

A new random permutation test in ANOVA models

Byoung Cheol Jung¹, Myoungshic Jhun², Seuck Heun Song² *

¹ Department of Statistics, University of Seoul, Seoul 130-743, Korea;
(e-mail: bcjung@ous.ac.kr)

² Department of Statistics, Korea University, Seoul 136-701, Korea;
(e-mail: {jhun, ssong}@korea.ac.kr)

Received: April 21, 2004; revised version: January 24, 2005

Key words Analysis of variance, Random permutation tests, Simulation.

Abstract This paper developed an exact method of random permutations when testing both interaction and main effects in the two-way ANOVA model. The method of this paper can be regarded as a much improved model when compared with those of the previous studies such as Still and White (1981) and ter Braak (1992). We further conducted a simulation experiment in order to check the statistical performance of the proposed method. The proposed method works relatively well for small sample sizes compare with the existing methods.

This work was supported by Korea Science and Engineering Foundation Grant (R14-2003-002-01000)

* Corresponding Author

1 Introduction

Since Fisher (1935) proposed randomization and permutation tests, these tests have been applied to a great number of topics, including two-sample problems, linear regression and correlation problems, and experimental designs (Good, 1994; Kennedy, 1995; Edgington, 1995; Manly, 1997).

Several different methods of permutations have been proposed in the literature in order to test the significance of the effects in the ANOVA model. Welch (1990) studied a general method for constructing permutation tests for various experimental designs based on invariance and sufficiency. Still and White (1981) and Good (1994) suggested randomization methods on the residuals of the restricted model when testing for the existence of interaction effects in the models, whereas ter Braak (1992) proposed another method of randomizing the residuals of the full model. In addition, by using simulations on relatively small data sets, Gonzalez and Manly (1998) compared the efficiency of the competing methods. More recently, Anderson and ter Braak (2003) set a guideline on permutation tests for the multi-factorial ANOVA models used.

However, one statistical drawback for the approach of Still and White (1981) and ter Braak (1992) has been pointed out. That is, the observed distribution of the random permutation differs from that of the original data. This may be evidence against the intuition behind random permutation tests. As an attempt to remedy this statistical issue, we propose a new test methodology of random permutations in this paper. The proposed method differs from those of the previous studies in that we attempted to obtain uncorrelated residuals by removing the correlations present among the residuals of the restricted models. In this process, we will apply the transformation using the decomposition of the idempotent matrix to correlated residuals in order to obtain uncorrelated residuals. That is, we are basing our randomization method on uncorrelated residuals. In this study, we found that the

probability distribution of the permuted observations is the same as that of the original observations when no interaction effect is present. In this sense, we claim that the proposed permutation method is statistically exact compared with the existing methods. This approach shares the same idea with the study on multiple linear regression by Huh and Jhun (2001).

The structure of the paper is as follows. Section 2 describes the model and presents the existing tests on random permutations. In Section 3, we derive a new test methodology on random permutations. Then in Section 4, we compare the efficiency of the proposed method with those of existing studies by using simulations. In Section 5, we provide a numerical example as an application of our model, and finally, Section 6 summarizes and concludes the paper.

2 Model and Existing Random Permutation Tests

Consider the two-way ANOVA model,

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}, \quad i = 1, \dots, a, \quad j = 1, \dots, b, \quad \text{and} \quad k = 1, \dots, n, \quad (1)$$

where μ is a fixed unknown parameter, α_i is the effect of level i of the treatment factor A , β_j is the effect of level j of the treatment factor B , $(\alpha\beta)_{ij}$ is an interaction term of treatment factor A and B , and ε_{ijk} is a remainder error term which follows a *iid* F , where F is a probability distribution with the mean 0 variance σ^2 . The null and alternative hypotheses for the interactions of the model are

$$H_0 : (\alpha\beta)_{11} = \dots = (\alpha\beta)_{ab} = 0 \quad \text{vs.} \quad H_1 : \text{not } H_0. \quad (2)$$

Under the normality assumption of error terms, the traditional F test statistic is

$$F_{AB} = \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2 / (a-1)(b-1)}{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (y_{ijk} - \bar{y}_{ij.})^2 / ab(n-1)}, \quad (3)$$

where $\bar{y}_{ij.} = (\sum_{k=1}^n y_{ijk})/n$, $\bar{y}_{i..} = (\sum_{j=1}^b \sum_{k=1}^n y_{ijk})/bn$, $\bar{y}_{.j.} = (\sum_{i=1}^a \sum_{k=1}^n y_{ijk})/an$, and $\bar{y}_{...} = (\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n y_{ijk})/abn$. Under the null hypothesis H_0 in (2), F_{AB} is distributed as $F((a-1)(b-1), ab(n-1))$.

