Correlation with Fuzzy Data and Its Applications in the 12-Year Compulsory Education in Taiwan

Berlin Wu
Dept. of Mathematics Sciences, National ChengChi University, Taiwan
berlin@nccu.edu.tw

Wentsung Lai
Dept. of Ed. Policy and Ad., National ChiNan University, Taiwan
s100407904@mail1.ncnu.edu.tw

Ching-Ling Wu
Dept. of Ed. Policy and Ad., National ChiNan University, Taiwan
clwu@ncnu.edu.tw

Tsung-kuo Tienliu
Dept. of Ed. Policy and Ad.& Center for General Ed., National ChiNan University, Taiwan
tkliu@ncnu.edu.tw

Abstract

In the study, we developed a new model of fuzzy correlation and empirically examined the fuzzy correlation with student test scores in subjects of Chinese and Mathematics. Our sample comprised 419 Taiwanese students in the 12-year compulsory education. We applied fuzzy theory and conducted simulations with the data from Normal, Uniform, and Cauchy distributions to illustrate the efficiency of our proposed methods.

Keywords: *fuzzy correlation,* Chinese*,* Mathematics, *12-year compulsory education policy*

1. Introduction

Taiwan has officially implemented 12-year compulsory education. Starting from year 2011, 7^{th} grade students in junior high school are required to take a national examination in their 9^{th} grade. In the examination, the Chinese subject will include composition writing, the Mathematics subject will include non-multiple-choice questions, and the English subject will include listening ability. It is understandable to add English listening since listening is basic for language learning. However, why are Mathematics and Chinese subjects specifically emphasized? Do the two famous proverbs "Chinese is the predecessor of learning" and "Mathematics is the predecessor of science" serve as adequate reasons for adding Mathematics and Chinese subjects? This study analyzes correlations among examination subjects using student interval fuzzy scores and contributes to the policy of 12-year compulsory education by providing empirical research results of student performance correlations among Chinese, Mathematics, and English subjects.

Generally, Pearson product-moment correlation is employed to estimate correlation coefficients. Because the new definition of fuzzy correlation intervals developed in Xie and Wu (2012) provided a more authentic and relevant fuzzy correlation interval, it enhanced the authenticity level of correlation estimation. This study employs this interval definition to calculate fuzzy correlation intervals among subjects. Regarding the new definition provided by this study, because, according to the definition of fuzzy correlation presented in Xie and Wu, fuzzy correlation interval type is an interval value (e.g., $(r, r+\delta)$); further application requires conversion into defuzzified values for use. For example, when comparing the strength or intensity of fuzzy correlation coefficients, a single value must be employed. To continue research on the fuzzy correlation defined by Xie and Wu (2012), this study further employs the defuzzified value of interval fuzzy numbers proposed by Wu (2005) to define the correlation coefficient r_f of the defuzzified value for fuzzy correlation intervals. This represents another important motivation and contribution of this study.

In the development of the social sciences, fuzzy correlation coefficients have gradually received increasing emphasis. Previous studies have investigated methods of calculating correlation coefficients. For example, Chiang and Lin (1999) discussed the membership functions of fuzzy sets and established the membership degree as a concrete observation value to define fuzzy correlation coefficients. Chaudhuri and Bhattacharya (2001) investigated fuzzy grade correlation coefficients to calculate the correlations of two fuzzy sets. Lin (2004) proposed that the fuzzy correlation coefficients measure the similarity and correlation of fuzzy data. Chen (2005) employed random verification or examination methods to examine hypotheses of whether correlation coefficients equal zero. He used the extension principle proposed by Zadeh (1978) and adopted the recommendation of Liu and Kao (2002) regarding the α-cut of fuzzy sets to obtain the α-cut of fuzzy correlation coefficients. Based on these cuts, the membership function of fuzzy test statistics can be obtained. Hong (2006) also proposed a calculation method for fuzzy correlation. Hung and Wu (2007) argued that multiple correlations and three or more partial correlations of fuzzy sets are critically important; thus, they proposed a calculation method using the concept of the multiple correlation model and the multiple and partial correlation of fuzzy data.

Xie and Wu (2012) developed fuzzy correlation coefficients based on correlation coefficients proposed by Liu and Kao (2002) and obtained fuzzy correlation intervals based on interval fuzzy sample data. Educational policies and administrative fields often employ traditional binary logic to process quantified data. Statistics generally employs Pearson correlation coefficients to express the intensity of the linear correlation between two variables and the direction of the correlation. The data that Pearson correlation coefficients process is concrete real values. However, when data represent fuzzy numbers, how to calculate general fuzzy correlation coefficients becomes an important problem. This study investigates interval fuzzy sample data values to obtain fuzzy correlation coefficients and proposes general fuzzy correlation coefficients. The proposed definition of fuzzy correlation interval can be applied to situations in which two sets of data values are real numbers or one of the sets is real numbers. This can be used to explain more correlation phenomena that occur in practical situations. However, utility problems remain during application, such as the range of correlation coefficients, calculating convenience, and problems when comparing intensities or strengths of fuzzy correlation intervals. This study intends to continue relevant research to develop a new model and increase the utility of the fuzzy correlation interval.

2. Research Methods

The research methodology for this study is the application of fuzzy theory. The first research step is to newly define student interval fuzzy scores (Definition 2.1). The second is to calculate the fuzzy correlation interval (Definition 2.2 and Definition 2.3). The third step is to defuzzify the fuzzy correlation interval (Definition 2.4), and the final step is to perform trial calculations and analysis using actual examples; the research samples are general tests for 419 students who are in their third year in junior high school. For statistical methods, we use statistical software, MINITAB16.0, for the quantified statistical analysis. Because the samples from the research are not necessarily distributed normally, using nonparametric analysis is more appropriate to actual situations (Wu, 2005; Wu and Xie, 2010; Hung, Vladik, Wu, and Gang, 2011; Hung and Wu, 2006). Figure 2.2 shows the research process framework.

