

hps-Pros 2019 Generating Fuzzy Interval Data and Its Application to Find The Relation Between Math

by Hari Susanto

Submission date: 07-Jul-2023 03:14PM (UTC+0700)

Submission ID: 2127629105

File name: I_Data_and_Its_Application_to_Find_The_Relation_Between_Math.pdf (665.47K)

Word count: 3400

Character count: 15846

Generating Fuzzy Interval Data and Its Application to Find The Relation Between Math Anxiety with Self Efficacy Using Correlations Analysis

H P Susanto¹, T Sutarti², and A S Hafidah³

College of Teacher Training and Education PGRI Pacitan, East Java, Indonesia
Cut Nya' Dien Road 4A Ploso Pacitan, East Java, Indonesia

haripurnomosusanto@gmail.com

Abstract. In this article, adding a dimension that has not yet been explored in published literature, we investigate the process of generating fuzzy interval data using a crisp data, applying the process to the students' anxiety and self-efficacy crisp data in learning mathematics, and determining the two variable correlations using Cheng and Yang fuzzy correlations. We take Fuzzy correlation developed by Cheng and Yang because of being easy to understand, and its development adopts the formula of Pearson correlation. The data used are fuzzy interval data and trapezoid fuzzy. In this case, the crisp data taken using traditional questionnaires cannot be employed. The findings of this research reveal that (1) the crisp data fuzzification process to be interval fuzzy provides a new alternative in obtaining fuzzy interval data; (2) fuzzy correlations using fuzzification results provide the conception that the anxiety has a negative correlation at a moderate effect level. The correlation coefficient shows the same conclusions with Pearson correlation and fuzzy correlation using original fuzzy data obtained from data collection.

1. Introduction

Fuzzy correlation is an expansion and development form of statistical correlation analysis methods to determine the relationship between two variables. The concept of fuzzy correlation was first developed [1] by determining the correlation between two fuzzy membership functions which meets several assumptions. We selected this dimension after exploring the previous researches over the past years. Chiang and Lin [2] employed the membership levels to obtain the correlation coefficients between fuzzy sets and finite sets. Hung and Wu [2] developed a fuzzy correlation with the expected-interval method. Hanafy, Salama, and Mahfouz [3] developed a fuzzy correlation on the neutrosophic sets with the Centroid method. Cheng and Yang [4] determined the fuzzy correlation coefficient by adopting statistical concepts using fuzzy interval data. The correlation coefficients produced in each of the fuzzy correlation concepts mentioned above meet the interval $[-1, 1]$.

Not all fuzzy correlation formulas use the same type of fuzzy data. Some use fuzzy interval data [2] [4][5]. Trapezoid fuzzy data, triangles fuzzy data [2] and membership values from the crisp data [2][1]. Fuzzy correlations with the interval data, trapezoid, and triangles fuzzy data determine the correlation of two variables based on the membership function which is formed from the data obtained from the data collection. This concept refers to [1]. Further, fuzzy correlation with the fuzzy in the form of membership value determines the fuzzy correlation coefficient by using fuzzy concepts [6].

Over past years, Cheng and Yang [4] extended the formula of Pearson correlation to fuzzy correlation with the aim that fuzzy correlations can be used to easify the data analysis, even for researchers or practitioners of the data analysis. They don't have to understand fuzzy in depth. The disadvantages of Cheng and Yang fuzzy correlations are that the data obtained from data collection cannot be used directly, because the data used are fuzzy interval data, trapezoid, or triangles. This weakness can only be overcome by data collection instrument modification so that the researcher can obtain the fuzzy data.



Content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Published under licence by IOP Publishing Ltd

Thus, the innovation in this paper is to generate fuzzy interval data using crisp data so that it can determine the two variables of fuzzy correlation using Cheng and Yang correlation methods. The Generating process is based on the concept δ -cut from a fuzzy set. Fuzzy interval data obtained from intervals δ -cut then reduced by the new fuzzy interval data with the center as the maximum or minimum limit of the interval data.

