One-Factor ANOVA Model Using Trapezoidal Fuzzy Numbers Through Alpha Cut Interval Method

P. Gajivaradhan[1] and S. Parthiban[2 *]

1. Department of Mathematics, Pachaiyappa's College, Chennai-600 030, Tamil Nadu, India. Email:

2. Research Scholar, Department of Mathematics, Pachaiyappa's College, Chennai-600 030, Tamil Nadu, India.

Abstract

Most of our traditional tools in descriptive and inferential statistics is based on crispness (preciseness) of data, measurements, random variable, hypotheses, and so on. By crisp we mean dichotomous that is, yes-or-no type rather than more-or-less type. But there are many situations in which the above assumptions are rather non-realistic such that we need some new tools to characterize and analyze the problem. By introducing fuzzy set theory, different branches of mathematics are recently studied. But probability and statistics attracted more attention in this regard because of their random nature. Mathematical statistics does not have methods to analyze the problems in which random variables are vague (fuzzy).

In this regard, a simple and new technique for testing the hypotheses under the fuzzy environments is proposed. Here, the employed data are in terms of trapezoidal fuzzy numbers (TFN) which have been transformed into interval data using α-cut interval method and on the grounds of the transformed fuzzy data, the one-factor ANOVA test is executed and decisions are concluded. This concept has been illustrated by giving two numerical examples.

Keywords: Fuzzy set, α-cut, Trapezoidal fuzzy number (TFN), Test of hypotheses, One-factor ANOVA model, Upper level data, Lower level data.

1. Introduction

Fuzzy set theory [32] has been applied to many areas which need to manage uncertain and vague data. Such areas include approximate reasoning, decision making, optimization, control and so on.

In traditional statistical testing [17], the observations of sample are crisp and a statistical test leads to the binary decision. However, in the real life, the data sometimes cannot be recorded or collected precisely. The statistical hypotheses testing under fuzzy environments has been studied by many authors using the fuzzy set theory concepts introduced by Zadeh [32].

The application by using fuzzy set theory to statistics has been widely studied in Manton et al. [21] and Buckley [8] and Viertl [27]. Arnold [6] proposed the fuzzification of usual statistical hypotheses and considered the testing hypotheses under fuzzy constraints on the type I and type II errors. Saade [24], Saade and Schwarzlander [23] considered the binary hypotheses testing and discussed the fuzzy likelihood functions in the decision making process by applying a fuzzified version of the Baye's criterion. Grzegorzewski [14] and Watanabe and Imaizumi [28] proposed the fuzzy test for testing hypotheses with

vague data and the fuzzy test produced the acceptability of the null and alternative hypotheses. The statistical hypotheses testing for fuzzy data by proposing the notions of degrees of optimism and pessimism was proposed by Wu [31]. Viertl [26] investigated some methods to construct confidence intervals and statistical tests for fuzzy data. Wu [30] proposed some approaches to construct fuzzy confidence intervals for the unknown fuzzy parameter. Arefi and Taheri [5] developed an approach to test fuzzy hypotheses upon fuzzy test statistic for vague data. A new approach to the problem of testing statistical hypotheses is introduced by Chachi et al. [9]. Mikihiko Konishi et al. [22] proposed a method of ANOVA for the fuzzy interval data by using the concept of fuzzy sets. Hypothesis testing of one factor ANOVA model for fuzzy data was proposed by Wu [29] using the h-level set and the notions of pessimistic degree and optimistic degree by solving optimization problems. Dubois and Prade [12] defined any of the fuzzy numbers as a fuzzy subset of the real line. Chen and Chen [11] presented a method for ranking generalized trapezoidal fuzzy numbers. The symmetric triangular approximation was presented by Ma et al. [20]. Chanas [10] derived a formula for determining the interval approximations under the Hamming distance. The trapezoidal approximation was proposed by Abbasbandy et al. [1-3]. Grzegorzewski et al. [15] proposed the trapezoidal approximation of a fuzzy number, which is considered as a reasonable compromise between two opposite tendencies: to lose too much information and to introduce too sophisticated form of approximation from the point of view of computation.

In this paper, we propose a new statistical fuzzy hypothesis testing of ANOVA model for finding the significance among more than two population means when the data of their samples are in terms of trapezoidal fuzzy data. We provide the decision rules which are used to accept or reject the fuzzy null and alternative hypotheses. In the proposed technique, we convert the given fuzzy hypothesis testing of one factor ANOVA model with fuzzy data into two hypothesis testing of one factor ANOVA models with crisp data namely, upper level model and lower level model then, we test the hypothesis of each of the one factor ANOVA models with crisp data and obtain the results and then we obtain a decision about the population means on the basis of the proposed decision rules using the results obtained. In the decision rules of the proposed testing technique, we are not using degrees of optimism, pessimism and h-level set which are used in Wu [29]. In fact we would like to counter an argument that α-cut interval method is general enough to deal with one-factor ANOVA method under fuzzy environments which fits better when compared to the similar problems involved under non-fuzzy data. For better understanding, the proposed fuzzy hypothesis testing technique of ANOVA model for fuzzy data is illustrated with numerical examples.

2. Preliminaries

Definition 2. 1 Generalized fuzzy number

A generalized fuzzy number A is described as any fuzzy subset of the real line $\mathbb{R}$, whose membership function $\mu_A(x)$ satisfies the following conditions:

i. $\mu_A(x)$ is a continuous mapping from $\mathbb{R}$ to the closed interval $[0, \omega],\ 0 \le \omega \le 1$,

ii. $\mu_A(x) = 0$, for all $x \in (-\infty, a]$,

iii. $\mu_L(x) = L(x)$ is strictly increasing on $[a, b]$,

iv. $\mu_A(x) = \omega$, for all $[b, c]$, as ω is a constant and $0 < \omega \le 1$,

v. $\mu_R(x) = R(x)$ is strictly decreasing on $[c, d]$,

vi. $\mu_A(x) = 0$, for all $x \in [d, \infty)$.

where a, b, c, d are real numbers such that $a < b \le c < d$.