For the permutation testing of (2), there exist several permutation methods. Still and White (1981) suggested that in order to test for interaction it is appropriate to randomize the observations after adjusting them to remove the overall effects of the factors. The adjusted values can be written as $y_{ijk}^* = y_{ijk} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}$. Therefore, the F statistic can be obtained as

$$F_{AB}^* = \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (\bar{y}_{ij.}^* - \bar{y}_{i..}^* - \bar{y}_{.j.}^* + \bar{y}_{...}^*)^2 / (a-1)(b-1)}{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (y_{ijk}^* - \bar{y}_{ij.}^*)^2 / ab(n-1)}. \quad (4)$$

It is easy to see that F_{AB}^* is equal to F_{AB} which is obtained directly from the raw data. Hence, the statistical significance of H_0 can be evaluated from the Monte Carlo distribution of

$$\tilde{F}_{AB}^* = \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (\tilde{y}_{ij.}^* - \tilde{y}_{i..}^* - \tilde{y}_{.j.}^* + \tilde{y}_{...}^*)^2 / (a-1)(b-1)}{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (\tilde{y}_{ijk}^* - \tilde{y}_{ij.}^*)^2 / ab(n-1)}, \quad (5)$$

where \tilde{y}_{ijk}^* is an ijk th element of $\tilde{\mathbf{y}}^*$ which is an $abn \times 1$ randomly permuted vector of $\mathbf{y}^* = (y_{111}^*, \dots, y_{11n}^*, \dots, y_{abn}^*)$. Thus, the significance of H_0 can be assessed by p -value = $P[\tilde{F}_{AB}^* \geq F_{AB}]$. This test is denoted by RPT1. Since, RPT1 is based on the correlated residuals $y_{ijk}^* = y_{ijk} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}$, it will not be exact for finite samples. However, following Good (2002), the studentized correlations between the residuals converge to a common value as the sample size increases; thus, the residuals are asymptotically exchangeable, and RPT1 is asymptotically exact.

Ter Braak (1992) proposed a random permutation method using the usual residuals $y_{ijk}^\dagger = y_{ijk} - \bar{y}_{ij.}$ instead of $y_{ijk}^* = y_{ijk} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}$. By using the freely randomized residuals y_{ijk}^\dagger , \tilde{F}_{AB}^\dagger can be calculated from equation (3). Then, the significance of H_0 can be assessed by p -value = $P[\tilde{F}_{AB}^\dagger \geq F_{AB}]$. This test is denoted by RPT2. It also will not be exact for a finite sample since RPT2 is based on the correlated residuals $y_{ijk}^\dagger = y_{ijk} - \bar{y}_{ij.}$.

Last, we consider a random permutation test based on freely randomizing observations. Let $\tilde{\mathbf{y}}^\dagger$ be the randomly permuted vector of \mathbf{y} , then \tilde{F}_{AB}^\dagger can be calculated from equation (3) by using $\tilde{\mathbf{y}}^\dagger$. Then, the significance of H_0 can be assessed by p -value = $P[\tilde{F}_{AB}^\dagger \geq F_{AB}]$. This test is denoted by RPT3.

3 Exact Random Permutation test

The permutation tests RPT1, RPT2, and RPT3 discussed in Section 2 are not exact, since the probability distribution of the permuted observations is different from that of the original observations even when no interaction effect exists. In the same spirit as Huh and Jhun (2001), we propose an exact permutation test that is an improvement of Still and White (1981). This is based on the uncorrelated residuals obtained from the transformation of the correlated residuals $y_{ijk}^* = y_{ijk} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}$ of Still and White's (1981) method. To obtain a new random permutation test, we first consider the matrix representation of model (1). In matrix form, model (1) can be written as

$$\mathbf{y} = X_0\mu + X_a\alpha + X_b\beta + X_{ab}\alpha\beta + \varepsilon, \quad (6)$$

