Chapter 10

HOLD A MIRROR UP TO NATURE: A NEW APPROACH ON CORRELATION EVALUATION WITH FUZZY DATA AND ITS APPLICATIONS IN ECONOMETRICS

Chih Ching Yang
Department of Statistics
National Chengchi University, Taiwan
96354502@nccu.edu.tw

Yu-Ting Cheng
Department of Statistics
National Chengchi University, Taiwan
ting@nccu.edu.tw

Berlin Wu
Department of Math Sciences,
National Chengchi University, Taiwan
berlin@nccu.edu.tw

Songsak Sriboonchitta
Faculty of Economics
Chiangmai University, Thailand
songsak@econ.cmu.ac.th

How to evaluate an appropriate correlation with fuzzy data is an important topic in the economics. Especially when the data illustrated is an uncertain, inconsistent and incomplete type. 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 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.

Keywords: Fuzzy correlation; fuzzy data; evaluation; psychometrics.

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 utilised to account for the full range of possible values. Especially, when dealing with psychometric measures, fuzzy statistics provides a powerful research tool. Since Zadeh (1965) 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 (Lowen, 1990; Ruspini, 1991; Dubois and Parde, 1991) and a fuzzy time series model to overcome the bias of stock markets was developed (Wu and Hsu, 2002).

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 (Bustince and Burillo, 1995; Yu, 1993; Liu and Kao, 2002; Hong, 2006). For example, Hong and Hwang (1995) and Yu (1993) define a correlation formula to measure the interrelation of intuitionist fuzzy sets. However, the range of their defined correlation is from zero to one, which contradicts with the conventional awareness of correlation which should range from -1 to 1. Wang and Li (1999)'s article also has the same problems of lying the correlations between zero and one for the interval valued fuzzy numbers. In order to overcome this issue, Chiang and Lin (1999) take 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. Liu and Kao (2002) 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 be 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. In this study, we have proposed three kinds of fuzzy correlation which are based on the Neyman Person's correlation as well as the extension principle Definition 2.1, 2.2 and 2.3, the advantages are that we can compute various samples with fuzzy type, such as interval, triangle and trapezoid, the type for the continuous sample.

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 by 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 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 is the association. This also indicates that when ρ is close to one, 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 is the association, 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 zero, 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 r_{xy} 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, $i = 1, 2, 3 \ldots\ldots.., n; \bar{x}$, $\bar{y}$ are sample mean 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 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 working hours they desired, then the scatter plot for these two sets of fuzzy interval numbers would approximate as shown in Fig. 1.

For the interval valued fuzzy number, we need to take out samples from population X and Y. Each fuzzy interval data for sample X centroids has x_i, and for sample Y has centroids y_i. For the interval data, we also have to consider whether the length of

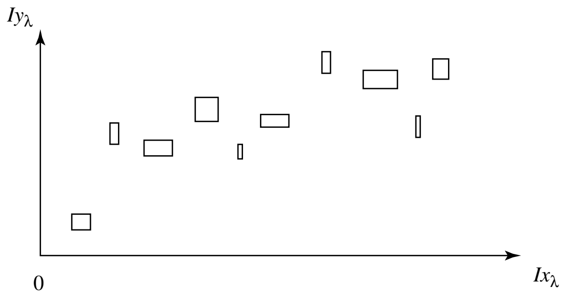

Fig. 1. Fuzzy correlation with interval data.

