.598	0.607
(0, 1, 2)	(0,0,0.9)	0.542	0.944	0.925	0.929	0.486	0.840	0.815	0.825
	(0,0,1.2)	0.837	0.994	0.991	0.992	0.774	0.955	0.949	0.952
	(0,0,1.5)	0.971	0.999	0.999	0.999	0.930	0.995	0.993	0.994
	(0,0,0.3)	0.034	0.383	0.334	0.347	0.026	0.334	0.302	0.311
	(0,0,0.6)	0.203	0.712	0.669	0.679	0.166	0.617	0.581	0.600
(1, 1, 2)	(0,0,0.9)	0.562	0.932	0.916	0.922	0.471	0.835	0.812	0.819
	(0,0,1.2)	0.869	0.996	0.994	0.993	0.768	0.953	0.944	0.947
	(0,0,1.5)	0.981	1	1	1	0.935	0.996	0.995	0.995
	(0,0,0.3)	0.065	0.501	0.459	0.465	0.185	0.383	0.339	0.352
	(0,0,0.6)	0.237	0.736	0.698	0.708	0.377	0.498	0.470	0.479
(0, 0, 3)	(0,0,0.9)	0.549	0.915	0.884	0.896	0.542	0.632	0.608	0.616
	(0,0,1.2)	0.820	0.986	0.982	0.984	0.656	0.728	0.701	0.714
	(0,0,1.5)	0.960	0.998	0.997	0.998	0.798	0.817	0.793	0.804
	(0,0,0.3)	0.031	0.464	0.422	0.437	0.038	0.348	0.310	0.323
	(0,0,0.6)	0.176	0.771	0.730	0.743	0.181	0.603	0.572	0.590
(0, 2, 2)	(0,0,0.9)	0.516	0.954	0.937	0.945	0.451	0.812	0.787	0.795
	(0,0,1.2)	0.825	0.997	0.996	0.996	0.719	0.944	0.932	0.935
	(0,0,1.5)	0.971	1	1	1	0.904	0.994	0.989	0.992
	(0,0,0.3)	0.043	0.363	0.318	0.334	0.036	0.322	0.295	0.308
	(0,0,0.6)	0.200	0.663	0.618	0.637	0.143	0.576	0.536	0.554
(2, 2, 2)	(0,0,0.9)	0.508	0.899	0.882	0.886	0.414	0.801	0.778	0.788
	(0,0,1.2)	0.814	0.985	0.979	0.982	0.696	0.932	0.925	0.929
	(0,0,1.5)	0.958	0.999	0.997	0.998	0.905	0.991	0.985	0.987

Table 4: Type 1 error rates of the modified tests by trimmed mean and variance

	(5, 10, 15)				(10, 10, 10)			
(o_1, o_2, o_3)	CF	GF	PB	W	CF	GF	PB	W
(0,0,0)	0.053	0.417	0.374	0.387	0.089	0.395	0.353	0.268
(0,0,1)	0.013	0.368	0.319	0.337	0.026	0.323	0.291	0.302
(0,1,1)	0.013	0.322	0.292	0.297	0.014	0.272	0.234	0.249
(1,1,1)	0.007	0.202	0.167	0.174	0.013	0.215	0.195	0.202
(0,0,2)	0.007	0.372	0.336	0.355	0.025	0.314	0.286	0.302
(0,1,2)	0.009	0.326	0.279	0.298	0.008	0.236	0.201	0.212
(1,1,2)	0.005	0.214	0.186	0.196	0.009	0.200	0.173	0.183
(0,0,3)	0.011	0.381	0.353	0.361	0.084	0.319	0.276	0.294
(0,2,2)	0.005	0.293	0.245	0.264	0.008	0.235	0.210	0.219
(2,2,2)	0.009	0.250	0.212	0.222	0.006	0.205	0.180	0.186
(0,0,0)	0.053	0.417	0.374	0.387	0.089	0.395	0.353	0.368
(0,0,1)	0.005	0.399	0.340	0.362	0.013	0.344	0.311	0.324
(0,1,1)	0.004	0.300	0.269	0.277	0.006	0.307	0.272	0.280
(1,1,1)	0.001	0.237	0.202	0.213	0.001	0.246	0.210	0.217
(0,0,2)	0.009	0.406	0.364	0.380	0.011	0.365	0.324	0.339
(0,1,2)	0.002	0.316	0.284	0.300	0.004	0.317	0.285	0.299
(1,1,2)	0	0.228	0.197	0.204	0.001	0.273	0.246	0.254
(0,0,3)	0.001	0.396	0.355	0.377	0.063	0.305	0.260	0.269
(0,2,2)	0	0.347	0.314	0.329	0.001	0.338	0.291	0.312
(2,2,2)	0.011	0.238	0.187	0.198	0	0.302	0.268	0.280
(0,0,0)	0.053	0.417	0.374	0.387	0.089	0.395	0.353	0.368
(0,0,1)	0.004	0.402	0.354	0.371	0.011	0.373	0.337	0.352
(0,1,1)	0.001	0.382	0.345	0.353	0.002	0.366	0.328	0.339
(1,1,1)	0	0.266	0.227	0.236	0.001	0.322	0.290	0.295
(0,0,2)	0	0.397	0.352	0.373	0.003	0.383	0.345	0.359
(0,1,2)	0	0.370	0.338	0.351	0.002	0.377	0.338	0.354
(1,1,2)	0	0.269	0.237	0.240	0	0.351	0.319	0.334
(0,0,3)	0	0.384	0.354	0.369	0.030	0.273	0.235	0.246
(0,2,2)	0	0.369	0.334	0.348	0	0.368	0.327	0.342
(2,2,2)	0.005	0.223	0.161	0.175	0	0.371	0.343	0.348