Throughout this paper, $\mathbb{R}$ stands for the set of all real numbers, $F(\mathbb{R})$ represents the set of fuzzy numbers, A expresses a fuzzy number and $A(x)$ its membership function $\forall x \in \mathbb{R}$.

Definition 2. 2

A fuzzy set A is called **normal** fuzzy set if there exists an element (member) 'x' such that $\mu_A(x) = 1$. A fuzzy set A is called **convex** fuzzy set if $\mu_A(\alpha x_1 + (1 - \alpha)x_2) \ge \min\{\mu_A(x_1), \mu_A(x_2)\}$ where $x_1, x_2 \in X$ and $\alpha \in [0, 1]$. The set $A_\alpha = \{x \in X / \mu_A(x) \ge \alpha\}$ is said to be the α - cut of a fuzzy set A.

Definition 2. 5

A fuzzy subset A of the real line $\mathbb{R}$ with **membership function** $\mu_A(x)$ such that $\mu_A(x): \mathbb{R} \to [0, 1]$, is called a fuzzy number if A is normal, A is fuzzy convex, $\mu_A(x)$ is upper semi-continuous and $\mathrm{Supp}(A)$ is bounded, where $\mathrm{Supp}(A) = \mathrm{cl}\{x \in \mathbb{R} : \mu_A(x) > 0\}$ and 'cl' is the closure operator.

It is known that for fuzzy number A, there exists four numbers $a, b, c, d \in \mathbb{R}$ and two functions $L_A(x), R_A(x): \mathbb{R} \to [0, 1]$, where $L_A(x)$ and $R_A(x)$ are non-decreasing and non-increasing functions respectively. Now, we can describe a membership function as follows:

$\mu_A(x) = L_A(x)$ for $a \le x \le b$; 1 for $b \le x \le c$; $R_A(x)$ for $c \le x \le d$; 0 otherwise. The functions $L_A(x)$ and $R_A(x)$ are also called the left and right side of the fuzzy number A respectively ([12, 13]).

In this paper, we assume that $\int_{-\infty}^{\infty} A(x)dx < +\infty$ and it is known that the α - cut of a fuzzy number is $A_\alpha = \{x \in \mathbb{R} / \mu_A(x) \ge \alpha\}$, for $\alpha \in (0, 1]$ and $A_0 = \mathrm{cl}\left(\bigcup_{\alpha \in (0, 1]} A_\alpha\right)$, according to the definition of a fuzzy number, it is seen at once that every α - cut of a fuzzy number is a closed interval. Hence, for a fuzzy number A, we have $A(\alpha) = [A_L(\alpha), A_U(\alpha)]$ where $A_L(\alpha) = \inf\{x \in \mathbb{R} : \mu_A(x) \ge \alpha\}$ and $A_U(\alpha) = \sup\{x \in \mathbb{R} : \mu_A(x) \ge \alpha\}$.

The left and right sides of the fuzzy number A are strictly monotone, obviously, A_L and A_U are inverse functions of $L_A(x)$ and $R_A(x)$ respectively.

Another important type of fuzzy number was introduced in [7] as follows:

Let $a, b, c, d \in \mathbb{R}$ such that $a < b \le c < d$. A fuzzy number A defined as $\mu_A(x): \mathbb{R} \to [0, 1]$,

$\mu_A(x) = \left(\frac{x - a}{b - a}\right)^n$ for $a \le x \le b$; 1 for $b \le x \le c$; $\left(\frac{d - x}{d - c}\right)^n$ for $c \le x \le d$; 0 otherwise. where $n > 0$, is denoted by $A = (a, b, c, d)_n$. And $L(x) = \left(\frac{x - a}{b - a}\right)^n$ and $R(x) = \left(\frac{d - x}{d - c}\right)^n$ can also be termed as left and right spread of the TFN [Dubois and Prade in 1981].

If $A = (a, b, c, d)_n$, then

$$A_\alpha = [A_L(\alpha), A_U(\alpha)] = \left[a + (b - a)\sqrt[n]{\alpha},\ d - (d - c)\sqrt[n]{\alpha}\right];\ \alpha \in [0, 1].$$

When $n = 1$ and $b = c$, we get a triangular fuzzy number. The conditions $r = 1, a = b$ and $c = d$ imply the closed interval and in the case $r = 1, a = b = c = d = t$ (some constant), we can get a crisp number 't'. Since a trapezoidal fuzzy number is completely characterized by $n = 1$ and four real numbers $a \le b \le c \le d$, it is often denoted as $A = (a, b, c, d)$. And the family of trapezoidal fuzzy numbers will be denoted by $F^T(\mathbb{R})$.

Now, for $n = 1$ we have a normal trapezoidal fuzzy number $A = (a, b, c, d)$ and the corresponding α - cut is defined by $A_\alpha = [a + \alpha(b - a),\ d - \alpha(d - c)];\ \alpha \in [0, 1]$. Now, we need the following results which can be found in [17, 19].

Result 2. 1

Let $D = \{[a, b],\ a \le b \text{ and } a, b \in \mathbb{R}\}$, the set of all closed, bounded intervals on the real line $\mathbb{R}$.

Result 2. 2

Let $A = [a, b]$ and $B = [c, d]$ be in D. Then $A = B$ if $a = c$ and $b = d$.

Result 2. 3

If s^2 is the variance of a sample of size 'n' drawn from the population with variance σ^2, then $E\left(\frac{ns^2}{n - 1}\right) = \sigma^2$, that is $\frac{ns^2}{n - 1}$ is an **unbiased estimator** of σ^2.

