***Yayın Geliş Tarihi**: 25.02.2016* *Dokuz Eylül Üniversitesi*
***Yayın Kabul Tarihi**: 02.06.2017* *İktisadi ve İdari Bilimler Fakültesi Dergisi*
***Online Yayın Tarihi:** 04.12.2017* *Cilt:32, Sayı:2, Yıl:2017, ss. 1-13*

Tests of Location Equality under Non-Identical Distributions

İrem YILMAZ[1] A. Fırat ÖZDEMİR[2] Gözde NAVRUZ[3]

Abstract

The ANOVA-F test is the most known procedure for comparing at least three population means. However, this conventional test might give misleading results when it's underlying assumptions are violated. In this study, Welch's test with trimmed mean, Welch's test with trimmed mean and a bootstrap-t, newly proposed B^2_{tk} *test and ANOVA-F test were compared in terms of actual Type I error rates under not only non-normality and heteroscedasticity, but also with non-identical distribution shapes. The newly proposed method outperformed ANOVA-F and other alternatives under various situations.*

Keywords: *ANOVA-F test,* B^2_{tk} *test, Welch test, Non-Identical Distribution Shapes, Type I error.*

JEL Classification Codes: *C12, C15.*

Aynı Olmayan Dağılımlar Altında Konum Eşitlik Testleri

Öz

ANOVA-F testi, en az üç kitle ortalamasının karşılaştırılması için en çok bilinen yöntemdir. Ancak, bu geleneksel yöntem varsayımları ihlal edildiğinde yanıltıcı sonuçlar verebilir. Bu çalışmada budanmış ortalama ile Welch testi, budanmış ortalama ve bootstrap-t ile Welch testi, yeni önerilen B^2_{tk} *testi ve ANOVA-F testi gerçekleşen 1. Tip hata oranlarına göre, sadece normalliğin ve homojen varyanslılığın sağlanmadığı durumlarda değil, aynı zamanda aynı olmayan dağılım şekilleri altında da karşılaştırılmıştır. Yeni önerilen yöntem, ANOVA-F ve diğer alternatiflere göre farklı durumlar altında daha iyi sonuçlar vermiştir.*

Anahtar Kelimeler: *ANOVA-F Test,* B^2_{tk} *Test, Welch Test, Aynı Olmayan Dağılım Şekilleri, 1. Tip Hata.*

JEL Sınıflandırma Kodları: *C12, C15.*

[1] Doktora Öğrencisi, Dokuz Eylül Üniversitesi, Fen Fakültesi, İstatistik Bölümü, iremm-yilmazz@hotmail.com

[2] Dr., Dokuz Eylül Üniversitesi, Fen Fakültesi, İstatistik Bölümü, firat.ozdemir@deu.edu.tr

[3] Doktora Öğrencisi, Dokuz Eylül Üniversitesi, Fen Fakültesi, İstatistik Bölümü, gnavruz@gmail.com

1. INTRODUCTION

Testing two or more groups for location equality is a common statistical interest in many working areas such as economics and administrative sciences, medicine, agriculture etc. When there are more than two groups, the most known procedure is the analysis of variance (ANOVA) F test and to get accurate results from the ANOVA-F test, it's fundamental assumptions must be satisfied.

First assumption is the normality of population distributions. In fact, populations are almost never normally distributed (Micceri, 1989). Further they can have highly skewed and heavy tailed shapes. Actual Type I error rate of ANOVA-F test can be overly affected by non-normality that leads taking a wrong decision. Another assumption is equality of population variances which is called as homoscedasticity, and it is very common in wide range of study fields. When there are both non-normality and heteroscedasticity, ANOVA-F test's accuracy and validity can be strongly affected and the actual Type I error rate can highly deviate from the nominal level (Cribbie et al, 2012).

Moreover, the population distributions from which the observations are sampled might be non-identical. Non-identicality of population distributions is a common situation as well as non-normality and heteroscedasticity. The purpose of this study is to compare ANOVA-F, Welch's test with trimmed mean, Welch's test with trimmed mean and a bootstrap-t, and newly proposed B^2_{tk} test with respect to their actual Type I error rates under different experimental conditions.

