Journal of Mahani Mathematical Research Center
Print ISSN: 2251-7952 Online ISSN: 2645-4505

AN IMPLICATION OF FUZZY ANOVA IN VEHICLE BATTERY MANUFACTURING

A. Parchami*, Mashallah Mashinchi, and C. Kahraman

Dedicated to sincere professor Mehdi Radjabalipour on turning 75

Article type: Research Article
(Received: 20 March 2021, Revised: 15 July 2021, Accepted: 20 September 2021)
(Communicated by N. Shajareh Poursalavati)

Abstract. Analysis of variance (ANOVA) is an important method in exploratory and confirmatory data analysis when explanatory variables are discrete and response variables are continues and independent from each other. The simplest type of ANOVA is one-way analysis of variance for comparison among means of several populations. In this paper, we extend one-way analysis of variance to a case where observed data are non-symmetric triangular or normal fuzzy observations rather than real numbers. Meanwhile, a case study on the car battery length-life is provided on the basis on the proposed method.

Keywords: Fuzzy decision, Non-symmetric fuzzy data, Arithmetic fuzzy numbers, Analysis of variance.
2020 MSC: 62A86.

1. The place and importance of fuzzy ANOVA in manufacturing and literature

Analysis of variance (ANOVA) is concerned with analyzing variation in the means of several independent populations. The central point in classical ANOVA is a test about the significance of the difference among population means, which allows us to conclude whether or not the differences among the means of several populations are too deviated to be attributed to the sampling error [6]. There are several real-life populations in which imprecise values can be assigned to their experimental outcomes. In such cases, the fuzzy numbers are suitable models to formalize and handle these populations, such as the monthly income of a taxi driver, the desirability amount of life and the battery lifetime. The Fuzzy ANOVA (FANOVA) model can be realistically used in numerous industries and applications in which the data observed as fuzzy numbers. To show the importance of FANOVA, we present some applied examples as follows:

*Corresponding author, ORCID: 0000-0002-0593-7324
E-mail: parchami@uk.ac.ir
DOI: 10.22103/jmmrc.2021.17325.1137

How to cite: A. Parchami, M. Mashinchi and C. Kahraman, *An Implication of Fuzzy ANOVA in Vehicle Battery Manufacturing*, J. Mahani Math. Res. Cent. 2021; 10(2): 33-47.

(1) Comparing the amount of gas usage for the different vehicles, or the same vehicle under different fuel types, or road types,
(2) Comparison the quality of the produced product under different conditions (such as different weathers, different personnel, different material requirements, assemble lines in a workshop/faculty,
(3) Investigation on the pressure or chemical concentration on some chemical reaction (power reactors, chemical plants, etc),
(4) The amount of interest/satisfaction of the workers in different parts of a factory,
(5) Studying whether the different kinds of product advertisements can cause to different selling amounts, and
(6) Comparison the amount of exited gas volume from the volcano crater (per hour, or per day) for different active volcanoes.

Montenegro et al. [11] have presented an exact one-way ANOVA testing procedure for the case in which the involved fuzzy random variables are assumed normal as intended by Puri and Ralescu [19]. On the other hand, Montenegro et al. [10] demonstrated the convenience of incorporating bootstrap method to approximate the asymptotic one-sample tests with fuzzy random variables. A one-way ANOVA study has been developed by Cuevas et al. [2] for the functional data on a given Hilbert space. An introduction to the asymptotic multi-sample testing of means for simple fuzzy random variables is also was also sketched by Gil et al. [6], where a bootstrap approach to the multi-sample test of means for the significance of the difference among population means on the basis of the evidence supplied by a set of sample fuzzy data is studied. Also one-way ANOVA with fuzzy data was studied by Wu [23], where the cuts of fuzzy random variables, optimistic, pessimistic degrees and solving an optimization problem are used. Lee et al. [9] discussed on the analysis of variance with fuzzy data based on permutation method as a non-parametric approach.

In this paper a simple method for a fuzzy one-way ANOVA, as an extension for classical ANOVA, is presented where the observations are non-symmetric triangular and normal fuzzy numbers. The organization of this paper is as follows: In section 2, preliminaries on fuzzy concepts and some arithmetic operations are stated. In the section 3, fuzzy ANOVA is explained for fuzzy data. Meanwhile in section 3, the decision rule for testing hypothesis of equality of means of populations data are discussed for FANOVA model, where the observations are reported by non-symmetric triangular and normal fuzzy. In section 4 a real applied example on vehicle battery manufacturing is presented to reveal the ideas of this paper. Conclusion part is given in the final section.

2. Preliminaries

Let X be a universal set and $F(X) = \{A \mid A : X \to [0,1]\}$. Any $A \in F(X)$ is called a fuzzy set on X. In particular, let R be the set of all real numbers.

We will use $F_T(R)=\left\{\tilde{T}\mid a,s_a^l,s_a^r\in R;\ s_a^l,s_a^r>0\right\}$, where

$$(1)\ \ \tilde{T}(x)=T(a,s_a^l,s_a^r)(x)=\begin{cases}\frac{x-a+s_a^l}{s_a^l} & if \quad a-s_a^l\le x<a,\\ \frac{a+s_a^r-x}{s_a^r} & if \quad a\le x<a+s_a^r,\\ 0 & elsewhere.\end{cases}$$

Any $\tilde{T}\in F_T(R)$ is called a triangular fuzzy number (TFN), where a is center point, s_a^l and s_a^r are the left and right widths of TFN and it may written as $T(a,s_a^l,s_a^r)$. We assume $T(a,0,0)$ be $I_{\{a\}}$, the indicator function of a.

