

Analysis of variance in uncertain environments

A. Parchami¹ · M. Nourbakhsh² · M. Mashinchi¹

Received: 21 February 2017 / Accepted: 31 May 2017 / Published online: 10 June 2017
© The Author(s) 2017. This article is an open access publication

Abstract Testing one-way analysis of variance (ANOVA) is used for experimental data analysis in which there is a continuous response variable and a single independent classification variable. In this paper, we extend one-way ANOVA to a case where observed data are imprecise numbers rather than real numbers. Several fast computable formulas are calculated for symmetric triangular and normal fuzzy data. Similar to the classical testing ANOVA, the total observed variation in the response variable is explained as the sum of observed variation due to the effects of the classification variable and the observed variation due to random error. A real case is given to clarify the proposed method.

Keywords Symmetric triangular fuzzy number · Symmetric normal fuzzy number · Arithmetic operations · Testing hypotheses · Analysis of variance

Introduction and literature review

ANOVA

Analysis of variance (ANOVA) is concerned with analyzing variation in the means of several independent populations which developed by Ronald Fisher. The most important point in traditional ANOVA is a test about the significance of the difference among population means. This test permits the user to conclude whether the differences among the means of several populations are too deviated to be attributed to the sampling error or not [7]. The basic formulas of classical ANOVA model can be referred to any statistical book [1, 10, 19], but the classical ANOVA has been briefly reviewed in “Classical ANOVA” of this manuscript from [21].

In many applied sciences like geology, astrology, economics, medical sciences, engineering and environmental sciences, there exist actual cases where imprecise quantities can be assigned to experimental results. In such practical cases, non-precise numbers are suitable models to formalize and handle such imprecise quantities, which is the justification for require to fuzzy sets in ANOVA. The previous related works on fuzzy ANOVA are listed in the next subsection.

Fuzzy ANOVA

During the past three decades, several scientific works have been published in different fields of ANOVA using fuzzy set theory. These fields are as follow: one-way ANOVA testing procedure based on normal fuzzy random variables (FRVs) [18], asymptotic bootstrap ANOVA test for FRVs by one-sample [17], providing ANOVA method for the functional data in Hilbert space [2], asymptotic bootstrap ANOVA test for simple FRVs on the basis of multi-sample [7, 8], an optimization method for one-way ANOVA problem by the cuts

✉ A. Parchami
parchami@uk.ac.ir
M. Nourbakhsh
nourbakhsh@mathdep.iust.ac.ir
M. Mashinchi
mashinchi@uk.ac.ir

¹ Department of Statistics, Faculty of Mathematics and Computer, Shahid Bahonar University of Kerman, Kerman, Iran

² School of Mathematics, Iran University of Science and Technology, Narmak, Tehran, Iran

of FRVs when observations are imprecise [29], investigating on one-way fuzzy ANOVA and comparing it with a regression model [4], processing method for ANOVA for the vague data using the moment correction [15], providing one-way ANOVA for triangular fuzzy numbers on the basis of the extension principle [21], developing one-way ANOVA for the closed interval observations [9], providing R package ‘SAFD’ for ANOVA test based on defuzzification of the FRVs [16], extending the classical ANOVA test statistics to investigate on fuzzy means of imprecise observation by bootstrap approach [25], providing ANOVA test based on least squares approach for FRVs [12], and proposing a decision-rule approach to test fuzzy hypotheses in ANOVA considering triangular imprecise observation and vague hypotheses [13].

Unlike these studies, another simple method for one-way ANOVA, as a meaningful extension on the classical ANOVA, is presented in this paper where the observations are symmetric triangular or normal fuzzy numbers. The proposed ANOVA method is on the basis of a distance between symmetric fuzzy numbers which is easy to be used by professional clients.

The organization of this paper is as follows. After introducing symmetric triangular and normal fuzzy numbers, some arithmetic operations for imprecise numbers are reviewed in ‘‘Preliminaries’’. In ‘‘Classical ANOVA’’, classical ANOVA is briefly explained from [21]. In ‘‘Analysis of variance based on fuzzy observations’’, a new method for ANOVA test for fuzzy observations is provided and discussed. In ‘‘Fast computation formulas’’, several fast computable formulas are presented for symmetric triangular and normal fuzzy data. A case study on several brands of soap is presented in ‘‘A case study on the process of soap production’’ which is revealing the ideas of this paper. Conclusion part is given in the final section.

Preliminaries

In this section, we review briefly some definitions which will be referred through this paper. Let X be a universal set and $F(X) = \{A|A : X \rightarrow [0, 1]\}$. Any $A \in F(X)$ is called a fuzzy set on X . In particular, let $F(R)$ be the set of fuzzy numbers on R (the set of real numbers). The α -cut of A is the crisp set given by $A_\alpha = \{x \in X|A(x) \geq \alpha\}$, for any $\alpha \in (0, 1]$.

