Kafkas Univ Vet Fak Derg
16 (5): 711-716, 2010
DOI:10.9775/kvfd.2009.1507

RESEARCH ARTICLE

Comparison of ANOVA F and WELCH Tests with Their Respective Permutation Versions in Terms of Type I Error Rates and Test Power

Mehmet MENDEŞ * Erkut AKKARTAL **

* Çanakkale Onsekiz Mart Universitesi Ziraat Fakültesi Zootekni Bölümü Biyometri&Genetik Anabilim Dalı, TR-17100 Çanakkale - TÜRKİYE

** Yeditepe Universitesi Uluslararası Lojistik ve Taşımacılık Bölümü, TR-34755 İstanbul - TÜRKİYE

Makale Kodu (Article Code): KVFD-2009-1507

Summary

We compared Analysis of Variance (F) and the Welch test (W) with their respective permutation versions (PF and PW) in terms of Type I error rate (α) and test power (1-β) by Monte Carlo simulation technique. Simulation results showed that when the variances were homogeneous, the permutation versions of F and W tests displayed more reliable results in terms of protecting Type I error rate at nominal level, regardless of distribution shape and sample size. Violation of homogeneity of variances adversely affected all tests. Regardless of sample size and effect size, the PF test was slightly more powerful compared to the F test as long as the variances were homogeneous and the distributions were skewed (χ^2 (3) and Exp [0.75]). The PF and F tests had similar power levels when the distributions were symmetrical (Beta (5.5)). The W test was more powerful with homogenous variances, while the PW test was slightly superior with heterogonous variances except for unbalanced sample sizes (i.e., 5:10:15).

Keywords: *Analysis of variance, Permutation tests, Type I error, Test power, Welch test*

ANOVA F ve Welch Testi ile Bunların Permutasyon Versiyonlarının 1. Tip Hata ve Testin Gücü Bakımından Karşılaştırılması

Özet

Bu çalışmada Varyans analizi tekniği (F) ve Welch testi ile bunların permutasyon versiyonları (PF ve PW) 1.Tip hata ve testin gücü bakımından karşılaştırılmıştır. Söz konusu karşılaştırmalar Monte Carlo simulasyon tekniği kullanılmıştır. Yapılan simülasyon çalışmaları sonucunda varyanslar homojen iken bu testlerin permutasyon versiyonlarının 1. Tip hata olasılığını koruma bakımından daha güvenilir sonuçlar verdikleri görülmüştür. Diğer taraftan varyansların heterojenleşmesinden bütün testlerin olumsuz yönde etkilendikleri görülmüştür. Varyansların heterojen ve dağılımların da çarpık (χ^2 (3) ve Exp [0.75]), olması halinde örnek hacmi ve etki büyüklüğü ne olursa olsun PF testinin F testine göre biraz daha güçlü olduğu görülmüştür. Ancak dağılımlar simetrik iken (β (5.5)) PF ve F testlerinin güç değerleri benzerdir. W testi varyansların homojen olması halinde daha güçlü iken, PW testi varyansların homojen olmadığı ve örnek hacimlerinin dengesiz olduğu (mesela 5:10:15) durumda biraz daha güçlüdür.

Anahtar sözcükler: *Varyans analizi, Permutasyon testleri, 1. Tip hata, Testin gücü, Welch testi*

INTRODUCTION

Three solutions are generally recommended for the situations in which assumptions of analysis of variance (normality and homogeneity of variance) are not fulfilled. These are: a) trying to meet these assumptions by subjecting the data to an appropriate transformation b) using the non-parametric counterparts, Kruskal-Wallis test, of ANOVA, and c) using some parametric alternatives such as Welch and Brown-Forsythe tests [1-2]. Another solution would be to apply resampling methods like Permutation tests [3-9], which are considered as non-

İletişim (Correspondence)

☎ +90 286 2180018/1348

mmendes@comu.edu.tr

parametric [10]. These tests are not always as effective as their parametric counterparts [11,12]. However, when sample sizes are small or normality assumption is not satisfied, compared to the parametric counterparts, they can generally give more reliable results in terms of Type I error and power of test [9,13-15]. Furthermore, Bracken [16] and Tanizaki [17] informed that the permutation tests can also be applied successfully in the cases where the homogeneity of variances is not satisfied. Accordingly, Bohdan [18] reported that the permutation tests decrease the Type I error rate and increase the power of test in cases where data is not normal or error variances can not be taken as equal [19]. The main purpose of this study is to compare the performance (Type I error and power of test) of ANOVA F and Welch tests with their respective permutation versions.

