Analysis of Variance for Fuzzy Data Based on Permutation Method

Woo-Joo Lee¹, Hye-Young Jung², Jin Hee Yoon³, and Seung Hoe Choi⁴

¹Department of Mathematics, Yonsei University, Seoul, Korea

²Department of Statistics, Seoul National University, Seoul, Korea

³School of Mathematics and Statistics, Sejong University, Seoul, Korea

⁴School of Liberal Arts and Science, Korea Aerospace University, Goyang, Korea

ljfis

Abstract

This paper deals with analysis of variance with fuzzy data (ANOVAF) based on permutation method. The permutation method is a nonparametric method introduced by Heap and Johnson for the data when the normal distribution cannot be assumed. We proposed two different approaches to test hypothesis of fuzzy means using the empirical distribution. To compare the results, several distances are considered especially using ρ -distance. Applying Monte Carlo simulation, it is confirmed through the numerical examples that the significant probability ($\hat{p}$ -value) get approached true parameter (p -value) regardless of distances or testing method based on proposed method. In addition, the number of permutation samples required is determined in the example to satisfy specified given accuracy.

Keywords: Fuzzy number, ANOVA, Permutation method, Monte Carlo simulation

1. Introduction

Analysis of variance (ANOVA) is a widely used statistical techniques in various fields such as design of experiments, survey design, quality improvement and many other industries. In general, the purpose of ANOVA is to test for significant differences of means among more than two independent populations. In ANOVA, the involved experimental data are assumed to be drawn from a normal distribution having equal variances. But in many practical situations the underlying distribution is far from being normal. In these situations, nonparametric method can be used as an alternative approach. Also, the data sometimes cannot be collected precise values due to ambiguous information or linguistic structure. An appropriate way of solving problems like these is to applying the concept of fuzzy set theory.

ANOVA using fuzzy set theory concepts has been studied by several authors. De Garibay [1] considered one-way fuzzy ANOVA and compares it with other classical techniques. Konishi et al. [2] propose the method of ANOVA which can treat the fuzzy interval data using the moment correction. Wu [3] solved one-way fuzzy ANOVA using the h-level set and the notions of pessimistic degree and optimistic degree. ANOVA study was developed by Gonzalez-Rodriguez et al. [4] and Montenegro et al. [5]. They presented an exact one-way ANOVA testing procedure for the case with normal fuzzy random variables. On the other hand, they approached large sample tests for simple fuzzy random variables. A bootstrap approach to ANOVA for fuzzy valued sample data is introduced in Gil et al. [6] and Montenegro et al. [7]. Lubiano and Trutschnig [8] used the R package SAFD (Statistical Analysis of Fuzzy

Received: Mar. 13, 2017

Revised : Mar. 20, 2017

Accepted: Mar. 23, 2017

Correspondence to: Seung Hoe Choi
 (shchoi@kau.ac.kr)

©The Korean Institute of Intelligent Systems

©This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Data). Also, Jiryaei et al. [9] applied least squares method.

In this paper, we discuss ANOVA model with fuzzy data applying the permutation method to compare the fuzzy mean responses. The permutation tests are significant tests based on permutation resamples drawn from the original data. Since the permutation resamples are drawn without replacement, in contrast to bootstrap samples, which are drawn with replacement, the permutation resamples include all original data. If the number of all possible permutations is large number then for computational reasons, we can use random permutation test. In this paper we perform the permutation test based on the Monte Carlo simulation to determine how many permutation samples are required to provide a sufficiently accurate p -value.

This paper is organized as follows: In Section 2, some preliminary concepts which are required to develop the main results are presented. In Section 3, numerical example is presented to illustrate ANOVA model with fuzzy data applying permutation test under several distances, special cases of ρ -distance.

2. Preliminaries

In this section, we briefly describe the main theoretical tools that have been used in the proposing ANOVA model using permutation method.

2.1 Fuzzy Numbers

Some basic concepts of fuzzy theory are provided as follows:

Definition 1. A fuzzy subset $\tilde{A}$ of the set of real numbers $\mathbf{R}^n$ with membership function $u_{\tilde{A}} : \mathbf{R}^n \rightarrow [0, 1]$ is called a fuzzy number if satisfies (i) $\tilde{A}$ is normal, i.e. $u_{\tilde{A}}(x_0) = 1$ for some $x_0 \in \mathbf{R}^n$, (ii) $\tilde{A}$ is convex set, i.e. $\forall x_1, x_2 \in \mathbf{R}^n$ and $\lambda \in [0, 1]$: $u_{\tilde{A}}(\lambda x_1 + (1 - \lambda)x_2) \geq \min(u_{\tilde{A}}(x_1), u_{\tilde{A}}(x_2))$ (iii) $u_{\tilde{A}}$ is semicontinuous.