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? If they are combined with equal weight, it is possible that the combined correlation would exceed the boundaries of 1 or −1. 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 $(X_i = [a_i, b_i], Y_i = [c_i, d_i]; i = 1, 2, \ldots n)$ *be a sequence of paired trapezoid fuzzy sample on population* Ω *with its pair of center* (cx_i, cy_i) *and pair of area* $(\|x_i\| = area(x_i), \|y_i\| = area(y_i)$. *The adjust correlation for the pair of area will be*

Definition 2.1. *Let* $(X_i = [a_i, b_i], Y_i = [c_i, d_i]; i = 1, 2, \ldots n)$ *be a sequence of paired trapezoid fuzzy sample on population* Ω *with its pair of center* (cx_i, cy_i) *and pair of area* $\|x_i\| = area(x_i), \|y_i\| = area(y_i)$:

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

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

Then fuzzy correlation is defined as:

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

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

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

In this case, the correlation between the two centers is:

$$cr_{xy} = \frac{\sum_{i=1}^{n} (cx_i - 27.1)(cy_i - 1.7)}{\sqrt{\sum_{i=1}^{n} (cx_i - 27.1)^2}\sqrt{\sum_{i=1}^{n} (cy_i - 1.7)^2}} = -0.26,$$

and the correlation between the two length is:

$$ar_{xy} = \frac{\sum_{i=1}^{n} (\|x_i\| - 3.4)(\|y_i\| - 1.8)}{\sqrt{\sum_{i=1}^{n} (\|x_i\| - 3.4)^2}\sqrt{\sum_{i=1}^{n} (\|y_i\| - 1.8)^2}} = 0.05.$$

Table 2 is a list of combinations for choosing $\beta_1 \beta_2$. The fuzzy correlation will be computed by cr_{xy} and ar_{xy} with $0 \le \beta_1, \beta_2 \le 1$. Such as, when $\beta_1 = 0.7$ and $\beta_2 = 0.3$ then $FC = 0.7 \times (-0.26) + 0.3 \times 0.05 = -0.17$.

Table 1. Numerical example for interval-valued fuzzy data.

	X			Y		
Student	Data	Center	Length	Data	Center	Length
A	[23,25]	24	2	[1, 2]	1.5	1
B	[21,26]	23.5	5	[0, 3]	1.5	3
C	[29,35]	32	6	[0, 1]	0.5	1
D	[28,30]	29	2	[1, 4]	2.5	3
E	[26,28]	27	2	[2, 3]	2.5	1
(fuzzy) mean		27.1	3.4		1.7	1.8

Table 2. Different combinations of β_1, β_2.

	(1, 0)	(0.9, 0.1)	(0.8, 0.2)	(0.7, 0.3)	(0.6, 4)	(0.5, 0.5)	(0.4, 0.6)	(0.3, 0.7)	(0.2, 0.8)	(0.1, 0.9)	(0, 1)
FC	−0.26	−0.23	−0.20	−0.17	−0.14	−0.11	−0.08	−0.05	−0.01	0.02	0.05

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:

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

$$\text{where} \quad ar_{xy} = \frac{\sum_{i=1}^{n} (\|x_i\| - \|\bar{x}_i\|)(\|y_i\| - \|\bar{y}_i\|)}{\sqrt{\sum_{i=1}^{n} (\|x_i\| - \|\bar{x}_i\|)^2}\sqrt{\sum_{i=1}^{n} (\|y_i\| - \|\bar{y}_i\|)^2}}, \tag{3}$$

since $-1 \leq ar_{xy} \leq 1$, the range of λr_{xy} will be $0 < \lambda r_{xy} < 0.3069$. We will have the following definition for fuzzy correlation interval.

Definition 2.2. *Let $(X_i = [a_i, b_i],\ Y_i = [c_i, d_i];\ i = 1, 2, \ldots n)$ be a sequence of paired trapezoid fuzzy sample on population Ω with its pair of center (cx_i, cy_i) and pair of area $\|x_i\| = area(x_i)$, $\|y_i\| = area(y_i)$:*

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

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

$$\textit{where} \quad ar_{xy} = \frac{\sum_{i=1}^{n} (\|x_i\| - \|\bar{x}_i\|)(\|y_i\| - \|\bar{y}_i\|)}{\sqrt{\sum_{i=1}^{n} (\|x_i\| - \|\bar{x}_i\|)^2}\sqrt{\sum_{i=1}^{n} (\|y_i\| - \|\bar{y}_i\|)^2}}. \tag{4}$$