MATERIAL and METHODS

Random numbers generated with simulation techniques provided the material for this study. The IMSL library of Microsoft FORTRAN Developer Studio was used to generate random numbers [20]. Random samples of equal (n1:n2:n3=5:5:5, 10:10:10 and 15:15:15) and unequal (n1:n2:n3=3:4:5 and 5:10:15) sizes were generated from three populations with $\chi^2(3)$, Exp (0.75) and β(5.5) by using the RNCHI, RNEXP and RNBET functions of the IMSL library. Situations where the homogeneity of variances was met ($\sigma_1^2:\sigma_2^2:\sigma_3^2=1:1:1$), and not met ($\sigma_1^2:\sigma_2^2:\sigma_3^2=1:1:4$ and $\sigma_1^2:\sigma_2^2:\sigma_3^2=1:1:9$), were both taken into account. In this study, the permutation and simulation numbers were determined as 20.000 and 50.000 respectively. Type I error rates for F and W tests were obtained by dividing the number of mistakenly rejected H_0 hypotheses after 50.000 trials by the total trial number. Type I error rates in terms of permutation versions of these tests (PF and PW) were calculated by finding cases where values were equal to or bigger than F and W of the PF and PW test statistics. To calculate the test power, differences in standard deviation form between the population means were created (effect size or standardized mean differences; Δ=0.5, 1.0 and 1.5), and then the number of H_0 hypothesis rejected was divided by the total trial number.

ANOVA- F Test (F-test)

It is well known that the test statistic for one-way fixed effect ANOVA F test is

$$F=\frac{SSB/(k-1)}{SSW/(N-k)}=\frac{MSB}{MSW}$$

Where SSB is sum of squares between treatments, SSW is the sum of squares within treatments, (k-1) and (N-k) are the degrees of freedom, between and within treatments respectively [21].

WELCH Test (W-test)

Test statistics for Welch test can be expressed as below:

$$F_W=\frac{\sum_{i=1}^{k}W_i(\bar{X}_{i.}-X'_{..})/(k-1)}{\left[1+\frac{2}{3}(k-2)\Lambda\right]}$$

$$\text{Where: } W_i=\frac{n_i}{S_i^2},\quad X'_{..}=\frac{\sum_{i=1}^{k}W_i\bar{X}_i}{\sum_{i=1}^{k}W_i}\quad \text{and}\quad \Lambda=\frac{3\sum(1-W_i/\sum_{i=1}^{k}W_i)^2/(n_i-1)}{(i^2-1)}$$

The Fw statistic has F-distribution with (k-1) and (1/Λ) degrees of freedom [22]. The computation of the test statistics for permutation versions of these tests (PF and PW) are the same as F and W test statistics. However, PF and PW test statistics are computed based on permutation samples.

RESULTS

Type I error rate estimates for F and W tests and their permutation versions (PF and PW) were given in *Table 1*. When variances were homogeneous ($\sigma_1^2:\sigma_2^2:\sigma_3^2\quad 1:1:1$), regardless of the distribution shapes and sample size, the PF and PW tests gave more reliable results compared to F and W tests. Type I error rates for PF and PW tests were quite close to each other. But the Type I error rates for all tests deviated from nominal alpha level (5.0%) when variances were not homogeneous ($\sigma_1^2:\sigma_2^2:\sigma_3^2\quad 1:1:4$ and $\sigma_1^2:\sigma_2^2:\sigma_3^2\quad 1:1:9$). This case was more obvious especially when the variance ratios were 1:1:9, distribution was $\chi^2(3)$ and sample sizes were low.

Test power estimates for F, W, PF and PW depending upon distribution shapes, sample sizes, variance ratios and effect sizes were given in *Tables 2, 3* and *4*, respectively. When variances were homogeneous and effect size or mean difference was Δ = 0.5, the test power values for all tests were at low levels *(Table 2).* Under these experimental conditions, the test power values for the PF test were higher than those of the F test. The W and PW tests displayed similar power values, except that numbers of observations in the groups were clearly unbalanced (5:10:15).

