Y. Y. Ü. Fen Bilimleri Enstitüsü Dergisi
Yıl : 2005, Cilt : 10, Sayı : 1, Sayfa : 16 -22

Confidence Interval for Test Power in Welch, James-Second Order and Alexander-Govern Tests: A Simulation Study

Mehmet MENDEŞ[1] Ensar BAŞPINAR[2] Fikret GÜRBÜZ[2]
[1]*Çanakkale Onsekiz Mart Üniversitesi, Ziraat Fakültesi, Zootekni Bölümü, Çanakkale*
[2] *Ankara Üniversitesi, Ziraat Fakültesi, Zootekni Bölümü, Ankara*

Abstract: A simulation study was conducted to construct confidence interval of test power of three commonly recommended analysis of variance (ANOVA) F test alternatives (Welch test, James second-order test, and Alexander and Govern test) for testing mean differences under non-normality and variance heterogeneity. At the end of 50,000 simulation trials, it was seen that as variances heterogeneous, the test power was decreased and confidence intervals became larger. However, the confidence intervals were narrower as the sample size and effect size (δ) were increased. This case was obvious especially when sample sizes were 50 and more and effect size was 1. When the sample size and effect size were small and medium (δ=0.50 δ=0.75), the constructed of the confidence intervals were much more reliable and informative than given of only the lower or upper bounds of the test power. On the other hand, when the sample sizes and effect sizes were large, there will be no serious problem to give either the lower or upper bound of the test power

Key Words: Power of test, confidence interval, analysis of variance, homogeneity of variance

Welch, James-Second Order ve Alexander-Govern Testlerinde Testin Gücü İçin Güven Aralığı: Simülasyon Çalışması

Özet: Bu çalışmada, normallik ve varyansların homojenliği ön şartlarının sağlanmadığı durumlarda varyans analizi tekniğine alternatif olarak kullanılan üç testin (Welch test, James second-order test, and Alexander and Govern test) güç değerleri için güven aralıkları oluşturulmuştur. Yapılan 50,000 simülasyon denemesi sonucunda varyansların homojenliği ön şartının yerine gelmemesi durumunda testin gücünün düştüğü ve güven aralığının daha genişlediği görülmüştür. Diğer taraftan, örnek hacmi ve ortalamalar arası standardize edilmiş farkın (δ) artması durumunda güven aralığı daralmaktadır. Bu durum, örnek hacmi 50 ve ortalamalar arası fark 1 standart sapma iken daha belirgindir. Örnek hacminin küçük, ortalamalar arası farkın ise küçük veya orta düzeyde (δ=0.50 δ=0.75) olması durumunda testin gücü için güven aralığının oluşturulması, sadece alt ya da üst sınırının verilmesinden çok daha bilgilendirici olmaktadır. Diğer taraftan, örnek hacmi ve ortalamalar arası farkın büyük olduğu durumlarda testin gücü için sadece alt ya da üst sınırının verilmesinin pek bir sakınca teşkil etmediği görülmüştür.

Anahtar Kelimeler: Testin gücü, güven aralığı, varyans analizi, varyansların homojenliği

Introduction

Consider k independent groups with $\mu_1,...,\mu_k$ and variances $\sigma_1^2,...,\sigma_k^2$. One of the more common goals in applied research is testing $H_0 : \mu_1 = ... = \mu_k$ (1) (Wilcox, 1997; Zar, 1999; Mendeş, 2002). In testing hypothesis (1) different tests (e.g. ANOVA F test, Welch test, Brown-Forsythe test, James second-order test, Alexander-Govern test, Trimmed mean, etc.) are used depend on normality and homogeneity of variance assumptions. There are several simulation studies for comparing ANOVA F and its some parametric alternatives which are used in testing hypothesis (1) with respect to test power (Welch, 1951; James, 1951; Yuen, 1974; Brown and Forsythe, 1974; Levy, 1978; Tabatabia and Tan, 1986; Wilcox et al., 1986; Wilcox, 1988; Wilcox, 1994; Oshima and Algina, 1992a; Oshima and Algina, 1992b; Oshima et al., 1994; Alexander and Govern, 1994; Hsuing and Olejnik, 1996; Wilcox, 1997; Schneider and Penfield, 1997; Keselman et al., 1998; Wei-ming, 1999; Keselman et al., 2002; Mendeş, 2002; Mendeş and Başpınar, 2003; Mendeş and Pala, 2004). But, many of those studies were never considered lower and upper bound of test power. However, test power changes depending on sample size, variance ratio, Type I error rate (α), relationship between sample size and variance ratio, effect size (standardized mean difference, δ), distribution shape, relationship between effect size and variance ratio. Different test powers are obtained even when the studies are done in the same experimental conditions. Let's assume there are three groups with normally distributed to be compared with variance ratios $\sigma_1^2 : \sigma_2^2 : \sigma_3^2$=1:4:8 and sample sizes of $n_1=n_2=n_3=15$. Depending on the differences among the population means, three different test power values may be obtained. a) If the relationship among the means is $\mu_1 : \mu_2 : \mu_3 = 0{:}\ 0{:}\ 1$ the test powers obtained by the Welch, James second-order, and Alexander-Govern tests at the end of 50,000 simulation trials were 18.18 %, 17.73 %, and 17.65 %, respectively. b) If the relationship among the means is $\mu_1 : \mu_2 : \mu_3 = 1{:}\ 0{:}\ 0$ the test powers obtained by the Welch, James second-order, and Alexander-Govern tests were 37.66 %, 36.83 %, and 38.07 % respectively. c) If the relationship among the means is $\mu_1 : \mu_2 : \mu_3 = 0{:}\ 1{:}\ 0$ the test powers obtained by the Welch, James second-order, and Alexander-Govern tests were 28.42 %, 27.70 %, and 28.26 % respectively.