The α -cut of fuzzy number $\tilde{A}$ is a finite closed interval $A_\alpha = \{x \in \mathbf{R}^n | u_{\tilde{A}}(x) \geq \alpha, \alpha \in [0, 1]\} = [A_\alpha^L, A_\alpha^U]$ where $A_\alpha^L = \inf\{x \in \mathbf{R}^n | u_{\tilde{A}}(x) \geq \alpha\}$, $A_\alpha^U = \sup\{x \in \mathbf{R}^n | u_{\tilde{A}}(x) \geq \alpha\}$. The statistical management of fuzzy numbers usually requires suitable metric between them. The support function of any compact convex set $A \subset \mathbf{R}^n$ is defined as function $s_A : S^{n-1} \rightarrow \mathbf{R}$ given by, for all $x \in S^{n-1}$

$$s_A(x) = \sup_{a \in A} (x \cdot a),$$

where S^{n-1} is the $(n - 1)$ -dimensional unit sphere in $\mathbf{R}^n$ and $\cdot$ is the inner product in $\mathbf{R}^n$. A ρ -distance as a unification of

different distance applied in statistical inference with fuzzy data was introduced by Näther [10].

Definition 2. The ρ -distance for any two fuzzy numbers $\tilde{A}$ and $\tilde{B}$ is defined as

$$\rho^2(\tilde{A}, \tilde{B}) := \int \int \int \int_S \{s_{\tilde{A}_\alpha}(x) - s_{\tilde{B}_\alpha}(x)\} \times \{s_{\tilde{A}_\beta}(y) - s_{\tilde{B}_\beta}(y)\} dK(x, \alpha, y, \beta),$$

where $S = S^{n-1} \times [0, 1] \times S^{n-1} \times [0, 1]$ and K is a symmetric and positive definite kernel.

Let $\tilde{A} = (a^l, a^c, a^r)$ and $\tilde{B} = (b^l, b^c, b^r)$ be the triangular fuzzy numbers. a^c is the center of $\tilde{A}$ and a^l, a^r are left, right spread of $\tilde{A}$. Using the special kernel, ρ -distance reduces to the several distances as follows:

ρ -distance reduces to well-known Diamond distance [11].

$$d_D^2(\tilde{A}, \tilde{B}) = (a^l - b^l)^2 + (a^m - b^m)^2 + (a^r - b^r)^2. \quad (1)$$

The distance by Ming et al. [12]

$$d_M^2(\tilde{A}, \tilde{B}) = \int_0^1 \{(A_\alpha^L - B_\alpha^L)^2 + (A_\alpha^U - B_\alpha^U)^2\} d\alpha \quad (2)$$

is included in the ρ -distance.

Bertoluzza et. al. [13] defined the distance $d_B^2(\tilde{A}, \tilde{B})$

$$\int_{[0,1]} \int_{[0,1]} [f_{\tilde{A}}(\alpha, \lambda) - f_{\tilde{B}}(\alpha, \lambda)]^2 dW(\lambda) d\varphi(\alpha), \quad (3)$$

where $f_{\tilde{A}}(\alpha, \lambda) = \lambda \sup A_\alpha + (1 - \lambda) \inf A_\alpha$, and W, φ are weighting measures.

2.2 Traditional Analysis of Variance Model

Traditional ANOVA model, one-factor model can be stated as follows:

$$Y_{ij} = \mu_i + \epsilon_{ij}, \quad i = 1, \dots, r; \quad j = 1, \dots, n_i,$$

where Y_{ij} is value of the response variable in the j th trial for the i th factor level or treatment and assume random sampling with equal variances, independent errors, and a normal distribution. ANOVA is a statistical method used to test differences between group means. The relevant hypotheses are

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_r, \quad \text{vs.}$$

H_a : At least one of the μ'_i s is not equal.

