Proceedings of 2020 IEEE 10th International Conference on Intelligent Systems

Intuitionistic Fuzzy Two-Factor Analysis of COVID-19 Cases in Europe

Velichka Traneva
"Prof. Asen Zlatarov" University
1 "Prof. Yakimov" Blvd,
Burgas 8000, Bulgaria
Email: veleka13@gmail.com

Deyan Mavrov
"Prof. Asen Zlatarov" University
1 "Prof. Yakimov" Blvd,
Burgas 8000, Bulgaria
Email: dg@mavrov.eu

Stoyan Tranev
"Prof. Asen Zlatarov" University
1 "Prof. Yakimov" Blvd,
Burgas 8000, Bulgaria
Email: tranev@abv.bg

***Abstract*—In this paper we apply an intuitionistic fuzzy two-factor ANOVA (2-D IFANOVA), based on the concepts of intuitionistic fuzzy sets (IFSs) and index matrices (IMs), over a unique dataset of daily COVID-19 cases up to 24 June 2020 to explore how the number of COVID-19 cases depends on the "density" and "climate zone" factors for the continent of Europe. In the source data, some information may be missing, unclear or imprecise. To deal with the uncertainty in the data, we apply Intuitionistic fuzzy logic. We also present a new software utility, which performs 2-D IFANOVA by using an implementation of Index matrices. Finally, a comparative analysis of the results obtained by the classical ANOVA and IFANOVA is performed.**

***Keywords*—ANOVA; Climate zone; COVID-19; Density; Intuitionistic fuzzy sets; Index matrix; Software**

I. Introduction

In 2020, the global coronavirus pandemic, that originated in China in 2019, spread initially to South Korea and Japan, and after that to Asia and then the rest of the world. In particular, countries such as Italy, Spain, France, Germany, the United Kingdom and the United States have thus far had some of the latest numbers of confirmed cases and deaths caused by the coronavirus. Many recent papers have aimed to predict how many COVID-19 cases there will be in the future in light of the serious effect this disease has closed on public life. In [2], the authors outlined an approach for the prediction of the epidemic spread of the coronavirus, driven by Chinese New Year-related travel. Analysis of the spatial-temporal distribution of COVID19 in China and a prediction method based on the Logistic model were presented in [29]. Predictions of the impact of epidemic outbreaks on the global supply chains with a simulationbased analysis was presented in [6]. In [20], a propagation analysis and a prediction of the real COVID19 time series are described. A forecasting method for the COVID19 outbreak in China with good results was presented in [21]. Analysis of the spatial spread relationships of the coronavirus pandemic using self-organizing maps was invesigated in [26]. One-way and two-way ANOVA by "age" and "density" of group was applied to number of corona cases in India [25]. The data related to COVID-19 that is available for analysis can be incomplete or unclear. To cope with uncertainty, Zadeh [22] and Atanassov [13] have introduced respectively the concepts of fuzzy(FSs) and intuitionistic fuzzy sets (IFSs). Multiple neural network models with fuzzy aggregation for predicting the COVID19 time series in Mexico were proposed in [27]. A popular method for data analysis is ANOVA, originally developed by Fisher [4], which is concerned with comparing the means of several samples. If the data are not exactly known, the FSs based approaches are incorporated into the decision-making method ANOVA [3]. A bootstrap approach to fuzzy FANOVA (FANOVA) was introduced in [23]. ANOVA using a set of confidence intervals for variance was considered in [11]. FANOVA has proposed by [7] using the levels of pessimistic and optimistic of the triangular fuzzy data. FANOVA has presented in [24], [1] based on Zadehs extension principle. González-Rodrguez el al. [10] have developed an one-way ANOVA test for fuzzy observations in which the fuzzy observations are treated as functional data of a functional Hilbert space. Two-factor ANOVA using Trapezoidal Fuzzy Numbers is explored in [30]. Two-way IFANOVA by converting IFSs to fuzzy sets was proposed in [8]. To analyze vaguer numbers, such as those discussed here, we have extended the classical ANOVA [4] to the one-way and the two-way IFANOVA without replication (see [31], [32], [33]), which can work over Intuitionistic fuzzy (IF) data, based on Intuitionistic fuzzy sets (IFSs) and Index matrices (IMs).

We also introduce a command-line utility for the application of 2-D IFANOVA, which uses a software implementation of IMs to calculate the results. The utility is then applied to find the dependencies of the COVID-19 case notification rate per 100 000 people up to 24 June 2020 on the "density" and "climate zone" factors of the European countries.

The rest of this paper is structured as follows: Section 2 describes some basic definitions of the concepts of IMs and IF logic. Section 3 discusses the classical two-factor (2-F) ANOVA and its application over the actual data on COVID-19 cases. In Section 4, we describe the algorithm of 2-D IFANOVA [31] and its software realization, then use it to investigate the effect of the factors "density" and "climate zone". The results obtained from IFANOVA are compared with those from the classical ANOVA. Section 5 gives some conclusions and outlines aspects for future research.

II. Short remarks on IMs and intuitionistic fuzzy logic

This section provides some remarks on Intuitionistic fuzzy logic (see [9], [15], [17], [18], [19], [34]) and on IMs (see [16], [36]).

2.1. Short Notes on Intuitionistic Fuzzy (IF) Logic

An **Intuitionistic Fuzzy Pair (IFP)** is an object of the form $\langle a,b\rangle = \langle \mu(p),\nu(p)\rangle$, where $a,b \in [0,1]$ and $a+b \leq 1$, that is used as an evaluation of a proposition p (see [17], [18]). $\mu(p)$ and $\nu(p)$ respectively determine the "truth degree" (degree of membership) and "falsity degree" (degree of non-membership).