Also, we will show $F_N(R)=\left\{\tilde{N}\mid a,s_a^l,s_a^r\in R;\ s_a^l,s_a^r>0\right\}$, where

$$(2)\ \tilde{N}(x)=N(a,s_a^l,s_a^r)(x)=\begin{cases} exp\left\{-\left(\frac{a-x}{s_a^l}\right)^2\right\} & if \quad x\le a,\\ exp\left\{-\left(\frac{x-a}{s_a^r}\right)^2\right\} & if \quad x>a.\end{cases}$$

Any $\tilde{N}\in F_N(R)$ is called a normal fuzzy number (NFN), where a is center point, s_a^l and s_a^r are the left and right widths of NFN and it may written as $N(a,s_a^l,s_a^r)$. If $\tilde{N}$ is a fuzzy number, the α-cut of $\tilde{N}$ is a closed and bounded interval, for $\alpha\in(0,1]$, which denoted by $\tilde{N}_\alpha=[n_1(\alpha),n_2(\alpha)]$.

Definition 2.1 (24)**.** Let $\tilde{A},\tilde{B}\in F(R)$. Then

$$(3)\qquad \tilde{A}\ominus\tilde{B}=\left\{\int_0^1 g(\alpha)\left[\tilde{A}_\alpha(-)\tilde{B}_\alpha\right]^2\right]^{\frac{1}{2}}$$

is called distance between $\tilde{A}$ and $\tilde{B}$, in which for any $\alpha\in[0,1]$

$$(4)\qquad \tilde{A}_\alpha(-)\tilde{B}_\alpha=\left\{[a_1(\alpha)-b_1(\alpha)]^2+[a_2(\alpha)-b_2(\alpha)]^2\right\}^{\frac{1}{2}}$$

measured the distance between $\tilde{A}_\alpha=[a_1(\alpha),a_2(\alpha)]$ and $\tilde{B}_\alpha=[b_1(\alpha),b_2(\alpha)]$, $a_1(0)$, $b_1(0)$, $a_2(0)$, $b_2(0)$ are taken as finite real numbers and g is a real valued non-decreasing function on $[0,1]$ with $g(0)=0$ and $\int_0^1 g(\alpha)\,d\alpha=\frac{1}{2}$ (for instance $g(\alpha)=\frac{m+1}{2}\alpha^m$ where $m=1,2,3,\cdots$).

It must be mentioned that the distance $\ominus$ is a metric on the set of all fuzzy numbers which has been proved in Theorem 4.1 of [17].

Remark 2.2. In Definition 2.1, $a_1(\alpha)-b_1(\alpha)$ and $a_2(\alpha)-b_2(\alpha)$ are the distance between the left and the right end points of the α-cut of $\tilde{A}$ and $\tilde{B}$, respectively. The value of $g(\alpha)$ can be understood as the weight of $\left[\tilde{A}_\alpha(-)\tilde{B}_\alpha\right]^2$, and the non-decreasing property of g means that the higher the membership of α-cut, the more important it is in determining the distance between $\tilde{A}$ and $\tilde{B}$. This defined operation synthetically reflects the information on every membership degree. The advantage of this arithmetic operation on fuzzy numbers is that they can let different α-cuts have different weights.

Remark 2.3. The introduced distance in Definition 2.1 coincides to the absolute deviation of a and b, when two numbers $\tilde{A}$ and $\tilde{B}$ are real numbers a and b, respectively (see [16], for proof).

Theorem 2.4. *The distance between two TFNs $\tilde{A} = T(a, s_a^l, s_a^r)$ and $\tilde{B} = T(b, s_b^l, s_b^r)$ is*

$$\begin{aligned}\left[\tilde{A} \ominus \tilde{B}\right]^2 &= (a-b)^2 + \frac{1}{(m+2)(m+3)}\left[(s_a^l - s_b^l)^2 + (s_a^r - s_b^r)^2\right] \\ &\quad + \frac{a-b}{m+1}\left[(s_a^r - s_b^r) - (s_a^l - s_b^l)\right], \end{aligned} \tag{5}$$

where the weighted function g is defined by $g(\alpha) = \frac{m+1}{2}\alpha^m$ for $m = 1, 2, 3, \cdots$.

Proof. The α-cuts of TFNs $\tilde{A}$ and $\tilde{B}$ are $\tilde{A}_\alpha = \left[a - s_a^l(1-\alpha), a + s_a^r(1-\alpha)\right]$ and $\tilde{B}_\alpha = \left[b - s_b^l(1-\alpha), b + s_b^r(1-\alpha)\right]$ for any $\alpha \in [0,1]$. Therefore by Definition 2.1,

$$\left[\tilde{A}_\alpha(-)\tilde{B}_\alpha\right]^2 =$$

$$\begin{aligned} &= [a_1(\alpha) - b_1(\alpha)]^2 + [a_2(\alpha) - b_2(\alpha)]^2 \\ &= \left[a - s_a^l(1-\alpha) - b + s_b^l(1-\alpha)\right]^2 + \left[a + s_a^r(1-\alpha) - b - s_b^r(1-\alpha)\right]^2 \\ &= \left[\alpha(s_a^l - s_b^l) + (a-b) - (s_a^l - s_b^l)\right]^2 + \left[\alpha(s_a^r - s_b^r) + (a-b) - (s_a^r - s_b^r)\right]^2 \\ &= 2(a-b)^2 + (\alpha-1)^2(s_a^l - s_b^l) + 2(a-b)(s_a^l - s_b^l)(\alpha - 1) \\ &\quad +(\alpha-1)^2(s_a^r - s_b^r) + 2(a-b)(s_a^r - s_b^r)(\alpha-1) \end{aligned}$$

