Correlation Evaluation with Fuzzy Data and its Application in the Management Science

Berlin Wu, Wei-Shun Sha and Juei-Chao Chen

Abstract How to evaluate an appropriate correlation with fuzzy data is an important topic in the educational and psychological measurement. Especially when the data illustrate uncertain, inconsistent and incomplete type, fuzzy statistical method has some promising features that help resolving the unclear thinking in human logic and recognition. Traditionally, we use Pearson's Correlation Coefficient to measure the correlation between data with real value. However, when the data are composed of fuzzy numbers, it is not feasible to use such a traditional approach to determine the fuzzy correlation coefficient. This study proposes the calculation of fuzzy correlation with three types of fuzzy data: interval, triangular and trapezoidal. Empirical studies are used to illustrate the application for evaluating fuzzy correlations. More related practical phenomena can be explained by this appropriate definition of fuzzy correlation.

1 Introduction

Traditional statistics reflects the results from a two-valued logic world, which often reduces the accuracy of inferential procedures. To investigate the population, people's opinions or the complexity of a subjective event more accurately, fuzzy logic should be utilized to account for the full range of possible values. Especially, when

B. Wu (✉)
Department of Mathematical Sciences, National Cheng Chi University, Taipei, Taiwan
e-mail: berlin@nccu.edu.tw

W.S. Sha
Graduate Institute of Business Administration, Fu Jen Catholic University, New Taipei, Taiwan
e-mail: P581122@ms57.hinet.net

J.C. Chen
Department of Statistics and Information Science, Fu Jen Catholic University, New Taipei, Taiwan
e-mail: 006884@mail.fju.edu.tw

V.-N. Huynh et al. (eds.), *Econometrics of Risk*, Studies in Computational Intelligence 583, DOI 10.1007/978-3-319-13449-9_19

dealing with psychometric measures, fuzzy statistics provides a powerful research tool. Since Zadeh [1] developed fuzzy set theory, its applications have been extended to traditional statistical inferences and methods in social sciences, including medical diagnosis or stock investment systems. For example, a successive series of studies demonstrated approximate reasoning methods for econometrics [2–4] and a fuzzy time series model to overcome the bias of stock markets was developed [5].

Within the framework of classical statistical theory, observations should follow a specific probability distribution. However, in practice, the observations are sometimes described by linguistic terms such as “Very satisfactory”, “Satisfactory”, “Normal”, “Unsatisfactory”, “Very unsatisfactory”, or are only approximately known, rather than equating with randomness. How to measure the correlation between two variables involving fuzziness is a challenge to the classical statistical theory. The number of studies which focus on fuzzy correlation analysis and its application in the social science fields has been steadily increasing [6–9]. For example [9, 10] define a correlation formula to measure the interrelation of intuitionist fuzzy sets. However, the range of their defined correlation is from 0 to 1, which contradicts with the conventional awareness of correlation which should range from -1 to 1. An article [11] also has the same problems of lying the correlations between 0 and 1 for the interval valued fuzzy numbers. In order to overcome this issue, [12] takes random sample from the fuzzy sets and treat the membership grades as the crisp observations. Their derived coefficient is between -1 and 1; however, the sense the fuzziness is gone [8] calculated the fuzzy correlation coefficient based on Zadeh’s extension principles. They used a mathematical programming approach to derive fuzzy measures based on the classical definition of the correlation coefficient. Their derivation is quite promising, but in order to employ their approach, the mathematical programming is required.

In addition, most previous studies deal with the interval fuzzy data, their definitions cannot deal with triangle or trapezoid data. In addition, formulas in these studies are quite complicated or required some mathematical programming which really limited the access of some researchers with no strong mathematical background. In this study, we give a simple solution of a fuzzy correlation coefficient without programming or the aid of computer resources. In addition, the provided solutions are based on the classical definition of Pearson correlation which should quite easy and straightforward. The definitions provided in this study can also be used for interval-valued, triangular and trapezoid fuzzy data.