The number of cases for the i th factor level is denoted by n_i and the total number of cases in the study is denoted by N , where $N = \sum_{i=1}^r n_i$. The total sum of squares (SSTO), the treatment sum of squares (SSTR) and the error sum of squares (SSE) be defined respectively as

$$\sum_{i=1}^r \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{..})^2, \sum_{i=1}^r \sum_{j=1}^{n_i} (\bar{Y}_{i.} - \bar{Y}_{..})^2, \sum_{i=1}^r \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i.})^2,$$

where $\bar{Y}_{i.}$ is the mean of observations within factor level i , and $\bar{Y}_{..}$ is the grand mean (i.e., $\bar{Y}_{..} = \sum_{i=1}^r \sum_{j=1}^{n_i} \bar{Y}_{i.} / N$). The test statistic

$$F = \frac{SSTR/(r-1)}{SSE/(N-r)}$$

follow an F distribution with $r-1$ and $N-r$ degrees of freedom. There are two equivalent decision rules to test given hypotheses. Since we know that f is distributed as $F_{r-1, N-r}$ when H_0 holds and that large values of f lead to conclusion H_a , we have a decision rule to control the level of significant at α as follows:

If $f \leq F_{\alpha; r-1, N-r}$, conclude H_0 ,

If $f > F_{\alpha; r-1, N-r}$, conclude H_a ,

where $F_{\alpha; r-1, N-r}$ is the $(1-\alpha)100$ percentile of the F distribution and f is the observed F -statistic value. Another approach of decision rule is p -value approach. The p -value is calculated as the area under the appropriate null sampling distribution of F that is bigger than f . If the p -value is smaller the level of significance, reject the null hypothesis that all means are the same for the different groups.

2.3 Permutation Method

In many studies, it is assumed that the data were taken from normally distributed population. However, when the data size is small, it is not suitable to assume normal population. Even the case we have a big data set, it is not very common that the data are taken from exactly normally distributed population. Hence, nonparametric methods have been used as its alternatives. But some nonparametric methods make use of only ranks of the data, which results in loss of raw data information. To overcome this problem, Heap [14] and Johnson [15] proposed permutation methods which employ raw data information. Moreover these

methods are not dependent on distribution of population. A permutation method is a distribution free method, which can be used in case the test statistic does not follow a well-known distribution. For example, in case the test of difference of means of two populations, even though a permutation method use t -statistic, it is not assume that t -statistic follow t -distribution. It obtains an empirical distribution for the test statistic directly from dataset. In other words, it obtains an empirical distribution after calculating all possible permutations from observed data, then it calculates p -value after investigating where the observed test statistic is located in the empirical distribution. This method could not be widely used in the past because of huge amount of calculation, but it actively has been studied recently due to computer development. As the number of data size is increased, the possible number of permutation gets increased geometrically. Accordingly, it is almost impossible to apply exact permutation method which needs an empirical distribution with total amount of possible permutations. In this case, a random permutation method can be applied which needs randomly chosen permutations to generate an approximate empirical distribution.

The general permutation method proceeds as follows:

1. Set a statistic for the test of the hypothesis.
2. Calculate the test statistic based on observed data.
3. Generate permutation data ($b = 1, \dots, B$).
4. Calculate the test statistic based on each permutation data, and find the empirical distribution.
5. Calculate the p -value in the empirical distribution based on the test statistic using observed data.
6. Test the hypothesis using p -value.

3. Analysis of Variance Model with Fuzzy Data

Consider one-way ANOVA model with fuzzy data (ANOVAF).

$$\tilde{X}_{ij} = \tilde{\mu}_i + \epsilon_{ij}, \quad i = 1, 2, \dots, r, \quad j = 1, 2, \dots, n_i, \quad (4)$$

where $\tilde{\mu}_i$ is the fuzzy mean of the i th population. ϵ_{ij} is an error term without distribution assumption. Given r populations with fuzzy data, we construct hypothesis for the test as follows:

$$H_0 : \tilde{\mu}_1 = \tilde{\mu}_2 = \dots = \tilde{\mu}_r, \quad \text{vs.}$$

$$H_a : \text{At least one of the } \tilde{\mu}'_i \text{'s is not equal.}$$

Likewise in traditional ANOVA model, following properties hold in ANOVA_F. The sum of squares of the distance deviations about the fuzzy grand mean can be decomposed into two sums, the first of which represents the within-group sum of squares, and the second, the between-group sum of squares:

$$\begin{aligned} \sum_{i=1}^r \sum_{j=1}^{n_i} \rho^2(\tilde{X}_{ij}, \bar{\tilde{X}}_{..}) \\ = \sum_{i=1}^r \sum_{j=1}^{n_i} \rho^2(\tilde{X}_{ij}, \bar{\tilde{X}}_{i.}) + \sum_{i=1}^r \sum_{j=1}^{n_i} \rho^2(\bar{\tilde{X}}_{i.}, \bar{\tilde{X}}_{..}), \end{aligned}$$