and so, we will have $g(\alpha) = \frac{m+1}{2}\alpha^m$ and

$$\begin{aligned}\left[\tilde{A} \ominus \tilde{B}\right]^2 &= \int_0^1 \frac{m+1}{2}\alpha^m \left\{[a_1(\alpha) - b_1(\alpha)]^2 + [a_2(\alpha) - b_2(\alpha)]^2\right\} \\ &= (a-b)^2 + \frac{1}{(m+2)(m+3)}\left[(s_a^l - s_b^l)^2 + (s_a^r - s_b^r)^2\right] \\ &\quad + \frac{a-b}{m+1}\left[(s_a^r - s_b^r) - (s_a^l - s_b^l)\right], \end{aligned}$$

for $m = 1, 2, 3, \cdots$. □

Theorem 2.5. *The distance between two NFNs $\tilde{A} = N(a, s_a^l, s_a^r)$ and $\tilde{B} = N(b, s_b^l, s_b^r)$ is*

$$\begin{aligned}\left[\tilde{A} \ominus \tilde{B}\right]^2 &= (a-b)^2 + \frac{1}{2(m+1)}\left[(s_a^l - s_b^l)^2 + (s_a^r - s_b^r)^2\right] \\ &\quad + \sqrt{\frac{\pi}{m+1}}\frac{(a-b)\left[(s_a^l - s_b^l) - (s_a^r - s_b^r)\right]}{2}, \end{aligned} \tag{6}$$

where the weighted function g is defined by $g(\alpha) = \frac{m+1}{2}\alpha^m$ for $m = 1, 2, 3, \cdots$.

Proof. The α-cuts of NFNs $\tilde{A}$ and $\tilde{B}$ are $\tilde{A}_\alpha = \left[a - s_a^l\sqrt{-\ln\alpha}, a + s_a^r\sqrt{-\ln\alpha}\right]$ and $\tilde{B}_\alpha = \left[b - s_b^l\sqrt{-\ln\alpha}, b + s_b^r\sqrt{-\ln\alpha}\right]$ for any $\alpha \in [0,1]$. Therefore by Definition 2.1,

$$\begin{aligned}
\left[\tilde{A}_\alpha(-)\tilde{B}_\alpha\right]^2 &= [a_1(\alpha) - b_1(\alpha)]^2 + [a_2(\alpha) - b_2(\alpha)]^2 \\
&= \left[a - s_a^l\sqrt{-\ln\alpha} - b + s_b^l\sqrt{-\ln\alpha}\right]^2 \\
&\quad + \left[a + s_a^r\sqrt{-\ln\alpha} - b - s_b^r\sqrt{-\ln\alpha}\right]^2 \\
&= \left[(a-b) - \sqrt{-\ln\alpha}(s_a^l - s_b^l)\right]^2 + \left[(a-b) + \sqrt{-\ln\alpha}(s_a^r - s_b^r)\right]^2 \\
&= 2(a-b)^2 - \ln\alpha\left[(s_a^l - s_b^l)^2 + (s_a^r - s_b^r)^2\right] \\
&\quad -2(a-b)\sqrt{-\ln\alpha}\left[(s_a^l - s_b^l)^2 - (s_a^r - s_b^r)^2\right]
\end{aligned}$$

and so, we will have $g(\alpha) = \frac{m+1}{2}\alpha^m$ and

$$\begin{aligned}
\left[\tilde{A} \ominus \tilde{B}\right]^2 &= \int_0^1 \frac{m+1}{2}\alpha^m \left\{[a_1(\alpha) - b_1(\alpha)]^2 + [a_2(\alpha) - b_2(\alpha)]^2\right\} \\
&= (a-b)^2 + \frac{1}{2(m+1)}\left[(s_a^l - s_b^l)^2 + (s_a^r - s_b^r)^2\right] \\
&\quad + \sqrt{\frac{\pi}{m+1}}\frac{(a-b)\left[(s_a^l - s_b^l) - (s_a^r - s_b^r)\right]}{2},
\end{aligned}$$

for $m = 1, 2, 3, \cdots$. □

3. Fuzzy analysis of variance (FANOVA)

The classical ANOVA model can be consulted in any book on linear models as well as well-known references, e.g. [7, 11]. We have been reviewed classical ANOVA model in our previously works [14, 15, 16] with similar notations and we refer the readers to these references.

In real world, the fuzziness of an observed variable often happens in two cases. The first case is due to technical conditions of measurements where the response variable cannot be measured exactly and so in this case data cannot be recorded clearly with precise numbers but only in linguistic terms to justify the required tolerance of the errors in measurements. The second case is due to the fact that the response variable will be given in terms of linguistic forms, such as linguistic report of experts or report of a farmer about his products, which are not numeric. In both cases, the data could be represented by the notion of fuzzy sets to analyze the experiment [14]. To explain this situation, an example from [23] is quoted. The water level of a river cannot be measured in an exact way because of the fluctuation; therefore, fuzzy sets naturally provide an appropriate tool in processing the imprecise data. Under this consideration, the more appropriate way to describe the water level is to say that the water level is "around 3 meters". The phrase around 3 meters should be regarded as a fuzzy number $\tilde{3}$. So here, only the observed values of a classical random variable can be considered as a fuzzy number while the model for observed values is still precise. Note that a similar idea is used by authors in [5, 20]. In this case, observations and recorded data are considered as TFNs $\tilde{y}_{ij} = T(y_{ij}, s^l_{y_{ij}}, s^r_{y_{ij}})$ or NFNs $\tilde{y}_{ij} = N(y_{ij}, s^l_{y_{ij}}, s^r_{y_{ij}})$ where $\tilde{y}_{ij}$ is interpreted as "approximately y_{ij}".