Traditionally, if one wishes to understand the relationship between the variables x and y, the most direct and simple way is to draw a scatter plot, which can approximately illustrate the relationship between these variables: positive correlation, negative correlation, or zero correlation. The issue at hand is how to measure the relationship in a rational way. Statistically, the simplest way to measure the linear relationship between two variables is using Pearson’s correlation coefficient, which expresses both the magnitude and the direction of the relationship between the two variables with a range of values from 1 to -1. However, Pearson correlations can only be applied to variables that are real numbers and is not suitable for a fuzzy dataset.

When considering the correlation for fuzzy data, two aspects should be considered: centroid and data shape. If the two centroids of the two fuzzy dataset are close, the correlation should be high. In addition, if the data shape of the two fuzzy sets is similar, the correlation should also be high. An approach to dealing with these two aspects simultaneously will be presented later in this study. Before illustrating the approach of calculating fuzzy correlations, a review of fuzzy theory and fuzzy datasets are presented in the next section.

2 Fuzzy Theory and Fuzzy Data

Traditional statistics deals with single answers or certain ranges of the answer through sampling surveys, but it has difficulty in reflecting people's incomplete and uncertain thoughts. In other words, these processes often ignore the intriguing and complicated yet sometimes conflicting human logic and feeling. For example, we would like to investigate the a person's favorite topics. In this case, consider a fuzzy set of favorite topics for a person as shown in Table 1. Note that in the extreme cases when a degree is given as 1 or 0, that is "like" or "dislike", a standard "yes" and "no" are in a complementary relationship, as in binary logic. Let A_1 represent for "favorite topics", A_2 "dislike the topics".

Based on the analysis of binary logic, we can find that he likes culture, religions and finance but dislikes politics and recreation. On the other hand, the fuzzy statistical result can be represented as:

$$\begin{aligned}\mu_{A_1} &= 0I_{\text{politics}}(x) + 0.8I_{\text{culture}}(x)\\ &\quad + 0.6I_{\text{religions}}(x) + 0.9I_{\text{finance}}(x) + 0.3I_{\text{recreation}}(x);\\ \mu_{A_2} &= 1I_{\text{politics}}(x) + 0.2I_{\text{culture}}(x)\\ &\quad + 0.4I_{\text{religions}}(x) + 0.1I_{\text{finance}}(x) + 0.7I_{\text{recreation}}(x).\end{aligned}$$

This means that the person likes the topic of politics 0 %, culture 80 %, and religion 60 %. etc. He dislikes the topic of finance 10 %, dislikes culture 20 %, dislikes religion 40 %, and dislikes recreation with 70 %. The percentages for each category represent the degree of their perceptions based on their own concept.

Table 1 Comparing fuzzy numbers with crisp numbers

	Fuzzy	Logic	Binary	Logic
Favorite topics	A_1 = like	A_2 = dislike	A_1 = like	A_2 = dislike
Politics	0	1		V
Culture	0.8	0.2	V	
Religions	0.6	0.4	V	
Finance	0.9	0.1	V	
Recreation	0.3	0.7		V

Therefore, based on the binary (like or dislike) logic, we can see only the superficial feeling about people's favorite topics. With the information of fuzzy response we will see a more detailed data representation. In illustrating human feelings with degrees, we encounter the problems that measurement cast the uncertainty and fuzzy property. Hence, a precise explanation about fuzzy numbers is illustrative and convincing.

2.1 Continuous Fuzzy Data

Continuous fuzzy data has been widely used in many applications. It can be classified into several types, such as interval-valued numbers, triangular numbers, trapezoid numbers, and exponential numbers. Typically, the nomenclature is based on the shape of the membership function. Even though there are various types of fuzzy numbers, here we limit the discussion to three usual types: interval-valued numbers, triangular numbers and trapezoid numbers. The definitions of the three types of fuzzy data are given as follows.

