THE EFFECTS OF NON-NORMALITY ON TYPE III ERROR FOR COMPARING INDEPENDENT MEANS

Mehmet MENDES[1]

PhD, Associate Professor, Canakkale Onsekiz Mart University, Agriculture Faculty,
Animal Science Department, Biometry and Genetics Unit 17020, Canakkale, Turkey

E-mail: mmendes@comu.edu.tr

Abstract: *The major objective of this study was to investigate the effects of non-normality on Type III error rates for ANOVA F its three commonly recommended parametric counterparts namely Welch, Brown-Forsythe, and Alexander-Govern test. Therefore these tests were compared in terms of Type III error rates across the variety of population distributions, mean difference (effect size), and sample sizes. At the end of 100,000 simulation trials it was observed that the Type III error rates for four tests were affected by the effect size (δ) and sample size, whereas Type III errors were not affected from distribution shapes. Results of the simulation also indicated that increases in sample size and population mean difference decreased Type III error, and increased statistical test power. Across the all distributions, sample sizes and population mean differences (δ), the Alexander-Govern test obtained higher estimates for power, lower estimates of Type III error (γ).*

Key words: *Type I error rates, power of test, Type III error rates, normality, ANOVA*

1. Introduction

There are variety of alternatives to ANOVA F test under non-normality and homogeneity of variance. Among the most frequently cited parametric alternatives to ANOVA are Welch test, James second-order test, Marascuilo test, Brown-Forsythe test, K-statistic, Wilcox H_m test, Alexander-Govern test, trimmed mean (Welch, 1951; Brown and Forsythe, 1974; Alexander and Govern, 1994; Mehrota, 1997; Keselman, et al., 2002; Mendeş, 2002; Camdeviren and Mendes, 2005). Most of the studies related to test power were considered Type I and Type II error rates, however a third type of error has been suggested in the literature (Leventhal and Huynh, 1996; Leventhal 1999; MacDonald, 1999). Previous studies have investigated, in general, the power and Type I error rates for ANOVA F test and its various alternatives across variety of population distributions, variance patterns, sample sizes, and effect sizes (δ). Unfortunately in practice, Type III error is not taken into consideration. To date, the studies of the error rates control in statistical tests have not examined the probability of rejection in the wrong direction for ANOVA F test and its parametric alternatives when mean differences (effect size) exist. However, Type III error affects test power especially when sample sizes are small (Leventhal and Huynh, 1996; Leventhal 1999; MacDonald, 1999; Mendes, 2004).

The authors of previous studies have explicitly state that the tests for statistical significance would be two tailed. It is known that in a two-tailed hypothesis test, rejection of the null hypothesis means that the groups tested do not come from populations with a common μ. However, this information does not give an idea about the direction of the difference since conventional two-tailed tests evaluate non-directional statistical hypotheses and do not provide directional decisions (Shaffer 1972, Leventhal and Huynh, 1996; Finner, 1999). Recently, Leventhal and Huynh (1996), Leventhal (1999), Jones and Tukey (2000) have reviewed interest in the directional two-tailed t test, in part because, these authors maintain, knowing the null hypothesis is false implies that one of the alternative hypothesis is true, but not which true. The directional two-tailed test makes its contribution by telling us which directional alternative to accept (Leventhal, 1999, page 6). Camdeviren and Mendes (2005) had a simulation study for Type III error rates of some variance homogeneity tests

The power of a test is traditionally defined as the probability of rejecting a false null hypothesis (Cohen, 1988; Zar, 1999; Ferron and Sentovich, 2002). But, this definition is not always appropriate. Leventhal and Huyhn (1996) suggested that power can be defined as the probability of correctly rejecting a false null hypothesis and can be calculated as Power=1-β-γ. Type III error (γ) refers to correctly rejecting the null hypothesis, but incorrectly inferring the direction of the effect. Directional decisions on non-directional tests will overestimate power, underestimate sample size, and ignore the risk of Type III error under the definition of Leventhal and Huyhn (1996). By studying the Type III error rates for tests, one can evaluate, empirically, relative merits of using the statistical tests to analyze data. Correction of the power value adjusted to the Type III error rate is much lower than the power value classically calculated, especially in small samples (Muller and Lavange, 1992; Sansgiry and Akman, 2000).