where $\bar{\tilde{X}}_{i.}$ be the fuzzy mean value of i th group and $\bar{\tilde{X}}_{..}$ be the total fuzzy grand mean. The test statistic to be used is

$$F = \frac{\frac{1}{r-1} \sum \sum \rho^2(\bar{\tilde{X}}_{i.}, \bar{\tilde{X}}_{..})}{\frac{1}{N-r} \sum \sum \rho^2(\tilde{X}_{ij}, \bar{\tilde{X}}_{i.})}. \quad (5)$$

If error is assumed to be normal distribution, the test statistic follows an F distribution with $r - 1$ and $N - r$ degrees of freedom [5]. In this study, we use to test hypothesis permutation method that does not require the restrictive assumption such as normal distribution. The proposed method proceeds as follows:

1. Calculate the F -value from the observed fuzzy data.
2. Generate all possible permutations from the data $\tilde{X} = \{\tilde{X}_{11}, \dots, \tilde{X}_{1n_1}, \tilde{X}_{21}, \dots, \tilde{X}_{rn_r}\}$. Let $b = 1, 2, \dots, B$ be index a specific permutation of X , where $B = {}_{C_{N-r_1} \times {}_{C_{N-(n_1+n_2)} \times \dots \times {}_{C_{N-\sum_{i=1}^{r-1} n_i}}}}^{C_{n_r}}$. The b th permutation sample is denoted $\tilde{X}^b = \{\tilde{X}_{11}^b, \dots, \tilde{X}_{1n_1}^b, \tilde{X}_{21}^b, \dots, \tilde{X}_{rn_r}^b\}$.
3. Calculate the F -value from b th permutation sample and denote F^b . Repeat for $b = 1, \dots, B$.
4. Find empirical permutation distribution of test statistics.

Now, we present two different approaches for the hypothesis test. First approach uses the p -value calculated by the ratio of the number of cases which are greater than F -value obtained from observations and F -value obtained from permutation samples.

$$p - \text{value} = \frac{\sum_{b=1}^B I(F^b \geq F)}{B}, \quad (6)$$

where $I(\cdot)$ is the indicator function, equaling '1' if the condition in parentheses is true, and '0' otherwise. This permutation test is denoted by PT1.

Table 1. Fuzzy data for sales volumes

Package design i			
1	$\tilde{11}=(9,11,13)$	$\tilde{16}=(14,16,18)$	
2	$\tilde{15}=(11,15,19)$	$\tilde{15}=(10,15,20)$	$\tilde{13}=(11,13,15)$
3	$\tilde{18}=(15,18,21)$	$\tilde{17}=(14,17,20)$	$\tilde{20}=(17,20,23)$
4	$\tilde{19}=(15,19,23)$	$\tilde{24}=(21,24,27)$	

The second approach uses the p -value calculated from the ratio of the number of cases which are greater than the critical points of F distribution. This permutation test is denoted by PT2.

$$p - \text{value} = \frac{\sum_{b=1}^B I(F^b \geq F_{\alpha, r-1, N-r})}{B}. \quad (7)$$

If the p -value is small than the level of significance, reject the null hypothesis that all means are the same for the different groups.

If all possible permutation numbers are large, we randomly choose some of them for the test, which use estimated significant probability. Let φ be the true parameter of the estimated significant probability. The data obtained from permutation data follow $p_1, \dots, p_n \sim \text{Bernoulli}(\varphi)$ and $\text{Var}(\hat{p}) = \varphi(1 - \varphi)/n$. Therefore, the estimated variance of the significant probability is small enough if we use more than a specific number of permutations.