Definition 2.1 (i) An especial case of fuzzy numbers called symmetric triangular fuzzy number (STFN) is defined by

$$\tilde{T}(x) = \begin{cases} (x - a + s_a)/s_a & \text{if } a - s_a \leq x < a \\ (a + s_a - x)/s_a & \text{if } a \leq x < a + s_a \\ 0 & \text{elsewhere,} \end{cases} \quad (1)$$

and it is denoted with $T(a, s_a)$.

(ii) Another especial case of fuzzy numbers called normal fuzzy number (NFN) is defined by

$$\tilde{N}(x) = \exp \left[- \left(\frac{x - a}{s_a} \right)^2 \right], \quad x \in R, s_a > 0 \quad (2)$$

and it is shown with $N(a, s_a)$. Here, $a \in R$ and $s_a \in R^+$ are named the core/mean and the spread of STFNs and NFNs, respectively. Moreover, $F_T(R)$ and $F_N(R)$ denote the sets of all STFNs and NFNs on R [23].

The following distance will be used through Sect. 4 to solve the problem of testing ANOVA based on fuzzy observations.

Definition 2.2 [23,29] For any $\tilde{A}, \tilde{B} \in F(R)$,

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

is called the distance of \tilde{A} and \tilde{B} where

$$\begin{aligned} \tilde{A}_\alpha(-)\tilde{B}_\alpha &= \left\{ [a_1(\alpha) - b_1(\alpha)]^2 \right. \\ &\quad \left. + [a_2(\alpha) - b_2(\alpha)]^2 \right\}^{\frac{1}{2}} \text{ for any } \alpha \in (0, 1] \end{aligned} \quad (4)$$

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

Remark 2.1 Definition 2.2 is an extended version of the absolute deviation between two real numbers. Since, for two real numbers a and b by indicator functions $I_{\{a\}}$ and $I_{\{b\}}$ we have

$$\begin{aligned} I_{\{a\}} \ominus I_{\{b\}} &= \left\{ \int_0^1 g(\alpha) [(I_{\{a\}})_\alpha(-)(I_{\{b\}})_\alpha]^2 d\alpha \right\}^{\frac{1}{2}} \\ &= \left\{ \int_0^1 g(\alpha) [(a - b)^2 + (a - b)^2] d\alpha \right\}^{\frac{1}{2}} \\ &= \left\{ 2(a - b)^2 \int_0^1 g(\alpha)d\alpha \right\}^{\frac{1}{2}} \\ &= |a - b|. \end{aligned}$$

More details of Definition 2.2 is presented in [23].

Theorem 2.1 The distance between two STFNs $\tilde{A} = T(a, s_a) \in F_T(R)$ and $\tilde{B} = T(b, s_b) \in F_T(R)$ is

$$\tilde{A} \ominus \tilde{B} = \left\{ (a - b)^2 + \frac{2}{(m + 2)(m + 3)}(s_a - s_b)^2 \right\}^{\frac{1}{2}}, \quad (5)$$

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

Proof For any $\alpha \in (0, 1)$,

$\tilde{A}_\alpha = [a - s_a(1 - \alpha), a + s_a(1 - \alpha)]$ and $\tilde{B}_\alpha = [b - s_b(1 - \alpha), b + s_b(1 - \alpha)]$.

Hence, according to Definition 2.2,

$$\begin{aligned} [\tilde{A}_\alpha(-)\tilde{B}_\alpha]^2 &= [a_1(\alpha) - b_1(\alpha)]^2 + [a_2(\alpha) - b_2(\alpha)]^2 \\ &= [a - s_a(1 - \alpha) - b + s_b(1 - \alpha)]^2 \\ &\quad + [a + s_a(1 - \alpha) - b - s_b(1 - \alpha)]^2 \\ &= [\alpha(s_a - s_b) + (a - b) - (s_a - s_b)]^2 \\ &\quad + [\alpha(s_b - s_a) + (a - b) + (s_a - s_b)]^2 \\ &= 2\alpha^2(s_a - s_b)^2 - 4\alpha(s_a - s_b)^2 \\ &\quad + 2(a - b)^2 + 2(s_a - s_b)^2 \end{aligned}$$

and so, using $g(\alpha) = \frac{m+1}{2}\alpha^m$ in Definition 2.2, it can be calculated that

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

for $m = 1, 2, 3, \dots$ □

Theorem 2.2 *The distance between two NFNs $\tilde{A} = N(a, s_a) \in F_N(R)$ and $\tilde{B} = N(b, s_b) \in F_N(R)$ is*

$$\tilde{A} \ominus \tilde{B} = \left\{ (a - b)^2 + \frac{1}{m+1} (s_a - s_b)^2 \right\}^{\frac{1}{2}}, \tag{6}$$