Although decreasing power of the tests in the case of nonnormality caused by outlier, approximate tests are more powerful with the modification of trimmed mean and variance. However, the type 1 error rates are more inflated with this modification even exceed the level of 0.2. Modified tests obtained by trimmed mean and variance are not recommended for comparing group means under heteroskedasticity and nonnormality because of inflated type 1 error rates. Thus, other robust modifications will be examined in the following sections.

5.3. Performances of the modified tests by median and median absolute deviation. In this section, median and median absolute deviation are used instead of maximum likelihood estimators to estimate location and scale parameters in the test statistics for comparing group means under heteroskedasticity and nonnormality. The performance of the modified tests are examined with Monte-Carlo simulations in terms of power of the test and type 1 error rates.

Table 5: Powers of the modified tests by median and median absolute deviation

(n_1, n_2, n_3)		(5, 10, 15)				(10, 10, 10)			
(o_1, o_2, o_3)	(μ_1, μ_2, μ_3)	CF	GF	PB	W	CF	GF	PB	W
(0, 0, 0)	(0,0,0.3)	0.235	0.331	0.286	0.299	0.233	0.299	0.266	0.276
	(0,0,0.6)	0.539	0.620	0.561	0.576	0.521	0.562	0.526	0.534
	(0,0,0.9)	0.808	0.861	0.816	0.822	0.810	0.802	0.774	0.786
	(0,0,1.2)	0.956	0.975	0.961	0.965	0.940	0.941	0.925	0.931
	(0,0,1.5)	0.991	0.998	0.996	0.996	0.986	0.983	0.977	0.979
(0, 0, 1)	(0,0,0.3)	0.135	0.317	0.276	0.288	0.162	0.313	0.287	0.295
	(0,0,0.6)	0.377	0.581	0.522	0.536	0.371	0.523	0.481	0.498
	(0,0,0.9)	0.698	0.844	0.794	0.802	0.621	0.731	0.707	0.715
	(0,0,1.2)	0.889	0.970	0.951	0.961	0.825	0.888	0.872	0.876
	(0,0,1.5)	0.972	0.998	0.996	0.996	0.926	0.975	0.969	0.971
(0, 1, 1)	(0,0,0.3)	0.108	0.333	0.291	0.301	0.096	0.314	0.279	0.290
	(0,0,0.6)	0.338	0.601	0.543	0.557	0.283	0.503	0.473	0.489
	(0,0,0.9)	0.608	0.814	0.777	0.789	0.519	0.741	0.710	0.720
	(0,0,1.2)	0.817	0.955	0.938	0.943	0.741	0.904	0.884	0.892
	(0,0,1.5)	0.946	0.992	0.986	0.987	0.900	0.975	0.969	0.972
(1, 1, 1)	(0,0,0.3)	0.093	0.385	0.347	0.361	0.087	0.313	0.273	0.288