2. DESCRIPTION OF TESTS

2.1 Welch's Test with Trimmed Mean

When two normal populations are considered without assuming the equal population variances, a problem arises that is known as Behrens-Fisher in literature (Behrens, 1929; Fisher, 1935). The first attempt that aims to solve this problem is Welch's approximate degrees of freedom solution (Welch, 1947). It's degrees of freedom depends on both the sample sizes and the sample variances. Later, Welch developed a generalization of his procedure for comparing k independent groups

(Welch, 1951). Although Welch test is robust to heteroscedasticity, it behaves so sensitive to slight departures from normality due to using the sample mean and variance. If the sample trimmed mean and winsorized variance are used instead of the sample mean and variance, the test becomes robust to both non-normality and heteroscedasticity (Wilcox, 2012). In this study, this procedure is referred as Welch's test with trimmed mean. The hypothesis to test is

$$H_0 : \mu_{t1} = ... = \mu_{tk} = \mu_t \tag{1}$$

Let X_{ij} denotes j th observation of i th group where i=1,2,…,k, j=1,2,…,n. The sample trimmed mean, winsorized mean, and winsorized variance of the i th group are denoted by $\bar{X}_{ti}$, $\bar{X}_{wi}$,and s^2_{wi} respectively.

To calculate the trimmed mean, first trimming proportion (γ) is chosen where $0 \le \gamma \le 0.5$. The number of observations that will be trimmed from each tail is found as $\lambda = [\gamma n]$ where n is the sample size. The observations are put in ascending order and after trimming, $h = n - 2\lambda$ is the effective sample size. The sample trimmed mean is

$$\bar{X}_t = \frac{1}{h}\sum_{j=\lambda+1}^{n-\lambda} X_{(j)} \tag{2}$$

To compute the winsorized variance, winsorized mean is evaluated. After trimming the observations from each tail define,

$$Y_j = \begin{cases} X_{(\lambda+1)}, & X_j < X_{(\lambda+1)} \\ X_j, & X_{(\lambda+1)} < X_j < X_{(n-\lambda)} \\ X_{(n-\lambda)}, & X_j \ge X_{(n-\lambda)} \end{cases} \tag{3}$$

$$\bar{X}_w = \frac{1}{n}\sum_{j=1}^{n} Y_j \tag{4}$$

$$s^2_w = \frac{1}{n-1}\sum_{j=1}^{n} (Y_j - \bar{X}_w)^2 \tag{5}$$

The corresponding test statistic is

$$F_t = \frac{A}{B+1} \tag{6}$$

where

$$A = \frac{1}{k-1}\sum_{i=1}^{k} w_i(\bar{X}_{ti} - \tilde{X})^2 \tag{7}$$

and

$$B = \frac{2(k-1)}{k^2-1}\sum_{i=1}^{k}\frac{\left(1-\frac{w_i}{u}\right)^2}{h_i - 1} \tag{8}$$

where

$$\tilde{X} = \frac{\sum_{i=1}^{k} w_i \bar{X}_{ti}}{U} \tag{9}$$

$$U = \sum_{i=1}^{k} w_i \tag{10}$$

$$w_i = \frac{1}{d_i} \tag{11}$$

$$d_i = \frac{(n_i-1)s_{wi}^2}{h_i(h_i-1)} \tag{12}$$

When the null hypothesis is true F_t has approximately an F distribution with $k-1$ and ν_{wt} degrees of freedom

$$\nu_{wt} = \left(\frac{3}{k^2-1}\sum_{i=1}^{k}\frac{\left(1-\frac{w_i}{U}\right)^2}{h_i}\right)^{-1} \tag{13}$$

The critical value is $F_{k-1,\nu_{wt},\alpha}$ and H_0 is rejected if $F_t \ge F_{k-1,\nu_{wt},\alpha}$.

2.2 Welch's Test with Trimmed Mean and Bootstrap-t

When Welch's test is used in conjunction with 20% trimmed mean and bootstrap-t, actual Type I error rate can be controlled well (Westfall & Young, 1993; Wilcox, 2012).The hypothesis to test is given in equation (1).