Definition 1 A fuzzy number $X = [a, b, c, d]$ defined on the universe set U of real numbers R with its vertices $a \le b \le c \le d$, is called a trapezoidal fuzzy number if its membership function is given by

$$u_A(x) = \begin{cases} \dfrac{x-a}{b-a}, & a \le x \le b \\ 1, & b \le x \le c \\ \dfrac{d-x}{d-c}, & c \le x \le d \\ 0, & \text{otherwise} \end{cases}$$

When $b = c$, X is called a triangular fuzzy number; when $a = b$ and $c = d$, X is called an interval-valued fuzzy number.

2.2 Collecting Continuous Fuzzy Data

Respondents choose one single answer or certain range of the answer in traditional sampling surveys. But traditional methods are not able to truly reflect the complex thoughts of each respondent. If people can express the degree of their feelings by using membership functions, the answer presented will be closer to real human thoughts. But unfortunately scholars disagree in opinion about the construction of continuous fuzzy data. Many studies use continuous fuzzy without describing the construction method. The core of all the questions is fuzzy data determined by its membership function, but the construction of membership function is quite subjective. To reflect this, we ask the respondents to determine the membership function on Geometer's Sketchpad (GSP).

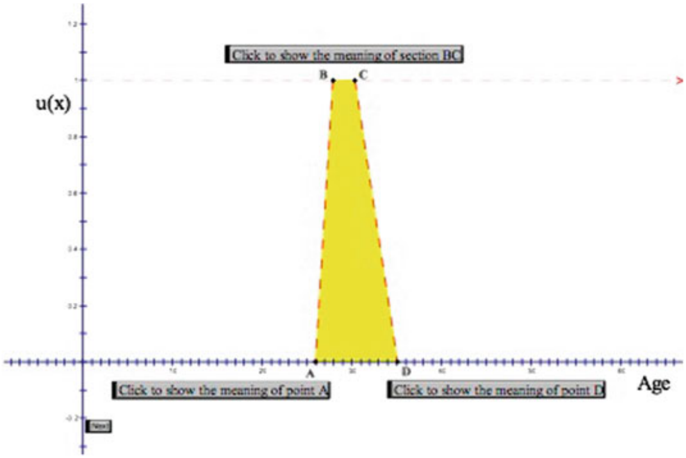

Fig. 1 A fuzzy answer for the expected marriage age

Figure 1 is the image of a fuzzy questionnaire item querying the prime time for marriage. Before answering the fuzzy questionnaire, respondents could click the three buttons to realize the meaning of each section and points. For example, people may decide: $\overline{AB}$ which represents the desire for marriage grows continuously for 2 years from 26 to 28, $\overline{BC}$ represents the desire for optimal marriage is 28–30, $\overline{CD}$ represents the desire for marriage falling continuously from 30 until it reaches 35.

Respondents can decide their own membership function of the prime time for marriage by moving the four points A, B, C, and D. By moving the four points, the age corresponding to the points will be changed automatically. There are probably three types of fuzzy data: The first is trapezoid; the second is triangular; the third is interval-valued type. Figure 2 illustrates these three kinds of fuzzy data. Triangular

