Hacettepe Journal of Mathematics and Statistics
Volume 41 (3) (2012), 361 – 364

A SIMULATION BASED APPROACH TO CALCULATE THE FUZZY CORRELATION COEFFICIENT OF FUZZY OBSERVATIONS

Cagdas Hakan Aladag*†, Erol Egrioglu‡ and Ufuk Yolcu§

,
Received 03 : 01 : 2012 : Accepted 15 : 08 : 2012

Abstract

Fuzzy set theory has been widely used in various fields of statistics in recent years. The correlation between fuzzy random variables can be measured by a fuzzy correlation coefficient. When the correlation of the fuzzy random variables has being calculated, mathematical programming and fuzzy arithmetic operations have been used in the literature. In this study, to calculate the fuzzy correlation coefficient, a new approach based on simulation is proposed. It is not necessary to employ mathematical programming or fuzzy arithmetic operations when the proposed method is used. The proposed approach is applied to fats and oils data to show the applicability of the method.

Keywords: Fuzzy number, Correlation, Fuzzy Correlation.

2000 AMS Classification: 62–04, 68 T 37, 03 B 52, 94 D 05; 62 G 86, 62 A 86.

1. Introduction

In statistical theory, the correlation coefficient is a measure for the linear relationship between two random variables. Given a sample of n independent pairs of observations

*Hacettepe University, Department of Statistics, Ankara, Turkey.
E-mail: aladag@hacettepe.edu.tr
†Corresponding Author.
‡Ondokuz Mayis University, Department of Statistics, Samsun, Turkey.
E-mail: erole@omu.edu.tr
§Giresun University, Department of Statistics, Giresun, Turkey.
E-mail: ufuk.yolcu@giresun.edu.tr

$(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$, the sample correlation coefficient $r_{X,Y}$ between X and Y is calculated as

$$r_{X,Y} = \frac{\sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n} (X_i - \overline{X})^2 \sum_{i=1}^{n} (Y_i - \overline{Y})^2}}, \tag{1.1}$$

where $\overline{X} = \sum_{i=1}^{n} X_i/n$ and $\overline{Y} = \sum_{i=1}^{n} Y_i/n$ are the sample means of X and Y, respectively. The formula given in ((**??**)) is used when the variables have crisp values. However, in real world applications, the observations are sometimes described using linguistic terms such as excellent, good, or bad; or are only approximately known, rather than known with certainty (Hong [6]).

Various methods have been proposed to calculate a correlation coefficient which is crisp for fuzzy observations. Some of these methods were proposed by Bustince and Burillo [2], Chiang and Lin [3], Gerstenkorn and Manko [4], Hong and Hwang [5], and Yu [8]. On the other hand, Liu and Kao [7], and Hong [6] calculated the correlation coefficient as a fuzzy number for a sample of n independent pairs of fuzzy observations by utilizing fuzzy sets, Zadeh [9]. Mathematical programming is used in the method proposed by Liu and Kao [7], while an extension principle based on T_W (the weakest t-norm) is employed in Hong's [6] method to calculate the fuzzy correlation coefficient. Although Hong's [6] method does not require mathematical programming or computer calculations, the possible value of the calculated fuzzy correlation coefficient can be greater than 1 for some α cut values in this approach. However, a correlation coefficient always lies between -1 and 1 since it is a standard measure for variation.

In this study, when the observations are fuzzy, a new approach based on simulation is proposed to calculate the correlation coefficient which is also fuzzy. In the proposed method, crisp data are derived from the fuzzy observations by simulation and a sample of correlation coefficients calculated from these crisp values is obtained. Then, an estimator of the fuzzy correlation coefficient is calculated by using the obtained sample for the correlation coefficient. The proposed approach does not require mathematical programming or fuzzy arithmetic operations. In addition, possible values of the calculated fuzzy correlation coefficient are in the interval [-1,1] for all α cut values.

The proposed approach is introduced in Section 2. The proposed method is applied to the data used by Hong [6] and the obtained results are given in Section 3. Section 4 concludes the paper.