3.1 Example 1

Consider the fuzzy data are given to be symmetric triangular fuzzy numbers [3]. We test whether or not the fuzzy mean are the same for the four package designs. The hypothesis for the test is given by:

$$H_0 : \tilde{\mu}_1 = \tilde{\mu}_2 = \tilde{\mu}_3 = \tilde{\mu}_4, \quad \text{vs.}$$

$$H_a : \text{At least one of the } \tilde{\mu}_i\text{'s is not equal.}$$

The fuzzy means for four package designs are $\bar{\tilde{Y}}_1 = (11.5, 13.5, 15.5)$, $\bar{\tilde{Y}}_2 = (10.67, 14.33, 18)$, $\bar{\tilde{Y}}_3 = (15.33, 18.33, 21.33)$ and $\bar{\tilde{Y}}_4 = (18, 21.5, 25)$. And fuzzy grand mean is $\bar{\tilde{Y}}_{..} = (13.7, 16.8, 19.9)$. As the special cases for ρ -distances, we also applied some distances introduced by Diamond [11], Ming et al. [12] and Bertoluzza et al. [13]. The F -value calculated from the observed fuzzy data are respectively 5.2410(F_D), 5.4330(F_M) and 5.5721(F_B) where F_D ,

Figure 1. Empirical permutation distribution in Example 1. (a) Diamond [11], (b) Ming et al. [12], and (c) Bertoluzza et al. [13].

Table 2. The illustration and the summary statistics for exact permutation method in Example 1

Replicate	Package design				F^b		
B	1	2	3	4	Diamond [11]	Ming [12]	Bertoluzza [13]
1 (F -value)	$\tilde{11}, \tilde{16}$	$\tilde{15}, \tilde{15}, \tilde{13}$	$\tilde{18}, \tilde{17}, \tilde{20}$	$\tilde{19}, \tilde{24}$	5.2410	5.4330	5.5721
2	$\tilde{11}, \tilde{16}$	$\tilde{15}, \tilde{15}, \tilde{13}$	$\tilde{18}, \tilde{17}, \tilde{19}$	$\tilde{20}, \tilde{24}$	7.0197	7.3946	7.6733
3	$\tilde{11}, \tilde{16}$	$\tilde{15}, \tilde{15}, \tilde{13}$	$\tilde{18}, \tilde{17}, \tilde{24}$	$\tilde{20}, \tilde{19}$	3.4223	3.4963	3.5487
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
25,200	$\tilde{19}, \tilde{24}$	$\tilde{18}, \tilde{17}, \tilde{20}$	$\tilde{15}, \tilde{15}, \tilde{13}$	$\tilde{11}, \tilde{16}$	5.2410	5.4330	5.5721
Critical point of empirical distribution					3.9103	4.0712	4.1893
p -value of EPT1					0.0203	0.0198	0.0203
Critical point of F					4.7571	4.7571	4.7571
p -value of EPT2					0.0254	0.0297	0.0325

F_M and F_B are F -value using the distance by Diamond, Ming and Bertoluzza. In this paper, we apply the permutation method to test the hypothesis. The number of total possible permutation is $25,200 (B = 10 \ C_2 \times 8 \ C_3 \times 5 \ C_3 \times 2 \ C_2)$. Figure 1 is the empirical distributions of F -value ($F^b; b = 1 \cdots, 25, 200$) based on permutation sample.

In exact PT1 (EPT1), the p -values which is greater than F -value calculated by observed data are 0.0203, 0.0198, 0.0203, respectively. And the null hypothesis is rejected because all p -values are smaller than the significance level $\alpha = 0.05$. Moreover, the critical points of distribution are 3.9103, 4.0712, 4.1893, which are greater than F -values obtained from observed data. Hence we conclude that the fuzzy means of package designs are different.

In EPT2, the critical point of F distribution is $F_{0.05;3,6} = 4.7571$ and F_D, F_M, F_B are greater than $F_{0.05;3,6}$, respectively. In addition, the p -values which are defined greater than $F_{0.05;3,6}$ are 0.0254, 0.0297 and 0.0325, respectively. Therefore, the null hypothesis is rejected at the significance level $\alpha = 0.05$. In practice, the number of all permutations can be too large to be tractable. In this case, we can test the hypothesis with randomly chosen permutation data. Table 3 shows Monte Carlo

simulation results through 5,000 iterations. Let the p -value calculated by EPT is the true parameter, the estimated p -value gets close to the true parameter as B , the number of permutation sample, increases. In other words, the sampling error gets small when B gets increased. In addition, there is no significant differences in errors regardless of distances (Diamond, Ming, Bertoluzza) and test methods (RPT1, RPT2). Moreover, when the alternative hypothesis (H_a) is true, the power of the test, increases, B increases. Further, the power of the test doesn't show any significant difference regardless of distances or test methods. Here, the power of the test is calculated based on the ratio of the number of rejection among 5,000 simulations. In addition, the ratio of the case when the true parameter is included in the confidence interval under 95% confidence level also gets increased, as B gets increased. Especially, the number of permutation sample should be large enough when more than 4,000 in order to satisfy 95% confidence level.