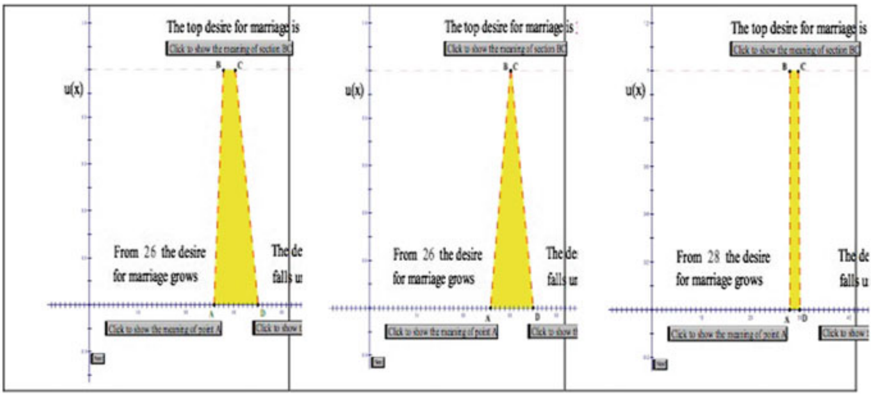

Fig. 2 Fuzzy observation for idea marriage year

data is a special case of trapezoid when point B equals to point C. It represents the prime time for marriage is only 30. The interval valued data shows the prime time for marriage is 28–30.

3 Fuzzy Correlation

The correlation coefficient is a commonly used statistics that presents a measure of how two random variables are linearly related in a sample. The population correlation coefficient, which is generally denoted by the symbol ρ is defined for two variables x and y by the formula:

$$\rho = \frac{\sigma_{X,Y}}{\sigma_X \sigma_Y} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}.$$

In this case, the more positive ρ is, the more positive the association is. This also indicates that when ρ close to 1, an individual with a high value for one variable will likely have a high value for the other, and an individual with a lower value for one variable will likely to have a low value for the other. On the other hand, the more negative ρ is, the more negative the association is, this also indicate that an individual with a high value for one variable will likely have a low value for the other when ρ is close to -1 and conversely. When ρ is close to 0, this means there is little linear association between two variables. In order to obtain the correlation coefficient, we need to obtain σ_X^2, σ_Y^2, and the covariance of x and y. In practice, these parameters for the population are unknown or difficult to obtain. Thus, we usually use r_{xy}, which can be obtained from a sample, to estimate the unknown population parameter. The sample correlation coefficient is expressed as:

$$r_{xy} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2}}, \tag{1}$$

where (x_i, y_i) is the ith pair observation value, $i = 1, 2, \ldots, n$, $\bar{x}$ and $\bar{y}$ are sample means for x and y respectively.

Pearson correlation is a straightforward approach to evaluate the relationship between two variables. However, if the variables considered are not real numbers, but fuzzy data, the formula above is problematic. For example, Mr. Smith is a new graduate from college; his expected annual income is 50,000 dollars. However, he can accept a lower salary if there is a promising offer. In his case, the annual income is not a definite number but more like a range. Mr. Smith's acceptable salary range is from 45,000 to 50,000. We can express his annual salary as an interval [45,000, 50,000]. In addition, when Mr. Smith has a job interview, the manager may ask how many hours he can work per day. In this case, Mr. Smith may not be able to provide

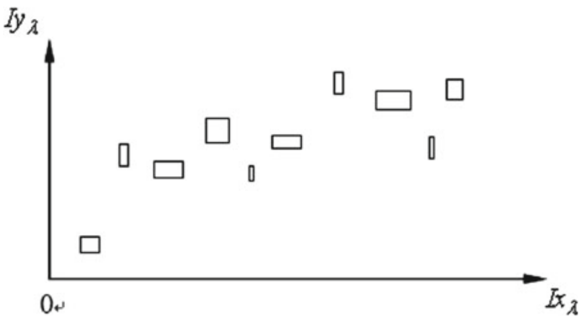

Fig. 3 Fuzzy correlation with interval data

a definite number since his everyday schedule is different. However, Mr. Smith may tell the manger that his expected working hours per day is an interval [8, 10].

We know Mr. Smith's expected salary ranges from [45,000, 50,000] and his expected working hours are [8, 10]. If we collect this kind of data from many new graduates, how can we use this data and calculate the correlation between expected salary and working hours? Suppose I_x is the expected salary for each new graduate, I_y is the number of working hours they desired, then the scatter plot for these two sets of fuzzy interval numbers would approximate that shown in Fig. 3.