2. The proposed method

Let $\tilde{X}_j = (x_j, \gamma_j)$ $j = 1, 2, \ldots, n$ and $\tilde{Y}_j = (y_j, \delta_j)$ $j = 1, 2, \ldots, n$ are random samples for the random variables X and Y, respectively. $\tilde{X}_j$ and $\tilde{Y}_j$ are symmetric triangular fuzzy numbers. In another words, the left and the right hand side values for $\tilde{X}_j$ and $\tilde{Y}_j$ are represented by γ_j and δ_j, respectively while x_j and y_j represent the center values of $\tilde{X}_j$ and $\tilde{Y}_j$, respectively. The proposed algorithm is given below.

Step 1. The iteration bound N is specified.

Step 2. α is generated using a Uniform (0,1) distribution.

Step 3. According to the determined α generated in the previous step, for fuzzy observations, α-cuts are calculated as follows:

$$\tilde{X}_j[\alpha] = \left[(x_j - \gamma_j) + (\gamma_j * \alpha),\ (x_j + \gamma_j) - (\gamma_j * \alpha)\right],\ j = 1, 2, \ldots, n, \tag{2.1}$$

$$\tilde{Y}_j[\alpha] = \left[(y_j - \delta_j) + (\delta_j * \alpha),\ (y_j + \delta_j) - (\delta_j * \alpha)\right],\ j = 1, 2, \ldots, n, \tag{2.2}$$

where n is the number of fuzzy observations.

Step 4. Taking the right and left hand side values of the α-cuts as the parameters of the Uniform distribution, samples $(X_1, X_2, \ldots, X_n)$ and $(Y_1, Y_2, \ldots, Y_n)$ are sampled from the Uniform distributions given as follows:

$$(2.3) \qquad X_j \sim Uniform((x_j - \gamma_j) + (\gamma_j * \alpha),\ (x_j + \gamma_j) - (\gamma_j * \alpha))$$

$$(2.4) \qquad Y_j \sim Uniform((y_j - \delta_j) + (\delta_j * \alpha),\ (y_j + \delta_j) - (\delta_j * \alpha)),$$

where $j = 1, 2, \ldots, n$. Then, the crisp correlation coefficient $r_{X,Y}$ is calculated by using ((**??**)).

Step 5. By turning back to Step 2, similar operations are made for N times. After N repetitions, the sample $(r_1, r_2, \ldots, r_N)$ is obtained.

Step 6. For the mean value of the sample of $(r_1, r_2, \ldots, r_N)$ obtained in the previous stage, confidence intervals $(1 - \alpha)100\%$ are calculated as follows (Buckley [1]):

$$(2.5) \qquad \left[\overline{r} - t_{\alpha/2}\ \left(s/\sqrt{n}\right), \overline{r} + t_{\alpha/2}\ \left(s/\sqrt{n}\right)\right],$$

where $t_{\alpha/2}$ is defined from the (Student's) t distributions, with $n-1$ degrees of freedom, so that the probability of exceeding it is $\alpha/2$, s is the standard deviation of the sample of correlation coefficients. The confidence interval obtained can be considered as an α-cut of the fuzzy correlation coefficient. According to the central limit theorem, the mean value of the correlation coefficients has a normal distribution since the sample size is big enough. Thus, the confidence interval given in ((**??**)) can be calculated. Then, the fuzzy correlation coefficient can be obtained from the confidence intervals calculated for different α values lying between 0.01 and 1, like in Buckley [1].

3. The implementation

The proposed method is applied to the fats and oils data which was also used in [6]. The fats and oils data is presented in Table 1. The correlation coefficient, which is a triangular shaped fuzzy number, calculated by using the proposed method and computed α cut values are given in Figure 1 and Table 2, respectively.