The test power obtained in condition (a) is the lower bound, while those obtained in condition (b) is the upper bound for the test power. This was because; a fixed number as a standard deviation form was added to the group with the highest variance for condition (a) and a fixed number as a standard deviation form was added to the group with the lowest variance for condition (b). On the other hand, the test powers obtained in condition (c) was in the middle in terms of test power. Those cases are valid for unbalanced sample sizes combinations. Therefore, constructing confidence intervals for test powers would be more informative and provides extra information about the test power. Ferron and Sentovich (2002) estimated statistical power for three randomization tests using multiplie-baseline designs. They stated that they used > 80 % as the sufficient power level for comparing the tests. Since, there is a relationship between Type I error and Type II error can be defined as β=4α. If Type I error rate (α) is assumed to be 0.05, power levels equal to or bigger than 80 % is sufficient.

The major goal of this study is forming the confidence interval for three commonly recommended parametric alternatives to ANOVA (Welch, James-Second Order and Alexander-Govern tests) when homogeneity of variance and normality assumptions is not met. For this aim, random samples from normal (μ, σ^2), Chi-square with 3 degrees of freedom, and Beta (3, 4) distributions were generated by Monte Carlo simulation technique. These particular types of non-normal distributions were selected since educational, medical, and psychological research data typically have skewed distributions, and those distributions are predominantly used in literature to study deviations from normality (Yuen, 1974; Rogan and Keselman, 1977; Levy, 1978;. Tiku and Balakrishnan, 1984; Tabatabia and Tan, 1986; Boos and Brownie, 1989; Sharma, 1991; Sawilowsky and Blair, 1992; Oshima and Algina, 1992a; Alexander and Govern, 1994; Wilcox, 1994; Wludyka and Nelson, 1999; Keselman, et al. 1998; Mendeş, 2002; Keselman et al., 2002; Mendeş and Başpınar, 2003; Mendeş and Pala, 2004).

1.1. Definition of Statistical Tests

1.1.1. Welch Test

The test statistic for Welch test is

$$W = \frac{\sum_k W_k (\overline{X}_k - X'_{..})^2 / (K-1)}{\left[1 + \frac{2}{3}(K-2)\Lambda\right]} \tag{2}$$

Where, $W_k = \frac{n_k}{S_k^2}$, and $X'_{..} = \frac{\sum_k W_k \overline{X}_{.k}}{\sum_k W_k}$ and

$$\Lambda = \frac{3\sum_k (1 - W_k / \sum_k W_k)^2 / (n_k - 1)}{(K^2 - 1)} \tag{3}$$

W statistic is approximately distributed as a central F variable with (K-1) and 1/ Λ degrees of freedom (Welch, 1951; Lix et al., 1996; Mendeş, 2002).

1.1.3. James Second-Order Test

The test statistic for James second order test is

$$J = \sum_{k=1}^{K} W_k (\overline{X}_{.k} - \overline{X}_{..})^2 \tag{4}$$

When H_0 is true, J is distributed asymptotically Chi-square with K-1 degrees of freedom. James second-order test is applied as follows. Let c be the 1-α quantile of a chi-square distribution with K-1 degrees of freedom, and let

$v_k = (n_k - 1)$, $R_{st} = \sum_{k=1}^{K} \frac{1}{(n_k - 1)^s} \left(\frac{W_k}{\sum_k W_k} \right)^t$, and