For instance, if true mean differences exist between population A and population B, or among population A, population B, and population C on some measures of interest (e.g., for two populations $\mu_A \rangle \mu_B$, and for three populations $\begin{matrix}\mu_A > \mu_B \\ \mu_A > \mu_C\end{matrix}$), it would be possible for a researcher to commit two types of errors:

a) Type II error, which is the acceptance of a false null hypothesis with the conditional probability β.

b) Type III error, which is the rejection of a false null hypothesis with the conditional probability of γ and concluding a mean difference in the wrong direction (e.g., for two populations $\mu_A < \mu_B$, and for three populations $\begin{matrix}\mu_A < \mu_B \\ \mu_A < \mu_C\end{matrix}$).

Note that we are only considering the case where one mean μ_A differs from the rest as opposed to general departure from equality when there are more than two groups. These two types of errors directly affect the power of a test. Under this definition of power, the probability of making a Type III error must be eliminated (Leventhal and Huyhn, 1996; Sharon and Carpenter 1999) for calculations of power and sample size. If the direction of an effect is known, results will be more informative.

Another way to understand the directional two-tailed test is to view it as a single test evaluating three statistical hypotheses: $H_{0,}$ H_1, and H_2. When testing the difference between two sample means, the hypotheses are

$H_0 : \mu_1 = \mu_2$, $H_1 : \mu_1 < \mu_2$ and $H_1 : \mu_1 > \mu_2$

Where H_0 is the null hypothesis, H_1, and H_2 are the alternative hypotheses.

Table 1. Relationship of the "Truth" and the decision about null hypothesis

Decision		Nature		
		H_1 true	H_0 true	H_2 true
Decision about nature	H_1 accept	Correct decision	Type I error (α)	Type III error (γ)
	H_0 accept	Type II error (β)	Correct decision	Type II error (β)
	H_2 accept	Type III error (γ)	Type I error (α)	Correct decision

Therefore, Type III error (γ) is only possible only when H_1 or H_2 is true. Two cells, accept H_2 when H_1 true and accept H_1 when H_2 true make different type of this error. There is no Type III error if null hypothesis is accepts. It can be seen that the non-directional two-tailed test does not provide for a directional decision and, hence cannot make a Type III error. Schaffer (1972) notified that a one-tailed test could make a Type III error by accepting directional alternative when the truth falls in the opposite direction. Therefore, in power studies, accordingly, with the revisited definition, the three-choice test's power is Power=1-- γ for a given state of nature.

In the simplest case, two groups with equal variance; the Type III error rate can be analytically derived from the non-central t distribution. The difference in means $\bar{X}_A - \bar{X}_B$ has standard error $\sqrt{2S^2/n}$ for two samples of size n so if t_0 is the left tail critical value for example, the probability of rejecting for a given δ becomes

$$\Pr\{\frac{\bar{X}_A - \bar{X}_B}{\sqrt{2S^2/n}} < t_0\} = \Pr\{\frac{(\bar{X}_A - \mu - \delta\sigma) - (\bar{X}_B - \mu) + \delta\sigma}{\sqrt{2S^2/n}} < t_0\} = \Pr\{\frac{Z + \delta\sqrt{n/2}}{\sqrt{S^2/\sigma^2}} < t_0\}$$

Where $Z = \dfrac{(\bar{X}_A - \mu - \delta\sigma) - (\bar{X}_B - \mu)}{\sqrt{2\sigma^2/n}}$.

This probability can be computed from the non-central t distribution with non-centrality parameter $\delta\sqrt{n/2}$.

The major objective of this study is to investigate the effects of non-normality on Type III error for ANOVA F, Welch, Brown-Forsythe, and Alexander-Govern tests.

1.1. Definition of Statistical Tests

Let X_{ik} be the i[th] observation in the k[th] group. Where i=... n_k and k=1...K; let $\sum n_k = N$. The X_{ik}'s are assumed to be independent and normally distributed with expected values μ_k and variances σ_k^2. The best linear unbiased estimates of μ_k and σ_k^2 are

$$\bar{X}_{.k} = \frac{\sum X_{ik}}{n_k} \text{ and } S_k^2 = \frac{\sum (X_{ik} - \bar{X}_{.k})^2}{(n_k - 1)} \text{ respectively.}$$

1.1.1. ANOVA F Test:

$$F=\frac{\sum_{k} n_k (\overline{X}_{.k}-\overline{X}_{..})^2/(K-1)}{\sum_{i}\sum_{k}(X_{ik}-X_{.k})\,/(N-K)} \qquad [1]$$

Where $X_{..}=\dfrac{\sum_{k} n_k \overline{X}_{.k}}{N}$, when population variances are equal. F is distributed as a central F variable with (K-1) and (N-K) degrees of freedom.