3.2 Example2

Some samples are selected to compare means of milk pouches which are produced by 4 different machines. The samples are

Table 3. The statistical results of random permutation method in Example 1

		B	50	100	1,000	3,000	5,000	10,000
Diamond [11]	RPT1	$\hat{p} - p$	0.0197	0.0097	0.0009	0.0002	0.0001	0.0001
		Power	0.7380	0.8584	1.0000	1.0000	1.0000	1.0000
		C.L	0.3612	0.4012	0.7768	0.9386	0.9830	0.9986
	RPT2	$\hat{p} - p$	0.0197	0.0097	0.0009	0.0003	0.0002	0.0001
		Power	0.6452	0.7584	1.0000	1.0000	1.0000	1.0000
		C.L	0.2708	0.5384	0.8002	0.9404	0.9816	0.9986
Ming et. al. [12]	RPT1	$\hat{p} - p$	0.0195	0.0097	0.0009	0.0002	0.0001	0.0001
		Power	0.7506	0.8698	1.0000	1.0000	1.0000	1.0000
		C.L	0.3718	0.4198	0.8050	0.9442	0.9834	0.9984
	RPT2	$\hat{p} - p$	0.0195	0.0097	0.0009	0.0003	0.0002	0.0001
		Power	0.5678	0.6650	0.9996	1.0000	1.0000	1.0000
		C.L	0.2182	0.4290	0.7658	0.9410	0.9772	0.9970
Bertoluzza et. al. [13]	RPT1	$\hat{p} - p$	0.0197	0.0098	0.0009	0.0003	0.0002	0.0001
		Power	0.7408	0.8600	1.0000	1.0000	1.0000	1.0000
		C.L	0.5600	0.3986	0.7714	0.9352	0.9852	0.9994
	RPT2	$\hat{p} - p$	0.0194	0.0096	0.0010	0.0003	0.0002	0.0002
		Power	0.5206	0.5956	0.9968	1.0000	1.0000	1.0000
		C.L	0.1894	0.3682	0.7630	0.9364	0.9808	0.9970

Table 4. Fuzzy data for milk pouches

Machine (i)			
1	(8,9,14)	(11,14,18)	
2	(7,11,15)	(8,10,14)	(8,11,14)
3	(12,15,18)	(12,14,19)	(14,17,23)
4	(12,15,19)	(19,21,24)	

observed as fuzzy numbers because the data could not recorded exactly due to unexpected situation. Table 4 shows the fuzzy observations which are non-symmetric triangular fuzzy numbers [16].

The fuzzy means for four package designs are $\bar{Y}_1 = (9.5, 11.5, 16)$, $\bar{Y}_2 = (7.67, 10.67, 14.33)$, $\bar{Y}_3 = (12.67, 15.33, 20)$ and $\bar{Y}_4 = (15.5, 18, 21.5)$. And fuzzy grand mean is $\bar{Y}_{..} = (11.1, 13.7, 17.8)$. The hypothesis for the test of means of numbers of milk pouches produced from each machine is as follow:

$$H_0 : \tilde{\mu}_1 = \tilde{\mu}_2 = \tilde{\mu}_3 = \tilde{\mu}_4, \quad \text{vs.}$$

$$H_a : \text{At least one of the } \tilde{\mu}_i\text{'s is not equal.}$$

F_D , F_M and F_B are 4.9019, 4.9399 and 5.0134, respectively .

The number of permutation sample is 25,200. In EPT1, critical point of significance level 0.05 from the empirical distribution are 4.3584, 4.4211 and 4.4866. The p -value are 0.0337, 0.0358 and 0.0354. Therefore, the null hypothesis is rejected. In EPT2, since F -value $> F_{0.05; 3, 6}$ and p -value $< \alpha$, H_0 is rejected. That is, we can conclude that means of the number of pouches are different. Table 6 shows the results from the simulation through 5,000 iterations. As B gets increased, the estimates of p -value get approached true parameter and the power of the test also get improved. In addition, confidence level increase as B increases. Especially, the number of permutation sample should be large enough when more than 4,000 in order to satisfy significance level $\alpha = 0.05$.