Let us have two IFPs $x = \langle a,b\rangle$ and $y = \langle c,d\rangle$. We will recall some basic operations:

$$\begin{gathered}\neg x = \langle b,a\rangle;\\ x \wedge_1 y = \langle \min(a,c), \max(b,d)\rangle;\\ x \vee_1 y = \langle \max(a,c)), \min(b,d)\rangle;\\ x \wedge_2 y = x+y = \langle a+c-a.c, b.d\rangle;\\ x \vee_2 y = x.y = \langle a.c, b+d-b.d\rangle;\\ \alpha.x = \langle 1-(1-a)^\alpha, b^\alpha\rangle (\alpha\in R);\\ x-y = \langle \max(0,a-c), \min(1,b+d,1-a+c)\rangle\\ x:y = \langle \min(1;a/c);\\ \min(\max(0;1-a/c);\max(0;(b-d)/(1-d)))\end{gathered} \tag{1}$$

and relations with IFPs

$$\begin{array}{ll} x\geq y \text{ iff } a\geq c \text{ and } b\leq d, & x\leq y \text{ iff } a\leq c \text{ and } b\geq d,\\ x\geq_\Box y \text{ iff } a\geq c, & x\leq_\Box y \text{ iff } a\leq c,\\ x\geq_\Diamond y \text{ iff } b\leq d, & x\leq_\Diamond y \text{ iff } b\geq d,\\ x=y & \text{iff } a=c \text{ and } b=d,\\ x\geq_R y & \text{iff } R_{\langle a,b\rangle}\leq R_{\langle c,d\rangle}, \end{array} \tag{2}$$

where following [9]

$$R_{\langle a,b\rangle} = 0,5(2-a-b)0,5(|1-a|+|b|+|1-a-b|).$$

2.2. Definition and Basic Operations over Intuitionistic Fuzzy Index Matrices

Let $\mathcal{I}$ be a fixed set. By **two-dimensional Intuitionistic fuzzy index matrix (2-D IFIM)** with index sets K and L ($K,L\subset\mathcal{I}$), we denote the object:

$$[K,L,\{\langle \mu_{k_i,l_j},\nu_{k_i,l_j}\rangle\}]$$

$$\equiv \begin{array}{c|ccccc} & l_1 & \dots & l_j & \dots & l_n\\ \hline k_1 & \langle \mu_{k_1,l_1},\nu_{k_1,l_1}\rangle & \dots & \langle \mu_{k_1,l_j},\nu_{k_1,l_j}\rangle & \dots & \langle \mu_{k_1,l_n},\nu_{k_1,l_n}\rangle\\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots\\ k_m & \langle \mu_{k_m,l_1},\nu_{k_m,l_1}\rangle & \dots & \langle \mu_{k_m,l_j},\nu_{k_m,l_j}\rangle & \dots & \langle \mu_{k_m,l_n},\nu_{k_m,l_n}\rangle \end{array},$$

where for every $1\leq i\leq m, 1\leq j\leq n$: $0\leq \mu_{k_i,l_j},\nu_{k_i,l_j},\mu_{k_i,l_j}+\nu_{k_i,l_j}\leq 1$.

Following [16], we recall some operations over two IMs $A = [K,L,\{\langle \mu_{k_i,l_j},\nu_{k_i,l_j}\rangle\}]$ and $B=[P,Q,\{\langle \rho_{p_r,q_s},\sigma_{p_r,q_s}\rangle\}]$.

Negation: $\neg A = [K,L,\{\langle \nu_{k_i,l_j},\mu_{k_i,l_j}\rangle\}]$.

Addition-$(\circ,*)$**:** $A \oplus_{(\circ,*)} B = [K\cup P, L\cup Q,\{\langle \phi_{t_u,v_w},\psi_{t_u,v_w}\rangle\}]$, where

$$\langle \phi_{t_u,v_w},\psi_{t_u,v_w}\rangle = \begin{cases} \langle \mu_{k_i,l_j},\nu_{k_i,l_j}\rangle, & \text{if } t_u=k_i\in K \text{ and } v_w=l_j\in L-Q\\ & \text{or } t_u=k_i\in K-P \text{ and } v_w=l_j\in L;\\ \langle \rho_{p_r,q_s},\sigma_{p_r,q_s}\rangle, & \text{if } t_u=p_r\in P \text{ and } v_w=q_s\in Q-L\\ & \text{or } t_u=p_r\in P-K\\ & \text{and } v_w=q_s\in Q;\\ \langle \circ(\mu_{k_i,l_j},\rho_{p_r,q_s}), & \text{if } t_u=k_i=p_r\in K\cap P\\ *(\nu_{k_i,l_j},\sigma_{p_r,q_s})\rangle, & \text{and } v_w=l_j=q_s\in L\cap Q;\\ \langle 0,1\rangle, & \text{otherwise}\end{cases}$$

and $\langle\circ,*\rangle\in\{\langle\max,\min\rangle,\langle\min,\max\rangle,\langle \text{ average,average}\rangle\}$.

