**Noname manuscript No.**

(will be inserted by the editor)

**Bootstrapping for Fuzzy Mediation and Moderated-Mediation Analysis using Fuzzy Lease Squares Estimation (FLSE) and Fuzzy Least Absolute Deviations (FLAD) with Evolutionary Algorithms**

**Da Jeong Kang1,†· Hyeon Gu Kang1,†· Geon Hee Lee2 · Sung Wook Baik3 · Zong Woo Geem2,\* · Jin Hee Yoon1,\***

Received: date / Accepted: date

|  |
| --- |
| **Abstract**  Mediation analysis is widely employed to explore how an independent variable (X) affects a dependent variable (Y) through a mediating variable. Traditional methods, such as the Baron and Kenny approach or Sobel test, often assume normal distributions and are inadequate when applied to ambiguous or fuzzy data, which are prevalent in behavioral and social sciences. Existing research has not adequately addressed the integration of fuzzy set theory with modern inferential tools like bootstrapping, leaving a gap in accurate analysis of uncertain, imprecise, or linguistically expressed data. This study aims to bridge that gap by introducing a novel methodology for conducting fuzzy mediation and moderated-mediation analysis using bootstrapping techniques. We propose an innovative framework combining fuzzy regression methods—Fuzzy Least Squares Estimation (FLSE) and Fuzzy Least Absolute Deviations (FLAD)—with bootstrapping for confidence interval estimation. FLSE uses convex optimization and has a closed-form solution, while FLAD, being non-differentiable, is optimized via Genetic Algorithm (GA) and Harmony Search (HS). Model performance is evaluated using fuzzy-specific error metrics, such as Fuzzy Root Mean Square Error (FRMSE) and Fuzzy Mean Absolute Error (FMAE). Empirical validation across three datasets—team dynamics, adolescent hate speech, and solar energy output—demonstrates the robustness of our approach. The proposed methods outperform traditional crisp-number models by capturing uncertainties more accurately and producing statistically significant estimates even in small or asymmetric datasets. This study pioneers the integration of bootstrapping into fuzzy mediation and moderated-mediation analysis, offering a statistically rigorous, interpretable, and scalable solution for handling ambiguity in empirical research. Our framework is particularly effective in fields like psychology, renewable energy, and social behavior, where linguistic and imprecise data are common.  **Keywords:** Fuzzy mediation analysis · Bootstrapping · Fuzzy least squares estimation (FLSE) · Fuzzy least absolute deviations (FLAD) · Evolutionary algorithm |

# 1 Introduction

Understanding how one phenomenon influences another has long been a central question in the social and behavioral sciences. Researchers have emphasized that human responses to external stimuli are rarely immediate; instead, they involve internal cognitive and emotional processing. This has led to significant interest in uncovering the mechanisms by which one variable affects another. To explain this causal relationship, scholars introduced the concepts of mediation and moderation by incorporating a third variable into statistical models. A mediator variable is one that logically intervenes between an independent and a dependent variable, helping to explain why or how the

1 Department of Mathematics and Statistics, Sejong University, Seoul 05006, South Korea ; {doramisister, mjsk1145}@naver.com

2 Department of Smart City, Gachon University, Seoul 13120, South Korea; [zwgeem@yahoo.com](mailto:zwgeem@yahoo.com); [ghlee490@gmail.com](mailto:ghlee490@gmail.com)

3 Department of Software, Sejong University, Seoul 05006,

Korea;[sbaik@sejong.ac.kr](mailto:sbaik@sejong.ac.kr)

**\***Correspondence: [jin9135@sejong.ac.kr](mailto:jin9135@sejong.ac.kr); zwgeem@yahoo.com

**†** These two authors contributed equally to this work

effect occurs. The purpose of mediation analysis is to identify an intermediate variable that accounts for the observed relationship between two variables. A moderator variable, on the other hand, affects the strength or direction of the relationship. Moderation analysis seeks to determine when, for whom, or under what conditions the relationship between variables is stronger or weaker.

Numerous studies have addressed both mediation and moderation effects (Baron and Kenny, 1986; Fairchild and MacKinnon, 2009; Hayes, 2017; MacKinnon, 2007, 2012; Sobel, 1982). Additional research has focused on integrating these two effects, leading to models such as moderated mediation (Baron and Kenny, 1986; Edwards and Lambert, 2007; Hayes, 2015, 2017; James and Brett, 1984; Morgan-Lopez and MacKinnon, 2006). The term "moderated mediation," first introduced by James and Brett (1984), refers to the condition in which the strength of the mediation effect varies based on the value of a moderator variable (Judd and Kenny, 1981).