The steps of bootstrap-t are

- Determine a bootstrap sample $(X^*_{i1}, X^*_{i2}, ..., X^*_{in})$ for i=1,2,…,k
- Set $C^*_{ij} = X^*_{ij} - \bar{X}_{ti}$, $j = 1, 2, ..., n_i$, where $\bar{X}_{ti}$ is the sample trimmed mean of i th original group
- Calculate F^*_t by using C^*_{ij} values
- Repeat this process B times
- Write the obtained B test statistics in ascending order ($F^*_{t(1)} \le ... \le F^*_{t(B)}$)
- Round $U = (1-\alpha)B$ to nearest integer
- If $F_t \ge F^*_{t(U)}$, reject the null hypothesis where F_t is the test statistic of original sample

2.3 B^2_{tk} Test

Newly proposed B^2_{tk} test is another alternative that works under both heteroscedasticity and non-normality (Özdemir & Yıldıztepe, 2013). The hypothesis to test is given in equation (1). Besides $\bar{X}_{ti}$, $\bar{X}_{wi}$, and s_{wi}, the standard error of the sample trimmed mean is denoted as $s_{\bar{X}_{ti}}$ where i=1,...,k.

$$s_{\bar{X}_{ti}} = \frac{s_{wi}}{(1-2\gamma)\sqrt{n_i}} \quad (14)$$

By using $s^2_{\bar{X}_{ti}}$, the weights of each group are calculated:

$$\omega_i = \frac{\frac{1}{s^2_{\bar{X}_{ti}}}}{\frac{1}{s^2_{\bar{X}_{t1}}} + \frac{1}{s^2_{\bar{X}_{t2}}} + \ldots + \frac{1}{s^2_{\bar{X}_{tk}}}} \tag{15}$$

Then, the weighted mean $X^+ = \omega_1 \bar{X}_{t1} + \omega_2 \bar{X}_{t2} + \ldots + \omega_k \bar{X}_{tk}$ and the following statistic are computed for each group:

$$T_{ti} = \frac{\bar{X}_{ti} - X^+}{s_{\bar{X}_{ti}}} \tag{16}$$

Bailey's normalizing transformation is applied to all T_{ti} (Bailey, 1980).

$$z_{ti} = \pm \frac{4\nu_i^2 + \frac{5(2z_c^2+3)}{24}}{4\nu_i^2 + \nu_i + \frac{(4z_c^2+9)}{12}} \nu_i^{1/2} \left\{ \ln\left(1 + \frac{T_{ti}^2}{\nu_i}\right) \right\} \tag{17}$$

where $\nu_i = h_i - 1$, $h_i = n_i - 2\lambda_i$ and z_c is the critical value for the related significance level under the standard normal distribution. The sign of the z_{ti} 's are the same as T_{ti} 's. The corresponding test statistic is calculated as given below

$$B_{tk}^2 = \sum_{i=1}^{k} z_{ti}^2 = \sum_{i=1}^{k} \left(\frac{4\nu_i^2 + \frac{5(2z_c^2+3)}{24}}{4\nu_i^2 + \nu_i + \frac{(4z_c^2+9)}{12}} \nu_j^{1/2} \left\{ \ln\left(1 + \frac{\left(\frac{\bar{X}_{ti} - X^+}{S_{\bar{X}_{ti}}}\right)^2}{\nu_i} \right) \right\}^{1/2} \right) \tag{18}$$

The critical value of the method is found by applying a bootstrap-t:

- Compute a bootstrap sample $X_{i1}^*, \ldots, X_{in_i}^*$ of size n_i and set $C_{ij}^* = X_{ij}^* - \bar{X}_{ti}$ $i = 1, \ldots, k; j = 1, 2, \ldots, n_i$

- Compute the test statistics B_{tk}^{2*} by using $C_{ij}^{*} = X_{ij}^{*} - \bar{X}_{ti}$
- Repeat this process B times yielding $B_{tk1}^{2*}, ..., B_{tkB}^{2*}$
- $B_{tk1}^{2*}, ..., B_{tkB}^{2*}$ values are put in ascending order, $B_{tk(1)}^{2*} \leq ... \leq B_{tk(B)}^{2*}$
- Set $U = (1-\alpha)B$, rounding to the nearest integer and find $B_{tk(U)}^{2*}$
- Reject H_0 if $B_{tk}^{2} \geq B_{tk(U)}^{2*}$ where B_{tk}^{2} is found from the original sample