For the interval valued fuzzy number, we need to take out samples from population X and Y. Each fuzzy interval data for sample X has centroids x_i, and for sample Y has centroids y_i. For the interval data, we also have to consider whether the length of interval fuzzy data are similar or not. In Mr. Smith's example, if the correlation between the expected salary and working hours are high, then we can expect two things: (1) the higher salary the new employee expects, the more working hours he can endure; (2) the wider the range of the expected salary, the wider the range of the working hours should be. However, how should one combine the information from both centroid and length? In addition, the effect of length should not be greater than the impact of centroids. In order to get the rational fuzzy correlations, we used natural logarithms to make some adjustments.

Let $(X_i = [a_i, b_i, c_i, d_i], Y_i = [e_i, f_i, g_i, h_i]; i = 1, 2, \ldots, n)$ be a sequence of paired trapezoid fuzzy sample on population Ω with its pair of centroids (cx_i, cy_i) and pair of areas $(\|x_i\| = area(x_i), \|y_i\| = area(y_i))$. The adjusted correlation for the pair of area will be

Definition 2 Let $(X_i = [a_i, b_i, c_i, d_i], Y_i = [e_i, f_i, g_i, h_i]; i = 1, 2, \ldots, n)$ be a sequence of paired trapezoid fuzzy sample on population Ω with its pair of centroids (cx_i, cy_i) and pair of areas $\|x_i\| = area(x_i)$, $\|y_i\| = area(y_i)$. Let

$$cr_{xy} = \frac{\sum_{i=1}^{n}(x_i - \overline{cx})(cy_i - \overline{cy})}{\sqrt{\sum_{i=1}^{n}(cx_i - \overline{cx})^2}\sqrt{\sum_{i=1}^{n}(cy_i - \overline{cy})^2}};$$

$$ar_{xy} = \frac{\sum_{i=1}^{n}(\|x_i\| - \overline{\|x\|})(\|y_i\| - \overline{\|y\|})}{\sqrt{\sum_{i=1}^{n}(\|x_i\| - \overline{\|x\|})^2}\sqrt{\sum_{i=1}^{n}(\|y_i\| - \overline{\|y\|})^2}}. \tag{2}$$

Then fuzzy correlation is defined as as

$$FC = \beta_1 cr_{xy} + \beta_2 ar_{xy}, \quad (\beta_1 + \beta_2 = 1).$$

We choose a pair of (β_1, β_2) depending on the weight of practical use. For instance, if we think the location correlation is much more important than that of area scale, $\beta_1 = 0.7$ and $\beta_2 = 0.3$ will be a good suggestion.

Example 1 Suppose we have the following data as shown in Table 2.

In this case, the correlation between the two centroids is

$$cr_{xy} = \frac{\sum_{i=1}^{n}(x_i - 26.62)(cy_i - 1.7)}{\sqrt{\sum_{i=1}^{n}(cx_i - 26.62)^2}\sqrt{\sum_{i=1}^{n}(cy_i - 1.7)^2}} = 0.17;$$

Table 2 Numerical example for interval-valued, triangular, and trapezoidal fuzzy data

X			
Student	Data	Centroid	Area (length)
A	[23, 25]	24	2
B	[21, 23, 26]	23.3	2.5
C	[26, 27, 29, 35]	28.3	5.5
D	[28, 30]	29	2
E	[25, 26, 28, 35]	28.5	6
(fuzzy) mean	[24.6, 25.12, 29, 30.2]	26.62	3.6
Y			
Student	Data	Centroid	Area (length)
A	[1, 2]	1.5	1
B	[0, 2, 3]	1.7	1.5
C	[0, 1]	0.5	1
D	[1, 2, 4]	2.3	1.5
E	[1, 2, 3, 4]	2.5	2
(fuzzy) mean	[0.6, 1.4, 2, 2.8]	1.7	1.4

$$ar_{xy} = \frac{\sum_{i=1}^{n}(\|x_i\| - 3.6)(\|y_i\| - 1.4)}{\sqrt{\sum_{i=1}^{n}(\|x_i\| - 3.6)^2}\sqrt{\sum_{i=1}^{n}(\|y_i\| - 1.4)^2}}.$$