Regression-based mediation analysis methods such as the Sobel test (Sobel, 1982), the Baron and Kenny approach (1986), and the Aroian and Goodman tests have been widely used. However, these methods have notable limitations: for instance, the Baron and Kenny method does not establish statistical significance; the Sobel, Aroian, and Goodman tests are complex and often lack statistical power; and none of them adequately address measurement errors. Bootstrapping has recently gained popularity as an alternative, as it does not rely on assumptions about normality or the sampling distribution (Davison and David, 1997; Efron, 1979; Hayes and Scharkow, 2013; Lee, 2014).

Traditional mediation analysis assumes that variables are expressed as precise numbers. However, in many real-world applications—particularly in psychology and social sciences—data are often ambiguous or linguistically expressed (e.g., "somewhat satisfied," "moderately likely"). Quantifying such responses into fixed numbers can lead to information loss and distortions. Individual perception scales vary, meaning the same numeric value may reflect different subjective evaluations. Zadeh (1965) proposed fuzzy numbers to better handle such uncertainty.

Building on this, Yoon (2020) introduced fuzzy mediation analysis. Kim et al. (2021) applied it to financial data. However, no prior studies have integrated bootstrapping into fuzzy mediation or fuzzy moderated-mediation models. While bootstrapping has become a standard approach to estimate indirect effects, its combination with fuzzy models remains underexplored.

This study proposes a new methodology that combines bootstrapping with fuzzy regression models—specifically Fuzzy Least Squares Estimation (FLSE) and Fuzzy Least Absolute Deviations (FLAD). FLSE leverages convex optimization, offering a closed-form global minimum solution. In contrast, FLAD handles non-differentiable functions using evolutionary algorithms such as Genetic Algorithm (GA) and Harmony Search (HS). Model accuracy is evaluated using fuzzy-specific metrics: Fuzzy Root Mean Squared Error (FRMSE) and Fuzzy Mean Absolute Error (FMAE).

While FLSE yields high performance through direct computation, FLAD approximates the solution due to non-differentiability. In more complex settings such as moderated mediation, FLSE's performance may sometimes be less optimal. Therefore, we employ both FLSE and FLAD, comparing their effectiveness across different datasets.

The primary goal of this study is to offer a more robust and interpretable framework for analyzing ambiguous or uncertain data.

Table 1 summarizes the key characteristics of major estimation methods commonly used in mediation analysis. While the Sobel test relies heavily on parametric assumptions such as normality and linearity, bootstrap-based inference offers a more flexible alternative that partially addresses nonlinearity and is moderately robust to outliers. However, both methods are designed for precise numeric data.  
In contrast, fuzzy regression methods such as FLSE (Fuzzy Least Squares Estimation) and FLAD (Fuzzy Least Absolute Deviations) are not only free from distributional assumptions, but also well-suited for handling uncertainty and linguistic ambiguity inherent in fuzzy data. Notably, FLAD also offers robustness to outliers through L1-norm optimization.

**Table 1** Comparison of Statistical and Fuzzy Estimation Methods Used in Mediation Analysis

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Method | Assumes Normality? | Handles Nonlinearity? | | Robust to Outliers? | Suitable for Fuzzy Data? |
| Sobel Test | Yes | | No | No | No |
| Bootstrap | No | | Partial | Partial | Yes (with fuzzy modeling) |
| FLSE | No | | Yes | No | Yes |
| FLAD | No | | Yes | Yes | Yes |

The original contributions of the Paper is as follows:

*1)Bootstrapping in Fuzzy Mediation Models:* This study is the first to integrate bootstrapping into fuzzy mediation and moderated-mediation analysis, eliminating the need for normality assumptions found in methods like the Sobel test.

*2) FLSE and FLAD Hybrid Estimation:* We employ convex optimization for FLSE and evolutionary algorithms (GA, HS) for FLAD to estimate coefficients in fuzzy mediation models.

*3) Performance Metrics*: The study introduces FRMSE and FMAE to assess the accuracy of fuzzy models.

*4) Empirical Validation:* We apply the model to three real-world datasets:

- Team Performance Data: Assesses how dysfunctional behavior impacts team performance.

* Hate Speech Data: Evaluates how exposure to hate speech influences perceived discrimination.
* Solar Energy Data: Analyzes the effects of weather conditions on solar power output.

Section 2 introduces conventional statistical tests and the bootstrap approach. Section 3 outlines the proposed fuzzy mediation and moderated-mediation models. Section 4 describes the estimation method using bootstrapping. Section 5 applies these methods to real-world data, while Section 6 compares optimization algorithms. Section 7 discusses the practical approach for the engineering field. Section 8 concludes the study and summarizes key findings.

# 2. Mediation and moderated-mediation analysis

**2.1 Baron & Kenny**

Baron & Kenny’s (Baron and Kenny 1986) research has made a clear definition of mediating effect and controlling factors and explained the logic of verification on mediating effect readily intelligibly and intuitively. It is the most widely cited method in the papers as a testing method of the mediating effect by verifying how the mediating effect can be proven.