The fuzzy interval data obtained in this paper is then applied to determine the correlation between mathematical anxiety and self-efficacy by using Cheng and Yang fuzzy correlation formulas. The generated correlation coefficients will be compared to fuzzy correlation coefficients obtained with anxiety and fuzzy self-efficacy data which is achieved from data collection using fuzzy instruments.

2. Method

In order to model one-dimensional fuzzy data, the best up-to-date mathematical model is so-called fuzzy numbers [7]. Gil, Lubiano, saa & Sinova [8] has written the fuzzy number definition as follows:

Definition 1. A fuzzy number x^* is a fuzzy subset of the space of real numbers, that is, a mapping $\tilde{U}: R \in [0,1]$ which is normal, convex and has compact levels (i.e., the α -level sets given by

$$\tilde{U}_\delta(x) = \{x \in R : \tilde{U}(x) \geq \delta \} \text{ if } \alpha \in (0,1) \tag{1}$$

$$\tilde{U}_0(x) = \text{cl}\{x \in R : \tilde{U}(x) > 0\} \tag{2}$$

$\tilde{U}_\delta(x)$ is called the δ -Cut of fuzzy number x^* . Fuzzy number applications are written as finite numbers of δ -Cut and written as a real number interval. These intervals are after this referred to as fuzzy interval numbers.

Definition 2. A fuzzy number is called a fuzzy interval if all its δ -cuts are non-empty closed bounded intervals.

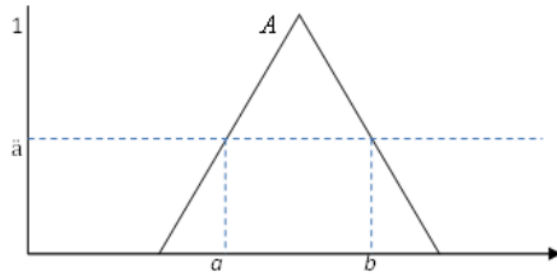


Figure 1. Fuzzy interval number of a^* and b^*

The above figure is given $a \in R$, in that case, the fuzzy numbers a^* has $C_\delta(a^*) = \{a \in R : A(a) \geq \delta \}$ if $\alpha \in (0,1)$ and written as $C_\delta(a^*) = [a, b]$. Otherwise, being given the fuzzy number $b \in R$, then, it will result in $C_\delta(a^*) = [a, b]$

2.1. The mathematical operation for fuzzy interval data

The addition of two fuzzy interval data more detail see [7][8]. Lets $x_1^* = [a_{11}, a_{21}]$ and $x_2^* = [a_{12}, a_{22}]$ be two fuzzy number. Then

$$x_1^* + x_2^* = [a_{11}, a_{21}] + [a_{12}, a_{22}] = [a_{11} + a_{12}, a_{21} + a_{22}] \tag{3}$$

The multiplication of two fuzzy interval data more detail see [10]. Lets $x_1^* = [a_{11}, a_{21}]$ and $x_2^* = [a_{12}, a_{22}]$ be two fuzzy number. Then

$$x_1^* \times x_2^* = [\min\{a_{11}a_{12}, a_{11}a_{22}, a_{21}a_{12}, a_{21}a_{22}\}, \max\{a_{11}a_{12}, a_{11}a_{22}, a_{21}a_{12}, a_{21}a_{22}\}] \tag{4}$$

2.2. Mean value of fuzzy interval data

Mean value of fuzzy interval data more detail see [7][8]. Lets n fuzzy $x_i^* = [a_{1i}, a_{2i}], i = 1, 2, \dots, n$ then the average of the data can be written as follows.

$$\bar{x}^* = \left[\frac{\sum_{i=1}^n a_{1i}}{n}, \frac{\sum_{i=1}^n a_{2i}}{n} \right] \quad (5)$$