where $\mathbf{y} = (y_{111}, \dots, y_{11n}, \dots, y_{abn})'$ is $abn \times 1$ response vector, and $\varepsilon = (\varepsilon_{111}, \dots, \varepsilon_{11n}, \dots, \varepsilon_{abn})'$ is $abn \times 1$ error vector, $X_0 = i_a \otimes i_b \otimes i_K$, $X_a = I_a \otimes i_b \otimes i_K$, $X_b = i_a \otimes I_b \otimes i_K$, $X_{ab} = I_a \otimes I_b \otimes i_K$. i_a, i_b and i_n are vectors of ones of dimension a, b and n , and I_a and I_b are identity matrices of dimension a and b respectively; \otimes denotes the Kronecker product, and $\alpha = (\alpha_1, \dots, \alpha_a)'$, $\beta = (\beta_1, \dots, \beta_b)'$, and $\alpha\beta = ((\alpha\beta)_{11}, \dots, (\alpha\beta)_{1b}, \dots, (\alpha\beta)_{ab})'$. In model (6), using the method of Kennedy (1995), the RPT1 of Still and White (1981) can be obtained in matrix notation as follows. Define $X_{ab} = (X_a \ ; \ X_b)$ and $H_{ab} = X_{ab}(X_{ab}'X_{ab})^{-1}X_{ab}'$, then it can easily be shown that $H_{ab} = I_a \otimes \bar{J}_b \otimes \bar{J}_n + \bar{J}_a \otimes I_b \otimes \bar{J}_n - \bar{J}_a \otimes \bar{J}_b \otimes \bar{J}_n$ using the generalized inverse of the partitioned matrix (see Marsaglia and Styan, 1974), where $\bar{J}_a = i_a i_a' / a$, $\bar{J}_b = i_b i_b' / b$, and

$\bar{J}_n = i_n i_n' / n$ are the matrix of dimension a , b , and n whose elements are all $1/a$, $1/b$, and $1/n$, respectively. Then, premultiplying $I_{abn} - H_{ab}$ to both sides of (6), we get

$$\mathbf{y}^* = X_{ab}^* \alpha \beta + \varepsilon^*, \quad (7)$$

with $\mathbf{y}^* = (I_{abn} - H_{ab})\mathbf{y}$, where the ijk^{th} element of $\tilde{\mathbf{y}}^*$ is $y_{ijk}^* = y_{ijk} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}$, $X_{ab}^* = (I_{abn} - H_{ab})X_{ab}$ and $\varepsilon^* = (I_{abn} - H_{ab})\varepsilon$. The idea is to remove the part of the design matrix corresponding to the parameters of not interests. The traditional F test statistic for the interaction is derived for the transformed model (7) as

$$F_{AB}^* = \frac{\mathbf{y}^{*'} \left(X_{ab}^* (X_{ab}^{*'} X_{ab}^*)^{-1} X_{ab}^{*'} \right) \mathbf{y}^* / (a-1)(b-1)}{\mathbf{y}^{*'} \left(I_{abn} - X_{ab}^* (X_{ab}^{*'} X_{ab}^*)^{-1} X_{ab}^{*'} \right) \mathbf{y}^* / (ab(n-1))}. \quad (8)$$

It can easily be shown that the F_{AB}^* will be same as F_{AB} in (3). Thus, the permutation method using $\tilde{\mathbf{y}}^*$, a randomly permuted vector of \mathbf{y}^* , is the same method as Still and White (1981).

However, from equation (7) for the permutation method of Kennedy (1995) or Still and White (1981), we note that

$$\varepsilon^* \sim (\mathbf{0}, \sigma^2 (I_{abn} - H_{ab})).$$

Consequently, $\tilde{\varepsilon}^*$, a randomly permuted vector of ε^* , is distributed differently from ε^* . Hence, even under the null hypothesis H_0 , the distribution of $\tilde{\mathbf{y}}^*$ is not equal to that of \mathbf{y}^* . This fact may hurt the intuitional rationale behind the random permutation test. We remedy this problem in the transformation using the decomposition of the idempotent matrix $I_{abn} - H_{ab}$. Since $I_{abn} - H_{ab}$ is an idempotent matrix of rank $(abn - a - b + 1)$, there exists an $abn \times (abn - a - b + 1)$ matrix V such that

$$VV' = I_{abn} - H_{ab} \quad V'V = I_{abn-a-b+1}. \quad (9)$$

Pre-multiplying V' to both sides of (6), we obtain

$$\mathbf{y}^{**} = X_{ab}^{**} \alpha \beta + \varepsilon^{**}, \quad (10)$$

where $\mathbf{y}^{**} = V'\mathbf{y}^*$, and X_{ab}^{**} and $\boldsymbol{\varepsilon}^{**}$ are analogously defined. Since, $\boldsymbol{\varepsilon}^{**}$ is distributed as $(\mathbf{0}, \sigma^2 I_{abn-a-b+1})$, $\tilde{\boldsymbol{\varepsilon}}^{**}$, a randomly permuted vector of $\boldsymbol{\varepsilon}^{**}$, is distributed also as