	(0,0,0.6)	0.297	0.643	0.605	0.621	0.265	0.508	0.468	0.483
	(0,0,0.9)	0.609	0.884	0.849	0.867	0.514	0.743	0.707	0.725
	(0,0,1.2)	0.844	0.978	0.967	0.972	0.745	0.894	0.877	0.886
	(0,0,1.5)	0.953	0.998	0.997	0.997	0.892	0.964	0.958	0.960
(0, 0, 2)	(0,0,0.3)	0.093	0.303	0.255	0.270	0.096	0.243	0.224	0.228
	(0,0,0.6)	0.290	0.546	0.484	0.500	0.287	0.465	0.430	0.445
	(0,0,0.9)	0.578	0.806	0.759	0.773	0.542	0.701	0.656	0.671
	(0,0,1.2)	0.828	0.952	0.925	0.929	0.789	0.863	0.844	0.851
	(0,0,1.5)	0.954	0.991	0.983	0.984	0.909	0.944	0.931	0.936
(0, 1, 2)	(0,0,0.3)	0.086	0.343	0.304	0.316	0.059	0.227	0.200	0.206
	(0,0,0.6)	0.280	0.581	0.520	0.543	0.203	0.462	0.411	0.431
	(0,0,0.9)	0.556	0.818	0.769	0.784	0.480	0.711	0.678	0.689
	(0,0,1.2)	0.791	0.942	0.922	0.925	0.751	0.874	0.857	0.863
	(0,0,1.5)	0.921	0.990	0.980	0.986	0.904	0.949	0.938	0.942
(1, 1, 2)	(0,0,0.3)	0.083	0.386	0.353	0.363	0.059	0.264	0.232	0.244
	(0,0,0.6)	0.268	0.635	0.589	0.604	0.217	0.471	0.430	0.439
	(0,0,0.9)	0.536	0.859	0.825	0.832	0.479	0.709	0.674	0.687
	(0,0,1.2)	0.799	0.964	0.955	0.958	0.737	0.883	0.861	0.871
	(0,0,1.5)	0.941	0.994	0.992	0.993	0.908	0.953	0.947	0.952
(0, 0, 3)	(0,0,0.3)	0.074	0.268	0.235	0.246	0.109	0.245	0.213	0.229
	(0,0,0.6)	0.236	0.517	0.453	0.463	0.276	0.400	0.370	0.383
	(0,0,0.9)	0.534	0.797	0.739	0.753	0.505	0.601	0.576	0.587
	(0,0,1.2)	0.800	0.936	0.914	0.920	0.717	0.760	0.734	0.744
	(0,0,1.5)	0.944	0.985	0.978	0.978	0.858	0.875	0.859	0.863
(0, 2, 2)	(0,0,0.3)	0.055	0.278	0.229	0.239	0.047	0.213	0.186	0.194
	(0,0,0.6)	0.211	0.549	0.479	0.503	0.199	0.429	0.394	0.406
	(0,0,0.9)	0.509	0.807	0.769	0.781	0.450	0.654	0.617	0.629
	(0,0,1.2)	0.792	0.952	0.929	0.936	0.702	0.867	0.838	0.852
	(0,0,1.5)	0.934	0.994	0.991	0.992	0.881	0.942	0.931	0.934
(2, 2, 2)	(0,0,0.3)	0.053	0.216	0.194	0.202	0.045	0.191	0.165	0.172
	(0,0,0.6)	0.230	0.486	0.443	0.458	0.180	0.402	0.378	0.390
	(0,0,0.9)	0.497	0.771	0.730	0.742	0.431	0.658	0.625	0.641
	(0,0,1.2)	0.783	0.936	0.920	0.923	0.680	0.836	0.816	0.819
	(0,0,1.5)	0.928	0.991	0.985	0.986	0.864	0.937	0.928	0.928

Table 6: Type 1 error rates of the modified tests by median and median absolute deviation