3. DESIGN OF THE SIMULATION STUDY

Actual Type I error rates are investigated under 80 different experimental conditions including non-identical distributions. The conditions are obtained by combining homoscedasticity and heteroscedasticity, equal and unequal sample sizes with positive (the largest sample size paired with the largest population variance and the smallest sample size paired with the smallest population variance) and negative (the largest sample size paired with the smallest population variance and the smallest sample size paired with the largest population variance) pairings of variances, symmetric and asymmetric distributions. Non-identicality of distributions are designed by using three symmetric, two symmetric and one asymmetric, one symmetric and two asymmetric, and three asymmetric distributions. Symmetric distributions are N(0, 1), t(4), U(-1, 1), g-h(0, 0.14), g-h(0, 0.2), and g-h(0, 0.22) whereas asymmetric ones are $\chi^2(4)$, exp(3), g-h(0.81, 0), and g-h(1, 0).

g-h distribution allows to observe how a distribution differs from normality with the skewness parameter g and kurtosis parameter h, and when g=h=0 the g-h distribution is equivalent to standard normal distribution.

Let Z be a standard normal random variable and the following two transformations are used to generate data from g-h distribution.

When $g \neq 0$,

$$X = \frac{(\exp(gZ)-1)\exp\left(\frac{hZ^2}{2}\right)}{g} \quad (19)$$

When g≠ 0,

$$X = Z\exp\left(\frac{hZ^2}{2}\right) \quad (20)$$

(Hoaglin, 1985).

ANOVA-F test, Welch's test with trimmed mean, Welch's test with trimmed mean and a bootstrap-t, and B_{tk}^2 test are represented as F, TW, BTW, and B_{tk}^2 respectively. Nominal significance level is set at 0.05 and the number of replications for all experimental conditions is 10000. All simulations are performed with statistical programming language R (version 3.1.2).

4. SIMULATION RESULTS

Bradley's conservative criterion of robustness ($0.9\alpha \le \hat{\alpha} \le 1.1\alpha$) was used for evaluating actual Type I error rates of the methods (Bradley, 1978). When the actual Type I error rate of a method falls within the interval 0.045 and 0.055, that method is considered as robust when the nominal significance level is 0.05.Cells colored dark grey represent liberal results whereas cells colored light grey represent the conservative results.

Table 1 shows the actual type I error rates of the methods under equal sample sizes and homoscedasticity. In this table, F and TW generally gave liberal results whereas B_{tk}^2 and BTW were good at controlling Type I error rates for almost all settings. Although the population variances are equal with equal sample sizes, non-identical distribution shapes caused deviations from nominal significance level for F and TW. Note that all methods could control the actual Type I error rates when all populations come from symmetric distributions.

Table 2 shows the actual Type I error rates under equal sample sizes and heteroscedasticity. According to this table, F gave liberal results under all settings and TW followed it. However, B_{tk}^2 and BTW saved the nominal Type I error rates in 14 settings. Since there is heteroscedasticity, F test could not control the actual Type I error rates even with three symmetric distributions. On the other side, TW gave acceptable results under three symmetric distributions due to insensitivity to heteroscedasticity. Table 3 shows the actual Type I error rates under unequal sample sizes and homoscedasticity. In this table, F and TW could not save the results for almost all cases. Even though the population variances are equal, F and TW are affected by unequal sample sizes and non-identical distribution shapes.

Table 4 shows the actual Type I error rates under unequal sample sizes with positive pairing of variances. In this table, F gave both conservative and liberal results. Although TW controlled actual Type I error rates when all distributions are symmetric, it gave liberal results under settings with asymmetric distributions. Furthermore, results of B_{tk}^2 and BTW fell within the interval (0.045, 0.055) in almost all settings. Table 5 shows the actual Type I error rates under unequal sample sizes with negative pairing of variances. Because of negative pairing, F and TW had liberal results. However, B_{tk}^2 and BTW had a good control over actual Type I error rates as in the other tables.

5. CONCLUSION

In experimental research, there is an important interest in obtaining an appropriate method under non-normality, heteroscedasticity and non-identical distribution shapes. In this study, F, TW, BTW and a newly proposed B_{tk}^2 test were compared in terms of saving actual Type I error rates. Out of 80 different settings, the nominal significance level was saved 9 times by F, 24 times by TW, 70 times by BTW, and 73 times by B_{tk}^2 .In general B_{tk}^2 had the best results and BTW followed it. The number of conservative, liberal and controlled results for all methods is shown in Figure 1.