$$\tilde{\boldsymbol{\varepsilon}}^{**} \sim (\mathbf{0}, \sigma^2 I_{abn-a-b+1}),$$

because $\tilde{\boldsymbol{\varepsilon}}^{**}$ can be written as

$$\tilde{\boldsymbol{\varepsilon}}^{**} = P\boldsymbol{\varepsilon}^{**},$$

where P is an $(abn - a - b + 1) \times (abn - a - b + 1)$ permutation matrix, so that

$$E(\tilde{\boldsymbol{\varepsilon}}^{**}) = PE(\boldsymbol{\varepsilon}^{**}) = \mathbf{0} \quad V(\tilde{\boldsymbol{\varepsilon}}^{**}) = PV(\boldsymbol{\varepsilon}^{**})P' = \sigma^2 PP' = \sigma^2 I_{abn-a-b+1}.$$

Therefore, using equation (10), we obtain the following F statistic F_{AB}^{**} for testing the hypotheses of interaction in (2).

$$F_{AB}^{**} = \frac{\mathbf{y}^{**'}(X_{ab}^{**}(X_{ab}^{**'}X_{ab}^{**})^{-1}X_{ab}^{**'})\mathbf{y}^{**}/(a-1)(b-1)}{\mathbf{y}^{**'}(I_{abn-a-b+1} - X_{ab}^{**}(X_{ab}^{**'}X_{ab}^{**})^{-1}X_{ab}^{**'})\mathbf{y}^{**}/(ab(n-1))}. \quad (11)$$

Since, $\mathbf{y}^{**'}X_{ab}^{**} = \mathbf{y}^{*'VV'X_B^*} = \mathbf{y}^{*'}(I_{abn} - H_{ab})X_{ab}^* = \mathbf{y}^{*'}X_B^*$, $X_{ab}^{**'}X_{ab}^{**} = X_{ab}^{*'}VV'X_{ab}^* = X_{ab}^{*'}(I_{abn} - H_{ab})X_{ab}^* = X_{ab}^{*'}X_{ab}^*$, and $\mathbf{y}^{**'}\mathbf{y}^{**} = \mathbf{y}^{*'}VV'\mathbf{y}^* = \mathbf{y}^{*'}(I_{abn} - H_{ab})\mathbf{y}^* = \mathbf{y}^{*'}\mathbf{y}^*$, F_{AB}^{**} is equal to F_{AB} in (3) or F_{AB}^* in (4).

Under H_0 , \mathbf{y}^{**} is not related to X_{ab}^{**} in (10), the statistical significance of H_0 can be evaluated from the Monte Carlo distribution of

$$\tilde{F}_{AB}^{**} = \frac{\tilde{\mathbf{y}}^{**'}(X_{ab}^{**}(X_{ab}^{**'}X_{ab}^{**})^{-1}X_{ab}^{**'})\tilde{\mathbf{y}}^{**}/(a-1)(b-1)}{\tilde{\mathbf{y}}^{**'}(I_{abn-a-b+1} - X_{ab}^{**}(X_{ab}^{**'}X_{ab}^{**})^{-1}X_{ab}^{**'})\tilde{\mathbf{y}}^{**}/(ab(n-1))}, \quad (12)$$

where $\tilde{\mathbf{y}}^{**}$ is an $(abn - a - b + 1) \times 1$ randomly permuted vector of \mathbf{y}^{**} . Thus, the significance of H_0 can be assessed by p - value = $P[\tilde{F}_{AB}^{**} \geq F]$. This test is denoted by RPT4.

RPT4 can be explained as follows. From model (7), since the obtained residuals are correlated with each other, uncorrelated residuals are obtained

using the transformation of the correlated residuals. To obtain the uncorrelated residuals, a transformation matrix using the decomposition of the idempotent matrix is used. The proposed randomization is based on these uncorrelated residuals. Since the probability distribution of the permuted observations $\tilde{\mathbf{y}}^{**}$ is same as that of \mathbf{y}^{**} , RPT4 satisfies the exchangeability property and is an exact permutation test.

4 Simulation study

To explore the performance of the procedures proposed in Section 3 for testing the hypothesis (2) in the two-way ANOVA model, a simulation study is carried out. For the simulation study, we generate the parameters of the model from some distributions instead of having fixed values. By doing so, we may get the average performance of the methods introduced. In order to obtain the estimated significance level and power of the various methods, for each (a, b, n) combinations, the following steps are used.