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

Proof One can obtain $\tilde{A}_\alpha = [a - s_a\sqrt{-\ln \alpha}, a + s_a\sqrt{-\ln \alpha}]$ and $\tilde{B}_\alpha = [b - s_b\sqrt{-\ln \alpha}, b + s_b\sqrt{-\ln \alpha}]$, for any $\alpha \in (0, 1)$. Therefore, considering Definition 2.2,

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

Hence, using $g(\alpha) = \frac{m+1}{2}\alpha^m$, $m = 1, 2, 3, \dots$ in Definition 2.2, it can be concluded that

$$\begin{aligned} [\tilde{A} \ominus \tilde{B}]^2 &= \lim_{\varepsilon \rightarrow 0^+} \int_\varepsilon^1 \frac{m+1}{2} \alpha^m \{ [a_1(\alpha) - b_1(\alpha)]^2 \\ &\quad + [a_2(\alpha) - b_2(\alpha)]^2 \} d\alpha \\ &= (m+1) \left[\lim_{\varepsilon \rightarrow 0^+} \int_\varepsilon^1 \alpha^m \{ (a - b)^2 \right. \\ &\quad \left. - (s_a - s_b)^2 \ln \alpha \} d\alpha \right] \\ &= (m+1) \left\{ \frac{(a-b)^2}{m+1} + \frac{(s_a-s_b)^2}{(m+1)^2} \right\} \\ &= (a - b)^2 + \frac{1}{(m+1)}(s_a - s_b)^2. \end{aligned} \tag{□}$$

Remark 2.2 Note that in some cases the integral used in Definition 2.2 leads to an improper integral as appeared in the proof of Theorem 2.2.

Classical ANOVA

Although the classical ANOVA model can be referred to many statistical books as well as well-known references, e.g., see [1, 10, 19], but for a short introduction and review the methodology of ANOVA we propose ‘‘Classical ANOVA’’ from [21] since the considered notations are same here. So, we only present some classical formulas in Table 1 known as ANOVA table. To test whether the factor level means μ_i ’s are the same in quantity or not, the classical ANOVA has been contemplate to accept/reject hypothesis ‘‘ $H_0: \mu_1 = \mu_2 = \dots = \mu_r$ ’’, against ‘‘ $H_1: \text{not all } \mu_i \text{’s are equal}$ ’’, based on observed random samples [11].

Analysis of variance based on fuzzy observations

Quantities of continuous variables are not precise numbers and a more appropriate way to describing them is using imprecise/fuzzy numbers [24, 26, 27]. To explain such situation, one can see several applied examples from references [11, 14, 28]. It must be noted that in this section, only the observed amounts of precise random variables can be regarded as imprecise numbers while the model for observed values is still precise as follows

$$Y_{ij} = \mu_i + \varepsilon_{ij}, \text{ for } i = 1, \dots, r \text{ and } j = 1, \dots, n_i. \tag{7}$$

Note that a similar idea is used in [6, 11, 27] and in a such situation, observations and recorded data can be looked as STFNS $\tilde{y}_{ij} = T(y_{ij}, s_{y_{ij}})$ or NFNS $\tilde{y}_{ij} = N(y_{ij}, s_{y_{ij}})$, where \tilde{y}_{ij} is interpreted as ‘‘approximately y_{ij} ’’, for $i = 1, \dots, r$ and $j = 1, \dots, n_i$.

Considering above discussion, in this section and hereafter, it is assumed that we are concerned with a classical analysis of variance, where the whole components of the problem such as random variables, hypothesis and parameters of populations are crisp as are introduced in ‘‘Classical

Table 1 Details of ANOVA

Source of variation	SS	Degrees of freedom	MS	F
Between treatments	$SSTR = \sum_{i=1}^r n_i (\bar{Y}_i - \bar{Y}_{..})^2$	$r - 1$	$MSTR = \frac{SSTR}{r-1}$	$F = \frac{MSTR}{MSE} \sim F_{r-1, n_t-r}$
Within treatments (error)	$SSE = \sum_{i=1}^r \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2$	$n_t - r$	$MSE = \frac{SSE}{n_t-r}$	
Total	$SST = \sum_{i=1}^r \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{..})^2$	$n_t - 1$		

ANOVA”. Therefore, the considered ANOVA model is $Y_{ij} = \mu_i + \varepsilon_{ij}$, where $\varepsilon_{ij} \sim N(0, \sigma^2)$ and so Y'_{ij} s are ordinary random variables. But, just one point that will departed from classical ANOVA assumptions in the model (7) is that the sampled observations are fuzzy numbers rather than being real numbers and nothing else is altered in the prior testing ANOVA.