Table 1. *Type I error rates after 50.000 simulation trials (%)*
Tablo 1. *50.000 Simülasyon denemesi sonucundaki 1. Tip hata olasılıkları (%)*

Distributions	n	$\sigma_1^2:\sigma_2^2:\sigma_3^2=1:1:1$				$\sigma_1^2:\sigma_2^2:\sigma_3^2=1:1:4$				$\sigma_1^2:\sigma_2^2:\sigma_3^2=1:1:9$			
		F	**PF**	**W**	**PW**	**F**	**PF**	**W**	**PW**	**F**	**PF**	**W**	**PW**
	5:5:5	4.0	4.7	4.4	4.7	7.3	8.1	6.6	7.8	10.9	11.8	8.3	10.4
	10:10:10	4.6	5.2	5.3	5.0	7.1	7.7	6.2	6.4	9.9	10.5	7.2	7.9
$\chi^2(3)$	15:15:15	4.6	5.0	5.5	5.1	6.6	7.1	6.0	5.9	8.6	9.1	6.4	6.8
	3:4:5	4.4	5.1	4.8	4.9	5.8	6.1	6.1	7.4	7.8	8.2	7.2	9.6
	5:10:15	4.5	4.9	6.9	4.9	3.4	3.5	6.0	4.7	3.6	3.7	6.4	5.7
	5:5:5	5.7	5.4	6.3	5.4	7.0	7.3	6.5	6.6	9.1	9.8	6.1	7.8
	10:10:10	5.4	5.3	5.6	5.5	6.5	6.7	5.6	5.4	8.1	8.5	5.7	6.1
β (5.5)	15:15:15	4.9	5.0	5.4	5.1	6.7	6.9	6.0	5.9	7.5	7.7	5.3	7.6
	3:4:5	5.2	5.1	6.5	5.1	4.8	4.9	6.4	6.3	5.4	5.9	6.6	7.5
	5:10:15	4.8	5.2	6.6	5.3	2.3	2.3	5.9	4.8	2.2	2.3	6.0	5.1
	5:5:5	3.9	4.9	3.9	5.0	5.7	7.1	5.3	7.0	7.8	9.3	6.1	8.5
	10:10:10	4.5	5.4	5.3	5.3	5.5	6.6	5.4	5.7	7.2	8.6	5.3	6.1
Exp(0.75)	15:15:15	4.7	5.2	5.6	5.2	6.0	6.9	5.0	4.8	8.3	9.7	5.3	5.5
	3:4:5	4.1	4.8	4.4	4.8	4.6	4.8	4.6	5.9	5.0	5.6	5.2	7.3
	5:10:15	4.3	4.8	6.9	4.9	2.2	2.4	6.9	4.4	2.0	2.2	6.4	5.2

F: Anova F test , **PF:** Permutation version of Anova F test, **W:** Welch test, **PW:** Permutation version of Welch test

Table 2. *Test powers when samples are taken from three -*$\chi^2(3)$*- distributions (%)*
Tablo 2. *Örnekler* $\chi^2(3)$ *dağılımlarından alındıklarında testin güç değerleri (%)*