- Step 1.** We generate the main effects $\alpha_i, i = 1, \dots, a$ and $\beta_j, j = 1, \dots, b$ from uniform(-50,50) and uniform(-20,20) distributions respectively with constraints $\sum_{i=1}^a \alpha_i = 0$ and $\sum_{j=1}^b \beta_j = 0$.
- Step 2.** In order to obtain the estimates for significance level and power, we generate interaction effects as follows. For the estimation of the significance level, $(\alpha\beta)_{ij}$ equals 0 for all $i = 1, \dots, a$ and $j = 1, \dots, b$; and for the estimation of the power, $(\alpha\beta)_{ij}$ is generated from uniform(-c,c) with constraints $\sum_{i=1}^a (\alpha\beta)_{ij} = \sum_{j=1}^b (\alpha\beta)_{ij} = 0$. Set $\bar{\theta}_{AB} = \sum_i^a \sum_j^b (\alpha\beta)_{ij}^2 / ab$ as a parameter for the gradual increase of the interaction.
- Step 3.** For the values obtained in steps 1 and 2, we generate the error term ε_{ijk} from four different distributions, including $N(0, 1)$, $\exp(1) - 1$, $t(4)$ and $U(-\sqrt{3}, \sqrt{3})$.
- Step 4.** Apply the testing procedures, traditional F, RPT1, RPT2, RPT3, and RPT4 at significance level $\alpha = 0.05$ for each generated sample, and

check whether the null hypothesis is rejected or not. The permutation tests are based on 1,000 Monte Carlo replications.

Step 5. Steps 1 to 4 are repeated 10,000 times independently in order to estimate the significance level and power of the testing procedures.

Comparison of estimated significance level

Table 1 gives the average estimated significance level for various (a, b, n) combinations and four different error distributions at nominal size $\alpha = 0.05$.

Table 1 Average estimated significant level for various (a, b, n) combinations at nominal size $\alpha = 0.05$ *

(a, b, n)	distribution	F	RPT1	RPT2	RPT3	RPT4
(2,2,2)	N(0,1)	0.0491	0.0348	0.0660	0.0856	0.0502
	Exp(1)-1	0.0477	0.0333	0.0666	0.0833	0.0480
	t(4)	0.0439	0.0304	0.0619	0.0782	0.0460
	$U(-\sqrt{3}, \sqrt{3})$	0.0554	0.0408	0.0748	0.0884	0.0540
(2,2,3)	N(0,1)	0.0507	0.0491	0.0515	0.0385	0.0522
	Exp(1)-1	0.0437	0.0429	0.0488	0.0313	0.0455
	t(4)	0.0454	0.0458	0.0469	0.0333	0.0464
	$U(-\sqrt{3}, \sqrt{3})$	0.0523	0.0501	0.0526	0.0409	0.0534
(2,3,2)	N(0,1)	0.0481	0.0463	0.0510	0.0436	0.0500
	Exp(1)-1	0.0499	0.0495	0.0594	0.0469	0.0506
	t(4)	0.0414	0.0418	0.0454	0.0393	0.0439
	$U(-\sqrt{3}, \sqrt{3})$	0.0559	0.0531	0.0608	0.0516	0.0551
(2,3,3)	N(0,1)	0.0492	0.0500	0.0505	0.0464	0.0501
	Exp(1)-1	0.0454	0.0486	0.0523	0.0422	0.0468
	t(4)	0.0458	0.0495	0.0482	0.0430	0.0460
	$U(-\sqrt{3}, \sqrt{3})$	0.0525	0.0513	0.0532	0.0500	0.0529
(2,4,2)	N(0,1)	0.0485	0.0483	0.0499	0.0442	0.0504
	Exp(1)-1	0.0528	0.0537	0.0610	0.0487	0.0527
	t(4)	0.0432	0.0443	0.0460	0.0380	0.0444
	$U(-\sqrt{3}, \sqrt{3})$	0.0538	0.0528	0.0567	0.0490	0.0539
(3,4,3)	N(0,1)	0.0479	0.0483	0.0498	0.0468	0.0490
	Exp(1)-1	0.0464	0.0481	0.0498	0.0450	0.0460
	t(4)	0.0469	0.0510	0.0500	0.0452	0.0482
	$U(-\sqrt{3}, \sqrt{3})$	0.0481	0.0488	0.0487	0.0471	0.0494

* RPT1, RPT2, RPT3 and RPT4 are based on 10,000 Monte Carlo replications.