Then fuzzy correlation is defined as:

(i) When $cr_{xy} \geq 0$, $\lambda ar_{xy} \geq 0$, fuzzy correlation $= (cr_{xy}, min(1, cr_{xy} + \lambda ar_{xy}))$;

(ii) When $cr_{xy} \geq 0$, $\lambda ar_{xy} < 0$, fuzzy correlation $= (cr_{xy} - \lambda ar_{xy}, cr_{xy})$;

Table 3. Numerical example for interval-valued fuzzy data.

	X			Y		
Student	Data	Centroid	Area(length)	Data	Centroid	Area(length)
A	[23,25]	24	2	[1,2]	1.5	1
B	[21,26]	23.5	5	[0,3]	1.5	3
C	[29,35]	32	6	[0,1]	0.5	1
D	[28,30]	29	2	[1,4]	2.5	3
E	[26,28]	27	2	[2,3]	2.5	1
(fuzzy) mean		27.1	3.4		1.7	1.8

(iii) When $cr_{xy} < 0, \lambda ar_{xy} \geq 0$*, fuzzy correlation* $= (cr_{xy}, cr_{xy} + \lambda ar_{xy})$*;*
(iv) When $cr_{xy} < 0, \lambda ar_{xy} < 0$*, fuzzy correlation* $= (\max(-1, c_{xy} - \lambda ar_{xy}), cr_{xy})$*.*

Example 2. Suppose, we have the following data as shown in Table 3.

In this case, the correlation between the two centroids is:

$$cr_{xy} = \frac{\sum_{i=1}^{n}(cx_i - 27.2)(cy_i - 1.7)}{\sqrt{\sum_{i=1}^{n}(cx_i - 27.2)^2}\sqrt{\sum_{i=1}^{n}(cy_i - 1.7)^2}} = -0.26.$$

Similarly, the correlation between two lengths is:

$$ar_{xy} = \frac{\sum_{i=1}^{n}(\|x_i\| - 3.6)(\|y_i\| - 1.4)}{\sqrt{\sum_{i=1}^{n}(\|x_i\| - 3.6)^2}\sqrt{\sum_{i=1}^{n}(\|y_i\| - 1.4)^2}} = 0.05,$$

$$\lambda ar_{xy} = 1 - \frac{\ln(1 + 0.05)}{0.05} = 0.02.$$

Since the centers correlation $ar_{xy} \geq 0$, and the area(length) correlation $\lambda ar_{xy} \geq 0$, thus, fuzzy correlation $= (r, r + \lambda ar_{xy})) = (-0.26, -0.26 + 0.02) = (-0.26, -0.24)$. This implied that the relationship between the X and Y are quite small.

Another interesting idea is taking all possible correlations into consideration. That is we calculate the correlations for all endpoints of intervals. Then we take the mean of all possible correlations as our center of the fuzzy correlation. While the range is chosen by the three standard deviation, that is $\frac{3(r_{\min}-r_{\max})^2}{12}$. Here we apply the idea of three standard deviations from quality control.

Definition 2.3. *Let* $X_{ji} = [a_{1i}, a_{2i}]$ *and* $Y_{ji} = [b_{1i}, b_{2i}]$ *be a sequence of paired fuzzy sample on population* Ω. *Let*

$$r_{jk} = \frac{\sum_{i=1}^{n} (a_{ji} - \bar{a}_j)(b_{ki} - \bar{b}_k)}{\sqrt{\sum_{i=1}^{n} (a_{ji} - \bar{a}_j)^2}\sqrt{\sum_{i=1}^{n} (b_{ki} - \bar{b}_k)^2}}, \quad j = 1, 2, \ k = 1, 2.$$

