Research Article

Innovative Correlation Coefficient Measurement with Fuzzy Data

Berlin Wu and Chin Feng Hung

Department of Mathematical Sciences, National Chengchi University, Taipei 116, Taiwan

Correspondence should be addressed to Chin Feng Hung; superma.hung@msa.hinet.net

Received 17 December 2015; Revised 21 April 2016; Accepted 28 April 2016

Academic Editor: László T. Kóczy

Copyright © 2016 B. Wu and C. F. Hung. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Correlation coefficients are commonly found with crisp data. In this paper, we use Pearson's correlation coefficient and propose a method for evaluating correlation coefficients for fuzzy interval data. Our empirical studies involve the relationship between mathematics achievement and other projects.

1. Introduction

Human thought processes are mainly based on cognitive awareness of the environment and social phenomena. Human knowledge is fuzzy because of humans' subjective awareness of time and space. Therefore, Wu [1] proposed fuzzy theory in reference to how humans perceive complex and uncertain environmental phenomena.

To determine the correlation between phenomena X and Y , a scatter plot is often used. Using a scatter plot, the correlation between phenomena X and Y can be determined to be positive, negative, or statistically independent.

In traditional statistical analysis, correlation coefficients are often found using crisp data. In this paper, we use Pearson's correlation coefficient to calculate correlation coefficients for fuzzy interval data. Fuzzy correlation coefficients are often applied in the fields of engineering or economics but have also been increasingly emphasized in social sciences.

Fuzzy correlations are referenced in the literature. For instance, Nguyen et al. [2, 3] provided the fundamentals of statistics with fuzzy data. Hong and Hwang [4] established the correlation coefficient of intuitionistic fuzzy sets in probability space by using the generalization of fuzzy sets by Zadeh [5]. Chiang and Lin [6] argued that membership degrees are concrete observational values based on the membership functions of fuzzy sets to define fuzzy correlation coefficients. Chaudhuri and Bhattacharya [7] investigated the correlation

of two fuzzy sets that were defined by the members of the supports, which were ranked to evaluate the correlation coefficients of two fuzzy sets. Hong [8] and Ni and Cheung [9] also suggested some methods for calculating fuzzy correlations. Based on correlation coefficients developed by Liu and Kao [10], Xie and Wu [11] and Yang [12] established fuzzy correlation coefficients and obtained fuzzy correlation intervals based on fuzzy interval sample data. R. Saneifard and R. Saneifard [13] calculated the correlation coefficient for fuzzy data by adopting the method from central interval. Cheng and Yang [14] proposed a method for determining fuzzy correlation coefficients and explained the application of fuzzy correlation. Hanafy et al. [15] evaluated the correlation coefficients of neutrosophic sets by centroid method. Wu et al. [16] developed a new approach for determining fuzzy correlation and applied this approach to 12-year compulsory education in Taiwan. Lin et al. [17] investigated some problems on marketing research by using a soft computing technique and a new statistics tool.

The main purpose of this paper is to develop fuzzy correlation coefficients for fuzzy interval data. We propose a functional formula for determining fuzzy correlation coefficients of two variables. We can find the maximum and minimum values by differentiating our proposed functional formula. However, the formula can be applied not only when the value of one of two data sets is a real number but also when both data sets are real numbers. Using this method

of research, we can provide information for researchers to explain related phenomena in practice.

2. Research Approaches

Let (x_i, y_i) , $i = 1, 2, \dots, n$, be a fuzzy sample set; then, the correlation coefficient between x and y is defined as

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

where $\bar{x}$ and $\bar{y}$ are sample means for x_i and y_i , respectively.

Definition 1 (fuzzy interval number). Let a fuzzy number $A = [a, b]$ be an interval over the real number R , let $c = (a + b)/2$ be the center of interval A , and let $r = (b - a)/2$ be the radius of interval A ; then, interval A can be expressed as $A = [a, b]$ or $A = (c; r)$. Consider interval A a fuzzy interval number.

Consider the fuzzy sample set (x_i, y_i) , $i = 1, 2, \dots, n$, where $x_i = [x_{i1}, x_{i2}]$ and $y_i = [y_{i1}, y_{i2}]$ are fuzzy interval numbers, as shown in Figure 1.

The algorithm of the correlation coefficient between x and y consists of the following five steps.

Step 1. For any fuzzy interval number, $x_i = [x_{i1}, x_{i2}]$ and $y_i = [y_{i1}, y_{i2}]$, $x_i \otimes y_i$ is defined by a rectangle. In addition, the rectangle $x_i \otimes y_i$ has four vertices, A_i , B_i , C_i , and D_i , the coordinates of which are (x_{i1}, y_{i1}) , (x_{i2}, y_{i1}) , (x_{i2}, y_{i2}) , and (x_{i1}, y_{i2}) , respectively, as shown in Figure 2.

Step 2. Choose a point E_i lying in the line segment $\overline{A_i B_i}$ such that two segments' proportion $A_i E_i : E_i B_i = s : (1 - s)$, where $0 \leq s \leq 1$, and the point coordinate $E_i(x_{i1} + s(x_{i2} - x_{i1}), y_{i1})$ is obtained, as shown in Figure 3.

Step 3. Choose a point G_i lying in the line segment $\overline{C_i D_i}$ such that $\overline{E_i G_i}$ parallels $\overline{A_i D_i}$. Next, choose a point F_i lying in the line segment $\overline{E_i G_i}$ such that two segments' proportion $\overline{E_i F_i} : \overline{F_i G_i} = t : (1 - t)$, where $0 \leq t \leq 1$, and the point coordinate $F_i(x_{i1} + s(x_{i2} - x_{i1}), y_{i1} + t(y_{i2} - y_{i1}))$ is obtained, as shown in Figure 4.

Definition 2. The domain set $\Omega = \{(s, t) \mid 0 \leq s \leq 1, 0 \leq t \leq 1\}$.