Using the normal approximation, if the true significance level is 0.05, there is less than a 5% chance that a simulated significance level based on 10,000 replications will be less than 0.046 or larger than 0.054.

For $(a, b, n) = (2, 2, 2)$, when the error term is normal the traditional F-test works well as expected. However, when the error term follows $t(4)$ distribution, the significance level of the F-test is underestimated. RPT1 underestimates the nominal significance level while RPT2 and RPT3 overestimate it for all considered distributions of the error terms. These phenomena are presumably due to the fact that the randomly permuted observations of y_{ijk}^* , y_{ijk}^\dagger and y_{ijk} , which are used for permutation in RPT1, RPT2 and RPT3 respectively, are correlated. However, the estimated significance level of RPT4, which is proposed in this paper, is not significantly different from the nominal size.

For $(a, b, n) = (2, 2, 3)$, RPT1 and RPT2 improve a little bit, but RPT3 underestimates the significance level especially for $\exp(1) - 1$ and $t(4)$ distributions. Over all, RPT4 works relatively well.

As the a , b and n increase, the correlations between the randomly permuted observations used for RPT1, RPT2, and RPT3 get weaker. Thus, we would expect that the difference between the procedures gets smaller as a , b and n increase (see Table 1).

Comparison of estimated power

Table 2 gives the estimated power for the various testing methods at significance level $\alpha = 0.05$ for $(a, b, n) = (2, 3, 2)$ when the error term is $N(0, 1)$, $\exp(1) - 1$, $t(4)$ and $U(-\sqrt{3}, \sqrt{3})$.

Table 2 Average estimated power for $(a, b, n) = (2, 3, 2)$.

distribution	θ_{AB}	F	RPT1	RPT2	RPT3	RPT4
N(0, 1)	0.0000	0.0481	0.0463	0.0510	0.0436	0.0500
	0.1663	0.0643	0.0606	0.0662	0.0593	0.0637
	0.3850	0.1261	0.1240	0.1316	0.1198	0.1260
	0.6134	0.2539	0.2460	0.2579	0.2391	0.2470
	0.8763	0.4695	0.4559	0.4702	0.4460	0.4502
	1.1196	0.6772	0.6636	0.6802	0.6546	0.6529
	1.3636	0.8440	0.8331	0.8445	0.8276	0.8209
	1.6152	0.9371	0.9318	0.9369	0.9286	0.9249
	1.9391	0.9802	0.9770	0.9768	0.9768	0.9733
Exp(1)-1	0.0000	0.0499	0.0495	0.0594	0.0469	0.0506
	0.1663	0.0705	0.0677	0.0828	0.0656	0.0703
	0.3850	0.1658	0.1616	0.1891	0.1580	0.1659
	0.6134	0.3342	0.3304	0.3649	0.3182	0.3305
	0.8763	0.5503	0.5443	0.5871	0.5324	0.5386
	1.1196	0.7271	0.7204	0.7619	0.7088	0.7073
	1.3636	0.8485	0.8469	0.8711	0.8372	0.8332
	1.6152	0.9116	0.9075	0.9302	0.9040	0.9015
	1.9391	0.9706	0.9677	0.9789	0.9648	0.9600
t(4)	0.0000	0.0414	0.0418	0.0454	0.0393	0.0439
	0.1663	0.0537	0.0536	0.0578	0.0496	0.0560
	0.3850	0.0940	0.0924	0.1019	0.0874	0.0936
	0.6134	0.1765	0.1741	0.1878	0.1676	0.1743
	0.8763	0.3074	0.3032	0.3230	0.2909	0.2970
	1.1196	0.4743	0.4681	0.4973	0.4519	0.4602
	1.3636	0.6254	0.6142	0.6411	0.6039	0.6069
	1.6152	0.7526	0.7420	0.7666	0.7319	0.7296
	1.9391	0.8649	0.8561	0.8785	0.8486	0.8459
U(- $\sqrt{3}$, $\sqrt{3}$)	0.0000	0.0559	0.0531	0.0608	0.0516	0.0551
	0.1663	0.0639	0.0611	0.0688	0.0601	0.0629
	0.3850	0.1255	0.1211	0.1328	0.1186	0.1250
	0.6134	0.2321	0.2241	0.2338	0.2219	0.2236
	0.8763	0.4344	0.4227	0.4304	0.4140	0.4167
	1.1196	0.6552	0.6411	0.6412	0.6310	0.6283
	1.3636	0.8450	0.8323	0.8306	0.8249	0.8190
	1.6152	0.9531	0.9476	0.9437	0.9421	0.9378
	1.9391	0.9949	0.9941	0.9909	0.9925	0.9888

From Table 2, note that the estimated powers of all tests increase as the parameter for the degree of the interaction effect $\bar{\theta}_{AB}$ increases. Also, note that the differences in the estimated power of all tests are rather small.