$\chi_{2s} = \frac{c^5}{\left[(K-1)(K+1)\ldots(K+2s-3)\right]}$. The critical value of J statistic is;

$$h(\alpha) = c + 1/2(3\chi_4 + \chi_2) \sum_{j=1}^{j} \left(1 - \frac{W_j}{W}\right)^2 \left[\sum_{j=1}^{j} \left(1 - \frac{W_j}{W}\right) / V_j \right]^2$$

$$+ [1/16(3\chi_4 + \chi_2)^2 (1 - (j-3)/c) \left[\sum_{j=1}^{j} \left(1 - \frac{W_j}{W}\right)^2 / V_j \right]^2$$

$$+ 1/2(3\chi_4 + \chi_2)((8R_{23} - 10R_{22} + 4R_{21} - 6R_{12}^2$$
$$+ 8R_{12}R_{11} - 4R_{11}^2) + (2R_{23} - 4R_{22} + 2R_{21}$$
$$- 2R_{12}^2 + 4R_{12}R_{11} - 2R_{11}^2)(\chi_2 - 1)$$
$$+ 1/4(-R_{12}^2 + 4R_{12}R_{11} - 2R_{12}R_{10} - 4R_{11}^2 + 4R_{11}R_{10}$$
$$- R_{10}^2)(3\chi_4 - 2\chi_2 - 1)) + (R_{23} - 3R_{22} + 3R_{21} - R_{20})$$
$$(5\chi_6 + 2\chi_4 + \chi_2)$$
$$+ 3/16(R_{12}^2 - 4R_{23} + 6R_{22} - 4R_{21} + R_{20})$$
$$(35\chi_8 + 15\chi_6 + 9\chi_4 + 5\chi_2)$$
$$+ 1/16(-2R_{22} + 4R_{21} - R_{20} + 2R_{12}R_{10} - 4R_{11}R_{10} + R_{10}^2)$$
$$(9\chi_8 - 3\chi_6 - 5\chi_4 - \chi_2) + 1/4(-R_{22} + R_{11}^2)$$
$$(27\chi_8 + 3\chi_6 + \chi_4 + \chi_2)$$
$$+ 1/4(R_{23} - R_{12}R_{11})(45\chi_8 + 9\chi_6 + 7\chi_4 + 3\chi_2)]$$

Where, $W = \sum W_k$.

The null hypothesis is rejected if J > h (α) (James, 1951).

1.1.4. Alexander-Govern Test

The test statistic for this test is $AG = \sum_{k=1}^{K} Z_k^2$ (5)

Where

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

$$a = v_k - 0.5\,, \qquad b = 48a^2\,, \qquad c = \sqrt{a.\ln\left(1 + \frac{t_k^2}{v_k}\right)}\,,$$

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

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

To compare Welch (W), James-second order (J) and Alexander-Govern (AG) test, we generated Monte Carlo studies by computer simulation for three different distributions using Fortran Power Station Developer IMSL (Anonymous, 1994). The distributions were normal (μ, σ^2), Chi-square distribution with 3 degrees of freedom, and Beta (3, 4) distributions. For each given set of parameter values, samples from these distributions were generated using the subroutines RNNOA, RNCHI, and RNBET available with the IMSL library functions. The parameter values were taken for k=3 group. In this study we considered both equal sample sizes (5, 15, 30, 50, 75) and unequal sample sizes ((5, 10, 15) and (10, 15, 25)). Variance ratios were $\sigma_1^2 : \sigma_2^2 : \sigma_3^2$=1:2:3 and 1:4:8. We computed Welch test (W) value and counted the frequency satisfying W > F (k-1, 1/ Λ) degrees of freedom for α=0.05. For James second-order test we computed J and counted the frequency satisfying J > h (α) for α=0.05, and for Alexander-Govern test we computed AG statistic and counted the frequency satisfying AG > χ^2 (K-1) degrees of freedom for α=0.05. For each test, we checked to see if the hypothesis, which is false, was rejected at α=0.05. The experiment was repeated 50,000 times and the proportion of observations falling in the critical regions was recorded for different δ, n, variance pattern and distributions. This proportion estimation is the test power if the means from the populations do differ.

Populations mentioned have been standardized as they have different means and variances. Thus, not having changed the shape of distributions handled it was provided that their means were 0 and their standard deviations were 1. To form heterogeneity among population variances, standardized random numbers in the samples were multiplied by specific constant numbers ($\sigma = 1, \sqrt{2}, \sqrt{3}, \sqrt{4}, \sqrt{8}$). To create a difference between the population means, specific constant numbers in standard deviation form (δ=0.50, 0.75, 1.0) were added to the random numbers of the populations. The effect sizes (standardized mean difference) of 0.8 and more standard deviation approximate those suggested by Cohen (1969, 1988) to represent large effect sizes. In this study, we used 1.0 standard deviation to represent large effect size. 80 % was assumed to be the sufficient power level in this study.