Regarding the introduced distance between two fuzzy numbers in Definition 2.2, the observed values of the statistics SST, SSTR, SSE, MSTR, MSE and F can be calculated by the following extended formulas based on fuzzy observations:

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

$$\tilde{sstr} = \sum_{i=1}^r n_i (\bar{\tilde{y}}_i \ominus \bar{\tilde{y}}_{..})^2 \tag{9}$$

and

$$\tilde{sse} = \sum_{i=1}^r \sum_{j=1}^{n_i} (\tilde{y}_{ij} \ominus \bar{\tilde{y}}_i)^2. \tag{10}$$

In testing ANOVA based on fuzzy numbers, the observed amount of the test statistic is precisely computable by

$$\tilde{f} = \frac{\widetilde{mstr}}{\widetilde{mse}} = \frac{(n_t - r)\tilde{sstr}}{(r - 1)\tilde{sse}}, \tag{11}$$

in which $\widetilde{mstr} = \frac{\tilde{sstr}}{r-1}$ and $\widetilde{mse} = \frac{\tilde{sse}}{n_t-r}$ are respectively treatment mean square and the mean of squares.

Remark 4.1 The major problem in testing ANOVA is calculating the sum of squares of fuzzy observations. In this paper, we have used a logical defuzzification in Eqs. (8–10), since:

- (i) Although the square of a LR fuzzy number can be computed by extension principle, the result is not necessarily in the class of LR fuzzy numbers too [23]. Hence, one is faced with calculating the membership function of the

addition of fuzzy numbers, with not necessarily the same shape functions, based on the Zadeh’s extension principle.

- (ii) Calculation of \tilde{sst} , \tilde{sstr} and \tilde{sse} by Zadeh’s extension principle, will lead us to a very vague observed statistic at the end of computations, since usually the experimenter is faced to $\sum_{i=1}^r n_i$ fuzzy numbers in any testing ANOVA. In other words, as $\sum_{i=1}^r n_i$ increases, then the support of \tilde{sst} , \tilde{sstr} and \tilde{sse} become larger and this means the vagueness of the observed fisher statistic. For instance, see the observed fisher statistic in Sect. 8.2 of [21] which is obtained via Zadeh’s extension principle approach.
- (iii) Following (ii), the truth level of decision making can be low in many applied situations. See definition “the fuzziness of the decision” from [11].

The decision rule Let \tilde{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 , H_0 is accepted if $\tilde{f} \leq F_{1-\alpha; r-1, n_t-r}$; otherwise H_1 is accepted.

Remark 4.2 Considering Remark 2.1, if the observed data for test are real numbers y_{ij} which can be formulated by indicator functions $I_{\{y_{ij}\}}$ for $i = 1, \dots, r$ and $j = 1, \dots, n_i$, then all the introduced extended statistics in Eqs. (8)–(11) coincide to statistics of classical ANOVA.

The rejection region in testing ANOVA is $F > F_{1-\alpha; r-1, n_t-r}$ and so under hypothesis H_0 the error probability is

$$P(F > F_{1-\alpha; r-1, n_t-r}) = \alpha,$$

where $F_{1-\alpha; r-1, n_t-r}$ is α th quantile of F distribution. Therefore, in testing ANOVA based on fuzzy numbers, the p-value can be calculated by

$$p\text{-value} = P(F > \tilde{f})$$

in which \tilde{f} is introduced by (11) as the observed amount of test statistic on the basis of fuzzy observations.

Fast computation formulas

ANOVA for symmetric triangular fuzzy observations

In this subsection, we assume all observations are STFNs. Then the observed values of statistics are obtained by the following theorems in testing ANOVA on the basis of STFNs.

Theorem 5.1 *Based on symmetric triangular fuzzy observations $\tilde{y}_{ij} = T(y_{ij}, s_{y_{ij}}) \in F_T(R)$, $i = 1, 2, \dots, r$, $j = 1, 2, \dots, n_i$, the observed values of \widetilde{sst} , \widetilde{sstr} and \widetilde{sse} in testing ANOVA are the following real numbers:*

$$\widetilde{sst} = sst_y + \frac{2}{(m+2)(m+3)}sst_{s_y}, \tag{12}$$

$$\widetilde{sstr} = sstr_y + \frac{2}{(m+2)(m+3)}sstr_{s_y} \tag{13}$$

and

$$\widetilde{sse} = sse_y + \frac{2}{(m+2)(m+3)}sse_{s_y}, \tag{14}$$

where $sst_y = \sum_{i=1}^r \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{..})^2$, $sstr_y = \sum_{i=1}^r n_i (\bar{y}_{i.} - \bar{y}_{..})^2$ and $sse_y = \sum_{i=1}^r \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i.})^2$ are the observed sum of squares for the mean values of \tilde{y}_{ij} 's, and also $sst_{s_y} = \sum_{i=1}^r \sum_{j=1}^{n_i} (s_{y_{ij}} - \bar{s}_{y..})^2$, $sstr_{s_y} = \sum_{i=1}^r n_i (\bar{s}_{y_{i.}} - \bar{s}_{y..})^2$ and $sse_{s_y} = \sum_{i=1}^r \sum_{j=1}^{n_i} (s_{y_{ij}} - \bar{s}_{y_{i.}})^2$ are the observed sum of squares for spreads of \tilde{y}_{ij} 's.