Testing for main effects

Suppose that the null hypothesis H_0 in (2) is not rejected. Then, model (1) can be rewritten as

$$y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk}, \quad i = 1, \dots, a, \quad j = 1, \dots, b, \quad \text{and } k = 1, \dots, n. \quad (13)$$

In model (13), consider hypotheses for the existence of the main effects.

$$H_0 : \alpha_1 = \dots = \alpha_a = 0 \quad \text{vs.} \quad H_1 : \text{not } H_0. \quad (14)$$

Under the normality assumption of error terms, the traditional F test statistic is

$$F_A = \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (\bar{y}_{i..} - \bar{y}_{...})^2 / (a-1)}{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (y_{ijk} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2 / (abn - a - b + 1)}. \quad (15)$$

Under the null hypothesis H_0^a , F_A is distributed as $F(a-1, abn - a - b + 1)$.

For the permutation test, the permutation methods described in Sections 2 and 3 can be easily extended for this problem. For RPT1, RPT2 and RPT3, the adjusted values used in permutations are $y_{ijk}^* = y_{ijk} - \bar{y}_{.j.}$, $y_{ijk}^\dagger = y_{ijk} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}$, and y_{ijk} itself, respectively.

For RPT4, since $\varepsilon_{ijk}^* = \varepsilon_{ijk} - \bar{\varepsilon}_{.j.}$ are correlated, the uncorrelated error term ε_{ijk}^{**} can be obtained by using the singular transformation used in Section 3. Therefore, distribution of F_A^{**} can be obtained as in (12) by using the random permutation of uncorrelated observations y_{ijk}^{**} . Then, the p-value of RPT4 can be calculated by comparing the observed F_A value to the distribution.

Three-way ANOVA model

Based on invariance and sufficiency, Welch (1990) considered testing for interaction effects in the three-way ANOVA model.

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + \varepsilon_{ijk} \quad (16)$$

$$i = 1, \dots, a, \quad j = 1, \dots, b, \quad k = 1, \dots, c,$$

with proper constraints for the parameters. In equation (16), the hypotheses for testing the interactions between factors B and C are as follows.

$$H_0 : (\beta\gamma)_{11} = \dots = (\beta\gamma)_{bc} = 0 \quad \text{vs.} \quad H_1 : \text{not } H_0. \quad (17)$$

Under the normality assumption of error terms, the traditional F test is

$$F_{BC} = \frac{\frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c (\bar{y}_{.jk} - \bar{y}_{.j.} - \bar{y}_{i.k} + \bar{y}_{i..})^2}{(b-1)(c-1)}}{\frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c (y_{ijk} - \bar{y}_{i..} - \bar{y}_{.j.} - \bar{y}_{i.k} + \bar{y}_{i.j.} + \bar{y}_{i.k} + \bar{y}_{.jk} - \bar{y}_{i..})^2}{(a-1)(b-1)(c-1)}}. \quad (18)$$

Under the null hypothesis H_0 , F_{BC} is distributed as $F((b-1)(c-1), (a-1)(b-1)(c-1))$. Now, the permutation methods described in Sections 2 and 3 can also be easily extended for this problem. In this model, the adjusted values used in permutations are $y_{ijk}^* = y_{ijk} - \bar{y}_{ij.} - \bar{y}_{i.k} + \bar{y}_{i..}$ for RPT1, $y_{ijk}^\dagger = y_{ijk} - \bar{y}_{i..} - \bar{y}_{.j.} - \bar{y}_{i.k} + \bar{y}_{i.j.} + \bar{y}_{i.k} + \bar{y}_{.jk} - \bar{y}_{i..}$ for RPT2, and y_{ijk} itself for RPT3. RPT4 for this model is similar to that for the two-way ANOVA model. Since $\varepsilon_{ijk}^* = \varepsilon_{ijk} - \bar{\varepsilon}_{ij.} - \bar{\varepsilon}_{i.k} + \bar{\varepsilon}_{i..}$ are correlated, the uncorrelated error term ε_{ijk}^{**} can be obtained by using the singular transformation used in (10). Therefore, the distribution of F_{BC}^{**} can be obtained by using the random permutation of uncorrelated observations y_{ijk}^{**} . Then, the p-value of RPT4 can be calculated by comparing the observed F_{BC} value to the distribution of F_{BC}^{**} .