	(5, 10, 15)				(10, 10, 10)			
(o_1, o_2, o_3)	CF	GF	PB	W	CF	GF	PB	W
(0,0,0)	0.099	0.206	0.176	0.186	0.118	0.186	0.162	0.174
(0,0,1)	0.041	0.200	0.166	0.181	0.069	0.207	0.179	0.193
(0,1,1)	0.049	0.229	0.206	0.213	0.052	0.214	0.191	0.195
(1,1,1)	0.033	0.271	0.251	0.260	0.054	0.240	0.216	0.226
(0,0,2)	0.031	0.207	0.179	0.189	0.041	0.170	0.154	0.158
(0,1,2)	0.023	0.236	0.196	0.209	0.021	0.163	0.132	0.144
(1,1,2)	0.023	0.278	0.261	0.266	0.023	0.177	0.152	0.165
(0,0,3)	0.023	0.204	0.165	0.180	0.038	0.163	0.139	0.151
(0,2,2)	0.012	0.157	0.134	0.143	0.015	0.137	0124	0.129
(2,2,2)	0.016	0.137	0.108	0.112	0.012	0.115	0.096	0.105
(0,0,0)	0.099	0.206	0.176	0.186	0.118	0.186	0.162	0.174
(0,0,1)	0.017	0.211	0.177	0.190	0.044	0.193	0.159	0.170
(0,1,1)	0.006	0.168	0.148	0.155	0.022	0.182	0.160	0.169
(1,1,1)	0.005	0.167	0.135	0.144	0.011	0.160	0.139	0.144
(0,0,2)	0.017	0.217	0.185	0.197	0.026	0.168	0.141	0.153
(0,1,2)	0.003	0.185	0.163	0.169	0.010	0.163	0.137	0.148
(1,1,2)	0.002	0.164	0.142	0.151	0.004	0.153	0.138	0.139
(0,0,3)	0.002	0.213	0.182	0.191	0.011	0.185	0.163	0.169
(0,2,2)	0.002	0.183	0.156	0.167	0.003	0.146	0.125	0.219
(2,2,2)	0.003	0.136	0.106	0.111	0.001	0.133	0.111	0.117
(0,0,0)	0.099	0.206	0.176	0.186	0.118	0.186	0.162	0.174
(0,0,1)	0.012	0.216	0.183	0.193	0.027	0.189	0.161	0.174
(0,1,1)	0.003	0.210	0.181	0.184	0.004	0.188	0.161	0.169
(1,1,1)	0.001	0.158	0.112	0.119	0.002	0.174	0.152	0.159
(0,0,2)	0.001	0.218	0.180	0.196	0.005	0.185	0.155	0.169
(0,1,2)	0.002	0.207	0.174	0.187	0.002	0.186	0.158	0.169
(1,1,2)	0	0.154	0.108	0.120	0	0.164	0.137	0.149
(0,0,3)	0	0.201	0.179	0.189	0.001	0.186	0.168	0.175
(0,2,2)	0	0.186	0.162	0.173	0	0.163	0.135	0.143
(2,2,2)	0	0.108	0.078	0.081	0	0.159	0.135	0.143

Similar to previous section, the power of the modified tests with median and median absolute deviation are higher but type 1 error rates are higher than the nominal level. Thus, this modification is also can not be recommended instead of approximate tests under heteroskedasticity and nonnormality. In the next section, Huber's M-estimator modification will be tried to obtain modified tests.

5.4. Performances of the modified tests by Huber's M-estimators. In this section, Huber's M-estimators are used instead of maximum likelihood estimators to estimate location and scale parameters in the test statistics for comparing group means under heteroskedasticity and nonnormality. The performance of the modified tests are examined with Monte-Carlo simulations in terms of power of the test and type 1 error rates.