Effect Size	n	$\sigma_1^2:\sigma_2^2:\sigma_3^2=1:1:1$				$\sigma_1^2:\sigma_2^2:\sigma_3^2=1:1:4$				$\sigma_1^2:\sigma_2^2:\sigma_3^2=1:1:9$			
		F	**PF**	**W**	**PW**	**F**	**PF**	**W**	**PW**	**F**	**PF**	**W**	**PW**
	5:5:5	10.9	12.2	12.0	12.8	6.6	8.2	5.6	6.9	8.1	9.6	5.6	8.1
	10:10:10	19.2	20.4	21.1	20.2	8.8	10.0	6.6	6.8	8.0	9.3	5.6	6.5
Δ=0.50	15:15:15	27.6	28.6	29.3	29.0	12.3	13.5	9.0	8.7	9.1	10.2	6.0	6.4
	3:4:5	9.6	10.3	12.2	12.2	3.6	4.2	5.0	5.8	4.7	5.2	5.3	7.3
	5:10:15	20.9	21.9	28.3	24.5	3.5	4.0	10.3	7.1	2.1	2.3	6.7	5.2
	5:5:5	33.0	34.8	35.2	35.0	12.9	16.0	8.5	10.5	9.3	11.8	5.8	8.5
	10:10:10	60.4	61.5	66.9	66.2	28.0	31.4	19.1	19.4	14.6	17.6	8.6	9.7
Δ=1.00	15:15:15	79.5	80.0	84.6	84.2	43.8	46.5	31.7	31.0	22.5	25.0	12.6	12.7
	3:4:5	27.9	28.9	30.6	30.2	7.3	8.2	8.3	9.6	4.5	5.4	5.1	7.5
	5:10:15	67.4	68.1	70.3	69.0	19.3	20.8	29.7	21.9	13.3	15.1	21.6	16.2
	5:5:5	62.6	63.0	65.8	64.7	28.8	34.1	18.4	22.1	13.6	18.0	7.5	11.4
	10:10:10	91.1	91.5	95.1	95.0	61.9	66.0	48.3	49.4	33.0	37.9	19.0	20.4
Δ=1.50	15:15:15	98.4	98.5	99.6	99.5	83.6	85.1	73.3	72.9	51.3	55.3	32.3	32.1
	3:4:5	56.1	56.5	56.9	53.9	17.6	20.1	16.0	18.2	7.0	8.3	7.1	9.8
	5:10:15	93.9	94.0	93.5	93.7	57.9	60.6	64.8	56.2	19.4	21.1	29.4	20.8

Under the same conditions, when variances were not homogeneous, the test power values in terms of all tests seem to have decreased to really low levels. This situation became clearer when it was studied with small sample sizes and effect size. It was observed that test power values in terms of all tests decreased to low levels when variance ratios increased to 1:1:9 under the same conditions. Test power values under these conditions were quite far from the power value of 80.0%, which was widely accepted as sufficient.

When distribution was β(5.5), variances were homogeneous and Δ = 0.5, F and PF showed similar power values, while the W test was generally slightly more powerful than the PW test (*Table 3*). Test power estimates under these experimental conditions, however, were lower than the test power estimated under $\chi^2(3)$. This was more obvious especially for W and PW tests. When homogeneity of variances assumption was not met, the test power values for all tests were at quite low levels and it became clearer when variance ratios were increased to $\sigma_1^2:\sigma_2^2:\sigma_3^2=1:1:9$.

Table 3. *Test powers when samples are taken from three β (5.5) distributions (%)*
Tablo 3. *Örnekler β (5.5) dağılımlarından alındıklarında testin güç değerleri (%)*

Effect Size	n	$\sigma_1^2:\sigma_2^2:\sigma_3^2=1:1:1$				$\sigma_1^2:\sigma_2^2:\sigma_3^2=1:1:4$				$\sigma_1^2:\sigma_2^2:\sigma_3^2=1:1:9$			
		F	PF	W	PW	F	PF	W	PW	F	PF	W	PW
Δ=0.50	5:5:5	10.0	9.8	10.7	9.6	9.8	10.2	8.2	8.4	10.1	11.2	7.9	9.5
	10:10:10	17.2	16.9	17.3	16.1	12.1	12.3	9.5	9.2	10.9	11.6	7.7	8.1
	15:15:15	25.0	24.9	24.9	23.9	16.1	16.4	11.7	11.5	12.9	13.3	8.6	8.7
	3:4:5	9.4	9.0	9.6	8.1	6.0	6.2	7.1	6.8	6.2	6.7	6.8	8.0
	5:10:15	19.6	19.6	18.4	15.8	7.1	7.2	10.7	8.6	3.7	3.7	7.9	6.8
Δ=1.00	5:5:5	27.6	27.2	25.7	23.6	17.9	18.8	13.5	14.0	14.4	15.8	9.3	11.8
	10:10:10	57.5	57.2	56.0	53.9	31.1	31.5	22.2	21.8	19.7	20.4	12.8	12.9
	15:15:15	78.9	78.9	77.7	76.8	43.8	44.2	33.0	32.5	28.1	28.8	18.4	18.5
	3:4:5	22.5	22.1	19.9	18.0	11.7	12.0	11.0	11.6	8.8	8.6	10.6	10.9
	5:10:15	62.9	63.0	58.0	54.1	23.3	23.5	26.7	22.8	11.2	11.6	15.9	13.5
Δ=1.50	5:5:5	55.4	54.8	50.9	48.2	31.7	33.3	21.7	23.0	22.2	24.1	13.6	16.7
	10:10:10	92.4	92.3	91.1	90.5	58.0	58.7	43.8	43.1	35.8	37.0	23.3	23.7
	15:15:15	99.0	99.0	98.8	98.7	76.9	77.2	63.5	62.9	49.3	49.9	33.6	33.4
	3:4:5	46.2	45.5	38.8	35.9	21.5	22.1	17.9	18.3	14.0	15.0	12.2	14.1
	5:10:15	94.9	94.9	92.3	90.8	53.2	53.4	54.5	48.5	25.7	26.1	30.2	25.4