5 A Numerical Example

We will work on the same data that were also given as an example in Montgomery (1997, pp 288-289). The data were collected from an article in the IEEE Transactions on Electron Devices (Nov. 1986, p. 1754), a study on polysilicon doping. In this study, the base current was the response variable, with two levels of polysilicon doping (factor A) and three levels of anneal temperature (factor B). Table 3 provides the experiment and its result.

Table 3 Experiment and the resulting observed data

Polysilicon Doping(ions)	Anneal Temperature($^{\circ}C$)		
	900	950	1000
1×10^{20}	4.60	10.15	11.01
	4.40	10.20	10.58
2×10^{20}	3.20	9.38	10.81
	3.50	10.02	10.60

In this example, we attempt to answer the question of whether interaction effects between polysilicon doping and temperature exist. Table 4 provides the results from a number of testing procedures for interaction effects.

Table 4 Results of testing procedures

Source	p-value				
	F	RPT1	RPT2	RPT3	RPT4
Interaction (A*B)	0.0645	0.0727	0.0665	0.0596	0.0581

We find that in Table 4, under the normality assumption of errors, the traditional F test for interaction effects proved to be significant at the 10 percent level. Thus, it is concluded that a mild degree of interaction exists between polysilicon doping and temperature. However, when the normality assumption is not satisfied, the result of the traditional F test cannot be justified. We therefore apply the permutation tests described in Section 2 and Section 3. These permutation tests are based on 100,000 Monte Carlo replications. We find that all the permutation tests in this experiment were still significant at about the 10 percent level. It is found that the p-value for RPT4, the new method of this study, was 0.0581, which was quite close to that of RPT3. The p-values of RPT3 and RPT4 turned out to be smaller than those of other methods.

6 Conclusion

In this study, we proposed a new method of random permutations for testing interaction effects in balanced two-way and three-way ANOVA models. The advantage of this study is that our permutation method is statistically exact compared with the existing methods. We carried out a restricted simulation and confirmed the fact that the proposed method outperformed the existing methods for small sample sizes.

References

1. Anderson, M.J. and ter Braak, C.J.F. (2003), Permutation tests for multi-factorial analysis of variance, *Journal of Statistical Computation and Simulation* 73, 85-113.
2. Edgington, E.S. (1995), *Randomization Tests*, Marcel Dekker, New York.
3. Fisher, R.A. (1935), *The Design of Experiments*, Edinburgh : Oliver & Boyd.
4. Gonzales, L. and Manly, B.F.J. (1998), Analysis of variance by randomization with small data sets. *Environmetrics* 9, 53-65.
5. Good, P. (2002), Extensions of the concept of exchangeability and their applications. *Journal of Modern Applied Statistical Methods* 1, 243-247.
6. Good, P. (1994), *Permutation Tests : A Practical Guide to Resampling Methods for Testing Hypotheses*, Springer-Verlag, New York.
7. Huh, M.H. and Jhun, M. (2001), Random permutation testing in multiple linear regression. *Communications in statistics - Theory and Methods* 30, 2023 - 2032.
8. Kennedy, P.E. (1995), Randomization tests in econometrics. *Journal of Business and Economic Statistics* 13, 85-94.
9. Manly, B.F.J. (1997), *Randomization, Bootstrap and Monte Carlo Methods in Biology*, Chapman and Hall, London.
10. Marsaglia, G and Styan, G.P.H. (1974), Rank conditions for generalized inverses of partitioned matrices. *Sankhya* 36, 437-442.
11. Montgomery, D.C. (1997), *Design and Analysis of Experiments*, John Wiley and Sons, New york.

12. Still, A.W. and White, A.P. (1981), The approximate randomization test as an alternative to the F-test in the analysis of variance. *British Journal of Mathematical and Statistical Psychology* 34, 243-252.
13. ter Braak, C.J.F. (1992), Permutation versus bootstrap significance tests in multiple regression and ANOVA, in Jockel, K.H. (ed.), *Bootstrapping and Related Techniques*, Springer, Berlin, 79-86.
14. Welch, W.J. (1990), Construction of permutation tests. *Journal of American Statistical Association* 85, 693-698.