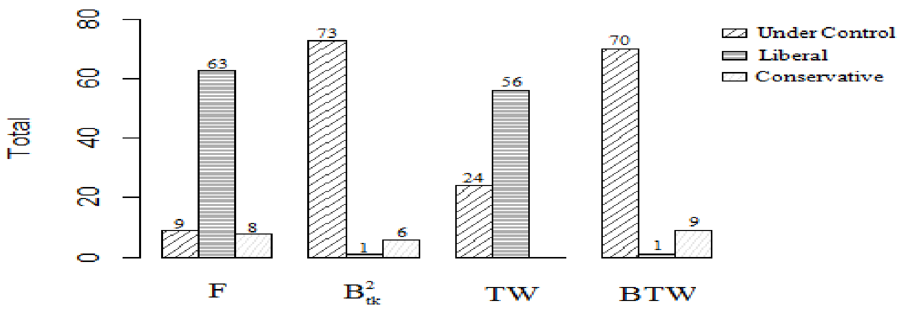

Figure 1. Number of The Controlled, Liberal and Conservative Results

As a consequence, the performances of newly proposed B^2_{tk} test and BTW on controlling actual Type I error rates were considerably well when normality and homogeneity of variances assumptions are violated and the distributions have non-identical shapes. According to this study, the newly proposed test B^2_{tk} and BTW can be recommended under similar conditions covered in this study.

Table 1. Actual Type I Error Rates under Equal Sample Sizes and Homoscedasticity

n=20 20 20; σ=1 1 1	F	B^2_{tk}	TW	BTW
3 symmetric				
Normal - t – Unif	0.0567	0.0475	0.0549	0.0489
Normal - Unif - gh(0, 0.2)	0.0523	0.0462	0.0517	0.0469
Normal - gh(0, 0.14) - gh(0, 0.2)	0.0482	0.0509	0.0533	0.0484
gh(0, 0.14) - gh (0, 0.2) - gh(0, 0.22)	0.0469	0.0490	0.0506	0.0488
2 symmetric, 1 asymmetric				
Normal - Unif – Chi	0.0844	0.0475	0.0554	0.0448
Normal - Unif – Exp	0.0639	0.0478	0.0606	0.0504
Normal - Unif - gh(1, 0)	0.1076	0.0455	0.0540	0.0453
Normal - Exp - gh(0, 0.2)	0.0624	0.0458	0.0543	0.0474
1 symmetric, 2 asymmetric				
Normal - Chi – Exp	0.0945	0.0456	0.0609	0.0483
Normal - Exp - gh(1, 0)	0.1088	0.0505	0.0648	0.0517
Chi - Exp - gh(0, 0.2)	0.0859	0.0471	0.0601	0.0483
gh(0, 0.2) - Chi - gh(0.81, 0)	0.0616	0.0498	0.0562	0.0510
3 asymmetric				
Chi - Chi – Exp	0.0633	0.0465	0.0661	0.0474
Chi - Exp - Exp	0.1044	0.0451	0.0551	0.0456
Chi - Exp - gh(0.81, 0)	0.0755	0.0518	0.0695	0.0537
gh(0.81, 0) - gh(0.81, 0) - gh(1, 0)	0.0449	0.0394	0.0460	0.0398

Table 2. Actual Type I Error Rates under Equal Sample Sizes and Heteroscedasticity

n=20 20 20; σ=1 2 4	F	B^2_{tk}	TW	BTW
3 symmetric				
Normal - t – Unif	0.0575	0.0474	0.0598	0.0488
Normal - Unif - gh(0, 0.2)	0.0786	0.0454	0.0525	0.0468
Normal - gh(0, 0.14) - gh(0, 0.2)	0.0668	0.0412	0.0481	0.0422
gh(0, 0.14) - gh (0, 0.2) - gh(0, 0.22)	0.0658	0.0469	0.0507	0.0489
2 symmetric, 1 asymmetric				
Normal - Unif – Chi	0.1018	0.0507	0.0655	0.0496
Normal - Unif – Exp	0.0577	0.0488	0.0587	0.0481
Normal - Unif - gh(1, 0)	0.1621	0.0515	0.0675	0.0518
Normal - Exp - gh(0, 0.2)	0.0800	0.0455	0.0523	0.0462
1 symmetric, 2 asymmetric				
Normal - Chi – Exp	0.0994	0.0463	0.0579	0.0479
Normal - Exp - gh(1, 0)	0.1650	0.0517	0.0684	0.0529
Chi - Exp - gh(0, 0.2)	0.0644	0.0490	0.0630	0.0534
gh(0, 0.2) - Chi - gh(0.81, 0)	0.0678	0.0506	0.0678	0.0519
n=20 20 20; σ=1 2 4	F	B^2_{tk}	TW	BTW
3 asymmetric				
Chi - Chi – Exp	0.0779	0.0507	0.0629	0.0534
Chi - Exp - Exp	0.0812	0.0475	0.0588	0.0500
Chi - Exp - gh(0.81, 0)	0.1107	0.0500	0.0702	0.0518
gh(0.81, 0) - gh(0.81, 0) - gh(1, 0)	0.1279	0.0571	0.0690	0.0591