1.1.2. Welch Test

The test statistic for this test is $F_w=\dfrac{\sum_{k} W_k (\overline{X}_{.k}-X'_{..})^2/(K-1)}{\left[1+\frac{2}{3}(K-2)\Lambda\right]}$ [2]

Where $W_k=\dfrac{n_k}{S_k^2}$, $X'_{..}=\dfrac{\sum_{k} W_k \overline{X}_{.k}}{\sum_{k} W_k}$ and $\Lambda=\dfrac{3\sum_{k}(1-W_k/\sum_{k} W_k)^2/(n_k-1)}{(K^2-1)}$ F_w statistic is approximately distributed as a central F variable with (K-1) and 1/∧ degrees of freedom.

1.1.3. Brown-Forsythe Test

Mehrota (1997) developed the following test

$$F_{BF}=\frac{\sum_{k} n_k (\overline{X}_{.k}-\overline{X})^2}{\sum_{k}(1-n_k/N)\,S_k^2} \qquad [3]$$

In attempt to correct a "flaw" in the original Brown-Forsythe test. The "flaw" in the Brown-Forsythe testing procedure, as identified by Mehrota (1997), is the specification of the numerator degrees of freedom. In this study, Brown-Forsythe method proposes by Mehrota (1997) was used instead of the usual Brown-Forsythe method. Specially, Brown-Forsythe used K-1 numerator degrees of freedom whereas Mehrota (1997) used a Box (1954) approximation to obtain the numerator degrees of freedom, v_1, where

$$v_1=\frac{\left[\sum_{i=1}^{K}(1-n_i/N)\,S_i^2\right]^2}{\sum_{i=1}^{K}S_i^4+\left[\sum_{i=1}^{K}n_i S_i^2/N\right]^2-2.\sum_{i=1}^{K}n_i S_i^4/N} \qquad [4]$$

and the denominator degrees of freedom; $v=\dfrac{\left[\sum_{i=1}^{K}n_i(1-n_i/N)\,S_i^2\right]^2}{\sum_{i=1}^{K}(1-n_i/N)^2 S_i^4/(n_i-1)}$ [5]

Under null hypothesis, F_{BF} is distributed approximately as an F variable with v_1 and v degrees of freedom (Mehrota, 1997).

1.1.4. Alexander-Govern Test

The test statistic for this test is $AG=\sum_{k=1}^{K} Z_k^2$ [6]

Where $Z_k = c + \frac{(c^3 + 3c)}{b} - \frac{(4c^7 + 33c^5 + 240c^3 + 855)}{(10b^2 + 8bc^4 + 1000b)}$

$a = v_k - 0.5$, $b = 48a^2$, $c = \sqrt{a * \ln(1 + \frac{t_k^2}{v_k})}$, $t_k = \frac{\overline{X}_k - X^+}{S_{\overline{X}_k}}$,

$X^+ = \sum_{k=1}^{K} W_k \overline{X}_k$, and $v_k = n_k - 1$.

AG statistic is approximately distributed as a chi-square distribution with (K-1) degrees of freedom (Alexander and Govern, 1994; Schneider and Penfield, 1997).

2. Material and Methods

Table 1. The characteristics of the distributions

Distributions	Mean	Variance	Skewness	Kurtosis
Normal (0, 1)	0.00	1.00	0.00	3.00
t (5)	0.00	1.67	0.00	6.00
χ^2 (3)	3.00	6.00	1.63	4.00
β (10, 10)	0.50	0.01	-0.001	-0.27
Exp (1.25)	1.25	1.60	2.00	6.08
W (1.5, 1)	0.92	0.37	1.07	1.38

N (0, 1): standard norlma dist., **t (5):** t-dist. with 5 df., χ^2 **(3):** Chi-square dist. with 3 df.
β (10, 10): Beta dist. with (10, 10) parameters, **Exp (1.25):** Exponential dist. with (1.25) parameter
W (1.5, 1): Weibull dist. with (1.5, 1) parameters