Step 4. Calculate the correlation coefficient function r_{xy} between x and y by using formula (1). In this case, the correlation coefficient function r_{xy} is a function of two variables s and t for the closed region bounded by Ω and is expressed as $r_{xy} = f(s, t)$.

Step 5. By the differentiation method, we can find the maximum and minimum values of the correlation coefficient function r_{xy} .

2.1. The Assumption of Corresponding Points of Each Rectangle. Our initial idea is to find the correlation coefficient for

FIGURE 1: The scatter plot with fuzzy interval data.

FIGURE 2: The graph of a rectangle $x_i \otimes y_i$.

FIGURE 3: The graph of a rectangle $x_i \otimes y_i$.

FIGURE 4: The graph of a rectangle $x_i \otimes y_i$.

FIGURE 6: The graph of Example 3 for the upper-right point coordinate of each rectangle.

FIGURE 5: The graph of Example 3 for the centroid of each rectangle.

FIGURE 7: The graph of Example 3 for the lower-right point coordinate of each rectangle.

corresponding points of each rectangle. For Example 3, we can find that the correlation coefficient $r_{xy} = 0$ for the centroid of each rectangle.

Example 3. Consider the rectangle sample data $x_1 \otimes y_1 = [0, 2] \otimes [4, 6]$, $x_2 \otimes y_2 = [4, 8] \otimes [8, 10]$, and $x_3 \otimes y_3 = [10, 12] \otimes [0, 10]$, as shown in Figure 5.

For Example 3, we also can find the correlation coefficient $r_{xy} = 0.918$ for the upper-right point coordinate, the correlation coefficient $r_{xy} = -0.397$ for the lower-right point coordinate, the correlation coefficient $r_{xy} = -0.596$ for the lower-left point coordinate, and the correlation coefficient $r_{xy} = 0.803$ for the upper-left point coordinate of each rectangle, as shown in Figures 6, 7, 8, and 9.

We also can find the correlation coefficient $r_{xy} = 0.466$ for interior corresponding point of each rectangle; for example, the coordinate $(s, t) = (1/3, 3/4)$ as shown in Figure 10.

We assume that the coordinate (s, t) of corresponding point of each rectangle is fixed; for example, the coordinate $(s, t) = (1/2, 1/2)$ for Figure 5 and the coordinate $(s, t) = (1, 1)$ for Figure 6. However, we do not consider the case that each

rectangle may have different coordinates (s, t) , for example, as shown in Figure 11.

The coordinate $(s, t) = (0, 1)$ of rectangle A , the coordinate $(s, t) = (1, 0)$ of rectangle B , and the coordinate $(s, t) = (0, 0)$ of rectangle C in Figure 11; their coordinates (s, t) are not different.

According to Figures 6–9, we cannot say the maximal value of the correlation coefficient $r_{xy} = 0.918$ and the minimal value of the correlation coefficient $r_{xy} = -0.596$, because there are infinitely many corresponding points for boundary points and interior points of each rectangle; we must evaluate every correlation coefficient for corresponding points of each rectangle. Therefore, we use Steps 1 to 5 and the differential rule of two variables to evaluate the maximal and the minimal values of the correlation coefficient.

Based on Example 3, three point coordinates, $F_1(2s, 4 + 2t)$, $F_2(4 + 4s, 8 + 2t)$, and $F_3(10 + 2s, 10t)$, are obtained. We then find the sample means $\bar{x} = (14 + 8s)/3$ and $\bar{y} = (12 + 14t)/3$,

FIGURE 8: The graph of Example 3 for the lower-left point coordinate of each rectangle.

FIGURE 9: The graph of Example 3 for the upper-left point coordinate of each rectangle.

respectively. Therefore, the correlation coefficient function between x and y is

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

$$= \frac{(-9 + 3s) + (16 - 2s)t}{2\sqrt{19 - s + s^2}\sqrt{3 - 6t + 4t^2}},$$

for the closed region bounded by Ω .

The first-order derivatives for s and t are, respectively,

$$f_s(s, t) = \frac{105 - 60t + 15s - 30st}{4\sqrt{3 - 6t + 4t^2} \cdot (19 - s + s^2)^{3/2}}, \quad (3)$$

$$f_t(s, t) = \frac{42 + 6s - 24t - 12st}{4\sqrt{19 - s + s^2} \cdot (3 - 6t + 4t^2)^{3/2}}.$$

Let $f_s(s, t) = f_t(s, t) = 0$; it follows that $(s+2)(2t-1) = 5$. There is no critical point for the equation $(s+2)(2t-1) = 5$ bounded by Ω . The reason is as follows: If $0 \leq t \leq 1$, then we obtain $-7 \leq s \leq -1/3$, and if $0 \leq s \leq 1$, then we obtain

FIGURE 10: The graph of Example 3 for the coordinate $(s, t) = (1/3, 3/4)$ of interior corresponding point of each rectangle.

FIGURE 11: There are different coordinates (s, t) in rectangles A, B, and C.

$4/3 \leq t \leq 7/4$. Hence, their critical points on the equation $(s+2)(2t-1) = 5$ do not belong to the set Ω .

The boundary of the region consists of the lines $s = 0$, $s = 1$, $t = 0$, and $t = 1$. Consideration of extrema on the boundary of the region along $s = 0$ leads to the function $f(0, t) = (-9 + 16t)/2\sqrt{19}\sqrt{3 - 6t + 4t^2}$, $0 \leq t \leq 1$. There is no critical point when setting the derivative with respect to t equal to zero. The endpoints are $(0, 0)$ and $(0, 1)$. Consideration of extrema on the boundary of the region along $s = 1$ leads to the function $f(1, t) = (-6 + 14t)/2\sqrt{19}\sqrt{3 - 6t + 4t^2}$, $0 \leq t \leq 1$. There is no critical point when setting the derivative with respect to t equal to zero. The endpoints are $(1, 0)$ and $(1, 1)$. Consideration of extrema on the boundary of the region along $t = 0$ leads to the function $f(s, 0) = (-9 + 3s)/2\sqrt{3}\sqrt{19 - s + s^2}$, $0 \leq s \leq 1$. There is no critical point when setting the derivative with respect to s equal to zero. The endpoints are $(0, 0)$ and $(1, 0)$. Consideration of extrema on the boundary of the region along $t = 1$ leads to the function $f(s, 1) = (7 + s)/2\sqrt{19 - s + s^2}$, $0 \leq s \leq 1$. There is no critical point when setting the derivative with respect to s equal to zero. The endpoints are $(0, 1)$ and $(1, 1)$. All candidates for the maximum and minimum values are listed in Table 1. We see that the minimum value is $f(0, 0) \approx -0.596$; the maximum