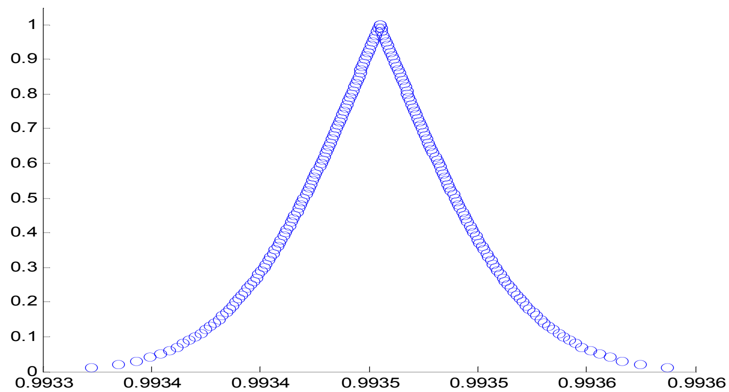
Table 1. The fuzzy fats and oils data

Fats and Oils	Specific gravity	Freezing Point	Lodine value	Soponificaion
Lenseed oil	[0.930,0.935]	[-27,-18]	[170,204]	[118,196]
Perilla oil	[0.930,0.937]	[-5,-4]	[192,208]	[188,197]

Table 2. The α-cuts of the fuzzy correlation coefficient between lenseed oil and perilla oil

α	0.1	0.2	0.3	0.4	0.5
Lower	0.99342	0.99343	0.99345	0.99346	0.993470
Upper	0.99359	0.99357	0.99355	0.99354	0.993540
α	0.7	0.8	0.9	1	0.6
Lower	0.99348	0.99349	0.99349	0.99350	0.993478
Upper	0.99352	0.99351	0.99351	0.99350	0.993532

Figure 1. Fuzzy correlation for fats and oils data set for $0.01 \leq \alpha \leq 1$.



4. Conclusion

In this study, a new method is proposed for calculating a correlation coefficient for variables consisting of fuzzy observations. The proposed method produces a correlation coefficient which is a triangular fuzzy number. Liu and Kao [7] and Hong [6] also suggested some approaches to calculating a fuzzy correlation coefficient for fuzzy observations. While Liu and Kao's [7] method employs mathematical programming, Hong's [6] method is based on fuzzy arithmetic operations. On the other hand, the proposed method is simpler than these methods available in the literature since it does not require mathematical programming or fuzzy arithmetic operations. Besides, in the Hong's [6] method, the calculated correlation coefficient can take values greater than 1 although the correlation coefficient value must be in the interval [-1 1]. However, the method proposed by us is always give a fuzzy correlation coefficient which is in the interval [-1 1]. In addition, the proposed method has produced much narrower fuzziness than Hong's method [6] for the fats and oils data. Therefore, for the analyzed data, it can be said that the fuzzy correlation coefficient calculated by the proposed method include less uncertainty than the correlation coefficient obtained from Hong's [6] method.

References

[1] Buckley, J. J. *Fuzzy Statistics*, Chapter 4 (Springer-Verlag, New York, 2004)

[2] Bustince, H. and Burillo, P. *Correlation of interval-valued intuitionistic fuzzy sets*, Fuzzy Sets and Systems **74**, 237–244, 1995.

[3] Chiang, D. A. and Lin, N. P. *Correlation of fuzzy sets,* Fuzzy Sets and Systems **102**, 221–226, 1999.

[4] Gerstenkorn, T. and Manko, J. *Correlation of intuitionistic fuzzy sets,* Fuzzy Sets and Systems **44**, 39–43, 1991.

[5] Hong, D. H. and Hwang, S. Y. *Correlation of intuitionistic fuzzy sets in probability space*, Fuzzy Sets and Systems **75**, 77–81, 1995.

[6] Hong, D. H. *Fuzzy measures for a correlation coefficient of fuzzy numbers under* T_w *(the weakest t-norm)-based fuzzy arithmetic operations*, Information Science **176**, 150–160, 2006.

[7] Liu, S. T. and Kao, C. *Fuzzy measures for correlation coefficient of fuzzy numbers*, Fuzzy Sets and Systems **128**, 267–275, 2002.

[8] Yu, C. *Correlation of fuzzy numbers,* Fuzzy Sets and Systems **55**, 303–307, 1995.

[9] Zadeh, L. A. *Fuzzy sets as a basis for a theory of possibility,* Fuzzy Sets and Systems **1**, 3–28, 1978.