Proof From Eq. (8), Definition 2.2 and Theorem 2.1, we have

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

where $\bar{y}_{..} = \frac{1}{n_t} \sum_{i=1}^r \sum_{j=1}^{n_i} y_{ij}$, $\bar{s}_{y..} = \frac{1}{n_t} \sum_{i=1}^r \sum_{j=1}^{n_i} s_{y_{ij}}$ and $g(\alpha) = \frac{m+1}{2}\alpha^m$. One can similarly prove Eqs. (13) and (14), by considering $\bar{y}_{i.} = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}$ and $\bar{s}_{y_{i.}} = \frac{1}{n_i} \sum_{j=1}^{n_i} s_{y_{ij}}$ for $i = 1, \dots, r$. \square

Theorem 5.2 *Under the same assumption of Theorem 5.1, the observed amount of ANOVA test statistic is*

$$\tilde{f} = \frac{(m+2)(m+3)mstr_y + 2mstr_{s_y}}{(m+2)(m+3)mse_y + 2mse_{s_y}}, \tag{15}$$

in which $mstr_y = \frac{sstr_y}{r-1}$ and $mse_y = \frac{sse_y}{n_t-r}$ are the observed mean squares of the mean values of \tilde{y}_{ij} 's, and also, $mstr_{s_y} = \frac{sstr_{s_y}}{r-1}$ and $mse_{s_y} = \frac{sse_{s_y}}{n_t-r}$ are the observed mean squares of spreads of \tilde{y}_{ij} 's.

Proof By Eqs. (13 and 14), mean squares can be obtained as

$$\widetilde{mstr} = mstr_y + \frac{2}{(m+2)(m+3)}mstr_{s_y} \tag{16}$$

and

$$\widetilde{mse} = mse_y + \frac{2}{(m+2)(m+3)}mse_{s_y}. \tag{17}$$

Therefore, Eq. (15) can be easily followed from $\tilde{f} = \frac{\widetilde{mstr}}{\widetilde{mse}}$. \square

ANOVA for normal fuzzy observations

In this subsection, we assume all observations are NFNs in testing ANOVA. Then the observed values of statistics in ANOVA model can be easily obtained by the following theorems.

Theorem 5.3 *Based on normal fuzzy observations $\tilde{y}_{ij} = N(y_{ij}, s_{y_{ij}}) \in F_N(R)$, $i = 1, 2, \dots, r$, $j = 1, 2, \dots, n_i$, the observed values of \widetilde{sst} , \widetilde{sstr} and \widetilde{sse} in testing ANOVA are the following real numbers:*

$$\widetilde{sst} = sst_y + \frac{1}{m+1}sst_{s_y}, \tag{19}$$

$$\widetilde{sstr} = sstr_y + \frac{1}{m+1}sstr_{s_y} \tag{20}$$

and

$$\widetilde{sse} = sse_y + \frac{1}{m+1}sse_{s_y}, \tag{21}$$

where sst_y , $sstr_y$ and sse_y are the observed sum of squares for the mean values of \tilde{y}_{ij} 's, and also sst_{s_y} , $sstr_{s_y}$ and sse_{s_y} are the observed sum of squares for spreads of \tilde{y}_{ij} 's.

Proof From Eq. (8) and Theorem 2.2, one can conclude that

$$\begin{aligned} \widetilde{sst} &= \sum_{i=1}^r \sum_{j=1}^{n_i} (\tilde{y}_{ij} \ominus \bar{y}_{..})^2 \\ &= \sum_{i=1}^r \sum_{j=1}^{n_i} [N(y_{ij}, s_{y_{ij}}) \ominus N(\bar{y}_{..}, \bar{s}_{y..})]^2 \end{aligned}$$

$$\begin{aligned}
 &= \sum_{i=1}^r \sum_{j=1}^{n_i} \left[(y_{ij} - \bar{y}_{..})^2 + \frac{1}{m+1} (s_{y_{ij}} - \bar{s}_{y_{..}})^2 \right] \\
 &= \sum_{i=1}^r \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{..})^2 + \frac{1}{m+1} \sum_{i=1}^r \sum_{j=1}^{n_i} (s_{y_{ij}} - \bar{s}_{y_{..}})^2 \\
 &= sst_y + \frac{1}{m+1} sst_{s_y},
 \end{aligned}$$

by considering $g(\alpha) = \frac{m+1}{2}\alpha^m, m = 1, 2, 3, \dots$ in Definition 2.2. Similarly, one can prove Eqs. (20) and (21). \square

Remark 5.1 As a result of Eq. (19) for ANOVA based on normal fuzzy observations, not only increasing sst_y can be caused the increase of \tilde{sst} , but also increasing sst_{s_y} can be caused the increase of \tilde{sst} . Similar results can be presented on the basis of formulas (19) and (20) in ANOVA based on normal fuzzy numbers, and also on the basis of formulas (12)–(14) for ANOVA based on symmetric triangular fuzzy numbers.