Then fuzzy correlation is $[-r_{low}, r_{up}]$ *with* $r_{low} = \bar{r} - s_r$ *and* $-r_{up} = \bar{r} + s_r$, *where*

$$\bar{r} = \frac{\sum_{j=1}^{2}\sum_{k=1}^{2} r_{jk}}{4} \quad and \quad s_r = \frac{\sum_{j=1}^{2}\sum_{k=1}^{2} (r_{jk} - \bar{r})^2}{4}.$$

***Example* 3.** Suppose we have the following data as shown in Table 4.

Since *the mean and Standard Deviation of* r_{jk} *are* -0.14 and 0.12, thus, fuzzy correlation $= (-0.26, -0.2)$. This implied that the relationship between the X and Y are small.

Table 4. Numerical example for interval-valued fuzzy data.

	X	Y	Correlation coefficient			
Student	$[a_1, a_2]$	$[b_1, b_2]$	r_{a1b1}	r_{a1b2}	r_{a2b1}	r_{a2b2}
A	[23,25]	[1,2]	−0.07	−0.07	−0.32	−0.09
B	[21,26]	[0,3]				
C	[32,35]	[0,1]				
D	[28,30]	[1,4]				
E	[26,28]	[2,3]				
interval	$\bar{r} = -0.14, s_r = 0.12$					

A correlation coefficient is a number between −1 and 1 which measures the degree to which two variables are linearly related. If there is perfect linear relationship with positive slope between the two variables, we have a correlation coefficient of one; if there is positive correlation, whenever one variable has a high value. Thus, based on the measure of evaluation, the degree of the population correlation coefficient, we will be considered for the correlation of fuzzy interval. As the correlation of fuzzy interval, $[r_{low}, r_{up}]$, is computed then the value of fuzzy correlation can be evaluated that is defined as:

(1) *When* $[r_{low}, r_{up}] \in [-0.1, 0.10]$, *the fuzzy correlation is not significant*.
(2) *When* $[r_{low}, r_{up}] \in [-0.39, -0.11]$ or $[0.11, 0.39]$, *the fuzzy correlation is low value*.
(3) *When* $[r_{low}, r_{up}] \in [-0.69, -0.40]$ or $[0.40, 0.69]$, *the fuzzy correlation is middle value*.
(4) *When* $[r_{low}, r_{up}] \in [-0.99, -0.70]$ or $[0.70, 0.99]$, *the fuzzy correlation is high value*.

3. Empirical studies

In this section, two empirical examples will be considered to study the relationship with three schemes. In the first part, we employ the fuzzy interval data to investigate the relationship between climate and the price of vegetable from 2009 to 2011 in Taiwan. In the second part, we apply the exchange rate and the price of agriculture in Thailand.

3.1. *Correlation between climate and agriculture price in Taiwan*

A total of 33 samples are collected from the Central Weather Bureau and Agriculture and Food Agency Council of Agriculture Executive Yuan in Taiwan to study the factors impacting the relationship between climate (X) and the price of vegetable (Y). The result presents the correlation for fuzzy data and in comparison with the price of vegetable.

Table 5. Correlations interval based on temperature and the price of vegetable in Taiwan.

Scheme	Correlation coefficient
Fuzzy Correlation by Definition 1	0.212
Fuzzy Interval by Definition 2	(0.339, 0.489)
Fuzzy Interval by Definition 3	(0.348, 0.480)

Based on Table 5, we have the following findings. First, besides the correlation of temperature and vegetable price is positive, this result presents that as the temperature is increases, the price of vegetable also increases. Second, the correlation coefficient of both new method and length and center are close. This means that there is almost middle relationship between temperature and vegetable price in Taiwan.

3.2. *Correlation of both pair agriculture price in Thailand*

A total of 17 samples are collected from Thailand bank and Agriculture and Food Agency Council in Thailand. The results show the correlation for the exchanges rate and various price of agriculture with three approaches of evaluation of correlation coefficient. The results are listed in Table 6.