Table 7: Powers of the modified tests by Huber's M-estimators

(n_1, n_2, n_3)		(5, 10, 15)				(10, 10, 10)			
(o_1, o_2, o_3)	(μ_1, μ_2, μ_3)	CF	GF	PB	W	CF	GF	PB	W
	(0,0,0.3)	0.165	0.170	0.123	0.133	0.168	0.152	0.120	0.129
	(0,0,0.6)	0.534	0.521	0.443	0.455	0.508	0.416	0.374	0.394
(0, 0, 0)	(0,0,0.9)	0.874	0.850	0.802	0.817	0.834	0.743	0.707	0.725
	(0,0,1.2)	0.989	0.982	0.975	0.979	0.981	0.935	0.917	0.921
	(0,0,1.5)	0.999	0.999	0.997	0.997	0.999	0.996	0.993	0.996
	(0,0,0.3)	0.080	0.147	0.115	0.124	0.076	0.123	0.108	0.112
	(0,0,0.6)	0.327	0.479	0.412	0.419	0.316	0.377	0.330	0.347
(0, 0, 1)	(0,0,0.9)	0.733	0.845	0.778	0.795	0.658	0.665	0.626	0.650
	(0,0,1.2)	0.955	0.978	0.954	0.963	0.895	0.890	0.874	0.878
	(0,0,1.5)	0.996	1	0.999	0.999	0.978	0.976	0.965	0.970
	(0,0,0.3)	0.040	0.151	0.114	0.123	0.054	0.131	0.104	0.115
	(0,0,0.6)	0.259	0.467	0.386	0.407	0.245	0.383	0.339	0.358
(0, 1, 1)	(0,0,0.9)	0.625	0.824	0.763	0.773	0.532	0.644	0.605	0.622
	(0,0,1.2)	0.893	0.969	0.948	0.951	0.821	0.878	0.857	0.868
	(0,0,1.5)	0.985	0.998	0.996	0.996	0.958	0.971	0.961	0.968
	(0,0,0.3)	0.051	0.147	0.120	0.129	0.038	0.131	0.109	0.117
	(0,0,0.6)	0.253	0.482	0.406	0.428	0.223	0.350	0.319	0.329
(1, 1, 1)	(0,0,0.9)	0.654	0.818	0.769	0.785	0.519	0.662	0.613	0.638
	(0,0,1.2)	0.911	0.973	0.952	0.960	0.807	0.871	0.851	0.861
	(0,0,1.5)	0.995	0.998	0.995	0.997	0.954	0.971	0.956	0.967
	(0,0,0.3)	0.066	0.166	0.123	0.132	0.123	0.139	0.105	0.118
	(0,0,0.6)	0.280	0.438	0.364	0.382	0.344	0.336	0.295	0.313
(0, 0, 2)	(0,0,0.9)	0.609	0.738	0.673	0.686	0.565	0.538	0.513	0.522
	(0,0,1.2)	0.852	0.915	0.884	0.887	0.748	0.710	0.685	0.697
	(0,0,1.5)	0.974	0.993	0.982	0.985	0.882	0.846	0.828	0.833
	(0,0,0.3)	0.054	0.166	0.124	0.136	0.095	0.132	0.104	0.113
	(0,0,0.6)	0.240	0.452	0.380	0.398	0.284	0.332	0.299	0.312
(0, 1, 2)	(0,0,0.9)	0.561	0.728	0.671	0.684	0.500	0.544	0.511	0.528
	(0,0,1.2)	0.821	0.918	0.888	0.890	0.709	0.711	0.684	0.699
	(0,0,1.5)	0.961	0.990	0.982	0.986	0.869	0.854	0.838	0.849
	(0,0,0.3)	0.054	0.175	0.147	0.149	0.094	0.125	0.109	0.114
	(0,0,0.6)	0.253	0.432	0.388	0.396	0.261	0.311	0.267	0.284
(1, 1, 2)	(0,0,0.9)	0.572	0.748	0.688	0.706	0.512	0.557	0.531	0.539
	(0,0,1.2)	0.845	0.923	0.894	0.904	0.715	0.772	0.707	0.712
	(0,0,1.5)	0.973	0.992	0.986	0.986	0.869	0.864	0.845	0.852
	(0,0,0.3)	0.122	0.193	0.150	0.158	0.314	0.192	0.154	0.169
	(0,0,0.6)	0.324	0.410	0.358	0.370	0.464	0.346	0.313	0.324
(0, 0, 3)	(0,0,0.9)	0.549	0.614	0.576	0.581	0.539	0.487	0.460	0.478
	(0,0,1.2)	0.760	0.801	0.764	0.764	0.606	0.572	0.556	0.564
	(0,0,1.5)	0.911	0.935	0.907	0.916	0.699	0.656	0.644	0.651
	(0,0,0.3)	0.075	0.194	0.156	0.168	0.112	0.158	0.125	0.136
	(0,0,0.6)	0.254	0.454	0.400	0.407	0.298	0.341	0.305	0.317
(0, 2, 2)	(0,0,0.9)	0.530	0.727	0.678	0.690	0.496	0.560	0.523	0.539
	(0,0,1.2)	0.795	0.912	0.878	0.888	0.698	0.719	0.694	0.701
	(0,0,1.5)	0.947	0.987	0.979	0.982	0.838	0.856	0.834	0.843
	(0,0,0.3)	0.093	0.195	0.160	0.170	0.124	0.186	0.155	0.166
	(0,0,0.6)	0.282	0.415	0.367	0.380	0.298	0.351	0.321	0.329
(2, 2, 2)	(0,0,0.9)	0.550	0.696	0.654	0.667	0.489	0.532	0.506	0.515
	(0,0,1.2)	0.801	0.888	0.864	0.868	0.695	0.706	0.682	0.689
	(0,0,1.5)	0.937	0.969	0.960	0.964	0.848	0.840	0.823	0.829