A computer simulation program was used Monte Carlo techniques to investigate the effects of non-normality on Type III error rates of ANOVA F and its three commonly recommended parametric alternatives across a variety of experimental conditions. The error rates of four tests were evaluated under six different population shapes (Normal (0, 1), t-distribution with 5 df, chi-square with 3 df, Exponential (1.25), Beta (10, 10), and Weibull (1.5, 1) and seven sample-size pairings (n_1, n_2, n_3) of (5, 5, 5), (10, 10, 10), (20, 20, 20), (30, 30, 30), (3, 4, 5), (5, 10, 15), and (10, 20, 30). Those distributions were selected since those distributions are predominantly used in literature (Alexander and Govern, 1994; Wilcox, 1994; Penfield, 1994; Keselman et al., 2002; Mendeş, 2002). Distributions were generated using random number generators from IMSL (functions RNNOA, RNSST, RNCHI, RNEXP, RNBET, and RNWIB) (Anonymous, 1994). Sawilowsky and Blair (1992) investigated effects of eight non-normal distributions, which were identified by Micceri (1989) on the robustness of Student's t test, and they found that only the distributions with the most extreme degree of skewness (e.g., skewness=1.64) affected Type I error control of the independent sample t statistics. In this study, maximum degree of skewness used was 2.00. In this study, only small sample size conditions were taken into consideration. Because in practice, researchers were studied with small sample sizes. The effects of Type III error on test power were more obvious, especially when sample sizes were small (MacDonald, 1999; Mendeş, 2004).

The populations were standardized because they have different means and variances. Shape of distributions was not changed while the means were changed to 0 and the standard deviations were changed to 1. The effect sizes (standardized mean differences (δ) of 0.8 and more standard deviation approximate those suggested by Cohen (1988) to represent large effect sizes. In this study, we used (0.25) standard deviation to represent small effect size, (0.75) standard deviation to represent medium effect size, and (1.0) standard deviation to represent large effect size. To make a difference between the population means in which generated samples were taken from, specific constant numbers in standard deviation form (δ=0.25, 0.75, 1.0) were added to the random numbers of the first population. We have done computations for many other parameter values as well; since the results are quite similar, for saving printing space, other results will not be given here. 100,000 runs were generated for each distribution and each given set of parameter values and frequencies of samples for the rejection regions were counted for the ANOVA F test, the Welch test, the Brown-Forsythe test, and the Alexander-Govern test. That is, for each pair of samples, ANOVA F (F), Welch (W), Brown-Forsythe (BF), and Alexander-Govern (AG) test statistics were calculated (for the F test we compute F and count the frequency satisfying F > F (k-1, N-k-1) d.f, for the Welch test we compute F_W values and count the frequency satisfying F_W > F (k-1, 1/$\wedge$) d.f, for Brown-Forsythe test, we compute F_{BF} and count the frequency satisfying F_{BF} > F (v_1, v) d.f., and for Alexander-Govern test we compute AG and count the frequency satisfying AG > χ^2 (k-1) and a check was made to see if the hypothesis which is actually true was rejected and which is actually false was rejected at α=0.05. The experiment was repeated 100,000 times and the proportion of observations falling in the critical regions was recorded. This proportion estimation is test power if the means from the populations do differ ($\mu_1 \neq \mu_2$). Type III error rate was obtained by counting how many times the highest mean population in real is smaller than the other population means (r_1) in the rejected sum of hypothesis and transforming this number into relative frequency ($\gamma=r_1$ / 100,000). That is, Type III (rejection of a false null hypothesis in the wrong direction) error rates were computed for conditions in which the null hypothesis was false. We wrote a FORTRAN program for Intel Pentium III processor to compute all tests.