4. Conclusion

In this study, a permutation method is applied to ANOVA model with fuzzy data based on the distances introduced by Diamond [11], Ming et. al. [12] and Bertoluzza et. al. [13]. A permutation method can be applied using an empirical distribution through resampling, which doesn't need any assumption regarding specific distribution. Regardless of distances or test methods, it is shown through Monte Carlo simulation that the

Table 5. The statistical results of exact permutation method in Example 2

Metric	F -value	C.P.	$F_{0.05; 3, 6}$	p -value of EPT1	p -value of EPT2
Diamond [11]	4.9019	4.3584	4.7571	0.0337	0.0395
Ming et. al. [12]	4.9399	4.4211	4.7571	0.0358	0.0406
Bertoluzza et. al. [13]	5.0134	4.4866	4.7571	0.0354	0.0419

Table 6. The statistical results of random permutation method in Example 2

		B	50	100	1,000	3,000	5,000	10,000
Diamond [11]	RPT1	$\hat{p} - p$	0.0193	0.0102	0.0013	0.0007	0.0005	0.0005
		Power	0.5080	0.5592	0.9928	1.0000	1.0000	1.0000
		C.L	0.1764	0.3368	0.7798	0.9378	0.9778	0.9982
	RPT2	$\hat{p} - p$	0.0191	0.0098	0.0010	0.0003	0.0002	0.0001
		Power	0.4190	0.4434	0.9268	0.9964	0.9996	1.0000
		C.L	0.4190	0.4434	0.7922	0.9288	0.9764	0.9986
Ming et. al. [12]	RPT1	$\hat{p} - p$	0.0192	0.0099	0.0010	0.0003	0.0002	0.0001
		Power	0.4732	0.5236	0.9810	1.0000	1.0000	1.0000
		C.L	0.4730	0.5236	0.7844	0.9382	0.9776	0.9988
	RPT2	$\hat{p} - p$	0.0189	0.0098	0.0010	0.0003	0.0002	0.0001
		Power	0.4082	0.4230	0.8932	0.9900	0.9990	1.0000
		C.L	0.4082	0.4230	0.7816	0.9392	0.9788	0.9974
Bertoluzza et. al. [13]	RPT1	$\hat{p} - p$	0.0192	0.0099	0.0010	0.0003	0.0002	0.0001
		Power	0.4790	0.5318	0.9852	1.0000	1.0000	1.0000
		C.L	0.4790	0.5318	0.7532	0.9306	0.9776	0.9984
	RPT2	$\hat{p} - p$	0.0188	0.0097	0.0010	0.0003	0.0002	0.0001
		Power	0.3912	0.3978	0.8470	0.9792	0.9968	1.0000
		C.L	0.3912	0.3978	0.7656	0.9388	0.9794	0.9982

power of the test increases and the significant probability ($\hat{p}$ -value) approach true parameter (p -value), moreover the confidence level increases, as the size of permutation data increases. Therefore, if the size of total possible permutation sample is large enough, the test results in accurate conclusion even if we use only randomly selected permutation sample. This permutation method can be expanded to two-way ANOVA model which will be discussed in our further studies.

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

References

- [1] V. G. De Garibay, "Behaviour of fuzzy ANOVA," *Kybernetes*, vol. 16, no. 2, pp. 107-112, 1987. <http://dx.doi.org/10.1108/eb005761>
- [2] M. Konishi, T. Okuda, and K. Asai, "Analysis of variance based on fuzzy interval data using moment correction method," *International Journal of Innovative Computing, Information and Control*, vol. 2, no. 1, pp. 83-99, 2006.
- [3] H. C. Wu, "Analysis of variance for fuzzy data," *International Journal of Systems Science*, vol. 38, no. 3, pp. 235-246, 2007. <http://dx.doi.org/10.1080/00207720601157997>
- [4] G. Gonzalez-Rodriguez, A. Colubi, and M. A. Gil, "Fuzzy data treated as functional data: a one-way ANOVA test approach," *Computational Statistics and Data Analysis*, vol. 56, no. 4, pp. 943-955, 2012. <http://dx.doi.org/10.1016/j.csda.2010.06.013>
- [5] M. Montenegro, G. Gonzalez-Rodriguez, M. A. Gil, A. Colubi, and M. R. Casals, "Introduction to ANOVA with fuzzy random variables," *Soft Methodology and Random*