2.3. The median value of fuzzy data interval

The addition of two fuzzy interval data more detail see [8]. Fuzzy $x_i^* = [a_{1i}, a_{2i}]$, $i = 1, 2, \dots, n$ thus, the average of the data can be written as follows.

$$Me^* = [Me\{a_{12}, a_{13}, \dots, a_{1n}\}, Me\{a_{21}, a_{22}, \dots, a_{2n}\}] \quad (6)$$

Fuzzy correlation developed in [4] uses the following definition

Let $x_{ji}^* = [a_{1j}, a_{2i}]$, $y_i^* = [b_{1j}, b_{2i}]$, be a sequence of paired fuzzy sample on populations Ω .

$$r_{jk} = \frac{\sum_{i=1}^n (a_{ji} - \bar{a}_j)(b_{ki} - \bar{b}_k)}{\sqrt{\sum_{i=1}^n (a_{ji} - \bar{a}_j)^2} \sqrt{\sum_{i=1}^n (b_{ki} - \bar{b}_k)^2}} \quad j = 1, 2 \text{ and } k = 1, 2 \quad (7)$$

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

$$\bar{r} = \frac{\sum_{j=1}^2 \sum_{k=1}^2 r_{jk}}{4} \text{ and } s_r = \frac{\sum_{j=1}^2 \sum_{k=1}^2 (r_{jk} - \bar{r})^2}{4} \quad (8)$$

The definition has a correlation coefficient value at the interval of [-1,1] and the correlation level of the inter-observed variables can be determined by the following definition.

- a. When $[r_{low}, r_{up}] \in [-0.10, 0.10]$, the fuzzy correlations are not significant
- b. When $[r_{low}, r_{up}] \in [-0.39, -0.11]$ or $[0.11, 0.39]$, the fuzzy correlations is low value.
- c. When $[r_{low}, r_{up}] \in [-0.69, -0.40]$ or $[0.40, 0.69]$, the fuzzy correlations is middle value.
- d. When $[r_{low}, r_{up}] \in [-0.99, -0.70]$ or $[0.70, 0.99]$, the fuzzy correlations is High value.

To apply the crisp data to determine fuzzy correlation, fuzzification innovation is carried out to obtain the fuzzy interval data. The followings are the steps to generate fuzzy interval data using the crisp data.

Given n the crisp data is stated as $X = (x_1, x_2, \dots, x_n)$, where $x_i \in R$ and X is observed variable.

Step 1. Categorizing the crisp data into j category that is A_1, A_2, \dots , and A_j .

$$\{A_1, A_2, \dots, A_{j-1}\} = \{[a_1, a_2], [a_2, a_3], \dots, [a_{j-1}, a_j]\},$$

Where $a_1 < a_2 < \dots < a_j$, $a_j \in R$, The categorization process can be adapted by the researchers who use empirical statistical techniques, hypothetical or other methods based on the expert's judgment related to the observed variables. The A_1 category has intervals of $[a_1, a_2]$ and so on. The interval width in each category does not have to be the same; A_i can be interpreted in language terms. For instance, given n crisp data from the students' mathematics learning anxiety. The anxiety data can be categorized using language terms to be A_1 = Very Low, A_2 = Low, A_3 = Moderate, A_4 = High, and A_5 = Very High.

Step 2. Defining the Fuzzy set based on step 1. The fuzzy set is generated based on the center of each crisp data categorization interval in step 1, so that the fuzzy set is obtained as $A_1^*, A_2^*, \dots, A_{j-1}^*$. Each category is defined as a fuzzy set with a triangular membership function. The interval width of each fuzzy set does not have to be the same, depending on step 1. The center can be specified with expert judgment in the field of observed variables.