3. Results and Discussion

The results are presented in Tablo 2-7. Table 2 contains the Type III error rates of four tests when distributions were normal. Across all sample sizes, the estimates of Type III error rates for W test ranged from 1.00 % to 1.52 % under small effect size (0.25), ranged from 0.09% to 0.80% under medium effect size (0.75), and ranged from 0.00% to 0.48% under large effect size (1.00). The estimates of Type III error rates for BF test ranged from 0.89% to 1.48% under small effect size (0.25), ranged from 0.09 % to 0.70 % under medium effect size (0.75), and ranged from 0.00% to 0.70% under large effect size (1.00). Under the same conditions, the estimates of Type III error rates for AG test ranged from 0.13% to 0.64%, from 0.05% to 0.26%, and from 0.00% to 0.17%, respectively, whereas the Type III error estimates for F test ranged from 1.31% to 2.20%, from 0.11% to 1.35%, and from 0.00% to 0.72% respectively. That result demonstrated that the alternative tests were more robust than the F test at controlling the probability of Type III error rates. On the other hand, it can be said that AG test is more robust than the others at controlling the probability of Type III error. As we expected, probability of a rejection in the wrong direction decreased as sample size and population mean differences increased. It was also seen that the effects of small sample sizes on Type III error is more pronounced. Leventhal and Huynh (1996) reported that the Type III error rate always less than 0,5α. Therefore, the difference between the tests is always less than 0,5α when two group means was compared. Similarly, Sarkar et al. (2002) stated that the chance of Type III error is less than that of Type I error (α). Results of this study are consistent with the reporting Leventhal and Huynh (1996), Sarkar et al. (2002), and the findings reported by MacDonald (1999) and Mendes (2004). However, nothing has been reported for the comparison of more than two group means. Neverlethess, Type III error rate might be found more than 0.5α under some experimental conditions. The reason for that might be the distribution shape, number of groups, variance ratio, and the relationship between the sample size and group variances (direct and inverse pairing).

When samples were drawn from three t (5) distributions, Type III error was higher for F test than that for W, BF, and AG test (Table 3). And, this was more obvious in small sample sizes and effect size (0.25). The Type III error rate was affected by total sample sizes rather than inequality in sample sizes. Under the same experimental conditions, when δ was 1.0, it was seen that the Type III error rate for F, W, BF, and AG test was found to be 0.94%, 0.65%, 0.56%, and 0.20% respectively, even if sample sizes were 5. Under this distribution, AG test is still superior to the others.

It can be seen that table 3, table 4, and table 6 gave similar results. Therefore, it can be said that the effects of t (5), χ^2 (3), and exponential (1.25) distributions on Type III error rates for all tests were similar. At he same time, the effect of table 2 and table 5 on Type III error were similar too.

When samples were drawn from three Weibull (1.5, 1) distributions, across all sample sizes, the estimates of Type III error rates for W test ranged from 1.02% to 1.43% for δ=0.25, ranged from 0.10% to 1.06% for δ=0.75, and ranged from 0.00% to 0.78% for δ=1.00. The estimates of Type III error rates for BF test ranged from 0.86% to 1.48% for δ=0.25, ranged from 0.10% to 0.88% for δ=0.75, and ranged from 0.00% to 0.61% for δ=1.00. Under the same conditions, the estimates of Type III error rates for AG test ranged from 0.24% to 0.83%, from 0.08% to 0.50%, and from 0.00% to 0.32%, respectively, whereas the Type III error estimates for F test ranged from 1.36% to 2.14%, from 0.10% to 1.40%, and from 0.00% to 1.02% respectively (Table 7). It can be seen that the results of Table 7 were similar to the results of the Table 2 and Table 5.

When Table 2-7 evaluated together, the superiority of the AG test can be seen for all distributions and sample sizes. Because, across the all distributions, sample sizes and population mean differences (δ), the AG test obtained higher estimates for power, lower estimates of Type III error (γ). Therefore, revisited version of test power of the AG test, Power=1- β- γ, will be higher than the others. Power of F test is smaller than the alternatives in general. Because, Type III error rates for F test were higher than W, BF, and AG test in general. On the other hand, simulation results suggested that Type III error rates for tests were not affected from distribution shape.

4.Implication

The results of the present simulation of the Type III error rates of the ANOVA F and its three commonly recommended parametric alternatives indicate that the AG test provides a considerable advantage over the F, W, and BF test in all experimental conditions. Because, in almost every experimental situation, the Type III error rates were lower for the AG test and the power of the AG test was higher than the others in many cases. On the other hand, simulation results indicated that the deviation from normality was not affect the Type III error even if distributions were exponential (1.25).