Lower and upper bounds of test power was estimated as follows:

For k=3 and variance ratios of $\sigma_1^2 : \sigma_2^2 : \sigma_3^2 = 1:2:3$ (1:4:8), the upper bound of test power was estimated as $\mu_1 : \mu_2 : \mu_3 = (\mu_1 + \delta) : 0 : 0$ and the lower bound of test power was estimated as $\mu_1 : \mu_2 : \mu_3 = 0 : 0 : (\mu_3 + \delta)$.

We wrote a FORTRAN 90 program for the Intel Pentium III processor to compute all the tests.

Table 1. The characteristics of the distributions

Distributions	Mean	Variance	Skewness	Kurtosis
Normal (0,1)	0.00	1.00	0.00	0.00
χ^2 (3)	3.00	6.00	1.63	4.00
β (3,4)	0.43	0.03	0.18	-0.55

Normal (0,1): Normal distribution, χ^2 (3): chi-square distribution with 3 d.f., β (3, 4): Beta distribution (3, 4).

3. Results and Discussion

Given in Table 2-7 are the lower and upper bounds of test power for Welch test (W), James second-order test (J), and Alexander-Govern (AG) test under different sample sizes, variance patterns, distributions and population mean difference.

When population distributions were normal and variance ratio was $\sigma_1^2 : \sigma_2^2 : \sigma_3^2$=1:2:3, the confidence intervals of the all tests were similar (Table 2). As sample sizes and effect sizes were increased, the confidence interval was narrowed. This case was obvious especially when sample sizes were 50 and more and effect size was 1. Thus, when n ≥50 and δ=1 the lower and upper bound of these test power were getting closer each other. On the other hand, when δ=0.75 (medium effect size) and sample sizes were between 5 and 30, the confidence interval constructed were larger than when δ=0.5.

Under same experimental conditions when variance ratio was $\sigma_1^2 : \sigma_2^2 : \sigma_3^2$=1:4:8, both test power was decreased and confidence interval was larger with respect to all tests (Table 3). This condition was valid even when δ=1 and sample sizes of 75. For example, when variance ratio was 1: 2: 3, δ=1 and n=75, the lower and upper bound of the test power of (W), (J), and (AG) test were 98.54-99.98 %, 98.54-99.98 % and 98.51-99.98 %, respectively. However, under same conditions when variance ratio was 1: 4: 8, the lower and upper bound of test power of those tests were 73.11-98.50 %, 73.10-98.50 % and 72.89-98.51 %, respectively.

It can be said that the lower and upper test power of the all tests were affected by Chi-square (3) distribution (Table 4 and Table 5). However, confidence interval with Chi-square (3) distribution has similarities to those obtained with normal distribution. This similarity was pronounced especially when sample size and effect size increased. When variance ratio was 1: 4: 8, the difference between lower and upper bound of test power of these tests was larger. That is, when variances were getting more heterogeneous, the confidence intervals were larger. When the populations were Beta (3, 4), the confidence interval constructed was very similar to results from normal populations (Table 6 and Table 7). This case was valid both variance ratios (1: 2: 3 and 1: 4: 8).

When Table 2-7 were evaluated altogether, we reached the following results:

a) As variances heterogeneous the test power was decreased and confidence intervals were larger. However, the confidence intervals were narrower as the sample size and effect size were increased. Thus, when the sample size and effect size were small and medium (δ=0.50 δ=0.75), the constructed confidence intervals were much more reliable and informative than given of only the lower or upper bounds of the test power. On the other hand, when the sample sizes and effect sizes were large, that was not cause serious problem to give either the lower or upper bound of the test power. Because, these two bounds were close each others.

b) Test power and shape of the confidence interval were changed depending on distribution shape.

c) It was seen that the three alternative tests showed very similar test powers in many instances. Powers of these tests were very similar. In contrast to ANOVA F test, the other tests were affected adversely by the distribution. While none of these tests are superior in all situations, Welch and Alexander-Govern test should be preferred over the others (ANOVA results were not given).

As expected, when heterogeneity of variances of the underlying populations increased, the power of test for all tests increased with increases in sample size and population mean differences (δ) regardless of the population distribution. The probability of test power decreased as heterogeneity of variances increased. The effect of heterogeneity and non-normality on test power obviously increased as sample size and δ decreased. The results are consistent with those of Welch, Brown-Forsythe, Wilcox and Alexander-Govern's. The findings also are consistent with those of Oshima et al. (1994), Mehrota (1997), Mendeş (2002), Mendeş and Başpınar (2003).