TABLE 1

(s, t)	$f(s, t)$
$(0, 0)$	-0.596
$(0, 1)$	0.803
$(1, 0)$	-0.397
$(1, 1)$	0.918

value is $f(1, 1) \approx 0.918$. Therefore, the fuzzy interval number of the correlation coefficient is $[-0.596, 0.918]$.

Theorem 4. If f is continuous on a closed, bounded region, then f has a maximum value and a minimum value on the region. These extrema occur either (1) where all first partial derivatives of f are zero, (2) where some first partial derivative of f does not exist, or (3) on the boundary of the region.

Proof. See [18]. $\square$

3. Case Studies

In this section, we discuss some cases of fuzzy correlation coefficients. First, we analyze a case in which maximal value 1 or minimal values -1 of the correlation coefficient occur for the closed region bounded by Ω . Second, we analyze a case in which maximal value 1 or minimal values -1 of the correlation coefficient do not occur for the closed region bounded by Ω .

Case 1. Maximal value 1 or minimal values -1 of the correlation coefficient occur for the closed region bounded by Ω .

Example 5. Consider the rectangle sample data $x_1 \otimes y_1 = [0, 2] \otimes [0, 2]$, $x_2 \otimes y_2 = [2, 6] \otimes [2, 6]$, and $x_3 \otimes y_3 = [6, 8] \otimes [6, 8]$, as shown in Figure 12.

Based on the previous discussion, three point coordinates, $F_1(2s, 2t)$, $F_2(2 + 4s, 2 + 4t)$, and $F_3(6 + 2s, 6 + 2t)$, are obtained. We then find the sample means $\bar{x} = (8 + 8s)/3$ and $\bar{y} = (8 + 8t)/3$, respectively. Therefore, the correlation coefficient function between x and y is

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} = f(s, t) \quad (4)$$

$$= \frac{(14 - s) + (-1 + 2s)t}{2\sqrt{7 - s + s^2}\sqrt{7 - t + t^2}},$$

for the closed region bounded by Ω .

The first-order derivatives for s and t are, respectively,

$$f_s(s, t) = \frac{27t - 27s}{2\sqrt{7 - t + t^2} \cdot (7 - s + s^2)^{3/2}}, \quad (5)$$

$$f_t(s, t) = \frac{27s - 27t}{2\sqrt{7 - s + s^2} \cdot (7 - t + t^2)^{3/2}}.$$

Let $f_s(s, t) = f_t(s, t) = 0$; it follows that $s = t$. Infinitely many critical points are found for the equation $s = t$ bounded by Ω . For example, there are two points, $(0, 0)$ or $(1, 1)$.

FIGURE 12: The graph of Example 5.

TABLE 2

(s, t)	$f(s, t)$
$(0, 0)$	1
$(0, 1)$	0.929
$(1, 0)$	0.929
$(1, 1)$	1

The boundary of the region consists of the lines $s = 0$, $s = 1$, $t = 0$, and $t = 1$. Consideration of extrema on the boundary of the region along $s = 0$ leads to the function $f(0, t) = (14 - t)/2\sqrt{7}\sqrt{7 - t + t^2}$, $0 \leq t \leq 1$. Setting the derivative with respect to t equal to zero gives the point $(0, 0)$. The endpoints are $(0, 0)$ and $(0, 1)$. Consideration of extrema on the boundary of the region along $s = 1$ leads to the function $f(1, t) = (13 + t)/2\sqrt{7}\sqrt{7 - t + t^2}$, $0 \leq t \leq 1$. Setting the derivative with respect to t equal to zero gives the point $(1, 1)$. The endpoints are $(1, 0)$ and $(1, 1)$. Consideration of extrema on the boundary of the region along $t = 0$ leads to the function $f(s, 0) = (14 - s)/2\sqrt{7}\sqrt{7 - s + s^2}$, $0 \leq s \leq 1$. Setting the derivative with respect to t equal to zero gives the point $(0, 0)$. The endpoints are $(0, 0)$ and $(1, 0)$. Consideration of extrema on the boundary of the region along $t = 1$ leads to the function $f(s, 1) = (13 + s)/2\sqrt{7}\sqrt{7 - s + s^2}$, $0 \leq s \leq 1$. Setting the derivative with respect to t equal to zero gives the point $(1, 1)$. The endpoints are $(0, 1)$ and $(1, 1)$. All candidates for the maximum and minimum values are listed in Table 2. We see that the minimum value is $f(0, 1) = 0.929$; the maximum value is $f(1, 1) = 1$. Therefore, the fuzzy interval number of the correlation coefficient is $[0.929, 1]$.

In this case, the center points of these three rectangles are positively correlated, and $r_{xy} = 1$. Moreover, these three rectangles are approximately symmetric to the straight line $y = x$. Hence, the tendency of positive correlation of these three rectangles is high. In other words, the fuzzy correlation coefficient r_{xy} may have a smaller range.

FIGURE 13: The graph of Example 6.

Example 6. Consider the rectangle sample data $x_1 \otimes y_1 = [1, 3] \otimes [1, 2]$, $x_2 \otimes y_2 = [3, 5] \otimes [0, 3]$, and $x_3 \otimes y_3 = [3, 9] \otimes [3, 4]$, as shown in Figure 13.