- Information Systems*, vol. 26, pp. 487-494, 2004. http://dx.doi.org/10.1007/978-3-540-44465-7_60
- [6] M. A. Gil, M. Montenegro, G. Gonzalez-Rodriguez, A. Colubi, and M. R. Casals, "Bootstrap approach to the multi-sample test of means with imprecise data," *Computational Statistics and Data Analysis*, vol. 51, no. 1, pp. 148-162, 2006. <http://dx.doi.org/10.1016/j.csda.2006.04.018>
- [7] M. Montenegro, A. Colubi, M. R. Casals, and M. A. Gil, "Asymptotic and bootstrap techniques for testing the expected value of a fuzzy random variable," *Metrika*, vol. 59, no. 1, pp. 31-49, 2004. <http://dx.doi.org/10.1007/s001840300270>
- [8] M. A. Lubiano and W. Trutschnig, "ANOVA for fuzzy random variables using the R-package SAFD," in *Combining Soft Computing and Statistical Methods in Data Analysis, Advances in Intelligent and Soft Computing Vol. 77*, C. Borgelt et al. Eds. Berlin: Springer, 2010, pp. 449-456. http://dx.doi.org/10.1007/978-3-642-14746-3_56
- [9] A. Jiryaei, A. Parchami, and M. Mashinchi, "One-way ANOVA and least squares method based on fuzzy random variables," *Turkish Journal of Fuzzy Systems*, vol. 4, no. 1, pp. 18-33, 2013.
- [10] W. Näther, "Random fuzzy variables of second order and applications to statistical inference," *Information Sciences*, vol. 133, no. 1-2, pp. 69-88, 2001. [http://dx.doi.org/10.1016/S0020-0255\(01\)00077-9](http://dx.doi.org/10.1016/S0020-0255(01)00077-9)
- [11] P. Diamond, "Fuzzy least squares," *Information Sciences*, vol. 46, no. 3, pp. 141-157, 1988. [https://doi.org/10.1016/0020-0255\(88\)90047-3](https://doi.org/10.1016/0020-0255(88)90047-3)
- [12] M. Ming, M. Friedman, and A. Kandel, "General fuzzy least squares," *Fuzzy Sets and Systems*, vol. 88, no. 1, pp. 107-118, 1997. [https://doi.org/10.1016/S0165-0114\(96\)00051-6](https://doi.org/10.1016/S0165-0114(96)00051-6)
- [13] C. Bertoluzza, N. Corral Blanco, and A. Salas, "On a new class of distances between fuzzy numbers," *Mathware & Soft Computing*, vol. 2, no. 2, pp. 71-78, 1995.
- [14] B. R. Heap, "Permutations by Interchanges," *The Computer Journal*, vol. 6, no. 3, pp. 293-298, 1963. <https://doi.org/10.1093/comjnl/6.3.293>
- [15] S. M. Johnson, "Generation of permutations by adjacent transposition," *Mathematics of Computation*, vol. 17, pp. 282-285, 1963. <https://doi.org/10.1090/S0025-5718-1963-0159764-2>
- [16] D. Kalpanapriya and P. Pandian, "Fuzzy hypothesis testing of Anova model with fuzzy data," *International Journal of Modern Engineering Research*, vol. 2, no. 4, pp. 2951-2956, 2012.

Woo-Joo Lee obtained a Ph.D. in Mathematics majoring mathematical statistics from Yonsei University, Seoul, Korea. His main research interests are data mining, fuzzy time series and intelligent systems.

Hye-Young Jung received her Ph.D. in Mathematics from Yonsei University in 2014. Currently, she is an associate teaching professor at the Faculty of Liberal Education, Seoul National University. Her research interests include statistical methods, fuzzy theory, and statistical genetics.

Jin Hee Yoon received Bachelor, Master and Ph.D. degree in Mathematics from Yonsei University. She is currently faculty of school of Mathematics and Statistics at Sejong University, Seoul, Korea. Her research interests are fuzzy regression analysis, fuzzy time series and F-transform. She is a board member of KIIS (Korean Institute of Intelligent Systems) and has been working as an editor, guest editor and editorial board member of The Scientific World Journal, Information, International Journal of Applied Economics and International Journal of Fuzzy logic and Intelligent Systems etc. Also, she has been working as an organizer and committee member of several international conferences such as 2012, 2013 NIMS Hot topic workshop, 2012, 2013 ICIS, 2014, 2015 FUZZ IEEE, 2015 ISIS, 2016 SCIS&ISIS, 2016 ICFMsquare 2016, ICMF 2016 etc.

Seung Hoe Choi obtained his Ph.D. degree in Mathematical Statistics from Yonsei University Korea in 1994. Since 1996, he is currently full professor of Korea Aerospace University. His main research interests are mathematical prediction method using the soft computing and statistical prediction in sports like soccer, baseball and basketball.