Table 4. *Test powers when samples are taken from three Exp (0.75) distributions (%)*
Tablo 4. *Örnekler Exp (0.75) dağılımlarından alındıklarında testin güç değerleri (%)*

Effect Size	n	$\sigma_1^2:\sigma_2^2:\sigma_3^2=1:1:1$				$\sigma_1^2:\sigma_2^2:\sigma_3^2=1:1:4$				$\sigma_1^2:\sigma_2^2:\sigma_3^2=1:1:9$			
		F	PF	W	PW	F	PF	W	PW	F	PF	W	PW
Δ=0.50	5:5:5	8.2	9.7	9.2	10.8	6.7	8.7	4.6	6.6	7.5	10.1	5.1	8.1
	10:10:10	13.2	14.6	14.6	14.5	11.0	13.2	8.1	8.4	11.2	13.7	6.8	7.8
	15:15:15	17.6	18.6	19.6	18.5	16.8	19.1	11.7	11.4	16.9	19.7	9.7	9.8
	3:4:5	6.6	7.4	9.2	10.2	3.4	4.1	4.4	6.1	3.9	4.8	4.8	6.9
	5:10:15	13.3	14.2	22.8	18.8	4.9	5.7	14.0	9.2	3.7	4.1	11.0	7.8
Δ=1.00	5:5:5	20.4	22.6	22.8	24.6	12.5	16.4	8.1	11.0	10.0	13.7	5.4	9.3
	10:10:10	39.3	41.3	45.1	44.6	28.6	33.5	19.4	20.0	21.8	26.6	11.9	13.6
	15:15:15	56.0	47.5	62.3	61.2	43.3	46.6	32.4	31.6	34.9	39.3	20.3	20.3
	3:4:5	18.6	20.0	23.1	23.4	6.3	7.7	7.3	9.4	4.9	6.2	5.4	8.3
	5:10:15	44.5	45.9	52.8	50.5	19.8	21.9	31.7	24.0	10.6	12.1	20.7	14.4
Δ=1.50	5:5:5	42.9	45.1	48.4	49.0	24.2	31.4	15.5	20.8	15.4	21.0	8.2	13.2
	10:10:10	72.3	73.6	79.9	79.6	55.8	60.8	43.3	44.8	38.8	45.7	22.6	24.4
	15:15:15	88.2	88.9	93.2	93.0	77.0	79.4	65.8	65.5	61.1	66.0	39.7	40.4
	3:4:5	37.2	38.5	41.4	41.1	14.7	17.7	14.3	17.0	7.3	9.2	7.1	10.6
	5:10:15	77.2	77.9	80.9	80.2	48.6	51.7	58.2	49.8	25.7	28.6	37.1	27.0

When distribution was Exp (0.75) and variances were homogeneous, the PF test was slightly more powerful than the F test *(Table 4).* The W test was also slightly more powerful than the PW test, except for very small sample sizes under small and moderate effect size (Δ = 0.5 and Δ = 1.0). For large effect size (Δ = 1.50), on the other hand, both tests produced similar power values. When variances were heterogeneous, the PF test was more powerful than the F test in general. When W and PW tests were compared, it was seen that the W test was more powerful under some conditions (for example when studied with 5:10:15 sample size combination), while, the PW test was powerful under some other conditions (for example when studied with 5:5:5, 10:10:10 and 3:4:5 sample size combination). In some conditions (for example when studied with 15:15:15 sample size combination) W and PW had similar power values.