3. One-Factor ANOVA Model

The Analysis of Variance (ANOVA) is a powerful statistical tool for tests of significance. The term "Analysis of Variance" was introduced by Prof. R. A. Fisher in 1920's to deal with problems in the analysis of agronomical data. Variation is inherent in nature. The total variation in any set of numerical data is due to a number of causes which may be classified as (i) Assignable causes and (ii) Chance causes.

The variation due to assignable causes can be detected and measured whereas the variation due to chance is beyond the control of human hand and cannot be traced separately. In general, ANOVA studies mainly the homogeneity of populations by separating the total variance into its various components. That is, this technique is to test the difference among the means of populations by studying the amount of variation within each of the samples relative to the amount of variation between the samples. Samples under employing in ANOVA model are assumed to be drawn from 'normal populations of equal variances'. The variation of each value around its own grand mean should be independent for each value. A one-factor ANOVA is used when the analysis involves only one factor with more than two levels and different subjects in each of the experimental conditions.

Let a sample of N values of a given random variable X drawn from a normal population with variance σ^2 which is subdivided into 'h' classes according to some factor of classification with which the classes are homogeneous, that is, there is no difference between various classes.

Now, let μ_i be the mean of i^{th} population class. The test of hypotheses are: Null hypothesis: $H_0 : \mu_1 = \mu_2 = \cdots = \mu_h$ against Alternative hypothesis: $H_A : \mu_1 \neq \mu_2 \neq \cdots \neq \mu_h$.

Let x_{ij} be the value of the j^{th} member of the i^{th} class, which contains n_i members. Let the general mean of all the N values be $\bar{x}$ and the mean of n_i values in the i^{th} class be $\bar{x}_i$. Now,

$$\sum_i \sum_j \left(x_{ij} - \bar{x}\right)^2 = \sum_i \sum_j \left\{\left(x_{ij} - \bar{x}\right) + \left(x_i - \bar{x}\right)\right\}^2 = \sum_i \sum_j \left(x_{ij} - \bar{x}_i\right)^2 + \sum_i n_i \left(\bar{x}_i - \bar{x}\right)^2$$
$$= Q_2 + Q_1$$

where $Q_1 = \sum_i n_i \left(\bar{x}_i - \bar{x}\right)^2$ is the sum of the squared deviations of class means from the general mean (variation between classes) and $Q_2 = \sum_i \sum_j \left(x_{ij} - \bar{x}_i\right)^2$ is the sum of the squared deviations of variates from the corresponding class means (variation within classes). Q is total variation.

Now, it is known from the theory of estimation that $\left(\frac{ns^2}{n-1}\right)$ is an unbiased estimate of σ^2, where s^2 is the variance of a sample of size 'n' drawn from a population with variance σ^2. That is, $E\left(ns^2 / n-1\right) = \sigma^2$. Since the items in the i^{th} class with variance $\frac{1}{n_i}\sum_{j=1}^{n_i}\left(x_{ij} - \bar{x}_i\right)^2$ may be considered as a sample of size n_i drawn from a population with variance σ^2. That is, $E\left\{\frac{n_i}{n_i - 1} \cdot \frac{1}{n_i}\sum_{j=1}^{n_i}\left(x_{ij} - \bar{x}_i\right)^2\right\} = \sigma^2$.

$$\text{i.e. } E\left[\sum_i \sum_j \left(x_{ij} - \bar{x}_i\right)^2\right] = \sum_{i=1}^{h}\left(n_i - 1\right)\sigma^2 \text{ i.e. } E\left[Q_2\right] = \left(N - h\right) \text{ i.e. } \sigma^2 E\left\{\frac{Q_2}{N - h}\right\} = \sigma^2$$

Hence, $\frac{Q_2}{N-h}$ is an **unbiased estimate** of σ^2 with $(N-h)$ degrees of freedom. Let us consider the entire group of N items with variance $\frac{1}{N}\sum_i \sum_j \left(x_{ij} - \bar{x}\right)^2$ as the sample of size N drawn from the same population. Now, $E\left\{\frac{N}{N-1} \cdot \frac{1}{N}\sum_i \sum_j \left(x_{ij} - \bar{x}\right)^2\right\} = \sigma^2$. That is, $E\left[\frac{Q}{N-1}\right] = \sigma^2$, this states that $\frac{Q}{N-1}$ is an unbiased estimate of σ^2 with $(N-1)$ degrees of freedom. Now,

$$E\left(Q_1\right) = E\left(Q\right) - E\left(Q_2\right) = \left(N-1\right)\sigma^2 - \left(N-h\right)\sigma^2 \Rightarrow E\left(\frac{Q_1}{h-1}\right) = \sigma^2.$$

Thus, $\frac{Q_1}{h-1}$ is also an **unbiased estimate** of σ^2 with $(h-1)$ degrees of freedom.

If we assume that the sample drawn from a normal population, then the estimates $\frac{Q_1}{h-1}$ and $\frac{Q_2}{N-h}$ are independent and hence the ratio $\frac{Q_1/(h-1)}{Q_2/(N-h)}$ follows F-distribution with $(h-1, N-h)$ degrees of freedom.

Choosing the ratio which is greater than one, we employ the F-test. For simplicity, let us choose, $M_1 = \frac{Q_1}{h-1}$ and $M_2 = \frac{Q_2}{N-h}$.

Aggregating the above results, the ANOVA table for one factor classification is given below ([16, 25]):

Source of Variation (S.V.)	Sum of Squares (S.S.)	Degrees of freedom (d.f.)	Mean Square (M.S.)	Variance Ratio (F-value)
Between Classes	Q_1	h - 1	$M_1 = \frac{Q_1}{(h-1)}$	$F = \left(\frac{M_1}{M_2}\right)^{\pm 1}$
Within Classes	Q_2	N - h	$M_2 = \frac{Q_2}{(N-h)}$	
Total	Q	N - 1	--	

The decision rules of F-test are given below:

(i) If $M_2 < M_1$ and $F = \frac{M_1}{M_2} < F_t$ where F_t is the tabulated value of F with $(h-1,\ N-h)$ degrees of freedom at 'k' level of significance, then we accept the null hypothesis H_0, otherwise the alternative hypothesis H_A is accepted.