Termwise subtraction-(max,min):

$$A -_{(\max,\min)} B = A\oplus_{(\max,\min)}\neg B.$$

Termwise multiplication-$(\min,\max)$:

$$A\otimes_{(\min,\max)} B = [K\cap P, L\cap Q,\{\langle \phi_{t_u,v_w},\psi_{t_u,v_w}\rangle\}],$$

where $\langle \phi_{t_u,v_w},\psi_{t_u,v_w}\rangle = \langle \min(\mu_{k_i,l_j},\rho_{p_r,q_s}),\max(\nu_{k_i,l_j},\sigma_{p_r,q_s})\rangle$.

Reduction: We use symbol "$\perp$" for lack of some component in the separate definitions. The operations $(k,\perp)$-reduction of a given IM A is defined by:

$$A_{(k,\perp)} = [K-\{k\}, L,\{c_{t_u,v_w}\}],$$

where $c_{t_u,v_w}=a_{k_i,l_j}$ for $t_u=k_i\in K-\{k\}$ and $v_w=l_j\in L$.

Projection: Let $M\subseteq K$ and $N\subseteq L$. Then,

$$pr_{M,N}A = [M,N,\{b_{k_i,l_j}\}],$$

where for each $k_i\in M$ and each $l_j\in N$, $b_{k_i,l_j}=a_{k_i,l_j}$.

Substitution: Let IM $A=[K,L,\{a_{k,l}\}]$ be given. Local substitution over the IM is defined for the couples of indices (p,k) and/or (q,l), respectively, by

$$\left[\frac{p}{k};\perp\right]A = [(K-\{k\})\cup\{p\}, L,\{a_{k,l}\}],$$

$$\left[\perp;\frac{q}{l}\right]A = [K,(L-\{l\})\cup\{q\},\{a_{k,l}\}].$$

Aggregation operations

Let $x\#_@ y = \langle average(a,c), average(b,d)\rangle$, where $x=\langle a,b\rangle$ and $y=\langle c,d\rangle$ are IFPs. Let $k_0\notin K$ be a fixed index. The aggregation operation by K is [16], [35]:

$$\alpha_{K,\#_@}(A,k_0) = \begin{array}{c|ccc} & l_1 & \dots & l_n\\ \hline k_0 & \underset{i=1}{\overset{m}{\#_@}}\langle \mu_{k_i,l_1},\nu_{k_i,l_1}\rangle & \dots & \underset{i=1}{\overset{m}{\#_@}}\langle \mu_{k_i,l_n},\nu_{k_i,l_n}\rangle\end{array}. \tag{3}$$

Aggregate global internal operation [36]: $AGIO_{\oplus_{(\max,\min)}}(A)$.

Internal subtraction of IMs' components [36]: $IO_{-_{(\max,\min)}}(\langle k_i,l_j,A\rangle,\langle p_r,q_s,B\rangle) = [K,L,\{\langle \gamma_{t_u,v_w},\delta_{t_u,v_w}\rangle\}]$.

III. Application of Nonreplicated 2-D ANOVA to the COVID-19 Cases in Europe

A. Nonreplicated 2-D ANOVA

Let x_{k_i,l_j} for $i = 1, 2, ..., m$ and $j = i_1, i_2, ...i_I (1 \leq i_I \leq n)$ denote the data from the k_i-th level of factor A and l_j-th level of factor B.

The sum of squared deviations about the mean - SST, between rows sum of squares (effect of factor A) - SSA, between columns sum of squares (effect of factor B) - SSB and error sum of squares SSE are calculated. The mean sum of squares MSA (factor A), the mean sums of squares MSB (factor B) and for error MSE are follow [4]:

$$MSA = \frac{SSA}{m-1}, MSB = \frac{SSB}{n-1}, MSE = \frac{SSE}{(m-1)(n-1)}.$$

Let $N* = (m-1)(n-1)$. If the test statistics

$$F_A = \frac{MSA}{MSE} \geq F_{(\alpha,m-1,N*)} \text{ or } \frac{1}{F_A} \leq \frac{1}{F_{(\alpha,m-1,N*)}} = F_{(1-\alpha,N*,m-1)} \quad (4)$$

and

$$F_B = \frac{MSB}{MSE} \geq F_{(\alpha,n-1,N*)} \text{ or } \frac{1}{F_B} \leq \frac{1}{F_{(\alpha,n-1,N*)}} = F_{(1-\alpha,N*,n-1)}, \quad (5)$$

where $F_{(\alpha,m-1,N*)}$ and $F_{(\alpha,n-1,N*)}$ are $\alpha-$quantile of $F-$distribution then the factors effect on significance level α (see [4], [28]).

B. 2-D ANOVA over COVID-19 Cases in Europe

Let us apply 2-D ANOVA using Excel on the dataset containing the COVID-19 case notification rate per 100 000 people of the European countries up to 24 June 2020 (see [37], [38]). Let x_{k_i,l_j} for $i = 1, 2, ..., 5$ and $j = i_1, i_2, ...i_I (1 \leq i_I \leq n)$ denote the data from the k_i-th level of factor "density" (people per km^2) and l_j-th level of factor "climate zone". There are five levels of the "density" factor ($[1-200]$, $[201-400]$, $[401-600]$, $[601-1000]$ and $[1001-20000]$) and four levels of the "climate zone" factor (subtropical, tempered, subpolar and mountainous areas) in our data. The following ANOVA Table I is obtained by 2-D ANOVA with $\alpha = 0,05$:

TABLE I
2-D ANOVA table by the factors "density" and "climate zone".