Table 2. Type III error (%) for different statistics when data are simulated from three Normal (μ_i, σ_i^2) distributions based on 100,000 simulations; α=0.05

	$\mu_1:\mu_2:\mu_3$=**0:0:0.25**				$\mu_1:\mu_2:\mu_3$=**0:0:0.75**				$\mu_1:\mu_2:\mu_3$=**0:0:1.0**			
	F	W	BF	AG	F	W	BF	AG	F	W	BF	AG
5:5:5	2.20	1.51	1.33	0.48	1.12	0.80	0.70	0.22	0.72	0.48	0.43	0.17
10:10:10	1.88	1.52	1.48	0.64	0.66	0.51	0.49	0.26	0.31	0.27	0.26	0.12
20:20:20	1.51	1.19	1.19	0.60	0.27	0.23	0.23	0.12	0.04	0.03	0.03	0.01
30:30:30	1.31	1.00	1.00	0.58	0.11	0.09	0.09	0.05	0.00	0.00	0.00	0.00
3:4:5	2.15	1.12	0.89	0.13	1.35	0.72	0.54	0.11	0.72	0.35	0.29	0.07
5:10:15	1.73	1.07	0.92	0.26	0.59	0.37	0.31	0.13	0.32	0.26	0.22	0.06
10:20:30	1.42	1.12	1.02	0.36	0.28	0.23	0.20	0.11	0.12	0.10	0.08	0.04

Table 3. Type III error (%) for different statistics when data are simulated from three t (5) distributions based on 100,000 simulations; α=0.05

	$\mu_1:\mu_2:\mu_3$=**0:0:0.25**				$\mu_1:\mu_2:\mu_3$=**0:0:0.75**				$\mu_1:\mu_2:\mu_3$=**0:0:1.0**			
	F	W	BF	AG	F	W	BF	AG	F	W	BF	AG
5:5:5	1.96	1.19	1.04	0.31	1.12	0.62	0.55	0.17	0.94	0.65	0.56	0.20
10:10:10	1.79	1.41	1.36	0.55	0.95	0.74	0.70	0.33	0.60	0.49	0.47	0.24
20:20:20	1.61	1.25	1.23	0.50	0.55	0.48	0.48	0.25	0.17	0.14	0.14	0.10
30:30:30	1.47	1.20	1.10	0.45	0.35	0.31	0.30	0.17	0.04	0.02	0.01	0.00
3:4:5	2.13	1.03	0.81	0.19	1.26	0.55	0.48	0.08	0.83	0.45	0.31	0.09
5:10:15	1.69	0.93	0.85	0.23	0.93	0.56	0.49	0.17	0.45	0.31	0.24	0.04
10:20:30	1.56	1.07	0.98	0.34	0.55	0.42	0.40	0.14	0.21	0.17	0.14	0.04

Table 4. Type III error (%) for different statistics when data are simulated from three Chi (3) distributions based on 100,000 simulations; α=0.05

	$\mu_1:\mu_2:\mu_3$=**0:0:0.25**				$\mu_1:\mu_2:\mu_3$=**0:0:0.75**				$\mu_1:\mu_2:\mu_3$=**0:0:1.0**			
	F	W	BF	AG	F	W	BF	AG	F	W	BF	AG
5:5:5	1.88	1.14	0.90	0.26	1.22	0.98	0.71	0.31	0.88	0.74	0.54	0.30
10:10:10	1.66	1.23	1.09	0.62	0.80	0.74	0.64	0.49	0.51	0.49	0.45	0.35
20:20:20	1.70	1.40	1.33	0.87	0.41	0.40	0.38	0.35	0.11	0.11	0.10	0.05
30:30:30	1.60	1.37	1.33	0.95	0.15	0.13	0.13	0.09	0.00	0.00	0.00	0.00
3:4:5	1.72	0.88	0.65	0.08	1.26	0.69	0.58	0.15	0.77	0.49	0.33	0.15
5:10:15	1.39	0.85	0.77	0.31	0.66	0.49	0.49	0.22	0.47	0.31	0.20	0.08
10:20:30	1.40	1.03	1.01	0.47	0.37	0.35	0.35	0.22	0.08	0.06	0.06	0.02

Table 5. Type III error (%) for different statistics when data are simulated from three Beta (10, 10) distributions based on 100,000 simulations; α=0.05