Based on the previous discussion, three point coordinates, $F_1(1 + 2s, 1 + t)$, $F_2(3 + 2s, 3t)$, and $F_3(3 + 6s, 3 + t)$, are obtained. We then find the sample means $\bar{x} = (7 + 10s)/3$ and $\bar{y} = (4 + 5t)/3$, respectively. Therefore, the correlation coefficient function between x and y is

$$r_{xy} = f(s, t) = \frac{(1 + 10s) + (2 - 4s)t}{2\sqrt{1 + 2s + 4s^2}\sqrt{7 - 8t + 4t^2}}, \quad (6)$$

for the closed region bounded by Ω .

The first-order derivatives for s and t are, respectively,

$$\begin{aligned} f_s(s, t) &= \frac{(9 - 6t) + (6 - 12t)s}{2\sqrt{7 - 8t + 4t^2} \cdot (1 + 2s + 4s^2)^{3/2}}, \\ f_t(s, t) &= \frac{(9 + 6s) + (-6 - 12s)t}{\sqrt{1 + 2s + 4s^2} \cdot (7 - 8t + 4t^2)^{3/2}}. \end{aligned} \quad (7)$$

Let $f_s(s, t) = f_t(s, t) = 0$; it follows that $(2t - 1)(2s + 1) = 2$. Infinitely many critical points can be found for equation $(2t - 1)(2s + 1) = 2$ bounded by Ω . For example, there are two points, $(1/2, 1)$ or $(1, 5/6)$.

The boundary of the region consists of the lines $s = 0$, $s = 1$, $t = 0$, and $t = 1$. Consideration of extrema on the boundary of the region along $s = 0$ leads to the function $f(0, t) = (1 + 2t)/2\sqrt{7 - 8t + 4t^2}$, $0 \leq t \leq 1$. There is no critical point when setting the derivative with respect to t equal to zero. The endpoints are $(0, 0)$ and $(0, 1)$. Consideration of extrema on the boundary of the region along $s = 1$ leads to the function $f(1, t) = (11 - 2t)/2\sqrt{7 - 8t + 4t^2}$, $0 \leq t \leq 1$. Setting the derivative with respect to t equal to zero gives the point $(1, 5/6)$. The endpoints are $(1, 0)$ and $(1, 1)$. Consideration of extrema on the boundary of the region along $t = 0$ leads to the function $f(s, 0) = (1 + 10s)/2\sqrt{7}\sqrt{1 + 2s + 4s^2}$, $0 \leq s \leq 1$. There is no critical point when setting the derivative with respect to s equal to zero. The endpoints are $(0, 0)$ and $(1, 0)$. Consideration of extrema on the boundary of the region along $t = 1$ leads to the function $f(s, 1) = (3 + 6s)/2\sqrt{3}\sqrt{1 + 2s + 4s^2}$, $0 \leq s \leq 1$. Setting

TABLE 3

(s, t)	$f(s, t)$
$(1, 5/6)$	1
$(1/2, 1)$	1
$(0, 0)$	0.189
$(0, 1)$	0.866
$(1, 0)$	0.786
$(1, 1)$	0.982

FIGURE 14: The graph of Example 7.

the derivative with respect to s equal to zero gives the point $(1/2, 1)$. The endpoints are $(0, 1)$ and $(1, 1)$. All candidates for the maximum and minimum values are listed in Table 3. We see that the minimum value is $f(0, 0) = 0.189$; the maximum value is $f(1, 1) = 5/6$. Therefore, the fuzzy interval number of the correlation coefficient is $[0.189, 1]$.

In this case, the center points of these three rectangles are positively correlated, but $0 < r_{xy} < 1$. Moreover, these three rectangles are not symmetric to any straight lines. Hence, the tendency of positive correlation of these three rectangles is not evident. In other words, the fuzzy correlation coefficient r_{xy} may be a large range.

Case 2. Maximal value 1 or minimal values -1 of the correlation coefficient do not occur for the closed region bounded by Ω .

Example 7. Consider the rectangle sample data $x_1 \otimes y_1 = [0, 2] \otimes [0, 2]$, $x_2 \otimes y_2 = [2, 6] \otimes [2, 4]$, and $x_3 \otimes y_3 = [7, 9] \otimes [0, 4]$, as shown in Figure 14.

Based on the previous discussion, three point coordinates, $F_1(2s, 2t)$, $F_2(2 + 4s, 2 + 2t)$, and $F_3(7 + 2s, 4t)$, are obtained. We then find the sample means $\bar{x} = (9 + 8s)/3$ and $\bar{y} = (2 + 8t)/3$, respectively. Therefore, the correlation coefficient function between x and y is

$$r_{xy} = f(s, t) = \frac{(-3 + 4s) + (12 - 2s)t}{2\sqrt{39 - 6s + 4s^2}\sqrt{1 - t + t^2}}, \quad (8)$$

for the closed region bounded by Ω .

TABLE 4

(s, t)	$f(s, t)$
$(0, 0)$	-0.240
$(0, 1)$	0.721
$(1, 0)$	0.082
$(1, 1)$	0.904

The first-order derivatives for s and t are, respectively,

$$f_s(s, t) = \frac{(147 - 42t) + (-42t)s}{2\sqrt{1-t+t^2} \cdot (39-6s+4s^2)^{3/2}},$$

$$f_t(s, t) = \frac{21 + (-6-6s)t}{4\sqrt{39-6s+4s^2} \cdot (1-t+t^2)^{3/2}}. \quad (9)$$

Let $f_s(s, t) = f_t(s, t) = 0$; it follows that $t(2s+2) = 7$. There is no critical point for equation $t(2s+2) = 7$ bounded by Ω . The reason is as follows: If $0 \leq t \leq 1$, then we obtain $s \geq 5/2$, and if $0 \leq s \leq 1$, then we obtain $7/4 \leq t \leq 7/2$. Hence, the critical points for equation $t(2s+2) = 7$ do not belong to the set Ω .