Simulation Results

Table 2. The lower and upper bounds of test power for Welch, James-second order and Alexander-Govern tests when distributions were normal and variance ratios 1:2:3

Variance ratio: 1:2:3		Welch		James 2nd-order		Alexander-Govern	
δ	n_1:n_2:n_3	L	U	L	U	L	U
0.50	5:5:5	6.68	7.65	5.87	6.67	6.48	7.54
	15:15:15	12.43	17.95	12.22	17.68	12.05	17.91
	30:30:30	22.24	33.92	22.18	33.84	21.87	33.98
	50:50:50	34.67	54.98	34.65	54.94	34.41	55.11
	75:75:75	50.04	74.02	50.03	74.01	49.78	74.06
	5:10:15	9.99	10.88	9.10	10.05	13.06	14.05
	10:15:25	15.27	16.70	15.01	16.44	17.81	19.08
0.75	5:5:5	8.90	11.25	7.83	9.81	8.40	11.15
	15:15:15	23.09	35.96	22.76	35.55	22.46	36.11
	30:30:30	44.64	67.37	44.55	67.25	44.03	67.50
	50:50:50	68.35	89.23	68.33	89.22	68.00	89.29
	75:75:75	86.13	98.05	86.11	98.04	85.95	98.05
	5:10:15	16.88	18.58	15.51	17.33	21.67	23.71
	10:15:25	30.20	33.00	29.83	32.60	34.17	36.81
1.0	5:5:5	12.36	17.06	11.07	15.05	11.69	17.03
	15:15:15	38.07	59.00	37.68	58.60	36.98	59.32
	30:30:30	70.09	90.69	70.00	90.62	69.48	90.75
	50:50:50	91.18	99.21	91.16	99.21	91.02	99.22
	75:75:75	98.54	99.98	98.54	99.98	98.51	99.98
	5:10:15	28.40	30.05	26.66	28.49	34.50	36.73
	10:15:25	51.04	54.24	50.53	53.82	55.43	58.64

Table 3. The lower and upper bounds of test power for Welch, James-second order and Alexander-Govern tests when distributions were normal and variance ratios 1:4:8.

Variance ratio: 1:4:8		Welch		James 2nd-order		Alexander-Govern	
δ	n_1:n_2:n_3	L	U	L	U	L	U
0.50	5:5:5	6.04	6.50	5.05	5.40	5.79	6.28
	15:15:15	7.95	12.32	7.65	11.90	7.73	12.31
	30:30:30	11.71	21.77	11.64	21.58	11.55	21.91
	50:50:50	16.62	34.42	16.57	34.35	16.50	34.60
	75:75:75	23.27	50.17	23.24	50.14	23.16	50.31
	5:10:15	7.14	8.95	6.40	8.00	8.57	10.51
	10:15:25	8.41	10.46	7.57	9.51	12.49	16.04
0.75	5:5:5	6.82	8.96	5.73	7.50	6.52	8.85
	15:15:15	12.05	22.78	11.60	22.21	11.65	22.97
	30:30:30	20.95	44.90	20.79	44.69	20.60	45.16
	50:50:50	32.95	68.13	32.88	68.05	32.63	68.34
	75:75:75	47.49	86.20	47.46	86.19	47.22	86.32
	5:10:15	9.91	14.47	8.93	13.09	11.73	16.49
	10:15:25	15.62	23.31	15.22	22.88	17.19	25.12
1.0	5:5:5	8.21	12.42	7.07	10.50	7.84	12.48
	15:15:15	18.18	37.66	17.73	36.83	17.65	38.07
	30:30:30	33.91	69.92	33.68	69.72	33.39	70.29
	50:50:50	54.51	91.36	54.44	91.33	54.13	91.48
	75:75:75	73.11	98.50	73.10	98.50	72.89	98.51
	5:10:15	14.98	21.64	13.59	20.00	17.40	25.33
	10:15:25	24.89	38.32	24.40	37.75	26.78	40.94

Table 4. The lower and upper bounds of test power for Welch, James-second order and Alexander-Govern tests when distributions were χ^2 (3) and variance ratios 1:2:3.