2.1 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 is 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 $\sigma_X{}^2$, $\sigma_Y{}^2$ 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:

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

where $(x_{i,}y_i)$ is the ith pair observation value, $i = 1,2,3........,n$. $\bar{x}, \bar{y}$ are sample mean for x and y respectively.

Yang, Wu and Sriboonchitta (2012) defined fuzzy correlation .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.Lai is a new graduate from college; his expected annual income is 48,000 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.Lai's acceptable salary range is from 46,000 to 51,000. We can express his annual salary as an interval [46000, 51000]. In addition, when Mr.Lai has a job interview, the manager may ask how many hours he can work per day. In this case, Mr.Lai may not be able to provide a definite number since his everyday schedule is different. However, Mr.Lai may tell the manger that his expected working hours per day is an interval [7, 9].

We know Mr. Lai's expected salary ranges from [46000, 51000] and his expected working hours are [7, 9]. 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 C_x is the expected salary for each new graduate, C_y is the working hours they desired, then the scatter plot for these two sets of fuzzy interval numbers would approximate that shown in Figure 2.1.

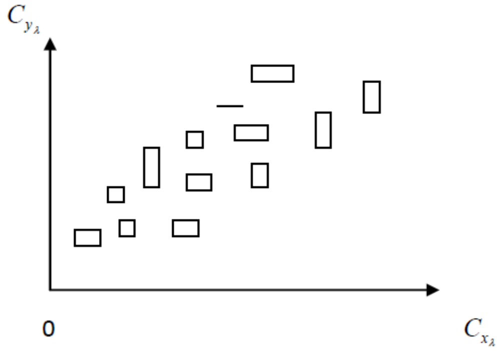

Figure 2.1 Fuzzy correlation with interval data

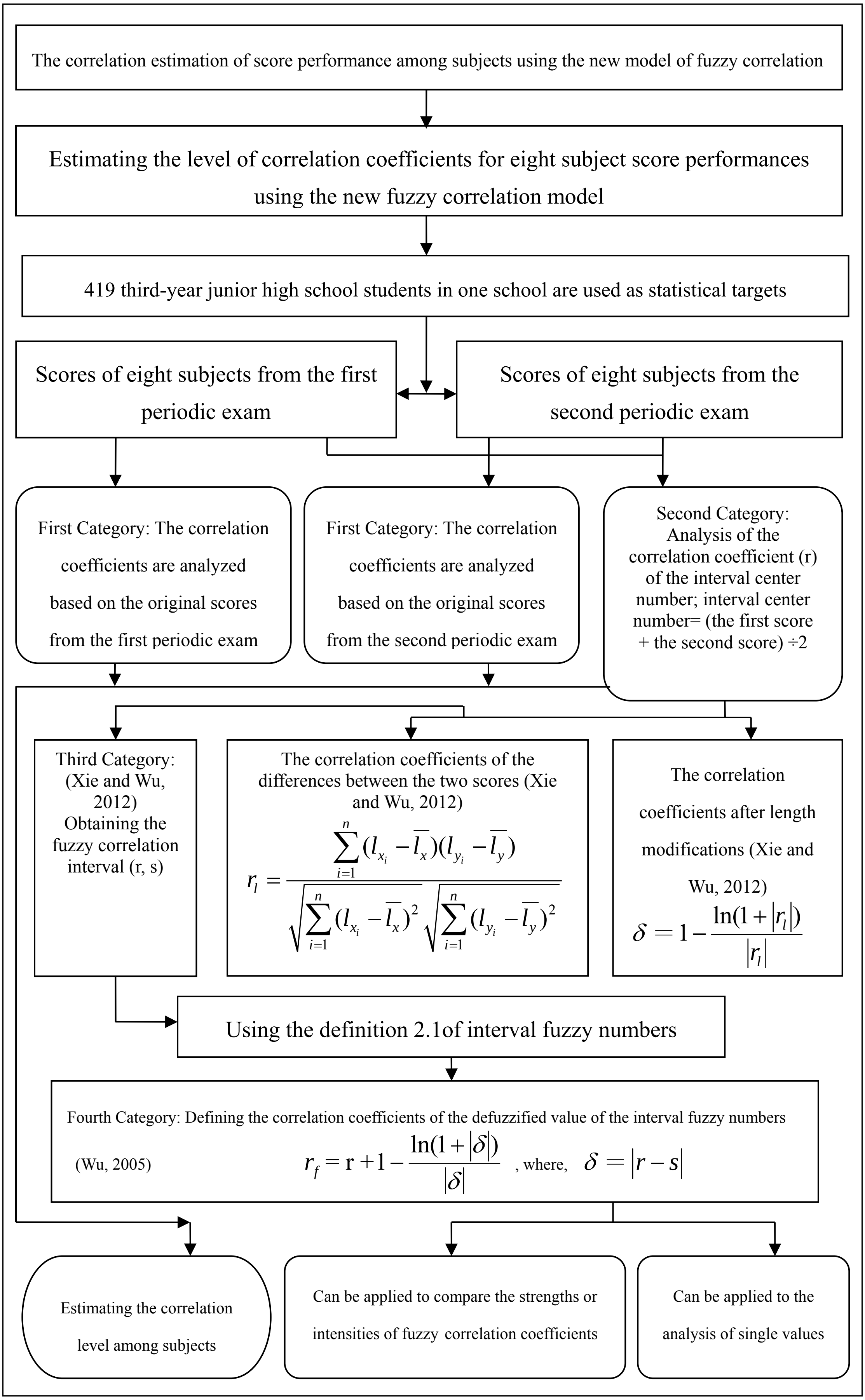

Figure 2.2 Research process frameworks

Definition 2.1: Student performance interval fuzzy scores

Definition 2.1 is student performance interval fuzzy scores (a, b). For this, a and b represent two different test scores for the same student, and $a \le b$. For example, considering a student's two recent Mathematics test scores, these scores are calculated using a one-hundred point system; the scores are 88 and 78. The student's Mathematics performance is represented using the interval fuzzy score (78, 88), which indicates that the student's Mathematics ability performance is between 78 and 88 points. Using the fuzzy score (78,88) to express this performance is more appropriate to the student's actual performance ability than employing the traditional method of calculating the average of the two numbers (78+88)/2= 83. This is because human ability is a fuzzy description. Using a single value for description is easier but not authentic or real.

Definition 2.2: Fuzzy correlation interval (adopting the interval center point and length methods; Xie and Wu, 2012)

When calculating the fuzzy correlation interval as defined by Xie and Wu (2012), suppose the length of the continuous interval sample x_i is l_{x_i}, and the length of the continuous interval sample y_i is l_{y_i}. Consequently, the correlation coefficient for length modification is as follows:

$$\delta = 1 - \frac{\ln(1+|r_l|)}{|r_l|}\text{; Where, } r_l = \frac{\sum_{i=1}^{n}(l_{x_i}-\overline{l}_x)(l_{y_i}-\overline{l}_y)}{\sqrt{\sum_{i=1}^{n}(l_{x_i}-\overline{l}_x)^2}\sqrt{\sum_{i=1}^{n}(l_{y_i}-\overline{l}_y)^2}}$$

Because $0<|r_l|<1$, and the range of δ is $0<\delta<0.3069$, set c_{x_i} and c_{y_i} as the fuzzy sample interval center point derived from the population *X*、*Y*, l_{x_i} and l_{y_i} as the interval length, r as the correlation coefficient of the center point, and δ as the correlation coefficient of length modification. Consequently, the correlation interval is defined as follows:

(1) If $r \geqq 0$, $r_l \geqq 0$,(r, min(1, $r+\delta$)) (2) If $r \geqq 0$, $r_l < 0$,($r-\delta$, r)

(3) If $r < 0$, $r_l \geqq 0$,(r, $r+\delta$) (4) If $r < 0$, $r_l < 0$,(max(-1, $r-\delta$), r)

Definition 2.3: Defuzzifying interval fuzzy numbers (Wu, 2005)

Make *X*= (a, b) the interval fuzzy number, $c = \frac{a+b}{2}$ its interval center, and *l*=|*b*-*a*| its interval range. Consequently, the defuzzified value of interval fuzzy numbers is

$x_f = c + 1 - \frac{\ln(1+|X|)}{|X|}$ (Formula 1). Where, $1 - \frac{\ln(1+|X|)}{|X|}$ (Formula 2).is the interval length defuzzified function.

If $a \to b$, then x_f approaches the range center value $\frac{a+b}{2}$.

Definition 2.4：The new model proposed by this study: The correlation coefficient r_f of the defuzzified value of the fuzzy correlation interval

Make *X*= (r, s) the definition of the fuzzy correlation interval calculated and defined by Xie and Wu (2012), and r the correlation coefficient of its interval center value. The interval value is c=（$r+s$）÷2, and $\delta = |r - s|$ is the interval range.

Consequently, the defuzzified value of the interval fuzzy number is r_f.

$$r_f = r + 1 - \frac{\ln(1+|\delta|)}{|\delta|} \quad \text{(Formula 3)}, \quad \text{where,} \quad 1 - \frac{\ln(1+|\delta|)}{|\delta|} \quad \text{(Formula 4)}$$

In addition, $r \leqq r_f \leqq \min(1, r+0.128)$. Because $-1 \leqq r \leqq 1$, the range of r_f is located between $-1 \leqq r_f \leqq 1$.

Definition 2.4.1：Calculating the range of the correlation coefficient r_f (Formula 3) of the defuzzified value of the interval fuzzy numbers

The range of r_f is calculated below.

Because $0 < \delta < 0.3069$, based on $r_f = r + 1 - \frac{\ln(1+|\delta|)}{|\delta|}$, the range of r_f is calculated below. The maximum value of $\frac{\ln(1+|\delta|)}{|\delta|}$ is 0.872; the minimum value is 0.

The equation is as follows:

$$\lim_{\delta \to 0} \frac{\ln(1+|\delta|)}{|\delta|} = \lim_{\delta \to 0} \frac{D_\delta \ln(1+|\delta|)}{D_\delta |\delta|} = \lim_{\delta \to 0} \frac{1}{1+|\delta|} = 1$$

$$\lim_{\delta \to 0.3069} \frac{\ln(1+|\delta|)}{|\delta|} = \frac{\lim_{\delta \to 0.3069} \ln(1+|\delta|)}{\lim_{\delta \to 0.3069} |\delta|} = \frac{\ln(1.3069)}{0.3069} = \frac{0.267657}{0.3069} = 0.872$$

Therefore, the minimum value of $1 - \frac{\ln(1+|\delta|)}{|\delta|}$ is 1-1=0; the maximum value is 1-0872 = 0.128.Consequently, the minimum value of $r_f = r + 1 - \frac{\ln(1+|\delta|)}{|\delta|}$ is r, and the maximum value is min (1, r +0.128).