(ii) If $M_1 < M_2$ and $F = \frac{M_2}{M_1} < F_t$ where F_t is the tabulated value of F with $(N-h,\ h-1)$ degrees of freedom at 'k' level of significance, then we accept the null hypothesis H_0, otherwise the alternative hypothesis H_A is accepted.

Note that here we use the notation for level of significance is to be "k" instead of "α" so as to avoid confusion with 'α - cut' value that can be seen in trapezoidal fuzzy numbers (TFN). For simplicity of calculations, the following formulae for Q, Q_1 and Q_2 are used:

$$Q = \sum_i \sum_j x_{ij}^2 - \frac{T^2}{N} \quad \text{where} \quad T = \sum_i \sum_j x_{ij}\ ; \qquad Q_1 = \sum_i \left(\frac{T_i^2}{n_i}\right) - \frac{T^2}{N} \quad \text{where} \quad T_i = \sum_j x_{ij} \quad \text{and}$$

$$Q_2 = Q - Q_1.$$

4. One-Factor ANOVA model with TFNs using α - cut method:

The fuzzy test of hypotheses of one-factor ANOVA model where the sample data are trapezoidal fuzzy numbers is proposed here. Using the relation, we transform the fuzzy ANOVA model to interval ANOVA model. Fetching the upper limit of the fuzzy interval, we construct upper level crisp ANOVA model and considering the lower limit of the fuzzy interval, we construct the lower level crisp ANOVA model. Thus, in this proposed approach, two crisp ANOVA models are designated in terms of upper and lower levels. Finally, we analyse lower level and upper level model using crisp one-factor ANOVA technique.

Let there be N values of samples for a given random variables 'X' which are subdivided into 'h' classes according to some kind of classification. Then the lower level data and upper level data for given trapezoidal fuzzy numbers using α - cut method can be assigned as follows:

Lower level data:

$a_{11}+\alpha(b_{11}-a_{11})$	$a_{12}+\alpha(b_{12}-a_{12})$	...	$a_{1j}+\alpha(b_{1j}-a_{1j})$
$a_{21}+\alpha(b_{21}-a_{21})$	$a_{22}+\alpha(b_{22}-a_{22})$	...	$a_{2j}+\alpha(b_{2j}-a_{2j})$
...			
$a_{i1}+\alpha(b_{i1}-a_{i1})$	$a_{i2}+\alpha(b_{i2}-a_{i2})$	...	$a_{ij}+\alpha(b_{ij}-a_{ij})$
where $0\le i\le h,\ 0\le j\le n$			

Upper level data:

$d_{11}-\alpha(d_{11}-c_{11})$	$d_{12}-\alpha(d_{12}-c_{12})$	...	$d_{1j}-\alpha(d_{1j}-c_{1j})$
$d_{21}-\alpha(d_{21}-c_{21})$	$d_{22}-\alpha(d_{22}-c_{22})$	...	$d_{2j}-\alpha(d_{2j}-c_{2j})$
...			
$d_{i1}-\alpha(d_{i1}-c_{i1})$	$d_{i2}-\alpha(d_{i2}-c_{i2})$	...	$d_{ij}-\alpha(d_{ij}-c_{ij})$
where $0\le i\le h,\ 0\le j\le n$			

The one-factor ANOVA formulae using α - cut can be tabulated as follows:

Lower level model	Upper level model
$Q^L=\sum_i\sum_j\left[a_{ij}+\alpha(b_{ij}-a_{ij})\right]^2-\frac{T^2}{N}$ where $0\le i\le h,\ 0\le j\le n$. $T_i=\sum_j\left[a_{ij}+\alpha(b_{ij}-a_{ij})\right];\ i=1,2,\ldots,h$ And $T=\sum_{r=1}^{h}T_r$, $Q_1^L=\sum_i\frac{T_i^2}{n_i}-\frac{T^2}{N}$ $Q_2^L=Q-Q_1^L$	$Q^U=\sum_i\sum_j\left[d_{ij}-\alpha(d_{ij}-c_{ij})\right]^2-\frac{T^2}{N}$ where $0\le i\le h,\ 0\le j\le n$. $T_i=\sum_j\left[d_{ij}-\alpha(d_{ij}-c_{ij})\right];\ i=1,2,\ldots,h$ And $T=\sum_{r=1}^{h}T_r$, $Q_1^U=\sum_i\frac{T_i^2}{n_i}-\frac{T^2}{N}$ $Q_2^U=Q-Q_1^U$

Let ‘k’ be the level of significance.

Now, the null hypothesis: $H_0:\ \mu_1=\mu_2=\cdots=\mu_h$ against the alternative hypothesis:

$$H_A:\ \mu_1\ne\mu_2\ne\cdots\ne\mu_h$$

$$\Rightarrow\quad [H_0]:[\mu_1]=[\mu_2]=\cdots=[\mu_h]\ \text{ against }\ [H_A]:\ [\mu_1]\ne[\mu_2]\ne\cdots\ne[\mu_h].$$

$$\Rightarrow\quad [H_0^L,\ H_0^U]:\ [\mu_1^L,\ \mu_1^U]=[\mu_2^L,\ \mu_2^U]=\cdots=[\mu_h^L,\ \mu_h^U]\ \text{ against}$$

$$[H_A^L,\ H_A^U]:\ [\mu_1^L,\ \mu_1^U]\ne[\mu_2^L,\ \mu_2^U]\ne\cdots\ne[\mu_h^L,\ \mu_h^U]$$

$\Rightarrow$ The following two set of hypotheses can be obtained.

(i) The null hypothesis $H_0^L:\ \mu_1^L=\mu_2^L=\cdots=\mu_h^L$ against the alternative hypothesis $H_A^L:\mu_1^L\ne\mu_2^L\ne\cdots\ne\mu_h^L$.