DISCUSSION

There are different statistical techniques which can be used to test the same hypothesis. However, none of these tests give reliable results in every trial condition. Therefore, the use of appropriate statistical methods for trial conditions according to the structure of the data set studied is very important in terms of the reliability of the results. In the determining the appropriate statistical methods, the two important criteria are keeping the Type I error rate at nominal level and having a high test power. As well known, analysis of variance is the most widely used statistical technique used when comparing group means in practice [29]. However, it is negatively affected by deviations from normality and homogeneity of variances. This effect becomes clearer when studied with small sample sizes [1,2,23]. In these situations, different solutions are applied. One of these solutions is the use of permutation versions of the mentioned tests.

It has been noted that the permutation versions (PF and PW) of ANOVA F and W test generally give more reliable results in terms of protection of Type I error rate when variances are homogeneous, regardless of the distribution shape and sample sizes. On the other hand, as the number of observations increased, the Type I error rates in terms of F and W tests gradually resembles PF and PW tests. Routledge [24], Ludbrook and Dudley [25], Corcoran and Mehta [26], Peres-Neto and Olden [5], Tussell [14], Maggini et al.[15], Balasubramani et al.[27] and Koşkan [9] have indicated that permutation tests give more reliable results than the variance analysis in terms of the protection of Type I error at the nominal level, especially when studied with small sample size, but both tests produce similar Type I error rates as parallel to the increase of sample size. Type I error rate in terms of all tests gradually deviated from 5.0% when variances were not homogeneous. This deviation is more obvious in the permutation versions of these tests. This situation may be accepted as an indicator that permutation tests are negatively affected by the heterogeneity of variances. Different findings have been reported in the studies aimed at determining whether permutation tests or variance analysis or Welch test give more reliable results in terms of protection of Type I error rate at the nominal level, when variances are not homogeneous. For example, while Bracken [16] and Tanizaki [17] indicated that permutation tests gave more reliable results when variances were heterogeneous, Huang et al.[28] reported that the use of permutation tests in testing H_0 hypothesis when the variances were not homogeneous might increase Type I error rate. Accordingly, Koşkan [9] indicated that the effect of non-heterogeneous variance is clearer in permutation tests. With regard to experimental conditions, when F and W tests and their permutation versions PF and PW tests are compared in terms of test power, it is seen that F test is more powerful than PF when variances are homogeneous and distribution shape is skew (χ^2(3) and Exp [0.75]), no matter what the sample size and effect size are. On the other hand, parallel to the increase of sample size and effect size, both tests have gradually produced similar test power values. When distribution shape is symmetric but not normal (Beta [5.5]), the test power values in terms of both F and PF tests are quite similar. When variances of populations from which the samples are taken are heterogeneous, the PF test is stronger than the F test. Therefore, it can be suggested that the PF test can be preferred to the F test. When W and PW tests are compared in terms of test power, it is seen that the W test is slightly more powerful when variances are homogeneous, but in the case that variances are heterogeneous, except in unbalanced sample sizes, the PW test is slightly more powerful than the W test. If PF and PW tests are compared; it is seen that, when variances are homogeneous and distribution shape is skewed, the PW test is generally more powerful than F, and especially when variances are heterogeneous and the distributions are symmetric, PF test is powerful. With a general evaluation (considering both Type I error rate and test power together), it can be suggested that the PF test can be preferred when assumptions of normality and homogeneity of variances are not met together and PW test can be preferred when homogeneity of variances is fulfilled but the normality assumption is not satisfied as long as the sample sizes are equal (n>=5). On the other hand, it should be remembered that the real effects of these tests can be further advanced by more inclusive studies, involving more details.

REFERENCES

1. Schneider PJ, Penfield DA: Alexander and Govern's approximation: Providing an alternative to ANOVA under variance heterogeneity. *J Exp Educ,* 65 (3): 271-286, 1997.

2. Mendeş M: The comparison of some parametric alternative tests to one-way analysis of variance in terms of Type I error rates and test power under non-normality and homogeneity of variance. *PhD Thesis.* Ankara University Graduate School of Natural and Applied Sciences Department of Animal Science (unpublished), 2002.