	$\mu_1:\mu_2:\mu_3$=0:0:0.25				$\mu_1:\mu_2:\mu_3$=0:0:0.75				$\mu_1:\mu_2:\mu_3$=0:0:1.0			
	F	W	BF	AG	F	W	BF	AG	F	W	BF	AG
5:5:5	2.22	1.55	1.38	0.50	1.10	0.71	0.62	0.24	0.84	0.60	0.53	0.21
10:10:10	1.79	1.40	1.36	0.51	0.72	0.57	0.56	0.26	0.30	0.24	0.24	0.11
20:20:20	1.40	1.16	1.15	0.51	0.34	0.32	0.32	0.21	0.03	0.03	0.03	0.01
30:30:30	1.36	1.20	1.20	0.58	0.08	0.07	0.07	0.04	0.00	0.00	0.00	0.00
3:4:5	2.37	1.30	1.06	0.26	1.28	0.79	0.64	0.16	0.73	0.40	0.33	0.10
5:10:15	1.75	1.15	1.01	0.31	0.66	0.44	0.34	0.09	0.34	0.22	0.20	0.03
10:20:30	1.48	1.18	1.10	0.48	0.31	0.24	0.22	0.10	0.08	0.07	0.06	0.04

Table 6. Type III error (%) for different statistics when data are simulated from three Exponential (1.25) distributions based on 100,000 simulations; α=0.05

	$\mu_1:\mu_2:\mu_3$=0:0:0.25				$\mu_1:\mu_2:\mu_3$=0:0:0.75				$\mu_1:\mu_2:\mu_3$=0:0:1.0			
	F	W	BF	AG	F	W	BF	AG	F	W	BF	AG
5:5:5	1.71	1.07	0.76	0.18	1.21	0.97	0.61	0.29	0.94	0.83	0.47	0.28
10:10:10	1.69	1.19	0.99	0.59	0.88	0.83	0.64	0.51	0.43	0.43	0.36	0.30
20:20:20	1.53	1.22	1.14	0.80	0.46	0.44	0.42	0.36	0.10	0.10	0.09	0.09
30:30:30	1.50	1.33	1.31	0.95	0.19	0.19	0.19	0.18	0.02	0.01	0.02	0.00
3:4:5	1.57	0.66	0.54	0.10	0.94	0.60	0.42	0.12	0.81	0.63	0.47	0.17
5:10:15	1.30	0.75	0.70	0.13	0.71	0.57	0.54	0.28	0.40	0.37	0.36	0.28
10:20:30	1.20	0.81	0.79	0.32	0.38	0.37	0.37	0.25	0.12	0.10	0.10	0.06

Table 7. Type III error (%) for different statistics when data are simulated from three Weibull (1.5, 1) distributions based on 100,000 simulations; α=0.05

	$\mu_1:\mu_2:\mu_3$=0:0:0.25				$\mu_1:\mu_2:\mu_3$=0:0:0.75				$\mu_1:\mu_2:\mu_3$=0:0:1.0			
	F	W	BF	AG	F	W	BF	AG	F	W	BF	AG
5:5:5	2.14	1.43	1.15	0.40	1.40	1.06	0.88	0.28	1.02	0.78	0.61	0.29
10:10:10	1.79	1.34	1.26	0.64	0.91	0.81	0.74	0.50	0.51	0.46	0.44	0.32
20:20:20	1.75	1.51	1.48	0.83	0.30	0.29	0.28	0.20	0.03	0.02	0.02	0.00
30:30:30	1.58	1.40	1.39	0.81	0.10	0.10	0.10	0.08	0.00	0.00	0.00	0.00
3:4:5	1.94	1.02	0.86	0.24	1.35	0.75	0.62	0.13	1.04	0.67	0.55	0.14
5:10:15	1.66	1.06	0.96	0.27	0.81	0.55	0.54	0.24	0.37	0.31	0.30	0.13
10:20:30	1.38	1.04	0.99	0.39	0.44	0.40	0.40	0.20	0.09	0.08	0.08	0.03

5.References

1. Alexander, R. A., and Govern, D.M. **A new and simpler approximation for ANOVA under variance heterogeneity,** Journal of Educational Statistics, 19, 1994, pp. 91-101
2. Anonymous **FORTRAN Subroutines for mathematical applications. IMSL MATH/LIBRARY**. Vol.1-2, Houston: Visual Numerics, Inc.1994
3. Box, G. E. P. **Some theorems on quadratic forms applied in the study of analysis variance problems, I. effect of inequatiley of variance in the one-way model,** The Annals of Mathematical Statistics, 25, 1954, pp. 290-302
4. Brown, M. B., and Forsythe, A. B. **The small sample behavior of some statistics which test the equality of several means,** Technometrics, 16, 1974, pp.129-132