(ii) The null hypothesis $H_0^U: \mu_1^U = \mu_2^U = \cdots = \mu_h^U$ against the alternative hypothesis $H_A^U: \mu_1^U \neq \mu_2^U \neq \cdots \neq \mu_h^U$.

Decision rules:

(i) If $F^L < F_t$ at 'k' level of significance with $(N - h,\ h - 1)$ degrees of freedom then the null hypothesis H_0^L is accepted for certain value of $\alpha \in [0, 1]$, otherwise the alternative hypothesis H_A^L is accepted.

(ii) If $F^U < F_t$ at 'k' level of significance with $(N - h,\ h - 1)$ degrees of freedom then the null hypothesis H_0^U is accepted for certain value of $\alpha \in [0, 1]$, otherwise the alternative hypothesis H_A^U is accepted.

Example 4. 1

A food company wished to test four different package designs for a new product. Ten stores with approximately equal sales volumes are selected as the experimental units. Package designs 1 and 4 are assigned to three stores each and package designs 2 and 3 are assigned to two stores each. We cannot record the exact sales volume in a store due to some unexpected situations, but we have the fuzzy data for sales volumes. The fuzzy data are given below [29]:

Package design (i)	**Store (Observation j)**		
	1	2	3
1	$(9,\ 10,\ 12,\ 13)$	$(14,\ 15,\ 17,\ 18)$	--
2	$(11,\ 13,\ 16,\ 19)$	$(10,\ 14,\ 16,\ 20)$	$(11,\ 12,\ 14,\ 15)$
3	$(15,\ 17,\ 19,\ 21)$	$(14,\ 16,\ 19,\ 20)$	$(17,\ 20,\ 21,\ 23)$
4	$(15,\ 18,\ 21,\ 23)$	$(21,\ 23,\ 25,\ 27)$	--

We test the hypothesis whether the fuzzy mean sales are same for four designs of package or not. Let μ_i be the mean sales for the i^{th} design. Then the null hypothesis $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$ against the alternative hypothesis $H_A: \mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4$.

Now, the interval model for the given trapezoidal fuzzy number using α - cut method is:

Package design (i)	**Store (Observation j)**		
	1	2	3
1	$[9 + \alpha,\ 13 - \alpha]$	$[14 + \alpha,\ 18 - \alpha]$	--
2	$[11 + 2\alpha,\ 19 - 3\alpha]$	$[10 + 4\alpha,\ 20 - 4\alpha]$	$[11 + \alpha,\ 15 - \alpha]$
3	$[15 + 2\alpha,\ 21 - 2\alpha]$	$[14 + 2\alpha,\ 20 - \alpha]$	$[17 + 3\alpha,\ 23 - 2\alpha]$
4	$[15 + 3\alpha,\ 23 - 2\alpha]$	$[21 + 2\alpha,\ 27 - 2\alpha]$	--

Now, the ANOVA tables for "lower level α - cut interval" and "upper level α - cut interval" are given below:

Lower level model:

Package design (i)	Store (Observation j)		
	1	2	3
1	$[9+\alpha]$	$[14+\alpha]$	--
2	$[11+2\alpha]$	$[10+4\alpha]$	$[11+\alpha]$
3	$[15+2\alpha]$	$[14+2\alpha]$	$[17+3\alpha]$
4	$[15+3\alpha]$	$[21+2\alpha]$	--

The null hypothesis $H_0^L:\ \mu_1^L=\mu_2^L=\mu_3^L=\mu_4^L$ against the alternative hypothesis $H_A^L:\mu_1^L\neq\mu_2^L\neq\mu_3^L\neq\mu_4^L$.

Upper level model:

Package design (i)	Store (Observation j)		
	1	2	3
1	$[13-\alpha]$	$[18-\alpha]$	--
2	$[19-3\alpha]$	$[20-4\alpha]$	$[15-\alpha]$
3	$[21-2\alpha]$	$[20-\alpha]$	$[23-2\alpha]$
4	$[23-2\alpha]$	$[27-2\alpha]$	--

The null hypothesis $H_0^U:\ \mu_1^U=\mu_2^U=\mu_3^U=\mu_4^U$ against the alternative hypothesis $H_A^U:\mu_1^U\neq\mu_2^U\neq\mu_3^U\neq\mu_4^U$

The ANOVA table for lower level model:

Source of Variance (S.V.)	Sum of Squares (S.S.)	Degrees of freedom (d.f.)	Mean Square (M.S.)	F-ratio (F_C^L)
Between Classes	Q_1^L	$h-1=(4-1)=3$	$M_1^L=\frac{Q_1^L}{3}$	$F_C^L=\frac{M_1^L}{M_2^L}$
Within Classes	Q_2^L	$N-h=(10-4)=6$	$M_2^L=\frac{Q_2^L}{6}$	

Here, $N=10$ and $n_i=\ 2,3,3,2$ for the package designs 1, 2, 3, 4 respectively.

$$T=137+21\alpha;\ \sum_i\frac{T_i^2}{n_i}=\frac{1}{6}\left[283\alpha^2+3540\alpha+11755\right]$$ and

$$\sum_i\sum_j\left[a_{ij}+\alpha\left(b_{ij}-a_{ij}\right)\right]^2=53\alpha^2+584\alpha+1995\ Q^L=\frac{1}{10}\left[89\alpha^2+86\alpha+1181\right];$$

$$Q_1^L=\frac{1}{60}\left[184\alpha^2+876\alpha+4936\right]$$ and

$Q_2^L = \frac{1}{60}\left[350\alpha^2 - 360\alpha + 2150\right]$.And $M_1^L = \frac{1}{180}\left[184\alpha^2 + 876\alpha + 4936\right]$;

$M_2^L = \frac{1}{360}\left[350\alpha^2 - 360\alpha + 2150\right]$ and $F_C^L = \left(\frac{4}{5}\right)\left[\frac{46\alpha^2 + 219\alpha + 1234}{35\alpha^2 - 36\alpha + 215}\right]$ where $0 \le \alpha \le 1$ and F_C^L is the calculated value of 'F' at lower level model. Now, the tabulated value of 'F' at k = 5% level of significance with $(h - 1, N - h) = (3, 6)$ degrees of freedom is $F_{t(at\ 5\%)} = 4.76$. Here, $F_C^L < F_t$ at $\alpha = 0.1$ and $F_C^L > F_t$ for $0.2 \le \alpha \le 1$.