The boundary of the region consists of the lines $s = 0$, $s = 1$, $t = 0$, and $t = 1$. Consideration of extrema on the boundary of the region along $s = 0$ leads to the function $f(0, t) = (-3 + 12t)/2\sqrt{39}\sqrt{1-t+t^2}$, $0 \leq t \leq 1$. There is no critical point when setting the derivative with respect to t equal to zero. The endpoints are $(0, 0)$ and $(0, 1)$. Consideration of extrema on the boundary of the region along $s = 1$ leads to the function $f(1, t) = (1 + 10t)/2\sqrt{37}\sqrt{1-t+t^2}$, $0 \leq t \leq 1$. There is no critical point when setting the derivative with respect to t equal to zero. The endpoints are $(1, 0)$ and $(1, 1)$. Consideration of extrema on the boundary of the region along $t = 0$ leads to the function $f(s, 0) = (-3 + 4s)/2\sqrt{39-6s+4s^2}$, $0 \leq s \leq 1$. There is no critical point when setting the derivative with respect to s equal to zero. The endpoints are $(0, 0)$ and $(1, 0)$. Consideration of extrema on the boundary of the region along $t = 1$ leads to the function $f(s, 1) = (9 + 2s)/2\sqrt{39-6s+4s^2}$, $0 \leq s \leq 1$. There is no critical point when setting the derivative with respect to s equal to zero. The endpoints are $(0, 1)$ and $(1, 1)$. All candidates for the maximum and minimum values are listed in Table 4. We see that the minimum value is $f(0, 0) \approx -0.240$; the maximum value is $f(1, 1) \approx 0.904$. Therefore, the fuzzy interval number of the correlation coefficient is $[-0.240, 0.904]$.

Based on the scatter plots of Examples 6 and 7, we intuitively think that the scatter plot of Example 7 is more dispersed than the scatter plot of Example 6. Therefore, the fuzzy correlation coefficient r_{xy} of Example 7 will have a larger range.

Example 8. Consider the rectangle sample data $x_1 \otimes y_1 = [0, 2] \otimes [0, 2]$, $x_2 \otimes y_2 = [2, 6] \otimes [2, 4]$, $x_3 \otimes y_3 = [7, 9] \otimes [0, 4]$, and $x_4 \otimes y_4 = [9, 13] \otimes [4, 6]$, as shown in Figure 15.

Based on the previous discussion, four point coordinates, $F_1(2s, 2t)$, $F_2(2+4s, 2+2t)$, $F_3(7+2s, 4t)$, and $F_4(9+4s, 4+2t)$, are obtained. We then find the sample means $\bar{x} = (9 + 6s)/2$

FIGURE 15: The graph of Example 8.

and $\bar{y} = (3 + 5t)/2$, respectively. Therefore, the correlation coefficient function between x and y is

$$r_{xy} = f(s, t) = \frac{(13 + 6s) + (5 - 2s)t}{\sqrt{53 + 8s + 4s^2}\sqrt{11 - 6t + 3t^2}}, \quad (10)$$

for the closed region bounded by Ω .

The first-order derivatives for s and t are, respectively,

$$f_s(s, t) = \frac{(266 - 126t) + (-28 - 28t)s}{\sqrt{11 - 6t + 3t^2} \cdot (53 + 8s + 4s^2)^{3/2}},$$

$$f_t(s, t) = \frac{(94 - 4s) + (-54 - 12s)t}{\sqrt{53 + 8s + 4s^2} \cdot (11 - 6t + 3t^2)^{3/2}}. \quad (11)$$

First, let $f_s(s, t) = f_t(s, t) = 0$; it follows that $t = (47 - 2s)/(27 + 6s) = (133 - 14s)/(63 + 14s)$. We obtain $(s, t) = (5/2, 1)$. This critical point for equation $t = (47 - 2s)/(27 + 6s) = (133 - 14s)/(63 + 14s)$ does not belong to the set Ω . Therefore, no local maximum or minimum is in Ω .

Second, the boundary of the region consists of the lines $s = 0$, $s = 1$, $t = 0$, and $t = 1$. Consideration of extrema on the boundary of the region along $s = 0$ leads to the function $f(0, t) = (13 + 5t)/\sqrt{53}\sqrt{11 - 6t + 3t^2}$, $0 \leq t \leq 1$. There is no critical point when setting the derivative with respect to t equal to zero. The endpoints are $(0, 0)$ and $(0, 1)$. Consideration of extrema on the boundary of the region along $s = 1$ leads to the function $f(1, t) = (19 + 3t)/\sqrt{65}\sqrt{11 - 6t + 3t^2}$, $0 \leq t \leq 1$. There is no critical point when setting the derivative with respect to t equal to zero. The endpoints are $(1, 0)$ and $(1, 1)$. Consideration of extrema on the boundary of the region along $t = 0$ leads to the function $f(s, 0) = (13 + 6s)/\sqrt{11}\sqrt{53 + 8s + 4s^2}$, $0 \leq s \leq 1$. There is no critical point when setting the derivative with respect to s equal to zero. The endpoints are $(0, 0)$ and $(1, 0)$. Consideration of extrema on the boundary of the region along $t = 1$ leads to the function $f(s, 1) = (18 + 4s)/2\sqrt{2}\sqrt{53 + 8s + 4s^2}$, $0 \leq s \leq 1$. There is no critical point when setting the derivative with respect to s equal to zero. The endpoints are $(0, 1)$ and $(1, 1)$. All candidates for the maximum and minimum values are listed in Table 5. We see

TABLE 5

(s, t)	$f(s, t)$
$(0, 0)$	0.538
$(0, 1)$	0.874
$(1, 0)$	0.711
$(1, 1)$	0.965

FIGURE 16: Scatter plot 1.

that the minimum value is $f(0, 0) \approx 0.538$; the maximum value is $f(1, 1) \approx 0.965$. Therefore, the fuzzy interval number of the correlation coefficient is $[0.538, 0.965]$.