Theorem 5.4 *Under the same assumption of Theorem 5.3, the observed value of ANOVA test statistic is*

$$\tilde{f} = \frac{(m+1)mstr_y + mstr_{s_y}}{(m+1)mse_y + mse_{s_y}}, \tag{22}$$

in which $mstr_y$ and mse_y are the observed mean squares of the mean values of \tilde{y}_{ij} 's, and also, $mstr_{s_y}$ and mse_{s_y} are the observed mean squares of spreads of \tilde{y}_{ij} 's.

Proof By Eqs. (11), (19) and (20), mean squares can be obtained as

$$\widetilde{mstr} = mstr_y + \frac{mstr_{s_y}}{m+1} \tag{23}$$

and

$$\widetilde{mse} = mse_y + \frac{mse_{s_y}}{m+1}. \tag{24}$$

Therefore, Eq. (22) can be easily followed from $\tilde{f} = \frac{\widetilde{mstr}}{\widetilde{mse}}$. \square

Remark 5.2 In ANOVA test based on STFNs and NFNs, all the introduced extended statistics in Eqs. (12)–(24) reduce to presented statistics in Sect. 3 of [21] for classical ANOVA, where spreads of \tilde{y}_{ij} 's are a fixed number for $i = 1, \dots, r$ and $j = 1, \dots, n_i$.

Remark 5.3 Testing ANOVA discussed here is different from Wu's approach [28]. His ANOVA method constructed on the basis of the cuts of FRVs, optimistic and pessimistic degrees by optimization. Also, the presented work in [21] is based on extension principle which can lead user to a very fuzzy decision by a very vague fisher statistic in practice. The major

advantages of the presented method in this paper are efficiency for large sampling data and its simplicity in use by clients in the real situations.

In the next section, the presented ANOVA method will be exemplified for fuzzy observations by a relevant study.

A case study on the process of soap production

The aim of this study is to investigate on the amount of solubility for three different kinds of soaps in tepid water. The treatment factor, soap, has been selected to have three levels: regular, deodorant, and moisturizing brands, from the same manufactory. This example is extracted from an experiment done by Suyapa Silvia [3] where the data are fuzzified. It must be stressed that the condition of the presented ANOVA test/problem in this paper is exactly same as the presented ANOVA test in [21], but their solutions are different. Therefore, we do not present the story of experiment process to shorten the length of paper and we refer the readers to Sect. 8.2 of [21] for more details about gathering data. Now, we are going to respond to this question: Is there a significant difference in the weight loss (between three kinds of soaps) because of dissolution in water when allowed to soak for the same time?

There are some unavoidable elements in the experiment reported by Suyapa Silvia [3], such as: (1) inability of the experimenter to cut exactly the same cubes with same weights, (2) limitation of digital laboratory scales of precision of 10 mg, and etc. Hence, a better plan to record the observations is to use STFNs where some of them are definitely non-positive. Therefore, in this real case, we rewrite the data by STFNs as in Table 2, to protect the unavoidable elements described in above. The spread of each observed symmetric triangular fuzzy number is considered as a function of its mean in Table 2. In other words fuzzy observations have been considered by STFNs $\tilde{y}_{ij} = T(y_{ij}, s_{y_{ij}})$, such that $s_{y_{ij}} = |y_{ij}|/10$, for $i = 1, 2, \dots, r$ and $j = 1, 2, \dots, n_i$. Note that in this study some observations are not reported by positive STFNs.

For the presented triangular fuzzy observations in Table 2, we wish to test whether the weight loss of three soaps are the same or not. In other words, we are going to test " $H_0 : \mu_1 = \mu_2 = \mu_3$ ", versus " H_1 : not all μ_i 's are equal, for $i = 1, 2, 3$ ".