Hence, the null hypothesis H_0^L is rejected at 5% level of significance for $0.2 \le \alpha \le 1$.

The ANOVA table for upper level model:

Source of Variance (S.V.)	Sum of Squares (S.S.)	Degrees of freedom (d.f.)	Mean Square (M.S.)	F-ratio (F_C^U)
Between Classes	Q_1^U	$h - 1 = (4-1) = 3$	$M_1^U = \frac{Q_1^U}{3}$	$F_C^U = \frac{M_1^U}{M_2^U}$
Within Classes	Q_2^U	$N - h = (10-4) = 6$	$M_2^U = \frac{Q_2^U}{6}$	

Here, $N = 10$ and $n_i = 2, 3, 3, 2$ for the package designs 1, 2, 3, 4 respectively.

$T = 199 - 19\alpha$; $\sum_i \frac{T_i^2}{n_i} = \frac{1}{6}\left[238\alpha^2 - 4580\alpha + 24407\right]$ and

$$\sum_i \sum_j \left[d_{ij} - \alpha\left(d_{ij} - c_{ij}\right)\right]^2 = 45\alpha^2 - 782\alpha + 4107$$

$Q^U = \frac{1}{10}\left[89\alpha^2 - 258\alpha + 1469\right]$; $Q_1^U = \frac{1}{30}\left[107\alpha^2 - 214\alpha + 3232\right]$ and

$Q_2^U = \frac{1}{6}\left[32\alpha^2 - 112\alpha + 235\right]$. And $M_1^U = \frac{1}{90}\left[107\alpha^2 - 214\alpha + 3232\right]$;

$M_2^U = \frac{1}{36}\left[32\alpha^2 - 112\alpha + 235\right]$ and $F_C^U = \left(\frac{2}{5}\right)\left[\frac{107\alpha^2 - 214\alpha + 3232}{32\alpha^2 - 112\alpha + 235}\right]$ where $0 \le \alpha \le 1$ and F_C^U is the calculated value of 'F' at upper level model. Now, the tabulated value of 'F' at k = 5% level of significance with $(h - 1, N - h) = (3, 6)$ degrees of freedom is $F_{t(at\ 5\%)} = 4.76$. Here, $F_C^U > F_t$ for all α where $0 \le \alpha \le 1$.

Hence we reject the null hypothesis H_0^U at 5% level of significance for all α $(0 \le \alpha \le 1)$. Thus, the rejection level of null hypotheses for lower and upper level data are given below:

H_0^L is rejected for all α; $0.2 \le \alpha \le 1$ and H_0^U is rejected for all α; $0 \le \alpha \le 1$.

Therefore, we accept the alternative hypothesis H_A of the fuzzy ANOVA model.

Conclusion 4. 1

The factor level fuzzy means μ_i are not equal. Hence, we conclude that there is a relation between package design and sales volumes.

Remark 4. 1

In this proposed method, the notions of pessimistic degree and optimistic degree are not used. The whole calculation technique is fully based on α - cut interval method [4]. And the decision obtained in the proposed fuzzy hypothesis testing using α - cut interval ANOVA method for example-1 fits better when compared with Wu [29].

Example 4. 2

In order to determine whether there is significant difference in the durability of 3 makes of computers, samples of size 5 are selected from each make and the frequency of repair during the first year of purchase is observed. The results are obtained in terms of fuzzy data due to different kinds of maintenance and usage. The results are as follows:

Makes		
A	**B**	**C**
$(3, 5, 7, 8)$	$(6, 8, 10, 13)$	$(4, 6, 8, 9)$
$(4, 6, 9, 10)$	$(8, 9, 11, 12)$	$(2, 4, 5, 7)$
$(6, 8, 10, 11)$	$(9, 11, 13, 15)$	$(2, 5, 7, 9)$
$(8, 10, 12, 14)$	$(9, 12, 14, 15)$	$(2, 5, 8, 10)$
$(5, 7, 9, 12)$	$(2, 4, 6, 9)$	$(1, 2, 4, 7)$

In view of the above data, the testing procedure is proposed to check "is there any significant difference in the durability of the 3 makes of computers?"

We test the hypothesis whether the fuzzy means of the 3 makes of computers differ or not.

Here, the null hypothesis $H_0 : \mu_1 = \mu_2 = \mu_3$ against the alternative hypothesis $H_A : \mu_1 \neq \mu_2 \neq \mu_3$.