Table 3. Actual Type I Error Rates under Unequal Sample Sizes and Homoscedasticity

n=20 25 30; σ=1 1 1	F	B^2_{tk}	TW	BTW
3 symmetric				
Normal - t – Unif	0.0633	0.0464	0.0493	0.0432
Normal - Unif - gh(0, 0.2)	0.0482	0.0517	0.0555	0.0488
Normal - gh(0, 0.14) - gh(0, 0.2)	0.0417	0.0483	0.0496	0.0463
gh(0, 0.14) - gh (0, 0.2) - gh(0, 0.22)	0.0481	0.0533	0.0529	0.0522
2 symmetric, 1 asymmetric				
Normal - Unif – Chi	0.0563	0.0500	0.0592	0.0505
Normal - Unif – Exp	0.0943	0.0466	0.0585	0.0463
Normal - Unif - gh(1, 0)	0.0880	0.0496	0.0572	0.0477
Normal - Exp - gh(0, 0.2)	0.0461	0.0480	0.0545	0.0509
1 symmetric, 2 asymmetric				
Normal - Chi – Exp	0.0906	0.0481	0.0596	0.0479
Normal - Exp - gh(1, 0)	0.0940	0.0489	0.0611	0.0486
Chi - Exp - gh(0, 0.2)	0.1081	0.0486	0.0575	0.0478
gh(0, 0.2) - Chi - gh(0.81, 0)	0.0650	0.0523	0.0566	0.0526
3 asymmetric				
Chi - Chi – Exp	0.0916	0.0486	0.0660	0.0488
Chi - Exp - Exp	0.1569	0.0481	0.0565	0.0488
Chi - Exp - gh(0.81, 0)	0.1056	0.0514	0.0650	0.0510
gh(0.81, 0) - gh(0.81, 0) - gh(1, 0)	0.0493	0.0426	0.0470	0.0428

Table 4. Actual Type I Error Rates under Unequal Sample Sizes and Heteroscedasticity (Positive Pairing)

n=20 25 30; σ=1 2 4	F	B^2_{tk}	TW	BTW
3 symmetric				
Normal - t – Unif	0.0423	0.0468	0.0532	0.0474
Normal - Unif - gh(0, 0.2)	0.0419	0.0486	0.0532	0.0481
Normal - gh(0, 0.14) - gh(0, 0.2)	0.0372	0.0442	0.0478	0.0467
gh(0, 0.14) - gh (0, 0.2) - gh(0, 0.22)	0.0375	0.0478	0.0465	0.0466
2 symmetric, 1 asymmetric				
Normal - Unif – Chi	0.0620	0.0528	0.0621	0.0522
Normal - Unif – Exp	0.0506	0.0498	0.0561	0.0484
Normal - Unif - gh(1, 0)	0.1155	0.0502	0.0639	0.0509
Normal - Exp - gh(0, 0.2)	0.0421	0.0453	0.0457	0.0439
n=20 25 30; σ=1 2 4	F	B^2_{tk}	TW	BTW
1 symmetric, 2 asymmetric				
Normal - Chi – Exp	0.0924	0.0497	0.0567	0.0474
Normal - Exp - gh(1, 0)	0.1132	0.0514	0.0635	0.0515
Chi - Exp - gh(0, 0.2)	0.0399	0.0467	0.0569	0.0473
gh(0, 0.2) - Chi - gh(0.81, 0)	0.0481	0.0520	0.0635	0.0528
3 asymmetric				
Chi - Chi – Exp	0.0894	0.0509	0.0635	0.0508
Chi - Exp - Exp	0.1045	0.0435	0.0530	0.0438
Chi - Exp - gh(0.81, 0)	0.0787	0.0477	0.0642	0.0495
gh(0.81, 0) - gh(0.81, 0) - gh(1, 0)	0.0885	0.0525	0.0579	0.0521