By the above discussion, in this section and hereafter, it is assumed that we are concerned with a classical ANOVA, where the entire*theoretical elements* of the model such as random variables, statistical hypothesis and populations parameter are crisp and hence the model is considered as $Y_{ij} = \mu_i + \varepsilon_{ij}$, with $\varepsilon_{ij} \simeq N(0, \sigma^2)$ in which Y_{ij}'s are ordinary random variables and the statistical hypotheses are considered as classical ones:

$$\begin{array}{lll} & H_0 & : \quad \mu_1 = \mu_2 = \cdots = \mu_r \\ (7) & H_1 & : \quad \text{not all } \mu_i\text{'s are equal (at least one pair with unequal means).} \end{array}$$

But, just one point that will departed from classical ANOVA assumptions in classical ANOVA model is that the sampled observations are STFNs or SNFNs rather than being real numbers and nothing else is altered in the ANOVA model prior to collecting the data. Regarding to the Definition 2.1, the observed values of the statistics SST, $SSTR$, SSE, $MSTR$, MSE and F can be obtained as follows:

$$\widetilde{sst} = \sum_{i=1}^{r}\sum_{j=1}^{n_i} \left(\tilde{y}_{ij} \ominus \overline{\tilde{y}_{..}}\right)^2, \tag{8}$$

$$\widetilde{sstr} = \sum_{i=1}^{r} n_i \left(\overline{\tilde{y}_{i.}} \ominus \overline{\tilde{y}_{..}}\right)^2, \tag{9}$$

and

$$\widetilde{sse} = \sum_{i=1}^{r}\sum_{j=1}^{n_i} \left(\tilde{y}_{ij} \ominus \overline{\tilde{y}_{i.}}\right)^2, \tag{10}$$

when $\tilde{y}_{ij}$'s are TFNs then

$$\overline{\tilde{y}_{i.}} = \frac{1}{n_i}\sum_{j=1}^{n_i} \tilde{y}_{ij} = T\left(\overline{y_{i.}}, \overline{s^l_{y_{i.}}}, \overline{s^r_{y_{i.}}}\right), \tag{11}$$

$$\overline{\tilde{y}_{..}} = \frac{1}{n_t}\sum_{i=1}^{r}\sum_{j=1}^{n_i} \tilde{y}_{ij} = T\left(\overline{y_{..}}, \overline{s^l_{y_{..}}}, \overline{s^r_{y_{..}}}\right), \tag{12}$$

and if $\tilde{y}_{ij}$'s are NFNs, then

$$\overline{\tilde{y}_{i.}} = \frac{1}{n_i}\sum_{j=1}^{n_i} \tilde{y}_{ij} = N\left(\overline{y_{i.}}, \overline{s^l_{y_{i.}}}, \overline{s^r_{y_{i.}}}\right), \tag{13}$$

$$\overline{\tilde{y}_{..}} = \frac{1}{n_t}\sum_{i=1}^{r}\sum_{j=1}^{n_i} \tilde{y}_{ij} = N\left(\overline{y_{..}}, \overline{s^l_{y_{..}}}, \overline{s^r_{y_{..}}}\right), \tag{14}$$

$\overline{y_{i.}} = \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i}$ and $\overline{y_{..}} = \frac{\sum_{i=1}^{r}\sum_{j=1}^{n_i} y_{ij}}{n_t}$. Then, we have

$$\widetilde{mstr} = \frac{\widetilde{sstr}}{r-1}, \tag{15}$$

$$\widetilde{mse} = \frac{\widetilde{sse}}{n_t - r}, \tag{16}$$

and

$$\widetilde{f} = \frac{\widetilde{mstr}}{\widetilde{mse}} = \frac{n_t - r}{r-1}\frac{\widetilde{sstr}}{\widetilde{sse}}, \tag{17}$$

The decision rule: Let $\widetilde{f}$ be the observed value of the test statistic and $F_{1-\alpha;r-1,n_t-r}$ be the α-th quantile of the Fisher distribution with $r-1$ and $n_t - r$ degrees of freedom. At the given significance level α, we accept the null hypothesis H_0 if $\widetilde{f} \leq F_{1-\alpha;r-1,n_t-r}$; otherwise we accept the alternative hypothesis H_1.

In testing ANOVA based on fuzzy numbers the p-value can be calculated by p-value= $P\left(F > \widetilde{f}\right)$ in which $\widetilde{f}$ is introduced by (17) as the observed value of the test statistic on the basis of fuzzy observations.

Remark 3.1. Considering Remark , when the observed data are precise numbers y_{ij}, that is they are indicator functions $I_{y_{ij}}$ for $i = 1, \cdots, r$ and $j = 1, \cdots, n_i$, then all the introduced extended statistics in equations (8)(17) coincide to statistics of classical ANOVA.

Remark 3.2. Although the observed data are considered fuzzy in this paper, but it must be highlighted that all fuzzy observed data are from a crisp random variable and so this fact causes to achieve the precise Fisher distribution with $r-1$ and $n_t - r$ degrees of freedom for the test statistic; see [16].

In continue of this section, we are going to obtain several fast computation formulas for FANOVA test.

3.1. **FANOVA for non-symmetric TFNs.** In this subsection, we assume all observations are TFNs. Then fuzzy statistics obtains as following theorems.

Theorem 3.3. *In ANOVA model, suppose the observed data are* $\tilde{y}_{ij} = T(y_{ij}, s^l_{y_{ij}}, s^r_{y_{ij}}) \in F_T(R)$, $i = 1, \cdots, r$, $j = 1, \cdots, n_i$ *then observed values of* $\widetilde{sst}$, $\widetilde{sstr}$ *and* $\widetilde{sse}$ *are as following real values:*

$$\widetilde{sst} = sst_y + \frac{1}{(m+2)(m+3)}\left[sst_{s^r_y} + sst_{s^l_y}\right], \tag{18}$$

$$\widetilde{sstr} = sstr_y + \frac{1}{(m+2)(m+3)}\left[sstr_{s^r_y} + sstr_{s^l_y}\right], \tag{19}$$

$$\widetilde{sse} = sse_y + \frac{1}{(m+2)(m+3)}\left[sse_{s^r_y} + sse_{s^l_y}\right], \tag{20}$$

where sst_y, $sstr_y$ and sse_y are the crisp values of sst, $sstr$ and sse for the centre points of $\tilde{y}_{ij} = T(y_{ij}, s^l_{y_{ij}}, s^r_{y_{ij}})$, and and the crisp values of sst, $sstr$ and sse for the left widths of $\tilde{y}_{ij} = T(y_{ij}, s^l_{y_{ij}}, s^r_{y_{ij}})$, and and the crisp values of sst, $sstr$ and sse for the right widths of $\tilde{y}_{ij} = T(y_{ij}, s^l_{y_{ij}}, s^r_{y_{ij}})$, respectively.