Now, the ANOVA model using α - cut interval method for given fuzzy data is tabulated below:

Make	Sample (Observation j)				
	1	**2**	**3**	**4**	**5**
A	$[3 + 2\alpha, 8 - \alpha]$	$[4 + 2\alpha, 10 - \alpha]$	$[6 + 2\alpha, 11 - \alpha]$	$[8 + 2\alpha, 14 - 2\alpha]$	$[5 + 2\alpha, 12 - 3\alpha]$
B	$[6 + 2\alpha, 13 - 3\alpha]$	$[8 + \alpha, 12 - \alpha]$	$[9 + 2\alpha, 15 - 2\alpha]$	$[9 + 3\alpha, 15 - \alpha]$	$[2 + 2\alpha, 9 - 3\alpha]$
C	$[4 + 2\alpha, 9 - \alpha]$	$[2 + 2\alpha, 7 - 2\alpha]$	$[2 + 3\alpha, 9 - 2\alpha]$	$[2 + 3\alpha, 10 - 2\alpha]$	$[1 + \alpha, 7 - 3\alpha]$

The ANOVA tables for "Lower level α - cut interval" and "Upper level α - cut interval" are given below:

Lower level α - cut interval:

Make	Sample (Observation j)				
	1	**2**	**3**	**4**	**5**
A	$[3 + 2\alpha]$	$[4 + 2\alpha]$	$[6 + 2\alpha]$	$[8 + 2\alpha]$	$[5 + 2\alpha]$
B	$[6 + 2\alpha]$	$[8 + \alpha]$	$[9 + 2\alpha]$	$[9 + 3\alpha]$	$[2 + 2\alpha]$
C	$[4 + 2\alpha]$	$[2 + 2\alpha]$	$[2 + 3\alpha]$	$[2 + 3\alpha]$	$[1 + \alpha]$

The null hypothesis $H_0^L: \mu_1^L = \mu_2^L = \mu_3^L$ against the alternative hypothesis $H_A^L: \mu_1^L \neq \mu_2^L \neq \mu_3^L$.

Upper level α - cut interval:

Make	Sample (Observation j)				
	1	2	3	4	5
A	$[8 - \alpha]$	$[10 - \alpha]$	$[11 - \alpha]$	$[14 - 2\alpha]$	$[12 - 3\alpha]$
B	$[13 - 3\alpha]$	$[12 - \alpha]$	$[15 - 2\alpha]$	$[15 - \alpha]$	$[9 - 3\alpha]$
C	$[9 - \alpha]$	$[7 - 2\alpha]$	$[9 - 2\alpha]$	$[10 - 2\alpha]$	$[7 - 3\alpha]$

The null hypothesis $H_0^U: \mu_1^U = \mu_2^U = \mu_3^U$ against the alternative hypothesis $H_A^U: \mu_1^U \neq \mu_2^U \neq \mu_3^U$

The ANOVA table for lower level model:

S.V.	S.S.	d.f.	M.S.	F-ratio (F_C^L)
Between Classes	Q_1^L	$h - 1 = (3-1) = 2$	$M_1^L = \frac{Q_1^L}{2}$	$F_C^L = \frac{M_1^L}{M_2^L}$
Within Classes	Q_2^L	$N - h = (15-3) = 12$	$M_2^L = \frac{Q_2^L}{12}$	

Here, $N = 15$ and $n_i = 5, 5, 5$ for the makes A, B, C respectively.

$$T = 71 + 31\alpha; \quad \sum_i \frac{T_i^2}{n_i} = \frac{1}{5}\left[321\alpha^2 + 1442\alpha + 1953\right] \text{ and}$$

$$\sum_i \sum_j \left[a_{ij} + \alpha\left(b_{ij} - a_{ij}\right)\right]^2 = 69\alpha^2 + 292\alpha + 445$$

$$Q^L = \frac{1}{15}\left[74\alpha^2 - 22\alpha + 1634\right]; \quad Q_1^L = \frac{1}{15}\left[2\alpha^2 - 76\alpha + 818\right] \text{and}$$

$$Q_2^L = \frac{1}{5}\left[24\alpha^2 + 18\alpha + 272\right]. \text{ And } \quad M_1^L = \frac{1}{30}\left[2\alpha^2 - 76\alpha + 818\right];$$

$M_2^L = \frac{1}{60}\left[24\alpha^2 + 18\alpha + 272\right]$ and $F_C^L = \left[\frac{2\alpha^2 - 76\alpha + 818}{12\alpha^2 + 9\alpha + 136}\right]$ where $0 \le \alpha \le 1$ and F_C^L is the calculated value of 'F' at lower level model. Now, the tabulated value of 'F' at k = 5% level of significance with $(h - 1, N - h) = (2, 12)$ degrees of freedom is $F_{t(at\ 5\%)} = 3.88$. Since, $F_C^L > F_{t(at\ 5\%)}$ $\forall \alpha, (0 \le \alpha \le 1)$, we reject the null hypothesis H_0^L.

$\Rightarrow$ **There is a significant difference in the durability of the 3 makes of computers at lower level of α - cut .**

The ANOVA table for upper level model:

S.V.	S.S.	d.f.	M.S.	F-ratio $\left(F_C^U\right)$
Between Classes	Q_1^U	$h - 1 = (3-1) = 2$	$M_1^U = \frac{Q_1^U}{2}$	$F_C^U = \frac{M_1^U}{M_2^U}$
Within Classes	Q_2^U	$N - h = (15-3) = 12$	$M_2^U = \frac{Q_2^U}{12}$	

Here, $N = 15$ and $n_i = 5, 5, 5$ for the makes A, B, C respectively.

$T = 161 - 28\alpha$; $\sum_i \frac{T_i^2}{n_i} = \frac{1}{5}\left[264\alpha^2 - 3000\alpha + 8885\right]$ and

$$\sum_i \sum_j \left[d_{ij} - \alpha\left(d_{ij} - c_{ij}\right)\right]^2 = 62\alpha^2 - 596\alpha + 1829$$

$Q^U = \frac{1}{15}\left[146\alpha^2 + 76\alpha + 1514\right]$; $Q_1^U = \frac{1}{15}\left[8\alpha^2 + 16\alpha + 734\right]$ and

$Q_2^U = \frac{1}{5}\left[46\alpha^2 + 20\alpha + 260\right]$. And $M_1^U = \frac{1}{30}\left[8\alpha^2 + 16\alpha + 734\right]$;

$M_2^U = \frac{1}{60}\left[46\alpha^2 + 20\alpha + 260\right]$ and $F_C^U = \left[\frac{8\alpha^2 + 16\alpha + 734}{23\alpha^2 + 10\alpha + 130}\right]$ where $0 \le \alpha \le 1$ and F_C^U is the calculated value of 'F' at upper level model. And the tabulated value of 'F' at k = 5% level of significance with $(h - 1, N - h) = (2, 12)$ degrees of freedom is $F_{t(\text{at } 5\%)} = 3.88$. Here, $F_C^U > F_{t(\text{at } 5\%)}$ $\forall\alpha, (0 \le \alpha \le 1)$, we reject the null hypothesis H_0^U.