Variance ratio: 1:2:3		Welch		James 2^{nd}-order		Alexander-Govern	
δ	$n_1:n_2:n_3$	L	U	L	U	L	U
0.50	5:5:5	4.94	12.76	4.01	11.04	5.31	12.77
	15:15:15	10.99	25.79	10.66	25.51	10.76	25.65
	30:30:30	20.91	40.43	20.78	40.35	20.52	40.46
	50:50:50	34.84	58.73	34.78	58.70	34.45	58.81
	75:75:75	52.08	75.87	52.07	75.87	51.81	75.92
	5:10:15	12.24	11.27	10.63	10.02	15.63	15.19
	10:15:25	17.84	19.50	17.48	19.15	20.14	22.25
0.75	5:5:5	6.88	19.89	5.58	17.71	7.20	19.49
	15:15:15	22.22	45.98	21.75	45.61	21.54	45.99
	30:30:30	46.47	72.08	46.34	71.98	45.73	72.22
	50:50:50	73.14	90.78	73.10	90.76	72.78	90.89
	75:75:75	90.02	97.97	90.02	97.97	89.93	98.00
	5:10:15	22.25	20.28	20.14	18.61	26.58	25.71
	10:15:25	36.10	37.77	35.56	37.16	39.76	41.58
1.0	5:5:5	10.48	29.36	8.57	26.74	10.48	28.53
	15:15:15	40.65	67.65	40.00	67.26	39.25	67.81
	30:30:30	76.47	91.83	76.38	91.79	75.89	91.94
	50:50:50	95.33	99.14	95.32	99.14	95.23	99.16
	75:75:75	99.54	99.95	99.54	99.95	99.53	99.96
	5:10:15	35.83	34.67	33.49	32.49	41.72	41.90
	10:15:25	59.02	62.02	58.57	61.47	63.53	66.03

Table 6. The lower and upper bounds of test power for Welch, James-second order and Alexander-Govern tests when distributions were Beta (3,4) and variance ratios 1:2:3.

Variance ratio: 1:2:3		Welch		James 2^{nd}-order		Alexander-Govern	
δ	$n_1:n_2:n_3$	L	U	L	U	L	U
0.50	5:5:5	7.35	7.59	6.43	6.63	7.05	7.45
	15:15:15	12.65	17.02	12.48	16.69	12.35	16.90
	30:30:30	21.64	33.38	21.57	33.31	21.29	33.39
	50:50:50	34.89	54.04	34.87	54.00	34.59	54.13
	75:75:75	49.84	73.23	49.83	73.21	49.56	73.30
	5:10:15	10.00	11.53	9.18	10.70	13.00	14.69
	10:15:25	15.42	16.79	15.07	16.48	18.07	19.39
0.75	5:5:5	9.93	10.98	8.94	9.50	9.55	10.97
	15:15:15	22.83	34.61	22.53	34.21	22.07	34.74
	30:30:30	44.15	66.47	44.08	66.36	43.65	66.64
	50:50:50	67.42	88.91	67.39	88.91	67.04	88.96
	75:75:75	85.55	98.01	85.55	98.01	85.38	98.02
	5:10:15	16.66	19.15	15.48	18.06	21.09	24.14
	10:15:25	30.31	31.86	27.11	31.52	34.14	35.75
1.0	5:5:5	12.68	16.12	11.41	14.20	12.09	16.23
	15:15:15	37.46	57.60	37.08	57.11	36.44	57.90
	30:30:30	69.40	90.52	69.28	90.46	68.74	90.60
	50:50:50	90.55	99.27	90.52	99.27	90.35	99.27
	75:75:75	98.37	99.98	98.36	99.98	98.34	99.98
	5:10:15	27.20	28.45	25.67	27.11	33.32	35.27
	10:15:25	48.82	52.82	48.45	52.44	53.39	57.21

Table 5. The lower and upper bounds of test power for Welch, James-second order and Alexander-Govern tests when distributions were χ^2 (3) and variance ratios 1:4:8.

Variance ratio: 1:4:8		Welch		James 2^{nd}-order		Alexander-Govern	
δ	$n_1:n_2:n_3$	L	U	L	U	L	U
0.50	5:5:5	5.98	14.98	5.00	13.35	6.21	14.75
	15:15:15	6.67	22.13	6.43	21.78	6.71	22.12
	30:30:30	9.21	30.45	9.10	30.31	9.10	30.54
	50:50:50	13.74	41.42	13.71	41.36	13.56	41.51
	75:75:75	21.26	54.76	21.24	54.75	21.10	54.90
	5:10:15	6.39	12.75	5.02	11.77	7.59	15.57
	10:15:25	8.74	18.27	8.35	17.94	9.40	19.88
0.75	5:5:5	6.07	20.21	5.03	18.24	6.27	20.00
	15:15:15	9.20	34.67	8.85	34.17	9.11	34.78
	30:30:30	17.57	52.49	17.41	52.35	17.28	52.66
	50:50:50	31.06	71.24	30.98	71.22	30.73	71.33
	75:75:75	47.73	86.03	47.71	86.02	47.45	86.13
	5:10:15	9.49	20.10	7.62	18.63	11.09	23.40
	10:15:25	15.26	30.75	14.70	30.25	16.35	32.93
1.0	5:5:5	6.46	26.60	5.25	24.23	6.75	26.32
	15:15:15	14.81	50.45	14.28	49.76	14.41	50.65
	30:30:30	32.17	73.63	31.86	73.46	31.45	73.86
	50:50:50	55.95	90.67	55.87	90.62	55.60	90.75
	75:75:75	77.67	97.94	77.65	97.93	77.42	97.96
	5:10:15	14.26	29.79	11.79	27.93	16.44	34.05
	10:15:25	25.83	48.25	25.15	47.69	27.65	50.65