Comparing the scatter plots of Examples 7 and 8, we intuitively think that the scatter plot of Example 8 is more concentrated and has a greater tendency of positive correlation than that of the scatter plot of Example 7. Moreover, when fuzzy interval data of the scatter plot increase, the fuzzy correlation coefficient will have a smaller range.

Lin et al. [17] proposed the formula of the fuzzy correlation coefficient r_{xy} as the following four situations:

$$r_{xy} = \begin{cases} [r_c, \min(1, r_c + \delta)] & \text{if } r_c \geq 0, r_l \geq 0 \\ [r_c - \delta, r_c] & \text{if } r_c \geq 0, r_l < 0 \\ [r_c, r_c + \delta] & \text{if } r_c < 0, r_l \geq 0 \\ [\max(-1, r_c - \delta), r_c] & \text{if } r_c < 0, r_l < 0, \end{cases} \quad (12)$$

where r_c is the correlation coefficient of the center point of each rectangle, r_l is the correlation coefficient of the interval lengths l_{x_i} and l_{y_i} of each of the fuzzy interval numbers x_i and y_i , and $\delta = 1 - \ln(1 + |r_l|)/|r_l|$.

Next, the four scatter plots are observed as follows.

Intuitively, the degrees of spread of the four scatter plots (refer to Figures 16, 17, 18, and 19) do not seem to be the same. Hence, the fuzzy correlation coefficient should not be equal. But the formula (12) of Wu et al. [16] shows that the four fuzzy correlation coefficients are equal, and $r_{xy} = [1, 1]$.

However, our proposed method obtains different results. The four fuzzy correlation coefficients (refer to Figures 16, 17, 18, and 19) obtained through our proposed method

FIGURE 17: Scatter plot 2.

FIGURE 18: Scatter plot 3.

are $[1, 1]$, $[0.976, 1]$, $[0.922, 1]$, and $[0.857, 1]$, respectively. Therefore, our proposed method produces results that are more consistent with our intuition.

4. Empirical Studies

In this section, we discuss some applications of fuzzy correlation coefficients. First, we analyze a case in which two data sets are fuzzy interval numbers. Second, we change the case to one in which one data set is a fuzzy interval number, and the other is a real number. Finally, we analyze a case in which both data sets are real numbers.

To understand the factors influencing mathematics achievement at a school, we investigate 10 students' data.

Example 9. Consider the rectangle sample data for 10 students: $x_1 \otimes y_1 = [80, 90] \otimes [1, 1.5]$, $x_2 \otimes y_2 = [80, 90] \otimes [6.5, 7]$, $x_3 \otimes y_3 = [60, 80] \otimes [4, 5]$, $x_4 \otimes y_4 = [90, 100] \otimes [1.5, 2.5]$, $x_5 \otimes y_5 = [40, 70] \otimes [16, 17]$, $x_6 \otimes y_6 = [70, 80] \otimes [15, 16]$, $x_7 \otimes y_7 = [60, 80] \otimes [15, 17]$, $x_8 \otimes y_8 = [80, 100] \otimes [1, 3]$, $x_9 \otimes y_9 = [80, 90] \otimes [0, 3]$, and $x_{10} \otimes y_{10} = [80, 90] \otimes [4.5, 5.5]$,

FIGURE 19: Scatter plot 4.

FIGURE 20: The graph of Example 9.

where x_i and y_i denote the mathematics score and weekly online time, respectively, of a student i , $i = 1, 2, \dots, 10$, as shown in Figure 20.

Based on the previous discussion, the correlation coefficient function between x and y is

$$r_{xy} = f(s, t) = \frac{(-639 + 197.5s) + (4 + 5s)t}{\sqrt{10}\sqrt{196 - 160s + 45s^2}\sqrt{372.7 - 14.2t + 5.54t^2}}, \quad (13)$$

for the closed region bounded by Ω .

The first-order derivatives for s and t are, respectively,

$$\begin{aligned} f_s(s, t) &= \frac{(-12410 + 1300t) + (12955 - 580t)s}{\sqrt{10}\sqrt{196 - 160s + 45s^2} \cdot (372.7 - 14.2t + 5.54t^2)^{3/2}}, \\ f_t(s, t) &= \frac{(-3046.1 + 3265.75s) + (3511.66 - 1129.65s)t}{\sqrt{10}\sqrt{196 - 160s + 45s^2} \cdot (372.7 - 14.2t + 5.54t^2)^{3/2}}. \end{aligned} \quad (14)$$

First, let $f_s(s, t) = f_t(s, t) = 0$; it follows that $(s, t) \approx (3.239, 51.066) \notin \Omega$ or $(3.239, 51.066) \notin \Omega$. Therefore, no local maximum or minimum is in Ω .

Second, based on the previous discussion, the fuzzy interval number of the correlation coefficient is $[-0.804, -0.748]$.

Clearly, x and y have a highly negative correlation. In other words, a higher mathematics score correlates with a lower weekly online time. The weekly online time of a student negatively influences the student's mathematics score.

Example 10. If Example 9 is adjusted to $x_1 \otimes y_1 = [85, 85] \otimes [1, 1.5]$, $x_2 \otimes y_2 = [85, 85] \otimes [6.5, 7]$, $x_3 \otimes y_3 = [70, 70] \otimes [4, 5]$, $x_4 \otimes y_4 = [95, 95] \otimes [1.5, 2.5]$, $x_5 \otimes y_5 = [55, 55] \otimes [16, 17]$, $x_6 \otimes y_6 = [75, 75] \otimes [15, 16]$, $x_7 \otimes y_7 = [70, 70] \otimes [15, 17]$, $x_8 \otimes y_8 = [90, 90] \otimes [1, 3]$, $x_9 \otimes y_9 = [85, 85] \otimes [0, 3]$, and $x_{10} \otimes y_{10} = [85, 85] \otimes [4.5, 5.5]$ (i.e., x_i is a real number, $i = 1, 2, \dots, 10$), then the correlation coefficient function between x and y is

$$\begin{aligned} r_{xy} &= f(s = 0.5, t) \\ &= \frac{-540.25 + 6.5t}{11.28\sqrt{10}\sqrt{372.7 - 14.2t + 5.54t^2}}, \end{aligned} \quad (15)$$

bounded by the set $\{t \mid 0 \leq t \leq 1\}$.