Considering the contribution of (area) length correlation to the fuzzy correlation, the idea of correlation interval is proposed. Suppose we fix the (area) length correlation by the following adjusted values

$$\lambda ar_{xy} = 1 - \frac{\ln(1 + |ar_{xy}|)}{|ar_{xy}|}.$$

Since $-1 \leq ar_{xy} \leq 1$, the range of λar_{xy} will be $0 < \lambda ar_{xy} < 0.3069$.

We will have the following definition for fuzzy correlation interval.

Definition 3 Let ($X_i = [a_i, b_i, c_i, d_i]$, $Y_i = [e_i, f_i, g_i, h_i]$; $i = 1, 2, \ldots, n$) be a sequence of paired trapezoid fuzzy sample on population Ω with its pair of centroids (cx_i, cy_i) and pair of areas $\|x_i\| = area(x_i)$, $\|y_i\| = area(y_i)$. Let

$$cr_{xy} = \frac{\sum_{i=1}^{n}(x_i - \overline{cx})(cy_i - \overline{cy})}{\sqrt{\sum_{i=1}^{n}(cx_i - \overline{cx})^2}\sqrt{\sum_{i=1}^{n}(cy_i - \overline{cy})^2}};$$

$$ar_{xy} = \frac{\sum_{i=1}^{n}(\|x_i\| - \overline{\|x\|})(\|y_i\| - \overline{\|y\|})}{\sqrt{\sum_{i=1}^{n}(\|x_i\| - \overline{\|x\|})^2}\sqrt{\sum_{i=1}^{n}(\|y_i\| - \overline{\|y\|})^2}},$$

and

$$\lambda ar_{xy} = 1 - \frac{\ln(1 + |ar_{xy}|)}{|ar_{xy}|}.$$

Then fuzzy correlation is defined as follows:

(i) When $cr_{xy} \geq 0$ and $\lambda ar_{xy} \geq 0$, fuzzy correlation $= (cr_{xy}, \min(1, cr_{xy} + \lambda ar_{xy}))$.
(ii) When $cr_{xy} \geq 0$ and $\lambda ar_{xy} < 0$, fuzzy correlation $= (cr_{xy} - \lambda ar_{xy}, cr_{xy})$.
(iii) When $cr_{xy} < 0$ and $\lambda ar_{xy} \geq 0$, fuzzy correlation $= (cr_{xy}, cr_{xy} + \lambda ar_{xy})$.
(iv) When $cr_{xy} < 0$ and $\lambda ar_{xy} < 0$, fuzzy correlation $= (\max(-1, cr_{xy} - \lambda ar_{xy}), cr_{xy})$.

Example 2 Suppose we have the following data as shown in Table 2.

In this case, the correlation between the two centroids is $cr_{xy} = 0.17$. Similarly, the correlation between two lengths is $ar_{xy} = 0.32$, so

$$\lambda ar_{xy} = 1 - \frac{\ln(1 + 0.32)}{0.32} = 0.13.$$

Since the centroids correlation $ar_{xy} \geq 0$, and the area (length) correlation $\lambda ar_{xy} \geq 0$, thus, fuzzy correlation $= (ar_{xy}, \min(1, ar_{xy} + \lambda ar_{xy})) = (0.17, \min(1, 0.30)) = (0.17, 0.30)$. This implied that the relationship between the X and Y are quite small.