Proof. From (8), Definition 2.1 and Theorem 2.4, and assumption $g(\alpha) = \frac{m+1}{2}\alpha^m$, $m = 1, 2, 3, \cdots$, we have

$$\begin{aligned}
\widetilde{sst} &= \sum_{i=1}^{r}\sum_{j=1}^{n_i} \left(\tilde{y}_{ij} \ominus \overline{\tilde{y}_{..}}\right)^2 \\
&= \sum_{i=1}^{r}\sum_{j=1}^{n_i} \left[T(y_{ij}, s^l_{y_{ij}}, s^r_{y_{ij}}) \ominus T(\overline{y_{..}}, \overline{s^l_{y_{..}}}, \overline{s^r_{y_{..}}})\right]^2 \\
&= \sum_{i=1}^{r}\sum_{j=1}^{n_i} \left\{(y_{ij} - \overline{y_{..}})^2 + \frac{1}{(m+2)(m+3)}\left[\left(s^l_{y_{ij}} - \overline{s^l_{y_{..}}}\right)^2 + \left(s^r_{y_{ij}} - \overline{s^r_{y_{..}}}\right)^2\right]\right. \\
&\quad \left. + \frac{(y_{ij} - \overline{y_{..}})}{m+2}\left[\left(s^l_{y_{ij}} - \overline{s^l_{y_{..}}}\right) + \left(s^r_{y_{ij}} - \overline{s^r_{y_{..}}}\right)\right]\right\} \\
&= \sum_{i=1}^{r}\sum_{j=1}^{n_i} (y_{ij} - \overline{y_{..}})^2 \\
&\quad + \frac{1}{(m+2)(m+3)}\left[\sum_{i=1}^{r}\sum_{j=1}^{n_i}\left(s^l_{y_{ij}} - \overline{s^l_{y_{..}}}\right)^2 + \sum_{i=1}^{r}\sum_{j=1}^{n_i}\left(s^r_{y_{ij}} - \overline{s^r_{y_{..}}}\right)^2\right] + 0 \\
&= sst_y + \frac{1}{(m+2)(m+3)}\left[sst_{s^r_y} + sst_{s^l_y}\right].
\end{aligned}$$

Similarly, one can prove (19) and (20). □

Remark 3.4. Under the same assumption of Theorem 3.3, the observed values of the mean squares $\widetilde{mstr}$, $\widetilde{mse}$ and the test statistic $\widetilde{f}$ are respectively as follows:

$$\widetilde{mstr} = \frac{\widetilde{sstr}}{r-1} = mstr_y + \frac{1}{(m+2)(m+3)}\left[mstr_{s^l_y} + mstr_{s^r_y}\right], \tag{21}$$

$$\widetilde{mse} = \frac{\widetilde{sse}}{n_t - r} = mse_y + \frac{1}{(m+2)(m+3)}\left[mse_{s^l_y} + mse_{s^r_y}\right], \tag{22}$$

and

$$\widetilde{f} = \frac{\widetilde{mstr}}{\widetilde{mse}} = \frac{mstr_y + \frac{1}{(m+2)(m+3)}\left[mstr_{s^l_y} + mstr_{s^r_y}\right]}{mse_y + \frac{1}{(m+2)(m+3)}\left[mse_{s^l_y} + mse_{s^r_y}\right]}, \tag{23}$$

3.2. **FANOVA for non-symmetric NFNs.** In this subsection, we assume all observations are TFNs. Then fuzzy statistics obtains as following theorems.

Theorem 3.5. *In ANOVA model, suppose the observed data are* $\tilde{y}_{ij} = N(y_{ij}, s^l_{y_{ij}}, s^r_{y_{ij}}) \in F_N(R)$ *,* $i = 1, \cdots, r,$ $j = 1, \cdots, n_i$ *then observed values of* $\widetilde{sst}$*,* $\widetilde{sstr}$ *and* $\widetilde{sse}$ *are as following real values:*

$$\widetilde{sst} = sst_y + \frac{1}{2(m+1)}\left[sst_{s^l_y} + sst_{s^r_y}\right], \tag{24}$$

$$\widetilde{sstr} = sstr_y + \frac{1}{2(m+1)}\left[sstr_{s^l_y} + sstr_{s^r_y}\right], \tag{25}$$

$$\widetilde{sse} = sse_y + \frac{1}{2(m+1)}\left[sse_{s^l_y} + sse_{s^r_y}\right]. \tag{26}$$