3. Mielke PW, Beny KJ: Permutation test for common locations among samples with unequal variances. *J Educ Behav Stat,* 19 (3): 217-236, 1994.

4. O'Gorman TW: An adaptive permutation test procedure for several common tests significance. *Comput Stat Data An,* 35, 335-350, 2001.

5. Peres-Neto PR, Olden JD: Assessing the robustness of randomization tests: Examples from behavioral studies. *Anim Behav,* 61, 79-86, 2001.

6. Anderson MJ, Robinson J: Permutation tests for linear models. *Aust NZ Stat,* 43 (1): 75-88, 2001.

7. Nichols TE, Holmes AP: Nonparametric permutation tests for functional neuro-imaging: A primer with examples. *Hum Brain Mapp,* 15, 1-25, 2001.

8. Önder H: Use and comparison of permutation tests in linear models. *PhD Thesis.* Çukurova University Graduate School of Natural and Applied Sciences Department of Animal Science (unpublished), 2005.

9. Koşkan O: Resampling approximation in group comparisons. *PhD Thesis.* Ankara University Graduate School of Natural and Applied Sciences Department of Animal Science (unpublished), 2007.

10. Feinstein AR: Permutation tests and statistical significance. *MD Comput: Comput Med Prac,* 10, 28-41, 1993.

11. Manly BFJ: Randomization, bootstrap and Monte Carlo methods in Biology, Second Ed., London: Chapman and Hall, 1997.

12. Opdyke JP: Fast permutation tests that maximize power under conventional Monte Carlo sampling for pairwise and multiple comparisons. *J Mod Appl Stat Meth,* 2 (1):27-49, 2003.

13. Anderson MJ, Legendre P: An empirical comparison of permutation methods for tests of partial regression coefficients in a linear model. *J Stat Comput Simul,* 62, 271-303, 1999.

14. Tussel F: A permutation test for randomness with power against smooth variation. *Stat Comput,* 11, 147-154.2001.

15. Maggini R, Guisan A, Cherix D: A stratified approach to modeling the distribution of a threatened ant species in the Swiss National Park. *Biodivers Conserv,* 11, 2117-2141, 2002.

16. Bracken MB: On stratification, minimization and protection against types 1 and 2 error. *J Clin Epidemiol,* 54, 104-107, 2001.

17. Tanizakı H: On small sample properties permutation tests: An independence test between two samples. *Int J Pure Appl Math,* 13 (2): 235-243, 2004.

18. Bohdan L: Resampling tests for regression coefficient. 2nd International Conference APLIMAT, 2003.

19. Mills D: Improved mathematical methods for drug design. http://www.cmd.port.ac.uk/immdd/doc/reports/permut.doc *Accessed:* 14.02.2005.

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

21. Winer BJ, Brown DR, Michels KM: Statistical Principles in Experimental Design. McGraw-Hill Book Company, New York, USA, 1991.

22. Welch BL: On the comparison of several mean values: an alternative approach. *Biometrika*, 38, 933-943, 1951.

23. Mendeş M, Başpınar E: Comparison of various test statistics calculated from the samples taken from nonnormal distributive populations. *J Agri Sci,* 9 (1): 23-28, 2003.

24. Routledge RD: P-values from permutation and F-tests. *Comput Stat Data Anal,* 24, 379-386, 1997.

25. Ludbrook J, Dudley H: Why permutation tests are superior to t and F tests in biomedical research. *Am Stat,* 52 (2): 127-132, 1998.

26. Corcoran CD, Mehta CR: Exact level and power of permutation, bootstrap and asymptotic tests of trend. *J Modern Appl Stat Meth,* 1, 42-51, 2002.

27. Balasubramani GK, Wisniewski SR, Zhank H, Eng HF: Development of an efficient SAS macro to perform permutation test for two independent samples. *Comput Meth Prog Bio,* 79, 179-187, 2005.

28. Huang Y, Xu H, Calian V, Hsu JC: To permute or not permute. *Bioinformatics,* 22 (18): 2244-2248, 2006.

29. Ergün, G, Aktaş, S: ANOVA modellerinde kareler toplamı yöntemlerinin karşılaştırılması. *Kafkas Univ Vet Fak Derg,* 15 (3): 481-484, 2009.