4 Empirical Studies

In this section, 11 samples (5 girls and 6 boys) are collected from a middle high school at Taipei city in Taiwan. We want to investigate which factors will impact their academic achievement. The results present the correlation for fuzzy data and in comparison with the traditional person correlation to demonstrate the difference. Suppose we are interesting in measuring the strength of the linear relationship between the students: sleeping hours per day (X), play hours on Internet per day (Y), studying hours in exercising mathematics per day (Z), and grades (range) of mathematical tests in last two months (T), as shown in Table 2.

The data set consists of interval-valued, triangular and trapezoidal fuzzy numbers. For example, for variable X, the data [8, 8.5, 9.5] represents a triangular fuzzy number, which represents that normal sleeping hours per day is 8.5 h, but the range of his/her sleeping hours is 8 to 9.5 h. Similarly, the data [9, 10.5, 11, 12] represented a trapezoidal fuzzy data, in this case, the normal sleeping hours is 10.5–11, and the range of sleeping time falls from 9 to 12 h (Table 3).

Table 3 Survey of fuzzy data

Sample	X	Y	Z	T
1	[8, 8.5, 9.5]	[1, 1.5]	[2, 2.5]	[90, 95]
2	[7, 7.5]	[1, 2, 3.5]	[2, 3.5, 4]	[92, 96]
3	[9, 10.5, 11, 12]	[1, 3]	[1, 2]	[85, 87]
4	[8, 8.5]	[1.5, 2.5]	[0, 0.5, 1]	[70, 72]
5	[6, 7.5]	[1, 1.5]	[2, 3]	[90, 97]
6	[10, 11, 13]	[1, 2, 4]	[0.5, 1]	[56, 63]
7	[7, 8]	[3, 3.5, 5]	[0, 1]	[35, 67]
8	[8, 10, 11]	[1, 2]	[1.5, 2]	[80, 85]
9	[6.5, 8]	[0, 1.5, 2, 2.5]	[2, 2.5, 3]	[92, 100]
10	[7.5, 8.5]	[2, 2.5, 4]	[0.5, 1]	[35, 55]
11	[8, 8.5]	[1, 2]	[1, 1.5]	[60, 67]
Fuzzy mean	8.50	2.02	1.58	75.9

Table 4 Correlation with fuzzy data

Fuzzy corr	X	Y	Z	T
X	1	[−0.01, 0.09]	[−0.38, −0.27]	[−0.18, −0.17]
Y		1	[−0.59, −0.49]	[−0.73, −0.66]
Z			1	[0.87, 0.96]
T				1

Table 5 Pearson correlations based on centroids

Pearson corr. (center)	X	Y	Z	T
X	1	−0.01	−0.38	−0.18
Y		1	−0.59	−0.73
Z			1	0.87
T				1

Based on Tables 4 and 5, we have the following findings. First, besides the correlation of studying hours in exercising mathematics per day (Z) and grades (range) of mathematical tests last two month (T) is positive, all of the other measures were negatively correlated.

Second, the correlation between X and Z is close to 0. This means there is almost no relationship between sleeping hours per day and studying hours in exercising mathematics per day. Third, the correlations between Y and Z and between Y and T are moderately negative. This means if the students spend more time on internet, then they will have less time study mathematics. In addition, the more time they spend on internet, the lower math grade will be. Fourth, the correlations between X and Z and between X and T are slightly negative. This means the relationship between student's sleeping hours and time study on mathematics are weakly related. The relationship between the sleeping hours and student math grades are also weakly related.