Proof. From (8), Definition 2.1 and Theorem 2.5, and assumption $g(\alpha) = \frac{m+1}{2}\alpha^m$, $m = 1, 2, 3, \cdots$, we have

$$\begin{aligned}
\widetilde{sst} &= \sum_{i=1}^{r}\sum_{j=1}^{n_i}\left(\tilde{y}_{ij} \ominus \overline{\tilde{y}_{..}}\right)^2 \\
&= \sum_{i=1}^{r}\sum_{j=1}^{n_i}\left[N(y_{ij}, s^l_{y_{ij}}, s^r_{y_{ij}}) \ominus N(\overline{y_{..}}, \overline{s^l_{y_{..}}}, \overline{s^r_{y_{..}}})\right]^2 \\
&= \sum_{i=1}^{r}\sum_{j=1}^{n_i}\Bigg\{\left(y_{ij} - \overline{y_{..}}\right)^2 + \frac{1}{2(m+1)}\left[\left(s^l_{y_{ij}} - \overline{s^l_{y_{..}}}\right)^2 + \left(s^r_{y_{ij}} - \overline{s^r_{y_{..}}}\right)^2\right] \\
&\quad + \sqrt{\frac{\pi}{m+1}}\frac{\left(y_{ij} - \overline{y_{..}}\right)\left[\left(s^l_{y_{ij}} - \overline{s^l_{y_{..}}}\right) + \left(s^r_{y_{ij}} - \overline{s^r_{y_{..}}}\right)\right]}{2}\Bigg\} \\
&= \sum_{i=1}^{r}\sum_{j=1}^{n_i}\left(y_{ij} - \overline{y_{..}}\right)^2 \\
&\quad + \frac{1}{2(m+1)}\left[\sum_{i=1}^{r}\sum_{j=1}^{n_i}\left(s^l_{y_{ij}} - \overline{s^l_{y_{..}}}\right)^2 + \sum_{i=1}^{r}\sum_{j=1}^{n_i}\left(s^r_{y_{ij}} - \overline{s^r_{y_{..}}}\right)^2\right] + 0 \\
&= sst_y + \frac{1}{2(m+1)}\left[sst_{s^l_y} + sst_{s^r_y}\right].
\end{aligned}$$

Similarly, one can prove (25) and (26). □

Theorem 3.6. *Under the same assumption of Theorem 3.5, the observed values of the mean squares* $\widetilde{mstr}$ *and* $\widetilde{mse}$ *are as follows:*

$$\widetilde{mstr} = mstr_y + \frac{1}{2(m+1)}\left[mstr_{s^l_y} + mstr_{s^r_y}\right], \tag{27}$$

$$\widetilde{mse} = mse_y + \frac{1}{2(m+1)}\left[mse_{s^l_y} + mse_{s^r_y}\right]. \tag{28}$$

Proof. It is obvious that

$$\begin{aligned}\widetilde{mstr} &= \frac{\widetilde{sstr}}{r-1} = \frac{sstr_y}{r-1} + \frac{1}{2(m+1)}\left[\frac{sstr_{s_y^l}}{r-1} + \frac{sstr_{s_y^r}}{r-1}\right]\\ &= mstr_y + \frac{1}{2(m+1)}\left[mstr_{s_y^l} + mstr_{s_y^r}\right],\end{aligned}$$

$$\begin{aligned}\widetilde{mse} &= \frac{\widetilde{sse}}{n_t-r} = \frac{sse_y}{n_t-r} + \frac{1}{2(m+1)}\left[\frac{sse_{s_y^l}}{n_t-r} + \frac{sse_{s_y^r}}{n_t-r}\right]\\ &= mse_y + \frac{1}{2(m+1)}\left[mse_{s_y^l} + mse_{s_y^r}\right].\end{aligned}$$

□

Remark 3.7. Under the same assumption of Theorem 3.5, the observed values of the test statistic $\widetilde{f}$ in analysis of variance is as following:

$$\widetilde{f} = \frac{\widetilde{mstr}}{\widetilde{mse}} = \frac{mstr_y + \frac{1}{2(m+1)}\left[mstr_{s_y^l} + mstr_{s_y^r}\right]}{mse_y + \frac{1}{2(m+1)}\left[mse_{s_y^l} + mse_{s_y^r}\right]}. \tag{29}$$

4. Case study: FANOVA application on the average length of car battery life

Under ideal conditions, the average length of car battery life manufactured by an Iranian factory is estimated to be about 3 to 4 years. In addition to the quality of battery building, factors such as the amount of battery usage, maintenance, ambient temperature, vibration, audio system usage, charging and discharging can also affect the length life of the batteries used. Consider this fact that the length of car battery life does not end at a moment and its useful life/capability will down slowly over the time. Therefore, the length of car battery life can be recorded by a non-precise / fuzzy number. On the other hand, on most car batteries, an visual marker is Embedded to measure the electrolyte concentration of the battery (which is in fact a simple hydrometer / acidometer) and shows the battery status to the user by displaying one of the three following colors:

(1) the green color means a healthy battery
(2) the black (or red) color means the battery is lower than the standard and needs to be charged, but the car is still able to start, and
(3) the white color means the battery is completely discharged and the car can not start.

Therefore, in this research, asymmetric triangular numbers are used to record the data on the battery lifetime. Three assembly lines at an Iranian factory are simultaneously producing car batteries with a similar brand. Recently, a claim has been made that the length of battery life is different in these three assembly lines. Therefore, in order to check the independency of the produced batteries lifetime from the production line number, the factory wish to test whether the batteries lifetime from the different production line number are the same or not. In other words, we are going to test

$$H_0 \ : \ \mu_1 = \mu_2 = \mu_3$$
$$H_1 \ : \ \text{not all } \mu_i\text{'s are equal, for i=1,2,3.}$$

To this lifetime test, 29 triangular fuzzy batteries lifetimes are selected as the experimental units in FANOVA test with sample sizes $n_1 = 10$, $n_2 = 8$ and $n_3 = 11$, respectively (see Table 1).