Definition 2.4.2：Determining the strength intervals of the correlation coefficients for the new model

Winpot software 3D graphics are used to estimate the location value of r_f at 20%, 40%, 60%, and 80%, as shown in Fig. 2.3 to the right. This shows the function diagram of Formula 3 as a virtually vertically inclined plane. Therefore, that this study performed five equal portion estimations for the Z-axis coordinate value (i.e., r_f) to transform it into the determination for correlation coefficient strength intervals is appropriate and acceptable. That is, every 20% section is approximately 1.128÷5 = 0.2256. Relevant details are listed in Table 2.1. $r_f = r + 1 - \frac{\ln(1+|\delta|)}{|\delta|}$ (Formula 3). Where, $-1 \leqq r \leqq 1$; $0 < \delta < 0.3069$.

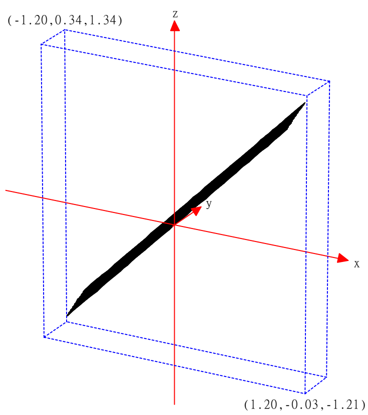

Figure 2.3

Definition 2.4.3：The intensity range of the correlation coefficients for the new model

Because, for various groups of boundary points, the explanation of strength and weakness for correlation coefficients is specious and difficult, the intensity of correlation coefficients applied in this study is transformed into triangular fuzzy numbers for consideration. Furthermore, defuzzification is conducted using the center of gravity method and simple sequencing is performed. Regarding the definition of triangular fuzzy numbers, as long as the left end point (a_1), the center point(a_2), and the right end point(a_3) of the triangular fuzzy numbers are decided, the triangular fuzzy number can be determined and represented as $\widetilde{A}=(a_1, a_2, a_3)$. The definition of the membership function is $\mu_{\widetilde{A}}(x)$, and depictions of the membership function are shown in Figs. 2.4 and 2.5:

$$\mu_{\widetilde{A}}(x)=\begin{cases} 0 & , \quad x<a_1 \\ (x-a_1)/(a_2-a_1) & , \quad a_1 \le x < a_2 \\ (a_3-x)/(a_3-a_2) & , \quad a_2 \le x < a_3 \\ 0 & , \quad x \ge a_3 \end{cases}$$

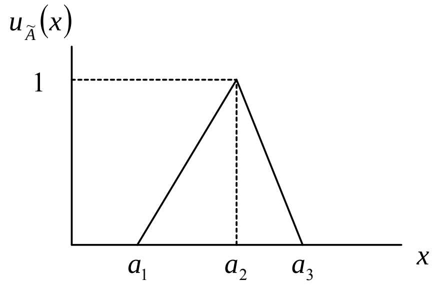

Figure 2.4

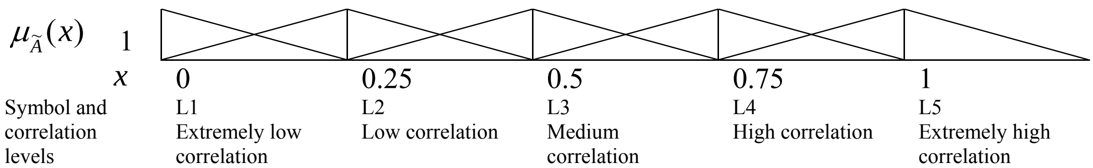

Figure 2.5

Table 2.1 the intensity range and meaning of the correlation coefficients for the new model

The intensity range of the correlation coefficients for the new model (triangular fuzzy number)	New model (defuzzification using the center of gravity method)	Correlation level of variables	Comparison with Pearson correlation coefficient range (absolute value)
(0.75, 1, 1)	(0.90 and above)	Extremely high correlation (L5)	(0.80 and above)
(0.5, 0.75, 1)	(0.68-0.90)	High correlation (L4)	(0.60-0.80)
(0.25,0.5, 0.75)	(0.45-0.68)	Medium correlation(L3)	(0.40-0.60)
(0, 0.25,0.5)	(0.23-0.45)	Low correlation (L2)	(0.20-0.40)
(0,0,0.25)	(0-0.23)	Extremely low correlation (L1)	(0-0.20)

3. Simulation Studies

In this section, we will use the MINITAB simulation to generate several sequence of fuzzy data set and then compare their correlations with different definition as proposed at the section 2. The distribution for the centric and area are generated by the normal, uniform, and Cauchy distribution respectively. The procedure to calculate correlation coefficient is described below: Table 3.1 illustrates the result

Step 1. Generate fuzzy set of sequence X from the normal, uniform, and Cauchy distribution respectively.

Step 2. Let Y=aX+e, calculate the fuzzy data set Y by the fuzzy data set X.

Step 3.Let interval score be (x , y), where x, y chose from the normal, uniform, and Cauchy distribution respectively.

Step 4. Find the correlation coefficient from the fuzzy data set by above definitions.

Table 3.1 provides the following results: (1) When interval fuzzy score (x, y) been combined by (Normal, Uniform) & (Cauchy, Normal), the correlation coefficient is very close. (2) When interval fuzzy score (x, y) been combined by (Uniform, Normal) & (Cauchy, Uniform), the correlation coefficient is very close. (3) When interval fuzzy score (x, y) been combined by (Cauchy, Uniform) & (Normal, Cauchy), the correlation coefficient is very close. (4) There are different advantages to test the correlation by definition 2.2 or definition 2.4. The interval correlation of Definition 2.2 is better to represent the interval property of statistics and more robustic. But the correlation coefficient of Definition 2.4 is a real value that is more robustic and useful to test the advanced statistics.

Table 3.1 the fuzzy correlation coefficient for various center and area model with definition 2.2, definition 2.4.