Considering Theorems 5.1 and 5.2, one can calculate the observed values of ANOVA statistics which are reported in Table 3, based on the given STFNs. For instance, the total sum of squares is calculated for $m = 1$ by Theorem 5.1 as follows

$$\begin{aligned}
 \tilde{sst} &= sst_y + \frac{2}{(m+2)(m+3)} sst_{s_y} \\
 &= \sum_{i=1}^3 \sum_{j=1}^4 (y_{ij} - \bar{y}_{..})^2 + \frac{2}{3 \times 4} \sum_{i=1}^3 \sum_{j=1}^4 (s_{y_{ij}} - \bar{s}_{y_{..}})^2
 \end{aligned}$$

Table 2 Weight loss for soaps

Type of soap (<i>i</i>)	Fuzzy weight loss (grams) (<i>j</i>)			
Regular	$\tilde{y}_{14} = T(0.40, 0.04)$	$\tilde{y}_{13} = T(-0.14, 0.014)$	$\tilde{y}_{12} = T(-0.10, 0.01)$	$\tilde{y}_{11} = T(-0.3, 0.03)$
Deodorant	$\tilde{y}_{24} = T(3.15, 0.315)$	$\tilde{y}_{23} = T(2.41, 0.241)$	$\tilde{y}_{22} = T(2.61, 0.261)$	$\tilde{y}_{21} = T(2.63, 0.263)$
Moisturizing brand	$\tilde{y}_{34} = T(1.82, 0.182)$	$\tilde{y}_{33} = T(2.26, 0.226)$	$\tilde{y}_{32} = T(2.03, 0.203)$	$\tilde{y}_{31} = T(1.86, 0.186)$

Table 3 Details of ANOVA for soap experiment

Source of variation	\tilde{ss}	Degrees of freedom	\tilde{ms}	\tilde{f}
Between treatments	4.036	2	2.018	
Within treatments (error)	0.695	9	0.077	26.103
Total	16.839	11		

$$\begin{aligned}
 &= \left[(-0.30 - 1.552)^2 + \dots + (1.82 - 1.552)^2 \right] \\
 &\quad + \frac{2}{3 \times 4} \left[(0.03 - 0.164)^2 + \dots + (0.182 - 0.164)^2 \right] \\
 &= 16.817 + \frac{2}{3 \times 4} \cdot 0.134 = 16.839.
 \end{aligned}$$

And, also by Theorem 5.2 the observed value of ANOVA test statistic for $m = 1$ is equal to

$$\begin{aligned}
 \tilde{f} &= \frac{(m + 2)(m + 3) \text{mstr}_y + 2\text{mstr}_{s_y}}{(m + 2)(m + 3) \text{mse}_y + 2\text{mse}_{s_y}} \\
 &= \frac{(3 \times 4 \times 2.0155) + (2 \times 0.016)}{(3 \times 4 \times 0.0772) + (2 \times 556 \times 10^{-6})} \\
 &= 21.103.
 \end{aligned}$$

By comparing the computed ANOVA test statistic, one can accept the alternative hypothesis H_1 at significance level 0.05. The critical value of ANOVA test is $F_{1-\alpha; r-1, n_t-r} = F_{0.95; 2, 9} = 4.256$ and also the computed p -value = 18×10^{-5} shows accuracy of H_1 strongly. Therefore, we conclude that there is a relation between soaps and weight loss based on recorded vague data in Table 2. One can see the solution of this applied case study, but by extension principle, in Sect. 8.2 of [21], which leads the decision maker to a very vague observed fisher statistic, and as presented in Remark 4.1, this matter can be caused lowering the truth level of decision.

Conclusions and future works

In most applied sciences, there are several situations for which non-precise values can be assigned to their experiments. Fuzzy/non-precise numbers are suitable models to formalize the observed values in such situations and the observed data could be presented by the notion of fuzzy sets

to analyze the experiment. The main contribution of this article is extending ANOVA test for fuzzy observations based on a distance between symmetric fuzzy numbers. Several fast computing formulas obtained where the observations are STFNs or NFNs. When all observations are real numbers, the presented ANOVA method reduces to the classical ANOVA approach, since the vagueness of the observed statistics is removed and what remained are only the central points of the statistics, as one can see in Theorems 5.1, 5.3 and Remark 4.2. The proposed ANOVA approach is a meaningful generalization for the classical ANOVA. Efficiency in large sampling is the major advantage of this method, and also this method is easy to be used by professional clients who are familiar with ANOVA, which are some of contributions of the presented manuscript.