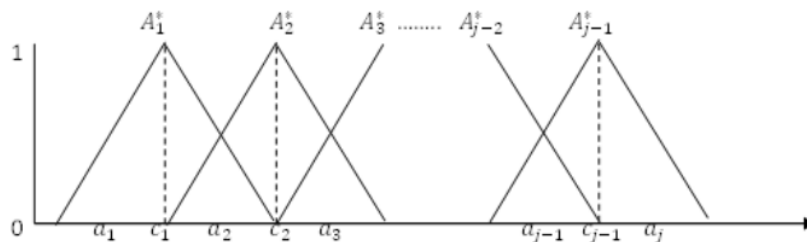


Figure 2. Fuzzy Set

Fuzzy data intervals are generated using concepts of δ -Cut, then $\forall x_i \in R, i = 1, 2, \dots, n$ obtained n fuzzy interval data $C_{\delta i}(x_i^*) = [a_i, b_i]$. To generate fuzzy interval data is then defined as,

$$x_i^* = \begin{cases} [2c_i - x_i, x_i], & \text{if } x_i \geq c_i \\ [c_i, c_i], & \text{if } x_i = c_i \\ [x_i, 2c_i - x_i], & \text{if } x_i \leq c_i \end{cases} \quad (9)$$

Where c_i is the center of each fuzzy set defined in step 2.

Step4. Reducing the fuzzy data, Fuzzy data reduction is an optional step because fuzzy interval data can already be applied in step 3. This step is done with the assumption that each x_t and x_s ($t, s \in i$,) applies $c_i = \frac{x_s + x_t}{2}$ with $x_t \leq c_i \leq x_s$, and c_i is the center from the fuzzy set A_i , obtained $\mu_{A_i}(x_t) = \mu_{A_i}(x_s)$ so that it will result in the same interval fuzzy data that is $[x_t, x_s]$. In this step, the new fuzzy data is obtained defined as.

$$x_i^* = \begin{cases} [x_i, c_i], & \text{for } x_i < c_i \\ [c_i, c_i], & \text{for } x_i = c_i \\ [c_i, x_i], & \text{for } x_i > c_i \end{cases} \quad (10)$$

Where c_i is the center of fuzzy set A_i^* . The smaller the distance x_i and c_i , then the closer $\mu(x_i)$ to 1 which means that the membership level of x_i on the fuzzy set A_i^* is getting stronger. Thus, the center becomes the primary reference for the formation of fuzzy data.

3. Result and Discussions

The data used in this study are mathematical anxiety and self-efficacy from 19 students in 11th grade. Both were taken using questionnaires based on modified fuzzy Likert [9]. The fuzzy questionnaire advantage used is to obtain two data at the same time, namely, the crisp and fuzzy interval data. The crisp and fuzzy data were continuously analyzed using Pearson correlation, as well as Cheng and Yang fuzzy correlations. Given Crisp data on mathematical self-efficacy and anxiety from 19 students in table 1.

Table 1. The Crisp and Fuzzy Data

No	Mathematical Anxiety		Self Efficacy	
	Crisp Data	Fuzzy Data	Crisp Data	Fuzzy Data
1	50	[44.1 50.9]	46	[40 45.8]
2	42	[37.9 45.4]	52	[51.5 56.2]
3	46	[43.9 48.3]	59	[53.2 60.4]
4	45	[38.9 47.8]	44	[38.5 46.7]
5	50	[45.2 52.8]	47	[43.9 50.8]
6	52	[47.6 52.7]	47	[42.9 49.3]
7	47	[41.8 53]	39	[32.1 40.6]
8	47	[43.1 54.2]	47	[41.6 52.6]
9	53	[44.7 58.6]	41	[32.2 46.5]
10	39	[35.7 42.1]	49	[46.2 51]
11	39	[35.3 39.2]	46	[42.3 47]
12	50	[36.5 57]	44	[36.5 47.5]
13	53	[50.7 55.1]	46	[47.7 47.7]
14	48	[42.9 50.2]	40	[31.8 41.7]
15	36	[28.7 39.8]	52	[51.4 60.1]
16	44	[39.9 46.1]	49	[45.2 51.9]

17	46	[43.2 48.9]	33	[26.7 36.1]
18	56	[49.1 59.5]	35	[28 37.1]
19	47	[42.8 47.7]	39	[37.5 43.9]

Step 1. Categorizing Crisp Data, Mathematical self-efficacy and learning anxiety are classified into five language terms, namely very high, high, moderate, low, and very low. The categorization of data in this study employs empirical statistics.