Simulation	Normal(0,1)	Uniform(0,1)	Cauchy(0,1)	Normal(0,1)
Interval (x, y)	(Normal, Uniform)		(Cauchy, Normal)	
Center value	$(x+y)/2$		$(x+y)/2$	
Pearson cor. coe.	-0.07			
Definition 2.2	(-0.13, -0.07)			
Definition 2.4	-0.04			
Simulation	Uniform(0,1)	Normal(0,1)	Cauchy(0,1)	Uniform(0,1)
Interval (x, y)	(Uniform, Normal)		(Cauchy, Uniform)	
Center value	$(x+y)/2$		$(x+y)/2$	
Pearson cor. coe.	-0.02			
Definition 2.2	(-0.02, 0.10)			
Definition 2.4	0.01			
Simulation	Cauchy(0,1)	Uniform(0,1)	Normal(0,1)	Cauchy(0,1)
Interval (x, y)	(Cauchy, Uniform)		(Normal, Cauchy)	
Center value	$(x+y)/2$		$(x+y)/2$	
Pearson cor. coe.	-0.13			
Definition 2.2	(-0.36, -0.13)			
Definition 2.4	-0.08			

4. Empirical Studies

The case study for this research employs 419 third-year junior high school students in one school as the subjects for statistical analysis. Correlation analysis is conducted using scores for eight subjects from the first and second periodic exams of the first semester in 2011. Because this study attempted to estimate the correlation coefficients for ideas regarding "Chinese is the predecessor of learning" and "Mathematics is the predecessor of science" without conducting cause-and-effect inferences, the correlation matrix of all eight subjects (Chinese (C), Mathematics (M), English (E), Natural Science (N), Biology (B), Geography (G), History (H), and Civic Studies (Z)) was analyzed. This study adopted the analysis of four data types for correlation analysis and divided the results into four types of correlation coefficients. The first type applies original data to directly conduct Pearson product-moment correlation and Spearman correlation analysis. It includes the data for the first periodic test and the second periodic test as shown in Tables 4.1 and 4.2. The second type employs the interval center value for the original scores of the two periodic tests to conduct Pearson product-moment correlation and Spearman correlation analysis. Using Mathematics as an example, the method for calculating the interval center value is the arithmetic mean of the subject score from the first and second periodic exam, that is, the interval center value c of Mathematics, as shown in Table 4.3. The third type calculates the fuzzy correlation intervals using the fuzzy correlation defined by Xie and Wu (2012), as shown in Tables 4.4 and 4.5. The fourth type is the new definition (Definition 2.4) proposed by this study. Because the third type is an interval value as determined by the definition of fuzzy correlation proposed by Xie and Wu (2012) (e.g., (r, r+δ)), for further application, conversion into a defuzzified value must be conducted. Therefore, this study continues research regarding the fuzzy correlation defined by Xie and Wu (2012) and employs the defuzzified value of interval fuzzy numbers proposed by Wu (2005) to define the defuzzified correlation coefficient r_f for fuzzy correlation intervals, as shown in Table 4.6.

4.1 Matrix of the correlation coefficients among subjects

(1) The correlation coefficients for the first type: Directly applying Pearson product-moment correlation and Spearman correlation analysis

Table 4.1 The first periodic exam (The lower triangle is the Pearson correlation coefficients; the upper triangle is Spearman correlation coefficients)

	C	E	N	M	B	G	H	Z
C	1	0.700	0.729	0.709	0.764	0.786	0.724	0.724
E	0.698	1	0.751	0.768	0.778	0.769	0.711	0.731
N	0.731	0.751	1	0.800	0.838	0.828	0.751	0.737
M	0.686	0.747	0.772	1	0.823	0.804	0.748	0.713
B	0.769	0.784	0.847	0.790	1	0.857	0.804	0.781
G	0.757	0.736	0.798	0.788	0.820	1	0.852	0.827
H	0.644	0.635	0.668	0.704	0.708	0.834	1	0.805
Z	0.621	0.618	0.641	0.690	0.655	0.800	0.774	1

Table 4.2 the second periodic exam (The lower triangle is Pearson correlation coefficients; the upper triangle is Spearman correlation coefficients)

	C	E	N	M	B	G	H	Z
C	1	0.759	0.747	0.769	0.808	0.822	0.750	0.795
E	0.753	1	0.730	0.758	0.734	0.736	0.677	0.751
N	0.737	0.739	1	0.780	0.802	0.790	0.734	0.761
M	0.769	0.743	0.753	1	0.800	0.794	0.719	0.776
B	0.802	0.717	0.768	0.790	1	0.884	0.794	0.838
G	0.817	0.725	0.772	0.782	0.865	1	0.816	0.860
H	0.705	0.635	0.659	0.694	0.761	0.777	1	0.755
Z	0.751	0.686	0.679	0.737	0.787	0.827	0.721	1

(2) The correlation coefficients for the second type: Using the interval center value for the original scores of the two exams to conduct Pearson product-moment correlation and Spearman correlation analysis

Table 4.3 Interval center values (The lower triangle is Pearson correlation coefficients; the upper triangle is Spearman correlation coefficients)

	C	E	N	M	B	G	H	Z
C	1	0.766	0.789	0.793	0.840	0.858	0.782	0.825
E	0.766	1	0.776	0.792	0.788	0.778	0.723	0.780
N	0.792	0.779	1	0.836	0.872	0.857	0.781	0.805
M	0.779	0.774	0.804	1	0.855	0.836	0.769	0.801
B	0.838	0.788	0.869	0.839	1	0.916	0.836	0.866
G	0.844	0.762	0.832	0.823	0.902	1	0.882	0.898
H	0.715	0.660	0.696	0.731	0.774	0.851	1	0.824
Z	0.743	0.686	0.709	0.756	0.783	0.864	0.791	1

(3) The correlation coefficients for the third type: Calculating fuzzy correlation intervals using the definition proposed by Xie and Wu (2012)

Table 4.4 Difference value (The lower triangle is Pearson correlation coefficients; the upper triangle is the δ value.)