Table 5. Actual Type I Error Rates under Unequal Sample Sizes and Heteroscedasticity (Negative Pairing)

n=20 25 30; σ=4 2 1	F	B^2_{tk}	TW	BTW
3 symmetric				
Normal - t – Unif	0.1041	0.0462	0.0552	0.0448
Normal - Unif - gh(0, 0.2)	0.1128	0.0454	0.0554	0.0452
Normal - gh(0, 0.14) - gh(0, 0.2)	0.0928	0.0477	0.0560	0.0463
gh(0, 0.14) - gh (0, 0.2) - gh(0, 0.22)	0.0946	0.0495	0.0530	0.0479
2 symmetric, 1 asymmetric				
Normal - Unif – Chi	0.0777	0.0479	0.0560	0.0483
Normal - Unif – Exp	0.1324	0.0463	0.0635	0.0469
Normal - Unif - gh(1, 0)	0.1026	0.0484	0.0566	0.0481
Normal - Exp - gh(0, 0.2)	0.1172	0.0427	0.0488	0.0420
1 symmetric, 2 asymmetric				
Normal - Chi – Exp	0.0880	0.0457	0.0608	0.0455
Normal - Exp - gh(1, 0)	0.1051	0.0472	0.0577	0.0470
Chi - Exp - gh(0, 0.2)	0.1574	0.0465	0.0577	0.0468
gh(0, 0.2) - Chi - gh(0.81, 0)	0.0812	0.0467	0.0554	0.0478

3 asymmetric				
Chi - Chi – Exp	0.1230	0.0453	0.0609	0.0433
Chi - Exp - Exp	0.1614	0.0518	0.0609	0.0471
Chi - Exp - gh(0.81, 0)	0.1584	0.0514	0.0633	0.0512
gh(0.81, 0) - gh(0.81, 0) - gh(1, 0)	0.1165	0.0541	0.0659	0.0549

REFERENCES

BAILEY, B. J. R. (1980), “Accurate normalizing transformations of student's t variate”, *Applied Statistics,* 29, 304-306.

BEHRENS, W. V. (1929), “Ein beitrag zur fehlerberechnung bei wenigen beobachtungen“, *LandwirischaftlicheJahrburcher,* 68, 807-837.

BRADLEY, J. V. (1978), “Robustness?”, *British Journal of Mathematical and Statistical Psychology,* 31, 144-152.

CRIBBIE, R. A., FIKSENBAUM, L., KESELMAN, H.J., WILCOX, R. R. (2012), “Effects of non-normality on test statistics for one-way independent group designs”, *British Journal of Mathematical and Statistical Psychology,* 65, 56-73.

FISHER, R. A. (1935), “The fiducial argument in statistical inference”, *Annals of Eugenics,* 6, 198-391.

HOAGLIN, D. C. (1985), *Summarizing shape numerically: The g-and-h distribution,* In: Haoglin D., Mostseller F., Tukey J. editors. *Exploring data tables trends and shapes.* Wiley New York, 461-515.

MICCERI, T. (1989), “The Unicorn, the normal curve, and other improbable creatures”, *Psychological Bulletin* 105: 156-166.

ÖZDEMİR, A. F., YILDIZTEPE, E. (2013), “Bir konum ölçüsü ile k bağımsız grubun karşılaştırılması için bir test”, *Anadolu University Journal of Science and Technology A-Applied Science and Engineering,* 14, 241-249.

WELCH, B. L. (1947), “The generalization of 'student's' problem when several population variances are involved”, *Biometrika,* 34, 28-35.

WELCH, B. L. (1951), “On the comparison of several mean values: An alternative approach”, *Biometrika,* 38, 330–336.

WESTFALL, P. H., YOUNG, S. S. (1993), *Resampling based multiple testing,* Wiley, New York.

WILCOX, R. R. (2012), *Introduction to robust estimation and hypothesis testing (third edition)*, Elsevier Academics Press San Diago, California.