Source	SS	df	MS	F	p-value	F crit
Rows	733519	4	183380	0,96	0,47	3,26
Columns	2431052	3	810351	4,22	0,03	3,49
Error	2303694	12	191974			

The conclusion of the ANOVA is that only the "climate zone" factor affects the number of COVID-19 cases. The COVID-19 case notification rates per 100 000 people are the highest in the subtropical climate, and they are the lowest in the subpolar climate. The number of cases is the highest in the countries in Europe with a density of $[401-600]$ and a subtropical climate. Fig. 1 shows a comparison between the average COVID-19 case notification rates of the European countries based on their climate zone and density:

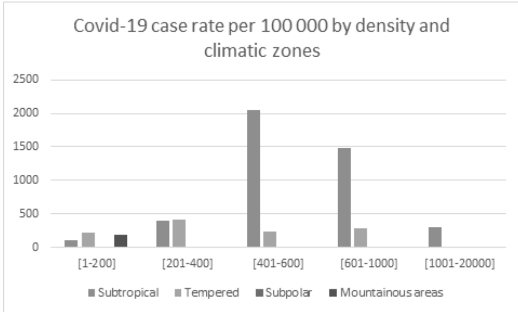

Fig. 1. COVID-19 case notification rate per 100 000 depend on "density" and "climatic zones".

A similar approach is applied in [25]. First, one-way ANOVA is used to analyse data from India and the conclusion made is that is that the population density of an area alone does not have a great effect on number of COVID-19 cases. On the other hand, two-way ANOVA shows that the density and the age group of the infected taken together have a much bigger impact on the number of cases.

IV. Intuitionistic Fuzzy Two-Way ANOVA over COVID-19 cases in Europe

In [31] a 2-D IFANOVA without replication is proposed, which uses the concepts of IMs and fuzzy logic. We describe this analysis shortly using pseudocode.

A. Two-factor IFANOVA

Step 1. The IF IM (IFIM) $X[K, L]$ is created, whose elements are the measured values according to the different levels of the studied two factors as follows:

$$\begin{array}{c|cccc} & l_1 & \ldots & l_n & \ldots \\ \hline k_1 & \langle \mu_{k_1,l_1}, \nu_{k_1,l_1} \rangle & \ldots & \langle \mu_{k_1,l_n}, \nu_{k_1,l_n} \rangle & \ldots \\ \vdots & \vdots & \ddots & \vdots & \ddots \\ k_m & \langle \mu_{k_m,l_1}, \nu_{k_m,l_1} \rangle & \ldots & \langle \mu_{k_m,l_n}, \nu_{k_m,l_n} \rangle & \ldots \\ Sr_1 & \langle \mu_{Sr_1,l_1}, \nu_{Sr_1,l_1} \rangle & \ldots & \langle \mu_{Sr_1,l_n}, \nu_{Sr_1,l_n} \rangle & \ldots \\ Sr & \langle \mu_{Sr,l_1}, \nu_{Sr,l_1} \rangle & \ldots & \langle \mu_{Sr,l_n}, \nu_{Sr,l_n} \rangle & \ldots \end{array}$$

$$\begin{array}{ccc} \ldots & Sr_2 & Sr \\ \hline \ldots & \langle \mu_{k_1,Sr_2}, \nu_{k_1,Sr_2} \rangle & \langle \mu_{k_1,Sr}, \nu_{k_1,Sr} \rangle \\ \ddots & \vdots & \vdots \\ \ldots & \langle \mu_{k_m,Sr_2}, \nu_{k_m,Sr_2} \rangle & \langle \mu_{k_m,Sr}, \nu_{k_m,Sr} \rangle \\ \ldots & \langle \mu_{Sr_1,Sr_2}, \nu_{Sr_1,Sr_2} \rangle & \langle \mu_{Sr_1,Sr}, \nu_{Sr_1,Sr} \rangle \\ \ldots & \langle \mu_{Sr,Sr_2}, \nu_{Sr,Sr_2} \rangle & \langle \mu_{Sr,Sr}, \nu_{Sr,Sr} \rangle \end{array},$$

where $\{k_1, k_2, \ldots, k_m\}$ are the factor A levels, $\{l_1, l_2, \ldots, l_n\}$ are the factor B levels and for $1 \leq i \leq m$, $1 \leq j \leq n$, $\{x_{k_i,l_j}, x_{k_i,Sr_2}, x_{k_i,Sr}, x_{Sr_1,l_j}, x_{Sr,l_j}\}$ are IFPs.

$x_{k_i,l_j} (1 \leq i \leq m,\ 1 \leq j \leq n)$ is the value according to k_i-th level of A and l_j-th level of B. We use symbol "$\perp$" for empty cells of IM X. The other elements of X at the beginning of the algorithm are equal to $\langle\perp,\perp\rangle$. Let us define the auxiliary IM $S=[K,L,\{s_{k_i,l_j}\}]$, such that $S=X$ i.e. $(s_{k_i,l_j}=x_{k_i,l_j}\ \ \forall k_i\in K,\forall l_j\in L)$.

Then we define these IMs:

$$S_1[K/\{Sr_1,Sr\},L/\{Sr_2,Sr\}]=pr_{K/\{Sr_1,Sr\},L/\{Sr_2,Sr\}}S,$$

$$S_2[K/\{Sr_1,Sr\},\{Sr_2\}]=\alpha_{L,\#_@}(S_1,Sr_2),$$

$$S_2^*[\{Sr_1\},L/\{Sr_2,Sr\}]=\alpha_{K,\#_@}(S_1,Sr_1).$$

A new form of IM S is obtained as follows:

$$S:=S\oplus_{(\max,\min)}S_2\oplus_{(\max,\min)}S_2^*,$$

go to *Step 2*.