	C	E	N	M	B	G	H	Z
C	1	0.033	0.019	0.052	0.006	0.034	0.009	0.027
E	0.069	1	0.037	0.068	0.096	0.055	0.026	0.098
N	0.040	0.078	1	0.102	0.016	0.104	0.089	0.053
M	0.112	0.149	0.235	1	0.069	0.076	0.102	0.103
B	0.012	0.219	0.032	0.151	1	0.073	0.009	0.089
G	0.071	0.119	0.242	0.168	0.161	1	0.100	0.127
H	0.019	0.053	0.201	0.236	0.018	0.230	1	0.102
Z	0.056	0.225	0.114	0.239	0.201	0.305	0.235	1

Table 4.5 Fuzzy correlation intervals

	C	E	N	M	B	G	H
E	(0.77,0.80)						
N	(0.80,0.82)	(0.78,0.82)					
M	(0.78,0.83)	(0.77,0.84)	(0.80,0.91)				
B	(0.83,0.84)	(0.79,0.88)	(0.87,0.89)	(0.84,0.90)			
G	(0.84,0.88)	(0.76,0.82)	(0.83,0.94)	(0.82,0.90)	(0.90,0.98)		
H	(0.72,0.73)	(0.66,0.69)	(0.70,0.79)	(0.73,0.83)	(0.77,0.79)	(0.85,0.95)	
Z	(0.74,0.77)	(0.69,0.78)	(0.71,0.76)	(0.76,0.86)	(0.78,0.87)	(0.86,0.99)	(0.79,0.89)

(4) The correlation coefficients for the fourth type: The new correlation coefficient model defined by this study

The fourth type of correlation coefficient in this study continues research regarding the definition of fuzzy correlation proposed by Xie and Wu (2012), and further employs the defuzzified value of interval fuzzy numbers proposed by Wu (2005) to define the defuzzified correlation coefficient r_f of the fuzzy correlation intervals. The critical calculation principles and processes are based on the following four steps. The first step is based on Definition 2.1: student performance interval fuzzy scores. The second step is based on Definition 2.2: fuzzy correlation interval (adopting the interval center points and length methods; Xie and Wu, 2012). The third step is based on Definition 2.3: defuzzifying interval fuzzy numbers (Wu, 2005). The fourth step is based on Definition 2.4 (the new model proposed by this research): the correlation coefficient r_f of the defuzzified value of the fuzzy correlation interval. Calculation values are shown in Table 4.6.

Table 4.6 the fuzzy correlation for the new model proposed by this study

	C	E	N	M	B	G	H	Z
C	1							
E	0.782	1						
N	0.802	0.797	1					
M	0.804	0.806	0.852	1				
B	0.841	0.833	0.877	0.872	1			
G	0.861	0.789	0.881	0.859	0.937	1		
H	0.720	0.673	0.738	0.779	0.778	0.898	1	
Z	0.756	0.732	0.735	0.804	0.825	0.923	0.839	1

(5) Overall integration of the correlation coefficients: The conclusion "Chinese is the predecessor of learning" shows high correlation; cause-and-effect inferences are not discussed. Shown in Table 4.7.

Table 4.7 Overall integration of the correlation coefficients for Chinese

C	E	N	M	B	G	H	Z
Type 1-1	0.698	0.731	0.686	0.769	0.757	0.644	0.621
Pearson	(L4)	(L4)	(L4)	(L4)	(L4)	(L4)	(L4)
Spearman	0.700	0.729	0.709	0.764	0.786	0.724	0.724
	(L4)	(L4)	(L4)	(L4)	(L4)	(L4)	(L4)
Type 1-2	0.753	0.737	0.769	0.802	0.817	0.705	0.751
Pearson	(L4)	(L4)	(L4)	(L5)	(L5)	(L4)	(L4)
Spearman	0.759	0.747	0.769	0.808	0.822	0.750	0.795
	(L4)	(L4)	(L4)	(L5)	(L5)	(L4)	(L4)
Type 2	0.766	0.792	0.779	0.838	0.844	0.715	0.743
Pearson	(L4)	(L4)	(L4)	(L5)	(L5)	(L4)	(L4)
Spearman	0.766	0.789	0.793	0.840	0.858	0.782	0.825
	(L4)	(L4)	(L4)	(L5)	(L5)	(L4)	(L4)
Type 3	(0.77,0.80)	(0.80,0.82)	(0.78,0.83)	(0.83,0.84)	(0.84,0.88)	(0.72,0.73)	(0.74,0.77)
Type 4	0.782	0.802	0.804	0.841	0.861	0.720	0.756
	(L4)	(L4)	(L4)	(L4)	(L4)	(L4)	(L4)

(6) Overall integration of the correlation coefficients: The conclusion "Mathematics is the predecessor of science" shows high correlation; cause-and-effect inferences are not discussed. Shown in Table 4.8.