Table 2. Data Categorization Based on Empirical Statistics

Language Terms	Interval Anxiety	Interval Self Efficacy	Center of Self-Efficacy	Center of Anxiety
Very High	$x > 54.4641$	$x > 54.354$	57.040	57.472
High	$49.313 < x \leq 51.888$	$48.118 < x \leq 54.354$	51.888	51.236
Moderate	$44.161 < x \leq 49.313$	$41.882 < x \leq 48.118$	46.737	45
Low	$39.010 < x \leq 44.161$	$35.646 < x \leq 41.882$	41.585	38.764
Very Low	$x \leq 39.010$	$x \leq 35.646$	36.434	32.528

Step 2. Forming a Fuzzy Set. The fuzzy set can be constructed from step 1. The membership function used in this article is the triangle membership function.

Table 3. Triangle Fuzzy Set of Self-efficacy and Mathematics Learning Outcomes Variable

Fuzzy Set	Fuzzy Set of Anxiety	Fuzzy Set of Self-efficacy
Very High	[51.888 57.040 62.191]	[51.236 57.472 63.708]
High	[46.737 51.888 57.040]	[45 51.236 57.472]
Moderate	[41.585 46.737 51.888]	[38.764 45 51.236]
Low	[36.434 41.585 46.737]	[32.528 38.764 45]
Very Low	[31.282 36.434 41.585]	[26.292 32.528 38.764]

Step 3. Based on fuzzy sets in table 3, the crisp data in table 1 can be fuzzified using formula (2) which is then obtained as fuzzy interval data (FID1) in table 4 below. Furthermore, **step 4.** the reduction of FID1 data is made by using formula (4) which is then obtained as fuzzy interval data (FID2) in table 4 below.

Table 4. Fuzzy data interval of each variable

No	FID1		FID2	
	Anxiety	self-efficacy	anxiety	self-efficacy
1	[50 53.777]	[44 46]	[50 51.888]	[45 46]
2	[41.171 42]	[52 56.708]	[41.585 42]	[51.236 52]
3	[46 47.474]	[49.708 59]	[46 46.737]	[57.472 59]
4	[45 48.474]	[44 46]	[45 46.737]	[44 45]
5	[50 53.777]	[43 47]	[50 51.888]	[45 47]
6	[50 53.777]	[43 47]	[50 51.888]	[45 47]
7	[46.474 47]	[39 41.64586]	[46.737 47]	[38.764 39]
8	[46.474 47]	[43 47]	[46.737 47]	[45 47]

9	[50.777 53]	[39.646 41]	[51.888 53]	[38.764 41]
10	[33.868 39]	[49 50.354]	[36.434 39]	[49 51.236]
11	[33.868 39]	[44 46]	[36.434 39]	[45 46]
12	[50 53.777]	[44 46]	[50 51.888]	[44 45]
13	[50.777 53]	[44 46]	[51.888 53]	[45.000 46]
14	[45.474 48]	[40 40.646]	[46.737 48]	[38.764 40]
15	[36 36.868]	[52 56.708]	[36 36.434]	[51.236 52]
16	[39.171 44]	[49 50.354]	[41.585 44]	[49 51.236]
17	[46 47.474]	[33 38.292]	[46 46.737]	[32.528 33]
18	[56 58.080]	[35 36.292]	[56 57.040]	[32.528 35]
19	[46.474 47]	[39 41.646]	[46.737 47]	[38.764 39]

Those fuzzy data are applied to determine the correlation between anxiety and self-efficacy. By using the fuzzy correlation formula (1), the following correlation coefficient is obtained.