Table 4 is the fuzzy correlation, and the correlations in Table 4 are fuzzy numbers. This overcomes the deficiency of those studies which the correlation coefficients calculated are crisp values, rather than the intuitively believed fuzzy numbers. Table 5 is the Pearson correlation, which calculated based on the centroids of two dataset. It is found that the results of Tables 4 and 5 are quite close, the difference is the correlations in Table 4 are fuzzy numbers, and in Table 5 are crisp values. This is because the calculation of fuzzy correlation considered not only the correlations of centroids, but also the correlation between the area (length) of two dataset, and the fuzzy correlation expands based on the direction of the two dataset's area correlation. For example, the Pearson correlation between two centroids of Y and Z is −0.59. However, after considering the area (length) of two fuzzy dataset, the fuzzy correlation becomes [−0.59, −0.49]. This is due to the area(length) correlation of two dataset are positive, and this positive effect push the actual fuzzy correlation to the positive side. On the other hand, the Pearson correlation between X and T is -0.18, and the fuzzy correlation is $[-0.18, -0.17]$. The range of this fuzzy correlation is quite narrow compare to the correlation between Y and Z. This is because the area correlation between two fuzzy dataset is quite small, thus the fuzzy correlation are mainly impacted by the correlations between two centroids.

5 Conclusions

This paper uses a simple way to derive fuzzy measures based on the classical definition of Pearson correlation coefficient which are easy and straightforward. Moreover, the range of the calculated fuzzy coefficient is a fuzzy number with domain $[-1, 1]$, which consist with the conventional range of Pearson correlation. In the formula we provided, when all observations are real numbers, the developed model becomes the classical Pearson correlation formula.

There are some suggestions for future studies. First, the main purpose of this study is to provide the formula of calculating fuzzy correlations. Only few samples are collected to illustrate how to employ the formula. Future interested researchers can use formula and collect a large-scale fuzzy questionnaires to make this formulas implement in practice. Second, when calculating the fuzzy correlation, we adopt λar_{xy} to adjust the correlations, but researchers can set up their own λar_{xy} values if there are defensible reasons. However, it is suggested that the impact of length correlation should not exceed the impact of centroid correlation. Third, this study only considered the fuzzy correlation for continuous data. It would be interested to investigate the fuzzy correlation for discrete fuzzy data.

In practice, many applications are fuzzy in nature. We can absolutely ignore the fuzziness and make the existing methodology for crisp values. However, this will make the researcher over confident with their results. With the methodology developed in this paper, a more realistic correlation is obtained, which provides the decision maker with more knowledge and confident to make better decisions.

References

1. Zadeh, L.A.: Fuzzy sets. Inf. Control **8**, 338–353 (1999)
2. Dubois, D., Prade, H.: Fuzzy sets in approximate reasoning, part 1: inference with possibility distributions. Fuzzy Sets Syst. **40**, 143–202 (1991)
3. Lowen, R.: A fuzzy language interpolation theorem. Fuzzy Sets Syst. **34**, 33–38 (1990)
4. Ruspini, E.: Approximate reasoning: past, present, future. Inf. Sci. **57**, 297–317 (1991)
5. Wu, B., Hsu, Y.: The use of Kernel set and sample memberships in the identification of nonlinear time series. Soft Comput. **8**(3), 207–216 (2002)
6. Bustince, H., Burillo, P.: Correlation of interval-valued intuitionistic fuzzy sets. Fuzzy Sets Syst. **74**, 237–244 (1995)
7. Hong, D.: Fuzzy measures for a correlation coefficient of fuzzy numbers under T_w—(the weakest t-norm)-based fuzzy arithmetic operations. Fuzzy Sets Syst. **176**, 150–160 (2006)

8. Liu, S., Kao, C.: Fuzzy measures for correlation coefficient of fuzzy numbers. Fuzzy Sets Syst. **128**, 267–275 (2002)
9. Yu, C.: Correlation of fuzzy numbers. Fuzzy Sets Syst. **55**, 303–307 (1993)
10. Hong, D., Hwang, S.: Correlation of intuitionistic fuzzy sets in probability space. Fuzzy Sets Syst. **75**, 77–81 (1995)
11. Wang, G., Li, X.: Correlation and information energy of interval-valued fuzzy numbers. Fuzzy Sets Syst. **103**, 169–175 (1999)
12. Chiang, D.A., Lin, N.P.: Correlation of fuzzy sets. Fuzzy Sets Syst. **102**, 221–226 (1999)