Step 2. We calculate the mean of the sample by the operation (3)

$$S_3[k_0,\{Sr_t\}]=\alpha_{K,\#_@}(S_2,k_0)(Sr_t\notin L/\{Sr_2,Sr\}).$$

for $i=1$ to m

$$\{S:=S\oplus_{(\max,\min)}\left[\frac{k_i}{k_0};\frac{Sr}{Sr_t}\right]S_3\}.$$

for $j=1$ to n

$$\{S:=S\oplus_{(\max,\min)}\left[\frac{Sr}{k_0};\frac{l_j}{Sr_t}\right]S_3\}.$$

We define

$$S_{4a}[K/\{Sr_1,Sr\},L/\{Sr_2,Sr\},a_{k_i,l_j}=\langle 0,1\rangle].$$

From each element of the matrix S_{4a}, subtract the means of the data of corresponding row and column, and add the total mean of the IM S:

for $j=1$ to n

for $i=1$ to m

$$\{S_{4a}:=IO_{-_{(\max,\min)}}\left(\langle k_i,l_j,S_{4a}\rangle,\langle Sr,l_j,S\rangle\right);$$

$$S_{4a}:=IO_{-_{(\max,\min)}}\left(\langle k_i,l_j,S_{4a}\rangle,\langle k_i,Sr,S\rangle\right)\}.$$

We define

$$S_{4b}[K/\{Sr_1,Sr\},L/\{Sr_2,Sr\},a_{k_i,l_j}=\langle 0,1\rangle].$$

for $j=1$ to n

for $i=1$ to m

$$\{S_{4b}:=IO_{-_{(\max,\min)}}\left(\langle k_i,l_j,S_{4b}\rangle,\langle Sr_1,l_j,S\rangle\right);$$

$$S_{4b}:=IO_{-_{(\max,\min)}}\left(\langle k_i,l_j,S_{4b}\rangle,\langle k_i,Sr_2,S\rangle\right)\}$$

$$S_4:=S_{4a}\oplus_{-}S_{4b}.$$

Go to *Step 3*.

Step 3. We calculate the mean sums MSA (for factor A) and MSB (for factor B) by the operations:

$$S_5[K/\{Sr_1,Sr\},\{Sr_2\}]=pr_{K/\{Sr_1,Sr\},\{Sr_2\}}S \text{ and}$$

$$S_6[K/\{Sr_1,Sr\},\{Sr\}]=pr_{K/\{Sr_1,Sr\},\{Sr\}}S;$$

Let $Dif_1\notin K\cup L$ then

$$S_7[K/\{Sr_1,Sr\},Dif_1]=\left[\perp;\frac{Dif_1}{Sr_2}\right]S_5-_{(\max,\min))}\left[\perp;\frac{Dif_1}{Sr}\right]S_6;$$

$$MSA=\frac{n}{m-1}AGIO_{\oplus_{(\max,\min))}}\left(S_7\otimes_{(\min,\max)}S_7\right);$$

$$S_8[\{Sr_1\},L/\{Sr_2,Sr\}]=pr_{\{Sr_1\},L/\{Sr_2,Sr\}}S \text{ and}$$

$$S_9[\{Sr\},L/\{Sr_2,Sr\}]=pr_{\{Sr\},L/\{Sr_2,Sr\}}S;$$

Let $Dif_2\notin K\cup L$ then

$$S_{10}[Dif_2,L/\{Sr_2,Sr\}]=\left[\frac{Dif_2}{Sr_1};\perp\right]S_8-_{(\max,\min))}\left[\frac{Dif_2}{Sr};\perp\right]S_9;$$

$$MSB=\frac{m}{n-1}AGIO_{\oplus_{(\max,\min))}}\left(S_{10}\otimes_{(\min,\max)}S_{10}\right);$$

Go to *Step 4*.

Step 4. We calculate:

$$\frac{1}{F_A}=\frac{MSE}{MSA},\frac{1}{F_B}=\frac{MSE}{MSB}.$$

The fuzzy estimators of the ANOVA key statistics $F_{t_A fuzzy(1-\alpha,N*,m-1)}$ and $F_{t_B fuzzy(1-\alpha,N*,n-1)}$ values are obtained using Pietraszeks approach ([12],2016).

Fig. 2 shows the fuzzy estimators of the bootstrapped F statistics:

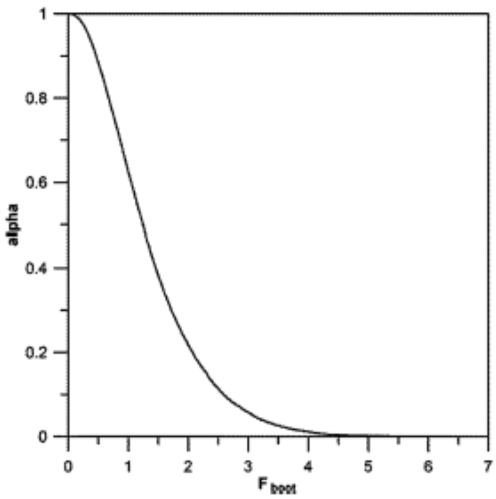

Fig. 2. Fuzzy estimator of the bootstrapped F statistics

If

$$\frac{1}{F_A}\leq F_{t_A fuzzy(1-\alpha,N*,m-1)} \text{ and}$$

$$\frac{1}{F_B}\leq F_{t_B fuzzy(1-\alpha,N*,n-1)}$$

in accordance with the relations (2) and (4), then the factors effect on significance level α.