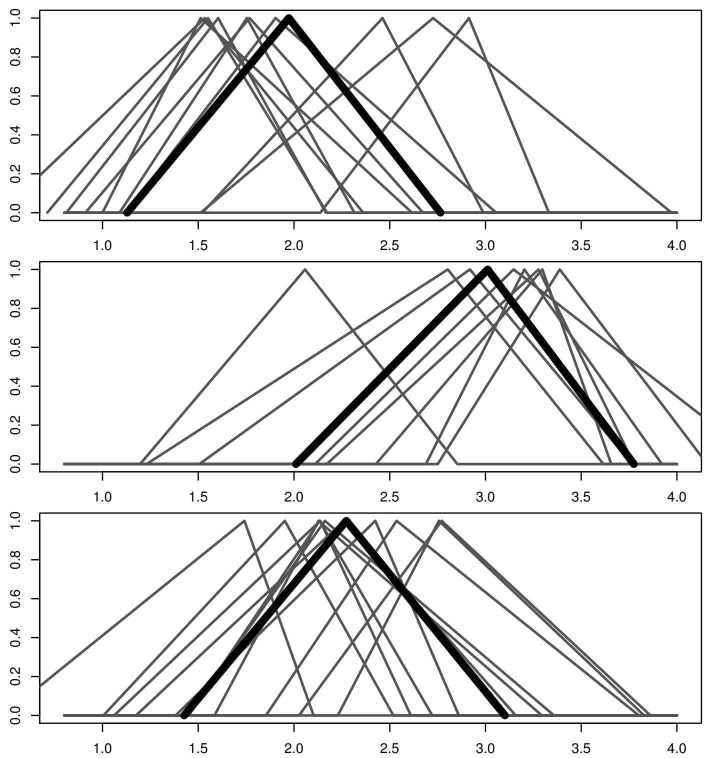

FIGURE 1. Triangular fuzzy observed data for the batteries lifetime in assembly lines 1, 2 and 3 (from above figure to bellow figure).

Considering Theorem 3.3 and Result 3.4, one can obtain the observed values of FANOVA statistics which are reported in Table 2, based on the given non-symmetric TFNs. For instance, the total sum of squares is calculated for $m = 1$ by Theorem3.3 in bellow

$$\widetilde{sst} \ = \ sst_y + \frac{1}{(m+2)(m+3)} \left[sst_{s_y^r} + sst_{s_y^l} \right]$$

Table 1. Fuzzy observed data for the length of car battery life in three factory assembly lines.

Observation #	Line 1	Line 2	Line 3
1	T(1.60, 0.79, 0.56)	T(2.06, 0.86, 0.80)	T(2.13, 0.54, 1.17)
2	T(2.73, 1.21, 1.25)	T(3.30, 0.87, 0.36)	T(2.76, 0.53, 1.07)
3	T(2.91, 0.78, 0.42)	T(3.28, 1.11, 0.50)	T(2.14, 1.08, 0.47)
4	T(1.90, 0.78, 1.15)	T(3.15, 1.04, 1.31)	T(2.16, 0.72, 1.20)
5	T(1.75, 0.66, 0.56)	T(2.80, 1.57, 0.81)	T(2.26, 1.09, 0.90)
6	T(1.55, 0.84, 0.62)	T(3.39, 0.64, 0.77)	T(2.54, 0.68, 1.27)
7	T(1.77, 0.86, 0.91)	T(2.92, 1.41, 0.85)	T(2.42, 1.04, 0.44)
8	T(2.46, 0.94, 0.53)	T(3.20, 0.51, 0.72)	T(1.95, 0.95, 0.57)
9	T(1.53, 1.07, 0.83)		T(2.77, 0.75, 1.09)
10	T(1.51, 0.51, 1.10)		T(2.13, 0.68, 0.59)
11			T(1.74, 1.26, 0.36)

$$
\begin{aligned}
&= \sum_{i=1}^{3}\sum_{j=1}^{n_i}(y_{ij}-\overline{y_{..}})^2+\frac{1}{12}\left[\sum_{i=1}^{3}\sum_{j=1}^{n_i}\left(s^l_{y_{ij}}-\overline{s^l_{y_{..}}}\right)^2+\sum_{i=1}^{3}\sum_{j=1}^{n_i}\left(s^r_{y_{ij}}-\overline{s^r_{y_{..}}}\right)^2\right]\\
&= \left[(1.60-2.37)^2+\cdots+(1.74-2.37)^2\right]\\
&\quad +\frac{1}{12}\left[\left\{(0.79-0.89)^2+\cdots+(1.26-0.89)^2\right\}\right.\\
&\quad \left.+\left\{(0.56-0.80)^2+\cdots+(036-0.80)^2\right\}\right]\\
&= 8.68.
\end{aligned}
$$

And, also by Eq. (23) from Result 3.4 the observed value of FANOVA test statistic is $\widetilde{f}=17.09$ for $m=1$ (see Table 2). Moreover, all computations of this paper are done by R software [20].

Table 2. Details of ANOVA for the length of car battery.

Source of variation	$\widetilde{ss}$	Degrees of freedom	$\widetilde{ms}$	$\widetilde{f}$
Between treatments	4.93	2	2.49	
Within treatments (error)	3.75	26	0.14	17.09
Total	8.68	28		

By comparing the computed FANOVA test statistic, one can accept the alternative hypothesis at significance level 0.05. The critical value of ANOVA test is $F_{1-\alpha;r-1,n_t-r}=F_{0.95;2,26}=3.37$ and also the computed p-value$=1.82\times10^{-5}$ indicates very weak evidence in favour of null hypothesis. Therefore, we conclude that there is a relation between produced batteries lifetime and the production line number at the considered significance level, based on the recorded fuzzy data in Table 1. It must be emphasised that the obtained p-value result is related to both randomness uncertainty and fuzziness uncertainty. In other words, the proposed decision rule is a function of the stochastic uncertainty due to the random sample, and also, is a function of the vagueness uncertainty due to the fuzzy observations.