Table 7. The lower and upper bounds of test power for Welch, James-second order and Alexander-Govern tests when distributions were Beta (3,4) and variance ratios 1:4:8.

Variance ratio: 1:4:8		Welch		James 2^{nd}-order		Alexander-Govern	
δ	$n_1:n_2:n_3$	L	U	L	U	L	U
0.50	5:5:5	6.93	7.00	6.04	6.02	6.65	6.87
	15:15:15	8.23	11.58	7.97	11.16	8.04	11.58
	30:30:30	12.08	20.59	12.00	20.48	11.94	20.68
	50:50:50	16.02	33.66	16.82	33.58	15.62	33.81
	75:75:75	23.15	50.13	23.14	50.11	23.03	50.27
	5:10:15	7.16	8.66	6.48	7.63	8.54	10.11
	10:15:25	9.53	12.13	9.35	11.74	10.60	13.25
0.75	5:5:5	7.92	8.49	6.83	7.18	7.60	8.40
	15:15:15	12.32	21.13	11.99	20.61	12.04	21.41
	30:30:30	21.30	43.29	21.14	43.04	20.94	43.57
	50:50:50	33.04	67.81	32.98	67.74	32.80	68.04
	75:75:75	47.70	86.02	47.68	86.00	47.50	86.12
	5:10:15	10.51	13.76	9.49	12.40	12.29	15.86
	10:15:25	15.69	22.34	15.29	21.88	17.35	23.92
1.0	5:5:5	9.21	11.16	7.93	9.43	8.71	11.23
	15:15:15	18.42	35.59	17.95	34.82	17.82	36.20
	30:30:30	34.21	69.50	33.99	69.27	33.72	69.78
	50:50:50	53.80	91.16	53.76	91.13	53.35	91.28
	75:75:75	72.76	98.55	72.73	98.55	72.53	98.56
	5:10:15	14.94	21.55	13.71	19.84	17.30	24.68
	10:15:25	25.09	36.76	24.69	36.19	27.22	39.36

References

Alexander, R.A., and Govern, D.M. (1994). A new and simple approximation for ANOVA under variance heterogeneity. Journal of education Statistics, 19, 91-101.

Anonymous, (1994). FORTRAN Subroutines for mathematical applications. IMSL MATH/LIBRARY. Vol.1-2, Houston: Visual Numerics, Inc.

Boos, D.D., and Brownie, C. (1989). Bootstrap methods for testing homogeneity of variances. Technometrics, 31(1), 69-82.

Brown, M. B., and Forsythe, A. B. (1974). Robust tests for the equality of variances. Journal of the American Statistical Association, 69, 364–367.

Cohen, J. (1969). Statistical power analysis for behavioral science. New York: Academic Press.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. Second Ed. New Jersey: Lawrence Erlbaum Associates, Hillsdale.

Fenstad, G.U. (1983). A comparison between U and V tests in the behrens-fisher problem. *Biometrika*, 70, 300-302.

Ferron J., and Sentovich, C. (2002). Statistical power of randomization tests used with multiple- baseline designs. *Journal of Experimental Education,* 70 (2), 165-178.

Hsiung, T.C., and Olejnik, S. (1996). Type I error rates and statistical power for James second-order test and the univariate F test in two-way ANOVA models under heteroscedasticity and/or non-normality. *The Journal of Experimental Education*, 65, 57-71.

James, G.S. (1951). The comparison of several groups of observations when the ratios of the population variances are unknown. Biometrika, 38, 324-329.

Keselman, H.J., Kowalchuk, R.K., and Lix, L.M. (1998). Robust nonorthogonal analyses revisited: An update based on trimmed means. *Psychometrika*, 63, 145-163.

Keselman, H.J., Wilcox, R.R., Othman, A.R., and Fradette, K. (2002). Trimming, transforming statistics, and bootstrapping: Circumventing the biasing effects of heteroscedasticity and non-normality. *Journal of Modern Applied Statistical Methods*, 1(2), 288-309.