End of algorithm.

B. Software utility for 2-D IFANOVA

In order to apply the 2D-IFANOVA algorithm to real data more quickly, a software utiity was developed. It is written in C++ and uses the index matrix template class (*IndexMatrix¡T¿*) described in [5], which implements the basic IM operations, such as addition, multiplication, termwise multiplication, termwise subtraction, projection and substitution, most of which are briefly detailed in Sect. 2.2. As this class was originally written to handle IMs of real numbers and integers, a new class representing intuitionistic fuzzy pairs (*IFPair*) was developed, with methods realising the operations on them (see Sect. 2.1). Several modifications and corrections had to be made to the IM class so that it could work properly with IFPs. Additional code was written for the $AGIO$, internal subtraction and average IFP aggregation operations, as these had not been defined at the time of the index matrix class' original development. Then the main program was written to calculate the results using IM objects containing IFPs (*IndexMatrix¡IFPair¿*), which thus represent IFIMs.
The utility takes its input in the form of a tab-separated text file, containing the data of a single IFIM. As of now, the input data must be pre-transformed to IFPs using the methods described previously, because the utility does not at this point handle that step. Column headings must be present in the first row, where the first cell is left empty. Each row after that must start with the row heading in the first cell, followed by a list of IFP values, where each IFP is represented by two decimal values separated by a semicolon (corresponding to the μ and ν of the IFP, respectively). The number of cells must correspond to the number of column headings; empty cells are read as $\langle 0, 1\rangle$ IFPs. The results, which include MSE, MSA, MSB, $\frac{1}{F_A}$ and $\frac{1}{F_A}$, are then printed on the console. Optionally, the user can use a "verbose" option to see the interim IMs used to calculate the final values.

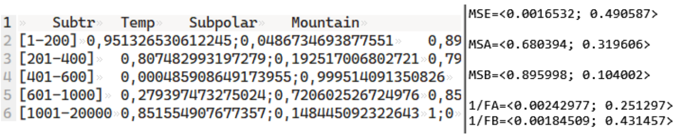
```
      Subtr    Temp    Subpolar    Mountain
[1-200] 0,951326530612245;0,0486734693877551    0,89
[201-400]    0,807482993197279;0,192517006802721 0,79
[401-600]    0,000485908649173955;0,999514091350826
[601-1000]   0,279397473275024;0,720602526724976 0,85
[1001-20000 0,851554907677357;0,148445092322643 1;0
```
```
MSE=<0.0016532; 0.490587>
MSA=<0.680394; 0.319606>
MSB=<0.895998; 0.104002>
1/FA=<0.00242977; 0.251297>
1/FB=<0.00184509; 0.431457>
```

Fig. 3. A snippet from an input file and the corresponding results from the software utility

C. Application of the software utility over COVID-19 cases in Europe

The effectiveness of the proposed 2-D IFANOVA method is tested by applying it to detect dependencies between the COVID-19 case rate per 100 000 people up to 24 June 2020 [37], [38], converted to IFPs, and the factors "density (people per km^2)" and "climate zone". At the beginning of the algorithm we transform the data values with COVID-19 cases into IFPs. Let us have the set of intervals $[i_1, i_I]$ for $1 \leq i \leq m$ and let $A_{min,i} = \min_{i_1 \leq j \leq i_I} x_{i,j} < \max_{i_1 \leq j \leq i_I} x_{i,j} = A_{max,i}$. For the interval $[i_1, i_I]$ we construct IFPs [15] as follows:

$$\mu_{i,j} = \frac{x_{i,j} - A_{min,i}}{A_{max,i} - A_{min,i}}, \nu_{i,j} = \frac{A_{max,i} - x_{i,j}}{A_{max,i} - A_{min,i}}. \quad (6)$$

The conditions $0 \leq \mu_{i,j}, \nu_{i,j} \leq 1, \leq 0 \leq \mu_{i,j} + \nu_{i,j} \leq 1$ are satisfied.

In case there is unclear or missing data about the number of COVID-19 cases, we can apply the expert approach described in detail in [15] to convert the data to IFPs.

The IFIM $X[K, L]$ is created, the elements of which are the values measured according to the different levels of the two studied factors using (6). The initial form of the IM X without the last two rows and columns is:

$$X[K/\{Sr_2, Sr\}, L/\{Sr_1, Sr\}]$$

	$Subtropical$	$Tempered$	$\dots$
$1-200$	$\langle 0.95, 0.048\rangle$	$\langle 0.89, 0.10\rangle$	$\dots$
$201-400$	$\langle 0.80, 0.19\rangle$	$\langle 0.79, 0.20\rangle$	$\dots$
$401-600$	$\langle 0.0004, 0.99\rangle$	$\langle 0.88, 0.11\rangle$	$\dots$
$601-1000$	$\langle 0.27, 0.72\rangle$	$\langle 0.85, 0.14\rangle$	$\dots$
$1000-2000$	$\langle 0.85, 0.14\rangle$	$\langle 1, 0\rangle$	$\dots$

$\dots$	$Subpolar$	$Mountainous\ \ areas$
$\dots$	$\langle 0.99, 0.0004\rangle$	$\langle 0.91, 0.08\rangle$
$\dots$	$\langle 1, 0\rangle$	$\langle 1, 0\rangle$
$\dots$	$\langle 1, 0\rangle$	$\langle 1, 0\rangle$
$\dots$	$\langle 1, 0\rangle$	$\langle 1, 0\rangle$
$\dots$	$\langle 1, 0\rangle$	$\langle 1, 0\rangle$