5. Cohen, J. **Statistical power analysis for the behavioral sciences, Second Ed.,** New Jersey : Lawrence Erlbaum Associates, Hillsdale, 1988
6. Camdeviren, H., and Mendes, M. **A Simulation study for Type III error rates of some variance homogeneity tests,** Jounal of Pakistan Statistics, 21(2), 2005, pp. 223-234
7. Ferron, J., and Sentovich, C. **Statistical power of randomization tests used with multiple-baseline designs,** Journal of Experimental Education, 70(2), 2002, pp.165-178
8. Finner, H. **Stepwise multiple test prodecures and control of directional errors**, Ann. statist., 27, 1999, pp. 274-289
9. Jones, L. V., and Tukey, J. W. **A sensible formulation of the significance test,** Psychological Methods, 5, 2000, pp. 411-414
10. Keselman, H. J., Wilcox, R. R., Othman, A. R., and Fradette, K. **Trimming, transforming statistics, and bootstrapping: Circumventing the biasing effects of heteroscedasticity and non-normality,** Journal of Modern Applied Statistical Methods, 1(2), 2002, pp. 288-309
11. Leventhal, L., and Huynh, C. L. **Directional decisions for two-tailed tests: Power, error rates and sample size,** Psychological Methods, 1996, pp.278-292
12. Leventhal, L. **Answering two criticisms of hypothesis testing**, Psychological Reports, 85, 1999, pp. 3-18
13. Mehrota, D. V. **Improving the Brown-Forsythe solution to generalized Behrens-Fisher problem,** Communication in Statistics, Series B 26, 1997, pp. 1139-1145
14. Mendes, M. **The comparison of some parametric alternative test to one-way Analysis of Variance in terms of Type I error rates and power of test under non-normality and heterogeneity of variance,** PhD. Thesis, Ankara University Graduates School of Natural and Applied Sciences Department of Animal Science (unpublished), 2002
15. Mendes, M. **Comparison of ANOVA-F and K-test in terms of Type III error rates,** Journal of Agriculture Sciences, 10(2), 2004, pp. 121-126
16. MacDonald, P. **Power, Type I, and Type III error rates of parametric and nonparametric statistical tests,** The Journal of Experimental Education, 67, 1999, pp. 367-379
17. Micceri, T. **The unicorn, the normal curve, and other improbable creatures,** Psychological Bulletin, 105 (1), 1989, pp. 156-166
18. Muller, K. E., and Lavange, L.M. **Power calculations for general linear multivariate models including repeated measures applications,** Journal of the American Statistical Association, 87, 1992, pp. 1209-1216
19. Penfield, D. A. **Choosing a two-sample location test,** The Journal of Experimental Education, 62 (4), 1994, pp. 343-360
20. Sansgiry, P., and Akman, O. **Transformations of the lognormal distribution as a selection model,** The American Statistician,. 54 (4), 2000, pp. 307-309
21. Sarkar, S. K., Sen, P. K., and Finner, H. **Improwing two results in multiple testing,** Temple University, University of North Carolina at chapel Hill and University of Duesseldorf, 2002
22. Sawilowsky, S. S., and Blair, R. C. **A more realistic look at the robustness and type II error properties of the t test to departures from population normality,** Psychological Bulletin III, 1992, pp.352-360
23. Schanieder, P. J., and Penfield, D. A. **Alexander and Govern's approximations: providing an alternative to ANOVA under variance heterogeneity,** The Journal of Experimental Education, 65 (3), 1997, pp. 271-286
24. Sharon, S., and Carpenter, K. M. **The rights answer for the wrong question: consequences of type III error for public health research,** American Journal of Public Health, 89(8), 1999, pp. 1175-1181
25. Shaffer, J. P. **Directional statistical hypothesis and comparisons among means,** Psychological Bulletin, 77, 1972, pp.195-197

26. Welch, B. L. **On the comparison of several mean values: An alternative approach,** Biometrika, 38, 1951, pp. 330-336
27. Wilcox, R. R. **A one-way random effects model for trimmed means,** Psychometrika, 59, 1994, pp. 289-306
28. Zar, J. H. **Biostatistical analysis,** New Jersey: Prentice –Hall Inc. Simon and Schuster/A Viacom Company, 1999

[1] Brief Description of the Fields of Statistical Expertise: Experimental Design, Statistical Data Analysis, Simulation Studies, Applied Regression analysis, Computer programming, Covariance analysis, Nonparametric Statistical Analysis, Statistical Package Programs, Multivariate Analysis Techniques, Analysis of Longitudinal data, Growth Curves.