Table 4.8 Overall integration of the correlation coefficients for Mathematics

M	C	E	N	B	G	H	Z
Type 1-1	0.686	0.747	0.772	0.790	0.788	0.704	0.690
Pearson	(L4)	(L4)	(L4)	(L4)	(L4)	(L4)	(L4)
Spearman	0.709	0.768	0.800	0.823	0.804	0.748	0.713
	(L4)	(L4)	(L5)	(L5)	(L5)	(L4)	(L4)
Type 1-2	0.769	0.743	0.753	0.790	0.782	0.694	0.737
Pearson	(L4)	(L4)	(L4)	(L4)	(L4)	(L4)	(L4)
Spearman	0.769	0.758	0.780	0.800	0.794	0.719	0.776
	(L4)	(L4)	(L4)	(L5)	(L4)	(L4)	(L4)
Type 2	0.779	0.774	0.804	0.839	0.823	0.731	0.756
Pearson	(L4)	(L4)	(L5)	(L5)	(L5)	(L4)	(L4)
Spearman	0.793	0.792	0.836	0.855	0.836	0.769	0.801
	(L4)	(L4)	(L5)	(L5)	(L5)	(L4)	(L5)
Type 3	(0.78,0.83)	(0.77,0.84)	(0.80,0.91)	(0.84,0.90)	(0.82,0.90)	(0.73,0.83)	(0.76,0.86)
Type 4	0.804	0.806	0.852	0.872	0.859	0.779	0.804
	(L4)	(L4)	(L4)	(L4)	(L4)	(L4)	(L4)

5. Discussion

This study adopts four types of data to conduct correlation analysis. The first type employs original data to directly conduct Pearson product-moment correlation and Spearman correlation analysis. The second type employs the interval center value for the original scores of the two periodic tests to conduct Pearson product-moment correlation and Spearman correlation analysis. The third type calculates the correlation intervals using Definition 2.3 proposed by Xie and Wu (2012) regarding fuzzy correlation intervals. The fourth type is the new definition (Definition 2.4) proposed by this study, which continues research regarding the fuzzy correlation interval defined by Xie and Wu (2012) and further uses the defuzzified value of interval fuzzy numbers proposed by Wu (2005) to define the defuzzified correlation coefficient r_f of the fuzzy correlation intervals. The fourth type of correlation coefficient is much closer to authentic correlations for human beings based on its definition method. The analysis results of Table 4.7 demonstrate that the correlation coefficient interval defuzzified values for Chinese in relation to English, Natural Science, Mathematics, Biology, Geography, History, and Civic Studies are 0.782, 0.82, 0.804, 0.841, 0.861, 0.720, and 0.756, respectively. According to Table 2.1, these values all represent high correlation. Consequently, based on estimation analysis, the conclusion that Chinese is the predecessor of learning demonstrates high correlation. However, high correlation does not represent that cause-and-effect relationships exist between two variables. The analysis results of Table 4.8 show that the correlation coefficient interval defuzzified values for Mathematics in relation to Chinese, English, Natural Science, Biology, Geography, History, and Civic Studies are 0.804, 0.806, 0.852, 0.872, 0.859, 0.779, and 0.804, respectively. Among these results, the correlation coefficient of Mathematics and Natural Science is 0.852 and the correlation coefficient of Mathematics and Biology is 0.872; both values evince high levels of correlation according to Table 2.1. Consequently, analyzing the conclusion that Mathematics is the predecessor of science demonstrates that a high correlation exists for this conclusion. However, high correlation does not represent that cause-and-effect relationships exist between two variables.

This study organized the strength and weakness changes for the four types of correlation coefficients, and determined, based on Table 4.7, that the category changes of the correlation coefficients for Chinese and Biology according to type are L4 (high correlation), L4, L5 (extremely high correlation), L5, L5, L5, and L4. The category changes of the correlation coefficients for Chinese and Geography according to type are L4, L4, L5, L5, L5, L5, and L4. Table 4.8 shows that the category changes of the correlation coefficients for Mathematics and Natural Science according to type are L4, L5, L4, L4, L5, L5, and L4. The category changes of the correlation coefficients for Mathematics and Biology according to type are L4, L5, L4, L5, L5, L5, and L4. The category changes of the correlation coefficients for Mathematics and Geography according to type are L4, L5, L4, L4, L5, L5, and L4. Finally, the category changes of the correlation coefficients for Mathematics and Civic Studies according to type are L4, L4, L4, L4, L5, and L4.

The study also indicates that almost every other subject also demonstrates high correlation or an even higher level of correlation. For example, Table 4.6 shows the fuzzy correlation matrix for the study's new model. The more unique results are as follows: the correlation coefficient of English and History (0.673) demonstrates a medium correlation, the correlation coefficient of Biology and Geography (0.937) demonstrates extremely high correlation, and the correlation coefficient of Geography and Civic Studies (0.923) demonstrates extremely high correlation. Therefore, subsequent cause-and-effect research warrants further attention.

6. Conclusion

The new definition (Definition 2.4) proposed by this research, that is, the defuzzified correlation coefficient r_f of the fuzzy correlation intervals, possesses developmental characteristics. Using four types of data to conduct correlation analysis and dividing the coefficients into four types of correlation coefficients provides for greater authenticity. The results of the study show that establishing the new fuzzy correlation model based on defuzzifying fuzzy correlation intervals can resolve the two purposes or objectives of this study. One innovative contribution of this study is innovation. Based on a literature review, the new model defined by this study is much more appropriate to the actual abilities of students and can improve the standard of correlation evaluations. The second contribution refers to applications. The correlation coefficients for this study are single values and can be used to directly compare intensity. The new model correlation coefficient Definition 2.4 is a new definition proposed by this study. Because according to the definition proposed by Xie and Wu (2012), the fuzzy correlation interval is an interval value (e.g., (r, r+δ)), for further application, it must be converted into a defuzzified value before it can be used to compare fuzzy correlation coefficients or for other purposes. Therefore, the new model of correlation coefficients proposed by this study possesses developmental characteristics.