For future research work, one can try to use the same approach of this paper to extend other experimental designs such as random block design, Latin square design, etc., where the observations are fuzzy numbers rather than being real numbers. Another interesting topic for future works is to extend the results of this paper for trapezoidal fuzzy numbers and non-symmetric fuzzy numbers, or in general for LR fuzzy numbers.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Cochran WG, Cox GM (1957) Experimental designs, 2nd edn. Wiley, New York
2. Cuevas A, Febrero M, Fraiman R (2004) An ANOVA test for functional data. *Comput Stat Data Anal* 47:111–122
3. Dean A, Voss D (1999) Design and analysis of experiments. Springer, New York

4. De Garibay VG (1987) Behaviour of Fuzzy ANOVA. *Kybernetes* 16(2):107–112
5. Dubois D, Prade H (1980) Fuzzy sets and systems: theory and application. Academic, New York
6. Filzmoser P, Viertl R (2004) Testing hypotheses with fuzzy data: the fuzzy p -value. *Metrika* 59:21–29
7. Gil MA, Montenegro M, González-Rodríguez G, Colubi A, Casals MR (2006) Bootstrap approach to the multi-sample test of means with imprecise data. *Comput Stat Data Anal* 51(1):148–162
8. González-Rodríguez G, Colubi A, Gil MA (2011) Fuzzy data treated as functional data: a one-way ANOVA test approach. *Comput Stat Data Anal* 56:943–955
9. Hesamian G (2016) One-way ANOVA based on interval information. *Int J Syst Sci* 47(11):2682–2690
10. Hocking RR (1996) Methods and applications of linear models: regression and the analysis of variance. Wiley, New York
11. Ivani R, Sanaei Nejad SH, Ghahraman B, Astaraei AR, Feizi H (2016) Fuzzy analysis of variance and its practical application in agriculture. In: Kahraman C, Kabak O (eds) Fuzzy statistical decision-making: theory and applications. Studies in fuzziness and soft computing. Springer, Switzerland, pp 315–327
12. Jiryaei A, Parchami A, Mashinchi M (2013) One-way ANOVA and least squares method based on fuzzy random variables. *Turk J Fuzzy Syst* 4(1):18–33
13. Kalpanapriya D, Pandian P (2012) Fuzzy hypothesis testing of ANOVA model with fuzzy data. *Int J Mod Eng Res* 2:2951–2956
14. Kaya I, Kahraman C (2011) Process capability analyses with fuzzy parameters. *Expert Syst Appl* 38(9):11918–11927
15. Konishi M, Okuda T, Asai K (2006) Analysis of variance based on fuzzy interval data using moment correction method. *Int J Innov Comput Inf Control* 2(1):83–99
16. Lubiano MA, Trutschnig W (2010) ANOVA for fuzzy random variables using the R-package SAFD. *Comb Soft Comput Stat Methods Data Anal Adv Intell Soft Comput* 77:449–456
17. Montenegro M, Colubi A, Casals MR, Gil MA (2004) Asymptotic and bootstrap techniques for testing the expected value of a fuzzy random variable. *Metrika* 59:31–49
18. Montenegro M, Gonzalez-Rodriguez G, Gil MA, Colubi A, Casals MR (2004) Introduction to ANOVA with fuzzy random variables. In: López-Díaz MC, Angeles Gil M, Grzegorzewski P, Hryniewicz O, Lawry J (eds) *Soft Methodol Random Inf Syst*. Springer, Berlin, pp 487–494
19. Montgomery DC (1991) Design and analysis of experiments, 3rd edn. Wiley, New York
20. Nguyen HT, Walker EA (2005) A first course in fuzzy logic, 3rd edn. Chapman Hall/CRC, Paris
21. Nourbakhsh M, Parchami A, Mashinchi M (2013) Analysis of variance based on fuzzy observations. *Int J Syst Sci* 44(4):714–726
22. Parchami A, Ivani R, Mashinchi M, Kaya İ (2017) An implication of fuzzy ANOVA: metal uptake and transport by corn grown on a contaminated soil. *Chemom Intell Lab Syst* 164:56–63
23. Parchami A, Sadeghpour-Gildeh B, Nourbakhsh M, Mashinchi M (2014) A new generation of process capability indices based on fuzzy measurements. *J Appl Stat* 41(5):1122–1136
24. Parchami A, Taheri SM, Mashinchi M (2012) Testing fuzzy hypotheses based on vague observations: a p -value approach. *Stat Pap* 53(2):469–484
25. Rodriguez G, Colubi A, Gil MA (2012) Fuzzy data treated as functional data: a one-way ANOVA test approach. *Comput Stat Data Anal* 56(4):943–955
26. Taheri SM, Arefi M (2009) Testing hypotheses based on fuzzy test statistic. *Soft Comput* 13:617–625
27. Viertl R (2011) Statistical methods for fuzzy data. Wiley, New York
28. Wu HC (2007) Analysis of variance for fuzzy data. *Int J Syst Sci* 38:235–246
29. Xu R, Li C (2001) Multidimensional least-squares fitting with a fuzzy model. *Fuzzy Sets Syst* 119:215–223
30. Zadeh LA (1975) The concept of a linguistic variable and its application to approximate reasoning—I. *Inf Sci* 8:199–249
31. Zadeh LA (1965) Fuzzy sets. *Inf Control* 8:338–359