The first-order derivative with respect to t is

$$\begin{aligned} f_t(s = 0.5, t) &= \frac{-1413.23 + 2946.84t}{11.28\sqrt{10} \cdot (372.7 - 14.2t + 5.54t^2)^{3/2}}. \end{aligned} \quad (16)$$

Let $f_t(s = 0.5, t) = 0$; it follows that $t \approx 0.48$, which is a critical point bounded by the set $\{t \mid 0 \leq t \leq 1\}$. Based on the previous discussion, the fuzzy interval number of the correlation coefficient is $[-0.786, -0.784]$.

Example 11. If Example 9 is adjusted to $x_1 \otimes y_1 = [85, 85] \otimes [1.25, 1.25]$, $x_2 \otimes y_2 = [85, 85] \otimes [6.75, 6.75]$, $x_3 \otimes y_3 = [70, 70] \otimes [4.5, 4.5]$, $x_4 \otimes y_4 = [95, 95] \otimes [2, 2]$, $x_5 \otimes y_5 = [55, 55] \otimes [16.5, 16.5]$, $x_6 \otimes y_6 = [75, 75] \otimes [15.5, 15.5]$, $x_7 \otimes y_7 = [70, 70] \otimes [16, 16]$, $x_8 \otimes y_8 = [90, 90] \otimes [2, 2]$, $x_9 \otimes y_9 = [85, 85] \otimes [1.5, 1.5]$, and $x_{10} \otimes y_{10} = [85, 85] \otimes [5, 5]$ (i.e., both x_i and y_i are real numbers, $i = 1, 2, \dots, 10$), then the correlation coefficient between x and y is $r_{xy} = f(s = 0.5, t = 0.5) = -0.786$.

According to the results of Examples 9, 10, and 11, we find that Example 9 is the generalized situation of Examples 10 and 11.

Example 12. Consider the rectangle sample data of 10 students: $x_1 \otimes y_1 = [80, 90] \otimes [8, 8.5]$, $x_2 \otimes y_2 = [80, 90] \otimes [7, 7.5]$, $x_3 \otimes y_3 = [60, 80] \otimes [9, 10.5]$, $x_4 \otimes y_4 = [90, 100] \otimes [8, 8.5]$, $x_5 \otimes y_5 = [40, 70] \otimes [6, 7.5]$, $x_6 \otimes y_6 = [70, 80] \otimes [10, 11]$, $x_7 \otimes y_7 = [60, 80] \otimes [7, 8]$, $x_8 \otimes y_8 = [80, 100] \otimes [8, 10]$, $x_9 \otimes y_9 = [80, 90] \otimes [6.5, 8]$, and $x_{10} \otimes y_{10} = [80, 90] \otimes [7.5, 8.5]$, where x_i and y_i denote the mathematics score and weekly sleeping time, respectively, of a student i , $i = 1, 2, \dots, 10$, as shown in Figure 21.

FIGURE 21: The graph of Example 12.

Based on the previous discussion, the correlation coefficient function between x and y is

$$r_{xy} = f(s, t) = \frac{(36 - 25s) + (-27 + 20s)t}{\sqrt{10}\sqrt{196 - 160s + 45s^2}\sqrt{12.6 - 0.9t + 2.4t^2}}, \quad (17)$$

for the closed region bounded by Ω .

The first-order derivatives for s and t are, respectively,

$$\begin{aligned} f_s(s, t) &= \frac{(-2020 + 1760t) + (380 - 385t)s}{\sqrt{10}\sqrt{12.6 + 0.9t + 2.4t^2} \cdot (196 - 160s + 45s^2)^{3/2}}, \\ f_t(s, t) &= \frac{(-648 + 481.5s) + (-148.5 + 102s)t}{2\sqrt{10}\sqrt{196 - 160s + 45s^2} \cdot (12.6 - 0.9t + 2.4t^2)^{3/2}}. \end{aligned} \quad (18)$$

First, let $f_s(s, t) = f_t(s, t) = 0$; it follows that $(s, t) \approx (4.697, -4.881) \notin \Omega$ or $(1.368, 1.216) \notin \Omega$. Therefore, no local maximum or minimum is in Ω .

Second, based on the previous discussion, the fuzzy interval number of the correlation coefficient is $[0.037, 0.229]$.

Based on the previous discussion, there is a minor positive correlation between mathematics score and weekly sleeping time. Therefore, a higher mathematics score correlates with a lower weekly sleeping time. The influence of weekly sleeping time on mathematics score is minor.

Example 13. Consider the rectangle sample data of 10 students: $x_1 \otimes y_1 = [80, 90] \otimes [70, 80]$, $x_2 \otimes y_2 = [80, 90] \otimes [80, 90]$, $x_3 \otimes y_3 = [60, 80] \otimes [70, 80]$, $x_4 \otimes y_4 = [90, 100] \otimes [80, 90]$, $x_5 \otimes y_5 = [40, 70] \otimes [60, 70]$, $x_6 \otimes y_6 = [70, 80] \otimes [60, 80]$, $x_7 \otimes y_7 = [60, 80] \otimes [70, 80]$, $x_8 \otimes y_8 = [80, 100] \otimes [80, 90]$, $x_9 \otimes y_9 = [80, 90] \otimes [60, 70]$, and $x_{10} \otimes y_{10} = [80, 90] \otimes [70, 80]$, where x_i and y_i denote the mathematics and Chinese scores, respectively, of a student i , $i = 1, 2, \dots, 10$, as shown in Figure 22.

FIGURE 22: The graph of Example 13.