Table 8: Type 1 error rates of the modified tests by Huber's M-estimators

	(5, 10, 15)				(10, 10, 10)			
(o_1, o_2, o_3)	CF	GF	PB	W	CF	GF	PB	W
(0,0,0)	0.037	0.053	0.036	0.044	0.070	0.067	0.053	0.055
(0,0,1)	0.009	0.047	0.031	0.032	0.033	0.068	0.056	0.062
(0,1,1)	0.013	0.063	0.041	0.044	0.018	0.069	0.057	0.062
(1,1,1)	0.010	0.047	0.040	0.042	0.016	0.059	0.042	0.047
(0,0,2)	0.010	0.063	0.050	0.054	0.048	0.074	0.064	0.065
(0,1,2)	0.011	0.061	0.045	0.049	0.029	60.068	0.051	0.058
(1,1,2)	0.010	0.075	0.054	0.062	0.024	0.063	0.047	0.052
(0,0,3)	0.028	0.076	0.058	0.063	0.224	0.126	0.099	0.104
(0,2,2)	0.019	0.074	0.055	0.060	0.056	0.090	0.068	0.073
(2,2,2)	0.040	0.128	0.099	0.107	0.063	0.115	0.085	0.096
(0,0,0)	0.037	0.053	0.036	0.044	0.070	0.067	0.053	0.055
(0,0,1)	0.002	0.048	0.039	0.042	0.011	0.058	0.041	0.049
(0,1,1)	0.002	0.055	0.035	0.041	0.008	0.056	0.045	0.047
(1,1,1)	0.003	0.044	0.033	0.036	0.003	0.047	0.038	0.043
(0,0,2)	0.009	0.064	0.045	0.054	0.027	0.077	0.061	0.063
(0,1,2)	0.003	0.059	0.044	0.048	0.014	0.082	0.054	0.062
(1,1,2)	0.002	0.044	0.027	0.033	0.008	0.057	0.046	0.050
(0,0,3)	0.003	0.065	0.050	0.052	0.219	0.107	0.077	0.091
(0,2,2)	0.001	0.069	0.051	0.059	0.017	0.080	0.059	0.061
(2,2,2)	0.043	0.104	0.072	0.080	0.015	0.097	0.071	0.082
(0,0,0)	0.037	0.053	0.036	0.044	0.070	0.067	0.053	0.055
(0,0,1)	0.002	0.051	0.041	0.044	0.006	0.059	0.045	0.052
(0,1,1)	0.001	0.052	0.031	0.034	0	0.059	0.044	0.050
(1,1,1)	0	0.040	0.029	0.032	0.001	0.050	0.038	0.038
(0,0,2)	0.001	0.058	0.040	0.048	0.003	0.072	0.056	0.062
(0,1,2)	0	0.049	0.039	0.040	0.002	0.083	0.064	0.072
(1,1,2)	0	0.042	0.023	0.029	0.002	0.063	0.047	0.049
(0,0,3)	0	0.058	0.041	0.043	0.184	0.088	0.066	0.072
(0,2,2)	0	0.057	0.040	0.044	0.002	0.077	0.058	0.061
(2,2,2)	0.027	0.070	0.041	0.047	0.002	0.091	0.072	0.077

The modified tests with Huber's M-estimators have expected performance in terms of power of the test and type 1 error rates as seen on **Table 7** and **Table 8**. Type 1 error rates of the modified tests with Huber's M-estimators are close to nominal level. Especially, modified generalized F-test are more powerful than others and its type 1 error rate is closest to nominal level. It is suggested that, the modified generalized F-test with Huber's M-estimators can be used for comparing group means under heteroskedasticity and nonnormality.

6. Results and conclusions

The motivation of this study is to propose a powerful test for comparing several group means under heteroskedasticity and nonnormality caused by outlier(s). There are many procedures developed for comparing group means under required assumptions. In the real life, these assumptions are not realistic. Most encountered violation is heteroskedasticity for comparing group means procedures. Handling with heteroskedasticity, various alternative procedures are proposed such as Welch F-test, Brown-Forsythe test, Parametric Bootstrap test and Generalized F-test called approximate tests. Although approximate tests show better performance under heteroskedasticity, it is negatively effected from the violation of the normality assumption. Powerful tests can be obtained by offering some modifications to the approximate tests. Cornerstone modification approach is provided by replacing maximum likelihood estimators with robust estimators in the test statistic. In this study Welch F-test, the Parametric Bootstrap test and the Generalized F-test are modified with trimmed mean and variance, median and median absolute deviation, Huber's M-estimators. The performance of the modified tests are compared in terms of power of the tests and type 1 error rates by various Monte-Carlo simulations.