After application of the IFANOVA, presented in Sect. IV above, we found that:

$$MSA = \langle 0.680394, 0.319606\rangle, MSB = \langle 0.895998, 0.104002\rangle;$$

$$MSE = \langle 0.0016532, 0.490587\rangle.$$

We then applied Pietraszeks approach ([12],2016), which in turn gave us the fuzzy estimators of the ANOVA Ft_A and Ft_B values: The classic value $Ft_A(0.95; 12; 4) = 0.31$ is tied to the fuzzy assessment $F_{fuzzy(0.95;12;4)} = \langle 0.95, 0\rangle$, while the value $Ft_B(0.95; 12; 3) = 0.29$ is tied to the fuzzy assessment $F_{fuzzy(0.95;12;3)} = \langle 0.96, 0\rangle$.

Therefore

$$\frac{MSE}{MSA} = \frac{1}{F_A} = \langle 0.00242977, 0.251297\rangle$$
$$\leq F_{fuzzy(0.95;12;4)} = \langle 0.95, 0\rangle$$
$$\text{and } \frac{MSE}{MSB} = \frac{1}{F_B} = \langle 0.00184509, 0.431457\rangle$$
$$\leq F_{fuzzy(0.95;12;3)} = \langle 0.96, 0\rangle$$

in accordance with the relation (2). From (4) and the above we can conclude that the "density" and "climate zone" factors do indeed have an effect on the number of COVID-19 cases.

A comparison of the results of 2-D IFANOVA with those obtained by 2-D ANOVA shows that they differ in the influence of the "density" factor, which could be explained by the high degree of the hesitancy $\pi = 0.75$ of the IFP

$$\langle 0.00242977, 0.251297 \rangle = \frac{1}{F_A}.$$

V. Conclusion

In this paper, a new software utility for 2-D IFANOVA without replication, based on the concepts of IMs and IFSs, was presented. The utility was then used to apply the 2-D IFANOVA to analyse the impact of the "density" and "climate zone" factors on COVID-19 cases in the continent of Europe using real data on cases per 100 000 people up to 24 June 2020. A comparative analysis was done on the results obtained by the classical ANOVA [4] and the IFANOVA. In the future, the outlined approach for 2-D IFANOVA and the software utility can be expanded to 2-D IFANOVA with replication [4].

Acknowledgment

The authors would like to express their gratitude for the support received through the "Index matrices as a tool for knowledge extraction" project, registered at Prof. Dr. Asen Zlatarov University under Ref. No. NIX-440/2020.

References

[1] A. Parchami, M. Nourbakhsh, M. Mashinchi, " Analysis of variance in uncertain environments", *Complex Intell. Syst.,* vol. 3 (3), 2017, pp. 189-196.

[2] C. Fan, L. Liu, W. Guo, A. Yang, C. Ye, M. Jilili, M. Ren, P. Xu, H. Long, Y. Wang, "Prediction of Epidemic Spread of the 2019 Novel Coronavirus Driven by Spring Festival Transportation in China: A PopulationBased Study," *Int. J. Environ. Res. Public Heal.,* vol. 17, 2020, pp. 1679-1685.

[3] C. Kahraman, K. Ozgur, *Fuzzy statistical decision-making: theory and applications,* Studies in fuzziness and soft computing, Cham: Springer, 2016.

[4] D. Doane, L. Seward, *Applied statistics in business and economics,* New York, USA: McGraw-Hill Education, 2016.

[5] D. Mavrov, *Software Implementation and Applications of Index Matrices,* Dissertation. Prof. Dr. Asen Zlatarov University, Burgas, 2016 (in Bulgarian).

[6] D. Ivanov, "Predicting the impacts of epidemic outbreaks on global supply chains: A simulationbased analysis on the coronavirus outbreak case," *Transp. Res. Part. E: Logist. Transp. Rev.,* vol. 136, 2020, pp. 101922.

[7] D. Kalpanapriya, P. Pandian, "Fuzzy hypotesis testing of ANOVA model with fuzzy data", *Int. J. Mod. Eng. Res.,* vol. 2 (4), 2012, pp. 29512956.

[8] D. Kalpanapriya, M. Unnissa, "Intuitionistic fuzzy ANOVA and its application using different techniques", *in: Madhu V., Manimaran A., Easwaramoorthy D., Kalpanapriya D., Mubashir Unnissa M. (eds),* Advances in Algebra and Analysis. Trends in Mathematics, Birkhuser, Cham, 2017, pp. 457-468.

[9] E. Szmidt, J. Kacprzyk, "Amount of information and its reliability in the ranking of Atanassovs intuitionistic fuzzy alternatives," *in: Rakus-Andersson, E., Yager, R., Ichalkaranje, N., Jain, L.C. (eds.),* Recent Advances in Decision Making, SCI, Heidelberg: Springer, vol. 222, 2009, pp. 719.

[10] G. González-Rodríguez, A. Colubi, A. Gil, "Fuzzy data treated as functional data: a one-way ANOVA test approach," *Comput. Stat. Data Anal.*, vol. 56, 2012, pp. 943955.

[11] J. Buckley, *Fuzzy probability and statistics,* Berlin Heidelberg: Springer, 2006.

[12] J. Pietraszek, M. Koomycki, etc. "The Fuzzy Approach to Assessment of ANOVA Results," *in: Nguyen NT., etc. (eds) Computational Collective Intelligence*, ICCCI 2016, LNCS, Springer, Cham, vol. 9875, 2016, pp. 260-268.