6.1 The new model.

Definition 2.4: The new model of fuzzy correlation coefficients proposed by this study

Making $X=(r,s)$ is the fuzzy correlation interval calculated and defined by Xie and Wu (2012). r is the correlation coefficient of the interval center value, and the interval center value is c=($r + s$)÷2. δ=| r - s | is the interval range. Consequently, the defuzzified value of interval fuzzy numbers is r_f .

$$r_f = r + 1 - \frac{\ln(1+|\delta|)}{|\delta|} \quad \text{(Formula 3), where,} \quad 1 - \frac{\ln(1+|\delta|)}{|\delta|} \quad \text{(Formula 4)}$$

In addition, $r \leqq r_f \leqq \min(1, r + 0.128)$. Because $-1 \leqq r \leqq 1$, the range of r_f is located between $-1 \leqq r_f \leqq 1$. The intensity range and significance for the new model correlation coefficients is shown in Table 2.1.

6.2 The study adopts four types of correlation coefficient analysis, providing for greater authenticity

This study employs four types of data to analyze correlation coefficients. The first type employs original data to directly conduct Pearson product-moment correlation and Spearman correlation analysis. The second type uses the interval center value of the original scores from the two periodic exams to conduct Pearson product-moment correlation and Spearman correlation analysis. The third type calculates the fuzzy correlation intervals using Definition 2.2 proposed by Xie and Wu (2012). The fourth type is the new correlation coefficient model 2.4 proposed by this study. This model continues research regarding the fuzzy correlation interval defined by Xie and Wu (2012) and further uses the defuzzified value of interval fuzzy numbers proposed by Wu (2005) to define the defuzzified correlation coefficient r_f of the fuzzy correlation interval. Examining the fourth type or the new correlation coefficient model (definition 2.4), based on both definition method and empirical results, as shown in Tables 4.7 and 4.8, it can be seen that this model is more appropriate to authentic correlations for human beings.

6.3 The correlation coefficients among the eight subjects approached high correlation or above

Through the overall integration of the correlation coefficients as shown in Tables 4.7 and 4.8, correlation coefficients of the fourth level can be seen to be more appropriate to the authentic correlations of human beings based on the definition method. The analysis results show that the correlation coefficients of Chinese and English, Natural Science, Mathematics, Biology, Geography, History, and Civic Studies all demonstrate high correlation. The analysis results also show that the correlation coefficients for Mathematics and Chinese, English, Natural Science, Biology, Geography, History, and Civic Studies also demonstrate high correlation. The results further show that the correlation coefficients among the eight subjects approach high levels of correlation or higher. However, high correlation does not represent cause-and-effect relationships.

Empirical analysis shows that the subjects, in which 12-year compulsory education has added extra non-multiple-choice questions, Chinese, Mathematics, and English, possess high correlation with the score performance of other subjects. However, high correlation does not represent cause-and-effect relationships. The fact that, within the 12-year compulsory education examination policy, extra composition writing tests are added for Chinese, calculation and proof questions are added for Mathematics, and listening ability tests are added for English, seems to weaken the other five subjects and warrants questions regarding the theoretical basis for these choices. Although this study employed four types of correlation coefficient analysis, and the results show that Chinese, Mathematics, and English have high correlations with the score performance for other subjects, the correlation coefficients of the other five subjects also show high correlations. Thus, this study provides the following suggestions for subsequent research: 1) how to separate Chinese, Mathematics, English, and the other five subjects to perform discussion and policy analysis should be studied further; 2) high correlation does not represent cause-and-effect relationships. Further research should develop more fully regarding the cause-and-effect relationships between subjects.

Reference

Chaudhuri, B. B. and Bhattacharya, A. (2001). On correlation between two fuzzy sets. *Fuzzy Sets and Systems. 118*：447-456.

Chen, C. C. (2005). *The research and application of statistical testing model with fuzzy data* (Unpublished doctoral dissertation). National Cheng Kung University, Tainan, Taiwan.

Chiang, D. A., and Lin, N. P. (1999). Correlation of fuzzy sets. *Fuzzy Sets and Systems. 102*(2)：221-226.

Hong, D. (2006). Fuzzy measures for a correlation coefficient of fuzzy numbers under TW (the weakest t-norm)-based fuzzy arithmetic operations. *Information Sciences. 176*(2)：150-160.

Hung, N., Vladik, K., Wu, B.and Gang, X. (2011). *Computing Statistics under Interval and Fuzzy Uncertainty*. Studies in Computational Intelligence. Springer Verlag.

Hung, T. N. and Wu, B. (2006). *Fundamentals of Statistics with Fuzzy ata*.Springer Verlag.

Hung W.L. and Wu J.W. (2007). Multiple and partial correlation coefficients of fuzzy sets. *QUALITY & QUANTITY. 41*(3)：333-340. (Doi: 10.1007/s11135-006-9002-1)

Lin, Y. H. (2004). Psychological measurement and statistics: Fuzzy correlation. *Journal of Education Research. 122*：148-149.

Liu, S. and Kao, C. (2002). Fuzzy measures for correlation coefficient of fuzzy numbers. *Fuzzy Sets and Systems. 128*(2)：267-275.

Wu, B. (2005). *Introduction of fuzzy statistics*. Taipei, Taiwan: Wu-Naan Book.

Wu, B. and Xie, M. C. (2010). *Modern educational and psychological statistics*. New Taipei, Taiwan: Airiti Press Book.

Xie, M.C.and Wu, B. (2012).The Relationship between High Schools Students Time Management and Academic Performance: An Application of Fuzzy Correlation. *Educational Policy Forum.15*(1)：157-176. (Doi : 10.396615608298201202150 1007)

Yang, C. C., Wu, B. and Sriboonchitta, S. (2012). A new approach on correlation evaluation with fuzzy data and its applications in econometrics. *International Journal of Intelligent Technologies and Applied Statistics.* 5(2):1-12.

Zadeh, L. A. (1978). Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems. 1*(1)：3-28.