5. Conclusion

In applied sciences such as economics, agriculture and social sciences, it may be confront with vague / fuzzy concepts, such as the threshold of patient tolerance, the degree of utility of life and the monthly income of a Taxi driver. In such situations, the classical ANOVA can not solve the vague test and it need to generalize based on fuzzy data. The proposed fuzzy ANOVA (FANOVA) model is a generalized version of the classical ANOVA using non-symmetric triangular and normal fuzzy observations, such that when all observations are real numbers, FANOVA reduces to ANOVA because vagueness of the fuzzy statistics are removed and what remain is only the center point of them. FANOVA is easy to be implemented as it is almost developed similar to classical ANOVA for its components, and it is easy to be used by professional clients who are familiar with ANOVA. For future works, one can try to use the approach of this paper to extend other experimental designs such as random block design, Latin square design and etc., for the case where the observations are fuzzy numbers rather than being numbers. As, the observations in this paper are assumed to be non-symmetric triangular / normal fuzzy numbers, another interesting issue is to extend the results of this paper to trapezoidal fuzzy numbers, and in general LR fuzzy numbers. Building user-friendly software packages for FANOVA based on fuzzy data is another field for future works.

6. Aknowledgement

The author would like to thank from the anonymous referees for their constructive suggestions and comments which helped to improve the manuscript.

References

[1] W.G. Cochran and G.M. Cox, *Experimental designs*, 2nd edition, Wiley, New York, 1957.

[2] A. Cuevas, M. Febrero and R. Fraiman, *An ANOVA test for functional data*, Comput. Statist. Data Anal. 47 (2004) 111–122.

[3] A. Dean and D. Voss, *Design and analysis of experiments*, Springer-Verlag, New York, 1999.

[4] D. Dubois, H. Prade, *Fuzzy sets and systems: Theory and application*, Academic, New York, 1980.

[5] P. Filzmoser and R. Viertl, *Testing hypotheses with fuzzy data: the fuzzy p-value* , Metrika 59 (2009) 21–29.

[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*, Comput. Statist. Data Anal. 51 (2006) 141–162.

[7] R.R. Hocking, *Methods and applications of linear models: Regression and the analysis of variance*, John Wiley & Sons, New York, 1996.

[8] A. Kaufmann and M.M. Gupta, *Introduction to fuzzy arithmetic: Theory and applications*, Van Nastrand, New York, 1985.

[9] W.J. Lee, H.Y. Jung, J.H. Yoon and S.H. Choi, *Analysis of variance for fuzzy data based on permutation method*, International Journal of Fuzzy Logic and Intelligent Systems 17 (2017) 43–50.

[10] 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 59 (2004) 31–49.

[11] M. Montenegro, G. Gonzalez-Rodriguez, M.A. Gil, A. Colubi and M.R. Casals, *Introduction to ANOVA with fuzzy random variables*, In: Lopez-Diaz, M., Gil, M.A., Grzegorzewski, P., Hryniewicz, O., Lawry, J. (Eds.), Soft Methodology and random information systems. Springer, (Berlin, 2004) 487–494.

[12] D.C. Montgomery, *Design and analysis of experiments*, Third edition, John Wiley & Sons, New York, 1991.

[13] H.T. Nguyen and E.A. Walker, *A first course in fuzzy logic*, Third edition, Chapman Hall/CRC, Paris, 2005.

[14] M.R. Nourbakhsh, M. Mashinchi and A. Parchami, *Analysis of variance based on fuzzy observations*, International Journal of Systems Science 44 (2013) 714–726.

[15] A. Parchami, R. Ivani, M. Mashinchi and . Kaya, *An implication of fuzzy ANOVA: Metal uptake and transport by corn grown on a contaminated soil*, Chemometrics and Intelligent Laboratory Systems 164 (2017) 56–63.

[16] A. Parchami, M.R. Nourbakhsh and M. Mashinchi, *Analysis of variance in uncertain environments*, Complex & Intelligent Systems 3 (2017, 189–196.

[17] A. Parchami, B. Sadeghpour-Gildeh, M. Nourbakhsh and M. Mashinchi, *A new generation of process capability indices based on fuzzy measurements*, Journal of Applied Statistics 41 (2014) 1122–1136.

[18] A. Parchami, M. Mashinchi and C. Kahraman, *A case study on vehicle battery manufacturing using fuzzy analysis of variance*, Proceedings of the INFUS 2020 Conference, July 21-23 (Istanbul, 2020), 916–923.

[19] M.L. Puri and D.A. Relescu, *Fuzzy random variables*, J. Math. Anal. Appl. 114 (1986) 409–422.

[20] M.L. Rizzo, *Statistical computing with R*, Chapman Hall/CRC, Paris, 2008.

[21] S.M. Taheri and M. Arefi, *Testing hypotheses based on fuzzy test statistic*, Soft Comput. 13 (2009) 617–625.

[22] R. Viertl and F.L. Boca Raton, *Statistical methods for non-precise data*, CRC Press, Paris, 1996.

[23] H.C. Wu, *Analysis of variance for fuzzy data*, International Journal of Systems Science 38 (2007) 235–246.

[24] R. Xu and C. Li, *Multidimensional least-squares fitting with a fuzzy model*, Fuzzy Sets and Systems 119 (2001) 215–223.

Abbas Parchami
Orcid number: 0000-0002-0593-7324
Department of Statistics
Faculty of Mathematics and Computer
Shahid Bahonar University of Kerman
Kerman, Iran
E-mail address: parchami@uk.ac.ir

Mashallah Mashinchi
Orcid number: 0000-0001-5848-8027
Department of Statistics
Faculty of Mathematics and Computer
Shahid Bahonar University of Kerman
Kerman, Iran
E-mail address: mashinchi@uk.ac.ir

Cengiz Kahraman
Orcid number: 0000-0001-6168-8185
Department of Industrial Engineering
Istanbul Technical University
Macka Istanbul, Turkey
E-mail address: kahramanc@itu.edu.tr