[13] K. Atanassov, "Intuitionistic Fuzzy Sets," VII ITKR Session, Sofia, 20-23 June 1983 (Deposed in Centr. Sci.-Techn. Library of the Bulg. Acad. of Sci., 1697/84) (in Bulgarian). Reprinted: *Int. J. Bioautomation,* vol. 20(S1), 2016, pp. S1-S6.

[14] K. Atanassov, "Generalized index matrices," *Comptes rendus de l'Academie Bulgare des Sciences,* vol. 40(11), 1987, pp. 15-18.

[15] K. Atanassov, *On Intuitionistic Fuzzy Sets Theory,* Heidelberg, Springer: STUDFUZZ, vol. 283, 2012.

[16] K. Atanassov, *Index Matrices: Towards an Augmented Matrix Calculus. Studies in Computational Intelligence*, Springer, Cham, vol. 573, 2014.

[17] K. Atanassov, "Intuitionistic Fuzzy Logics," *Studies in Fuzziness and Soft Computing,* Springer, vol. 351, 2017.

[18] K. Atanassov, E. Szmidt, J. Kacprzyk, "On intuitionistic fuzzy pairs," *Notes on Intuitionistic Fuzzy Sets,* vol. 19 (3), 2013, pp. 1-13.

[19] K. Atanassov, "Remark on an intuitionistic fuzzy operation division," *Annual of "Informatics". Section, Union of Scientists in Bulgaria* vol. 10, 2019 (in press).

[20] L. Li, Z. Yang, Z. Dang, C. Meng, J. Huang, H. Meng, D. Wang, G. Chen, J. Zhang, H. Peng, et al., "Propagation analysis and prediction of the COVID19," *Infect. Dis. Model,* vol. 5, 2020, pp. 282292.

[21] Q. Li, W. Feng, "Trend and forecasting of the COVID19 outbreak in China,".*J. Infect.,* vol. 80, 2020, pp. 469496.

[22] L. Zadeh, *Fuzzy Sets,* Information and Control, vol. 8 (3), 338-353, 1965.

[23] M. Gil,M. Montenegro, G. González-Rodríguez, A. Colubi, M. Casals, "Bootstrap Approach to the Multi-sample Test of Means with Imprecise Data," *Computer Statistics and Data Analysis,* vol. 51, 2006, pp. 14816.

[24] M. Nourbakhsh, A. Parchami, M. Mashinchi, "Analysis of variance based on fuzzy observations", *Int. J. Syst. Sci.,* vol. 44 (4), 2013, pp. 714-726.

[25] P. Anwla, "Introduction to ANOVA for Statistics and Data Science (with COVID-19 Case Study using Python)", *Analtics Vidhua*, https://www.analyticsvidhya.com/blog/2020/06/introduction-anova-statistics-data-science-covid-python, 2020.

[26] P. Melin, J. Monica, D. Sanchez, O. Castillo, "Analysis of spatial spread relationships of coronavirus (COVID-19) pandemic in the world using self organizing maps," *Chaos, solitons and fractals,* vol. 138, 2020, pp. 1-7.

[27] P. Melin, J. Monica, D. Sanchez, O. Castillo, "Multiple Ensemble Neural Network Models with Fuzzy Response Aggregation for Predicting COVID19 Time Series: The Case of Mexico," *Healthcare,* vol. 8 (181), 2020, pp. 1-13.

[28] R. Fisher, *Statistical Methods for Research Workers,* London, 1925.

[29] R. Huang, M. Liu, Y. Ding, "Spatialtemporal distribution of COVID19 in China and its prediction: A datadriven modeling analysis", *J. Infect. Dev. Ctries,* vol. 14, 2020, pp. 246253.

[30] S. Parthiban, P. Gajivaradhan, "A comparative study of Two factor ANOVA model under fuzzy environments using trapezoidal fuzzy numbers", *Intern. J. Fuzzy Mathematical Archive,* vol. 10 (1), 2016, pp. 1-25.

[31] V. Traneva, S. Tranev, "Intuitionistic Fuzzy Two-Factor Analysis of Variance of Movie Ticket Sales", *in: NAA 2020,* LNCS, 2020 (in press).

[32] V. Traneva, S. Tranev, "Inuitionistic Fuzzy Anova Approach to the Management of Movie Sales Revenue", *Studies in Computational Intelligence*, 2020 (in press).

[33] V. Traneva, S. Tranev, "Inuitionistic Fuzzy Analysis of Variance of Movie Ticket Sales", *in: INFUS 2020,* Advances in Intelligent Systems and Computing, Springer, Cham, 2020 (in press).

[34] V. Traneva, S. Tranev, V. Atanassova, "An Intuitionistic Fuzzy Approach to the Hungarian Algorithm," *in: G. Nikolov et al. (Eds.): NMA 2018,* LNCS 11189, Springer Nature , AG, 2019, pp. 19.

[35] V. Traneva, S. Tranev, M. Stoenchev, K. Atanassov, " Scaled aggregation operations over two- and three-dimensional index matrices," *Soft computing,* vol. 22, 2019, pp. 5115-5120.

[36] V. Traneva, S. Tranev, *Index Matrices as a Tool for Managerial Decision Making,* Bulgaria: Publ. House of the Union of Scientists, 2017 (in Bulgarian).

[37] https://github.com/owid/covid-19-data/tree/master/public/data, last accessed 24 June 2020.

[38] https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases, last accessed 24 june 2020.