In the Table 6, we have the following findings. First, besides the correlation of exchange rate and the price of agriculture is negative, and this result denotes that the exchange rate decreases then the price of agriculture increases. Second, the correlation coefficient is considered to be the high level for exchange rate and the price of corn, wheat, this means the price of corn and wheat have a lot of effect on the exchange rate. In addition, the price of sugar will be slightly affected by exchange rate, and the price of race cannot be influenced by exchange rate. Third, any both price of agriculture are positive, and there are at least middle relationships for any pair price of agriculture. This result show that one price of agriculture will affect other price of agriculture, such as the price of wheat can be affected by the price of rice.

Table 6. Correlations interval based on temperature and the price of agriculture in Thailand.

Fuzzy correlation	U$:TB	Sugar	Corn	Wheat	Rice
U$:TB	—	−0.373[1] (−0.540, −0.532)[2] (−0.560, −0.500)[3]	−0.551[1] (−0.850, −0.783)[2] (−0.868, −0.815)[3]	−0.545[1] (−0.893, −0.780)[2] (−0.902, −0.868)[3]	−0.013[1] (−0.019, 0.083)[2] (0.015, 0.143)[3]
Sugar		—	0.750[1] (0.684, 0.972)[2] (0.578, 0.705)[3]	0.585[1] (0.542, 0.781)[2] (0.468, 0.556)[3]	0.659[1] (0.648, 0.886)[2] (0.567, 0.663)[3]
Corn			—	0.797[1] (0.829, 1.000)[2] (0.744, 0.833)[3]	0.741[1] (0.767, 1.000)[2] (0.683, 0.765)[3]
Wheat				—	0.518[1] (0.561, 0.725)[2] (0.512, 0.554)[3]
Rice					—

Note: [1]Denote the value of Definition 1 under $\beta_1 = 0.7$, $\beta_2 = 0.3$.
[2]Denote the value of Definition 2.
[3]Denote the value of Definition 3.

4. Conclusions

Correlation between any two variables has wide applications in many applications. Previous studies have derived some solutions for calculating the correlation coefficient for fuzzy numbers. A common deficiency of those studies is that the correlation coefficients calculated are crisp values, instead of the intuitively believed fuzzy numbers. This chapter 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 large-scale fuzzy questionnaires to make this formula get implemented in practice. Second, when calculating the fuzzy correlation, we adopt $\lambda ar_{xy}$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. Therefore, 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 chapter, a more realistic correlation is obtained, which provides the decision maker with more knowledge and confident to make better decisions.

References

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

Chiang, DA and NP Lin (1999). Correlation of fuzzy sets. *Fuzzy Sets and Systems*, 102, 221–226.

Dubois, D and H Prade (1991). Fuzzy sets in approximate reasoning, Part 1: Inference with possibility distributions. *Fuzzy Sets and Systems*, 40, 143–202.

Hong, D and S Hwang (1995). Correlation of intuitionistic fuzzy sets in probability space. *Fuzzy Sets and Systems*, 75, 77–81.

Hong, D (2006). Fuzzy measures for a correlation coefficient of fuzzy numbers under Tw (the weakest t norm)-based fuzzy arithmetic operations. *Fuzzy Sets and Systems*, 176, 150–160.

Liu, S and C Kao (2002). Fuzzy measures for correlation coefficient of fuzzy numbers. *Fuzzy Sets and Systems*, 128, 267–275.

Lowen, R (1990). A fuzzy language interpolation theorem. *Fuzzy Sets and Systems*, 34, 33–38.

Ruspini, E (1991). Approximate reasoning: Past, present, future. *Information Sciences*, 57, 297–317.

Wu, B and Y Hsu (2002). The use of kernel set and sample memberships in the identification of nonlinear time series. *Soft Computing Journal*, 8(3), 207–216.

Zadeh, LA (1965). Fuzzy sets. *Information and Control*, 8, 338–353.

Yu, C (1993). Correlation of fuzzy numbers. *Fuzzy Sets and systems*, 55, 303–307.