It has been shown in Section 4.1, approximate tests did not perform enough under heteroskedasticity and nonnormality caused by outlier by the Monte-Carlo simulation studies. To overcome this problem, approximate tests are modified with replacing maximum likelihood estimators for location and scale parameters in the test statistics with robust alternatives. Trimmed mean and variance, median and median absolute deviation, Huber's M-estimators are considered for location and scale parameters respectively. As seen on Section 4, while modified tests with trimmed mean and variance, median and median absolute deviation achieve expected level of power, but the type 1 error rates of them are not close to nominal level. According to these results, using of modified tests with mentioned modifications can cause the wrong decisions when comparing the group means under heteroskedasticity and nonnormality. Huber's M-estimators modification to the approximate tests provide powerful test procedures and have nominal type 1 error rates.

Especially for small sample sizes, modified generalized F-test can be safely used for comparing group means under heteroskedasticity and nonnormality caused by outlier. It has better power and provides nominal type 1 error rates comparing with the alternatives.

References

[1] Barnett, V. and Lewis, T. *Outlier in Statistical Data*, Wiley and Sons, 1994.

[2] Ben, M.G. and Yohai, V.J. *Robust analysis of variance for randomized block design*, Communications in Statistics **21**6, 1779-1798, 1992.

[3] Braun, H.I. and McNeil, D.R. *Testing in robust ANOVA*, Communications in Statistics **10**2, 149-165, 1981.

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

[5] Chang, C.H., Lin, J.J., Pal, N. *Testing the equality of several gamma means: a parametric bootstrap method with applications*, Computational Statistics **76**, 26-55, 2011.
[6] Cavus, M., Yazici, B., Sezer, A. *Comparison of means under non-normality with outliers for unbalanced data*, 8th International Conference of the ERCIM Working Group on Computing and Statistics Proceedings, London, ENGLAND, 2015.
[7] Cavus, M., Yazici, B., Sezer, A. *Robust methods to compare means under violation of normality with outliers*, 12th German Probability and Statistics Days Proceedings, Bochum, GERMANY, 2016.
[8] Cavus, M., Yazici, B., Sezer, A. *Modified tests for comparison of group means in the presence of outlier(s)*, 2th International Researchers-Statisticians and Young Statisticians Congress Proceedings, Ankara, TURKEY, 2016.
[9] Fan, W. and Hancock, G. *Robust mean modelling: an alternative for hypothesis testing of independent means under variance heterogeneity and nonnormality*, Journal of Educational and Behavioral Statistics **37**, 137-156, 2012.
[10] Fisher, R.A. *Statistical methods and scientific inference*, Oliver and Boyd, 1925.
[11] Gamage, J. and Weerahandi, S. *Size performance of some tests in one-way ANOVA*, Communications in Statistics **27**3, 625-640, 1998.
[12] Gokpinar Yigit, E., Polat, E., Gokpinar, F., Gunay, S. *A new computational approach for testing equality of inverse Gaussian means under heterogeneity*, Hacettepe Journal of Mathematics and Statistics **42** 5, 581-590, 2013.
[13] Huber, P.J. *Robust estimation of a location parameter*, The Annals of Mathematical Statistics **69**, 73-101, 1964.
[14] Huber, P.J. and Ronchetti, E.M. *Robust Statistics*, John Wiley and Sons, New Jersey, 2009.
[15] Jian-Hong, S. and Jiang-Long, L. *A new generalized p-value for testing equality of inverse Gaussian means under heterogeneity*, Statistics and Probability Letters **82**, 96-102, 2012.
[16] Karagoz, D. *Modified Welch test statistic for ANOVA under Weibull distribution*, Hacettepe Journal of Mathematics and Statistics **45** 56, 2015.
[17] Karagoz, D. and Saracbasi, T. *Robust Brown-Forsythe and robust modified Brown-Forsythe ANOVA tests under heteroscedasticity for contaminated Weibull distribution*, Revista Colombiana de Estadistica **39**, 17-32, 2009.
[18] Krishnamoorthy, K., Lu F., Mathew, T. *A parametric bootstrap approach for ANOVA with unequal variances: Fixed and random models*, Computational Statistics and Data Analysis **51**, 5731-5742, 2007.
[19] Lee, H. and Fung K.Y. *Robust procedures for multi-sample location problems with unequal group variances*, Journal of Statistical Computation and Simulation **18**, 125-143, 1983.
[20] Lee, S. and Ahn, C.H. *Modified ANOVA for unequal variances*, Communications in Statistics **32** 4, 987-1004, 2003.
[21] Li, X. *A generalized p-value approach for comparing the means of several log-normal populations*, Statistics and Probability Letters **79**, 1404-1408, 2009.
[22] Ozkip, E., Yazici, B., Sezer, A. *Comparing the means for unbalanced several skewed data with unequal variances*, 5th International Conference of the ERCIM Working Group on Computing and Statistics Proceedings , 2012a.
[23] Ozkip, E., Yazici, B., Sezer, A. *Comparison of population means for skewed data with unequal variances*, Joint Statistical Meetings 2012 Proceedings , 2524-2529, 2012b.
[24] Ozkip, E., Yazici, B., Sezer, A. *Inference on the mean of unbalanced several lognormal population using generalized p-value and generalized confidence intervals*, 8th International Symposium of Statistics Proceedings , 2012c.
[25] Ozkip, E., Yazici, B., Sezer, A. *Performance of some tests compares for equality of means in skewed distribution data*, Joint Statistical Meetings Proceedings, Montreal, CANADA, 2013.
[26] Ozkip, E., Yazici, B., Sezer, A. *A simulation study on tests for the Behrens-Fisher problem*, Turkiye Klinikleri Biyoistatistik Dergisi,**6** 2, 59-66, 2014.
[27] Ozkip, E., Yazici, B., Sezer, A. *Comparison of tests for the ANOVA with unequal variance*, Joint Statistical Meetings Proceedings, Boston, USA, 2014.
[28] Reed, J.F. and Stark, D.B. *Robust analysis of variance: a simulation study*, Journal of Applied Statistics **22** 1, 87-104, 1995.