Table 5. Correlation Analysis Results

Correlation Method	The Used Data	Correlation coefficient	Correlation Direction	Correlation level
Pearson	Crisp Data	-0.45828	Negative	Medium
Cheng dan Yang	fuzzy Data from data collection	[-0.42853 -04194]	Negative	Medium
Cheng dan Yang	fuzzy Data from the 3 rd step	[-0.49873 -0.4956]	Negative	Medium
Cheng dan Yang	fuzzy Data from the 4 th step	[-0.48431 -0.48338]	Negative	Medium

Table 5 reveals the impressive results, where the correlation coefficient value from the generating fuzzy interval data results above concludes that it has no many different results with Pearson and Cheng-Yang correlations using fuzzy interval data from data collection.

4. Conclusion

These concepts of transformation the crisp data into fuzzy interval data can be the alternative for researchers whose data collection still use questionnaires that are not based on fuzzy concepts but want to analyze two variable correlation using Cheng and Yang fuzzy correlations. The application of generating process fuzzy interval data result on Cheng, and Yang Fuzzy correlation reveals that it has no many different results with Pearson correlations and fuzzy correlation using fuzzy data from data collection.

The step to generate fuzzy interval data in this article requires further constructive researches because the categorization process is carried out using empirical data in step 1, while this process can be carried out using hypothetical data or expert judgment. In this article, the center of the formed fuzzy set step 2 is the middle value of each category from step 1. The center of the fuzzy set can be determined separately and does not have to be the middle value of the category interval. For example, if there are consecutive top and bottom category which are very high and very low, in this case, it allows using a center that is not the middle value of the interval category. This center will result in the invalidity of the formula in step 3, so it must be redefined. Further, step 4 doesn't change at all.

References

- [1] C A Murthy, S K Pal, and D. Dutta Majumder 1985 Correlation between two fuzzy membership functions, *Fuzzy Sets and Systems* **17** pp 23-38

- [2] Hung WL and Wu JW 2001 A Note on The Correlation of Fuzzy Numbers By Expected Interval. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* **9** pp 517-523
- [3] Hanafy I M, Salama AA and Mahfouz K M 2013 Correlation Coefficients of Neutrosophic Sets by Centroid Method *Journal of International Probability and Statistics* **2** pp 9-12.
- [4] Cheng Y and Yang C C 2014 The Application of Fuzzy coefficient correlation with Fuzzy interval data. *International Journal of Innovative Management, Information & Production* **5** pp 67-68
- [5] Wu B and Hung C F 2016 *Innovative Correlation Measurement With Fuzzy Dat Mathematical Problems In Engineering* (Hindawi) 2016 pp 1-11
- [6] K Atanassov 1986 Intuitionistic Fuzzy Sets, *Fuzzy Sets and Systems* **20** pp 87-96.
- [7] Viertl R 2010 *Statistical methods for Fuzzy data*. Aptara Inc (New Delhi, India) pp 23-25
- [8] Gil M A, Lubiano M A, Saa SR, and sinova B 2015 Analyzing Data from Fuzzy rating scale-based Questionnaire. *A Case Study Psicothema* **27** pp 182-91
- [9] Merideth, M L 2007 The Correlation Between Self-Efficacy and Student Achievement in Male Student with a Specific Learning Disability. *ProQuest LLC.The USA*.

hps-Pros 2019 Generating Fuzzy Interval Data and Its Application to Find The Relation Between Math

ORIGINALITY REPORT

12%

SIMILARITY INDEX

8%

INTERNET SOURCES

8%

PUBLICATIONS

5%

STUDENT PAPERS

MATCH ALL SOURCES (ONLY SELECTED SOURCE PRINTED)

3%

★ gfzpublic.gfz-potsdam.de

Internet Source

Exclude quotes On

Exclude matches < 1%

Exclude bibliography On