Levy, K.J. (1978). Some empirical power results associated with Welch's robust analysis of variance technique. Journal of Statistical Computation and Simulation, 8, 43-48.

Lix, L.M., Keselman, J.C., and Keselman, H.J., (1996). Consequences of assumption violations revisited: A quantitative review of alternatives to the one-way analysis of variance F test. *Review of Educational Research*, 66, 579-619.

Mehrota, D.V. (1997). Improving the Brown-Forsythe solution to generalized Behrens-Fisher problem, *Communication in Statistics*, Series B, 26, 1139-1145.

Mendeş, M. (2002). 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. Ph.D. Thesis. Ankara University Graduates School of Natural and Applied Sciences Department of Animal Science (unpublished).

Mendeş, M., and Baspınar, E. (2003). Comparison of some tests statistics realized Type I error rate in the non-normal populations. *Journal of Agriculture Sciences*, 9 (1), 23-28.

Mendeş, M., and Pala, A. (2004). Evaluation of four tests when normality and homogeneity of variance assumptions are violated. *Pakistan Journal of Information and Technology*, 4 (1), 38-42.

Oshima, T.C.,and Algina, J. (1992a). Type I error rates for James's second-order test and Wilcox's H_m test under heteroscedasticity and non-normality. *British Journal of Mathematical and Statistical Psychology*, 45, 255-263.

Oshima,T.C., and Algina, J.(1992b). A SAS program for testing the hypothesis of equal means under heterocedasticity: James's second-order test. *Educational and Psychological Measurement*, 52, 117-118.

Oshima, T.C., Algina, R.A., and Lin, W.Y. (1994). Type I error rates for Welch's test and James's second-order test under non-normality and inequality of variance when there are two groups. *Journal of Educational and Behavioral Statistics*, 19, 275-291.

Rogan, J.C., and Keselman, H.J. (1977). Is the ANOVA F-test robust to variance heterogeneity when sample sizes are equal? An investigation via a coefficient of variation. *American Educational Research Journal,* 14, 493-498.

Sawilowsky, S.S., and Blair, R.C. (1992). A more realistic look at the robustness and type II error properties of the t test to departures from population normality. *Psychological Bulletin*, III, 352-360.

Schneider, P.J., and Penfield, D.A. (1997). Alexander and Govern's approximation: Providing an alternative to ANOVA under variance heterogeneity. *The Journal of Experimental Education*, 65, 271-286.

Sharma, S.C.(1991). A new jackknife test for homogeneity of variances. *Communications in Statistics-Simulation and Computation*, 20 (2-3), 479-495.

Tabatabia, M.A., and Tan, W.Y. (1986). Some Monte Carlo studies on the comparison of several means under heteroscedasticity and robustness with respect to departure from normality. *Journal of Biometry*, 7, 801-814.

Tiku, M.L., and Balakrishnan, N. (1984). Testing equality of population variances the robust way. *Communications in Statistics-Theory and Methods*, 13 (17), 2143-2159.

Tomarken, A. J.,and Serlin, R. C. (1986). Comparison of ANOVA alternatives under variance heterogeneity and specific noncentrality structures. *Psychological Bulletin*, *99*, 90-99.

Weerahandi, S. (1995). ANOVA under unequal error variances. *Biometrics*, 51, 589-599.

Welch, B.L.(1951). On the comparison of several mean values: an alternative approach. Biometrika, 38, 330-336.

Wilcox, R.R., Charlin, V.L., and Thompson, K.L. (1986). New Monte Carlo results on the robustness of the ANOVA F, W and F^* statistics. *Journal of Statistical Computation and Simulation*, 15, 33-943.

Wilcox, R.R. (1988). A new alternative to the ANOVA F and new results on James's second-order method. *British Journal of Mathematical and Statistical Psychology*, 41, 109-117.

Wilcox, R.R. (1994). A one-way random effects model for trimmed means. *Psychometrika*, 59, 289-306.

Wilcox, R.R (1997). A bootstrap modification of the Alexander-Govern ANOVA F method, plus comments on comparing trimmed means. Educational and Psychological Measurment, 57, 655-665.

Wludyka, P., and Nelson, P. R. (1999). Two non-parametric analysis-of-means type tests for homogeneity of variances. *Journal of Applied Statistics*, 26, 243–256.

Yuen, K.K. (1974). The two-sample trimmed t for unequal population variances. *Biometrika*, 61, 165-170.

Zar, J.H. (1999). Biostatistical analysis. New Jersey: Prentice – Hall Inc. Simon and Schuster/A Viacom Company.