Based on the previous discussion, the correlation coefficient function between x and y is

$$r_{xy} = f(s, t) = \frac{(60 - 10s) + (-2 - 5s)t}{\sqrt{196 - 160s + 45s^2}\sqrt{60 - 20t + 9t^2}}, \quad (19)$$

for the closed region bounded by Ω .

The first-order derivatives with for s and t are, respectively,

$$\begin{aligned} f_s(s, t) &= \frac{(2840 - 1140t) + (-1900 + 490t)s}{\sqrt{60 - 20t + 9t^2} \cdot (196 - 160s + 45s^2)^{3/2}}, \\ f_t(s, t) &= \frac{(480 - 400s) + (-520 + 140s)t}{\sqrt{196 - 160s + 45s^2} \cdot (60 - 20t + 9t^2)^{3/2}}. \end{aligned} \quad (20)$$

First, let $f_s(s, t) = f_t(s, t) = 0$; it follows that $(s, t) \approx (8.324, 4.416) \notin \Omega$ or $(1.596, -0.534) \notin \Omega$. Therefore, no local maximum or minimum is in Ω .

Second, based on the previous discussion, the fuzzy interval number of the correlation coefficient is $[0.553, 0.717]$.

Based on the previous discussion, x and y have a highly positive correlation. Therefore, a higher mathematics score correlates with a higher Chinese score. Students' Chinese scores positively influence their mathematics scores.

5. Conclusion

Scientists are accustomed to using binary logic to analyze information. Human logic is fuzzy and complex, and applying binary logic to analyze human thought processes causes some distortion. Fuzzy logic is based on human thought processes, and fuzzy logic has therefore been increasingly applied to social science.

Possible methods of calculating fuzzy correlation coefficients are proposed in the literature, but understanding most formulas used in the literature requires a strong mathematical background. In this paper, we use Pearson's correlation coefficient and the differentiation method to evaluate fuzzy correlation coefficients, which can be applied to cases in

which two data sets are fuzzy interval numbers, one of two data sets is a fuzzy interval number and the other is a real number, and both data sets are real numbers.

This paper discusses only fuzzy correlation coefficients of fuzzy interval number. However, we will extend the research method that we used to triangular or trapezoidal fuzzy numbers in the future.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

This work of Berlin Wu is partially supported by the National Science Council of the Republic of China under Contract 102-2410-H-004-182.

References

- [1] B. Wu, *Introduction of Fuzzy Statistics*, Wu-Naan Book, Taipei, Taiwan, 2005.
- [2] H. T. Nguyen, V. Kreinovich, B. Wu, and G. Xiang, *Computing Statistics under Interval and Fuzzy Uncertainty*, Springer, Heidelberg, Germany, 2011.
- [3] H. Nguyen and B. Wu, *Fundamentals of Statistics with Fuzzy Data*, Springer, Heidelberg, Germany, 2006.
- [4] D. H. Hong and S. Y. Hwang, "Correlation of intuitionistic fuzzy sets in probability spaces," *Fuzzy Sets and Systems*, vol. 75, no. 1, pp. 77–81, 1995.
- [5] L. A. Zadeh, "Fuzzy sets as a basis for a theory of possibility," *Fuzzy Sets and Systems*, vol. 1, no. 1, pp. 3–28, 1978.
- [6] D.-A. Chiang and N. P. Lin, "Correlation of fuzzy sets," *Fuzzy Sets and Systems*, vol. 102, no. 2, pp. 221–226, 1999.
- [7] B. B. Chaudhuri and A. Bhattacharya, "On correlation between two fuzzy sets," *Fuzzy Sets and Systems*, vol. 118, no. 3, pp. 447–456, 2001.
- [8] D. Hong, "Fuzzy measures for a correlation coefficient of fuzzy numbers under TW (the weakest t-norm)-based fuzzy arithmetic operations," *Information Sciences*, vol. 176, no. 2, pp. 150–160, 2006.
- [9] Y. Ni and J. Y. Cheung, "Correlation coefficient estimate for fuzzy data," in *Intelligent Systems Design and Applications*, vol. 23 of *Advances in Soft Computing*, pp. 2138–2144, Springer, Berlin, Germany, 2003.
- [10] S.-T. Liu and C. Kao, "Fuzzy measures for correlation coefficient of fuzzy numbers," *Fuzzy Sets and Systems*, vol. 128, no. 2, pp. 267–275, 2002.
- [11] M. C. Xie and B. Wu, "The relationship between high schools students time management and academic performance: an application of fuzzy correlation," *Educational Policy Forum*, vol. 15, no. 1, pp. 157–176, 2012.
- [12] C. C. Yang, "Correlation coefficient evaluation for the fuzzy interval data," *Journal of Business Research*, vol. 69, no. 6, pp. 2138–2144, 2016.
- [13] R. Saneifard and R. Saneifard, "Correlation coefficient between fuzzy numbers based on central interval," *Journal of Fuzzy Set Valued Analysis*, vol. 2012, Article ID jfsva-00108, 9 pages, 2012.
- [14] Y. T. Cheng and C. C. Yang, "The application of fuzzy correlation coefficient with fuzzy interval data," *International Journal of Innovative Management, Information & Production*, vol. 4, no. 1, pp. 65–71, 2013.
- [15] I. M. Hanafy, A. A. Salama, and K. M. Mahfouz, "Correlation coefficients of neutrosophic sets by centroid method," *International Journal of Probability and Statistics*, vol. 2, no. 1, pp. 9–12, 2013.
- [16] B. Wu, W. Lai, C. L. Wu, and T. K. Tienliu, "Correlation with fuzzy data and its applications in the 12-year compulsory education in Taiwan," *Communications in Statistics—Simulation and Computation*, vol. 45, no. 4, pp. 1337–1354, 2016.
- [17] H. Lin, C. Wang, J. C. Chen, and B. Wu, "New statistical analysis in marketing research with fuzzy data," *Journal of Business Research*, vol. 69, no. 6, pp. 2176–2181, 2016.
- [18] R. A. Hunt, *Calculus with Analytic Geometry*, Harper & Row, New York, NY, USA, 1998.