$\Rightarrow$ **There is a significant difference in the durability of the 3 makes of computers at upper level of** α - cut .

Conclusion 4. 2

Therefore, the null hypotheses H_0^L and H_0^U are rejected $\forall\alpha, (0 \le \alpha \le 1)$. We conclude in general that there is a significant difference between in the durability of the 3 makes of computers.

References

[1] S. Abbasbandy, B. Asady, The nearest trapezoidal fuzzy number to a fuzzy quantity, Appl. Math. Comput. 156 (2004) 381-386.
[2] S. Abbasbandy, M. Amirfakhrian, The nearest approximation of a fuzzy quantity in parametric form, Appl. Math. Comput. 172 (2006) 624-632.
[3] S. Abbasbandy, M. Amirfakhrian, The nearest trapezoidal form of generalized left right fuzzy number, Internat. J. Approx. Reason. 43 (2006) 166-178.
[4] S. Abbasbandy, T. Hajjari, Weighted trapezoidal approximation-preserving cores of a fuzzy number, Computers and Mathematics with Applications 59 (2010) 3066-3077.
[5] M. Arefi and S. M. Taheri, Testing fuzzy hypotheses using fuzzy data based on fuzzy test statistic, Journal of Uncertain Systems, Vol. 5, 2011, 45-61.
[6] B. F. Arnold, Testing fuzzy hypotheses with crisp data, Fuzzy Sets and Systems, Vol. 94, 1998, 323-333.

[7] S. Bodjanova, Median value and median interval of a fuzzy number, Inform. Sci. 172 (2005) 73-89.
[8] J. J. Buckley, Fuzzy statistics, Springer-Verlag, New York, 2005.
[9] J. Chachi, S. M. Taheri and R. Viertl, Testing statistical hypotheses based on fuzzy confidence intervals, Forschungsbericht SM-2012-2, Technische Universitat Wien, Austria, 2012.
[10] S. Chanas, On the interval approximation of a fuzzy number, Fuzzy Sets and Systems 102 (1999) 221-226.
[11] S. Chen and G. Li, Representation, ranking, and distance of fuzzy number with exponential membership function using graded mean integration method, Tamsui, Oxf. J. Math. Sci. 16 (2) (2000) 123-131.
[12] D. Dubois and H. Prade, Operations on fuzzy numbers, Int. J. Syst. Sci. 9 (1978) 613-626.
[13] D. Dubois and H. Prade, Fuzzy sets and systems: Theory and application, Academic Press, New York, 1980.
[14] P. Grzegorzewski, Testing statistical hypotheses with vague data, Fuzzy Sets and Systems, Vol. 112, 2000, 501-510.
[15] P. Grzgorzewski, E. Mrowka, Trapezoidal approximation of fuzzy numbers, Fuzzy Sets and Systems 153 (2005) 115-135.
[16] S. C. Gupta, V. K. Kapoor, Fundamentals of mathematical statistics, Sultan Chand & Sons, New Delhi, India.
[17] R. R. Hocking, Methods and applications of linear models: regression and the analysis of variance, New York: John Wiley & Sons, 1996.
[18] D. Kalpanapriya, P. Pandian, Fuzzy hypothesis testing of ANOVA model with fuzzy data, Int. Nat. J. Modern Engg. Research, Vol. 2, Issue 4 July-Aug. (2012) 2951-2956.
[19] George J. Klir and Bo Yuan, Fuzzy sets and fuzzy logic, Theory and Applications, Prentice-Hall, New Jersey, 2008.
[20] M. Ma, A. Kandel, M. Friedman, A new approach for defuzzification, Fuzy Sets and Systems 111 (2000) 351-356.
[21] K. G. Manton, M. A. Woodbury and H. D. Tolley, Statistical applications using fuzzy sets, New York: John Wiley & Sons, 1994.
[22] Mikihiko Konishi, Tetsuji Okuda and Kiyoji Asai, Analysis of variance based on fuzzy interval data using moment correction method, International Journal of Innovative Computing, Information and Control, Vol. 2, 2006, 83-99.
[23] J. J. Saade and H. Schwarzlander, Fuzzy hypothesis testing with hybrid data, Fuzzy Sets and Systems, Vol. 35, 1990, 197-212.
[24] J. J. Saade, Extension of fuzzy hypothesis testing with hybrid data, Fuzzy Sets and Systems, Vol. 63, 1994, 57-71.
[25] T. Veerarajan, Probability, statistics and random process, Tata McGraw Hill Education Pvt. Ltd., New Delhi, India.
[26] R. Viertl, Univariate statistical analysis with fuzzy data, Computational Statistics and Data Analysis, Vol. 51, 2006, 33-147.
[27] R. Viertl, Statistical methods for fuzzy data, John Wiley and Sons, Chichester, 2011.
[28] N. Watanabe and T. Imaizumi, A fuzzy statistical test of fuzzy hypotheses, Fuzzy Sets and Systems, Vol. 53, 1993, 167-178.
[29] H. C. Wu, Analysis of variance for fuzzy data, International Journal of Systems Science, Vol. 38, 2007, 235-246.
[30] H. C. Wu, Statistical confidence intervals for fuzzy data, Expert Systems with Applications, Vol. 36, 2009, 2670-2676.
[31] H. C. Wu, Statistical hypotheses testing for fuzzy data, Information Sciences, Vol. 175, 2005, 30-56.
[32] L. A. Zadeh, Fuzzy sets, Information and Control 8 (1965) 338-353.