[29] Rousseeuw, P. and Croux, C. *Alternatives to the median absolute deviation*, Journal of American Statistical Association **88** 424, 1273-1283, 1993.
[30] Sadooghi-Alvandi, S.M., Jafari, A.A., Mardani-Ford, H.A. *One-way ANOVA with unequal variances*, Communications in Statistics **41**, 4200-4221, 2012.
[31] Schrader, R.M. *Robust analysis of variance*, Communications in Statistics **6** 9, 879-894, 1977.
[32] So, Y.C. and Sen, P.K. *M-estimators based repeated significance tests for one-way ANOVA with adaptation to multiple comparisons*, Communications in Statistics **1** 2, 101-119, 1982.
[33] Tan, W.Y. and Tabatabai, M.A. *Some robust ANOVA procedures under heteroskedasticity and non-normality*, Communications in Statistics **14** 4, 1007-1026, 1985.
[34] Tiku, M. and Akkaya, A. *Robust Estimation and Hypothesis Testing*, New Age International, 2004.
[35] Tsui, K. and Weerahandi, S. *Generalized p-values in significance testing of hypotheses in the presence of nuisance parameters*, Journal of the American Statistical Association **84**, 602-607, 1989.
[36] Weerahandi, S. *Testing variance components in mixed models with generalized p-values*, Journal of the American Statistical Association **86**, 151-153, 1991.
[37] Weerahandi, S. *ANOVA under unequal error variances*, Biometrics **51**, 589-599, 1994.
[38] Weerahandi, S. *Exact Statistical Methods for Data Analysis*, Springer, 1995.
[39] Weerahandi, S. *Generalized Inference in Repeated Measures: Exact Methods in MANOVA and Mixed Models*, Wiley, 2004.
[40] Welch, B.L. *On the comparison of several group means values*, Biometrika **38**, 330-336, 1951.
[41] Wilcox, R.R. *Simulation results on solutions to the multivariate Behrens-Fisher problem via trimmed means*, The Statistician **44** 2, 213-225, 1995.
[42] Wilcox, R.R. *Introduction to Robust Estimation and Hypothesis Testing*, Elsevier, 2012.
[43] Wu, L.L. *Robust M-estimation of location and regression*, Sociological Methodology **15**, 316-388, 1985.
[44] Yazici, B., Sezer, A., Ozkip, E. *A simulation study to compare several population means under heteroskedasticity*, Abstract Book of 5th International Conference of the ERCIM Working Group on Computing and Statistics, Oviedo, SPAIN, 2012.
[45] Yazici, B., Sezer, A., Ozkip, E. *Generalized p-value test to compare several population means for unequal variances*, Joint Statistical Meetings Proceedings, San Diego, USA, 2012.