The method has recently encountered criticisms due to a bevy of problems. When estimating the size of the mediating effect, the conclusion on the mediating effect has been made indirectly by verifying with different figures in order not by verifying from the statistical reasoning to determine if the size has a significant meaning. An error can occur at in anytime, especially when examining a hypothesis. The probability of an error is inevitably getting higher as the number of hypotheses to be proven increases simultaneously. Hence, it has turned out that the reliability of the testing is weak due to excessive errors that occurred from the sequential testing of multiple hypotheses (Fritz and MacKinnon 2007; Hayes 2013). In addition, it is widely known that Baron & Kenny’s testing method analyzes the mediating effect based on the assumption that the effect of independent variables on the dependent variables should be statistically significant. However, it is not valid. The verification method of the mediating effect is under the criticism that it is not an accurate statistical method rather than it is not statistically close (Hayes 2017).

**2.2 Sobel test**

The core problem of Baron & Kenny’s verification method occurs indirectly in the verification process of mediating effect. Sobel’s method (Sobel 1982) can be considered an advanced approach in that the method calculates the magnitude of the effect directly. Researchers frequently cite the Sobel test since the method can be utilized comparatively simply than other methods in verifying the mediating effect. However, it is found that there are defects in Sobel’s verification method. When verifying the significance of the mediating effect with Sobel’s testing method, the assumption is that the sample distribution of the value forms the normal distribution. Unlike the assumption, however, the sampling distribution used widely by most researchers in mediating effect verification is mostly deflective, not showing the normal distribution Bollen and Robert 1990; Shrout and Niall 2002). Therefore, it can be deduced that Sobel’s method has limitation in telling the statistical significance of mediating effect (Fritz and MacKinnon 2007; Hayes 2013). Unlike traditional mediation tests such as Sobel’s method, which assume normality of the sampling distribution and are sensitive to small sample sizes, bootstrap methods provide a non-parametric alternative that does not rely on such assumptions. This makes bootstrapping particularly suitable for fuzzy data or linguistically coded variables, where classical assumptions are often violated. The bootstrap resampling approach enables more reliable estimation of indirect effects and their confidence intervals, especially under asymmetry or nonlinearity. As shown in recent literature (e.g., Hayes & Scharkow, 2013), it is considered more robust and statistically powerful than analytic approximations.

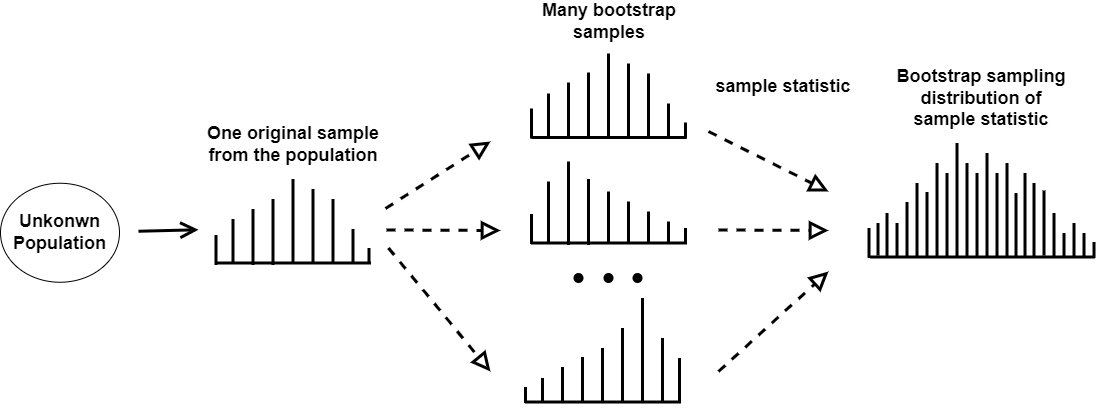
**2.3 Bootstrapping**

Generally, the confidence interval of the mediating effect has been calculated based on the assumption that the sampling distribution follows the normal distribution or a t distribution. However, the cases where the extracted samples do not follow a normal or t distribution are found. If the confidence interval is calculated with a normal distribution or a t distribution when the sampling distributions are not symmetrical, it can provide an approximation of the correct confidence interval. However, it cannot always provide a reliable approximate value as always. As an alternative to the weakness of the method, the bootstrapping method has recently become pervasive for researchers.

The bootstrapping method is a statistical method that estimates the sample distribution based on the empirical distribution utilizing sample data while the sample distribution is not informed. In other words, this method's strength lies in its ability to calculate the approximate standard error, confidence interval, and significance probability of the estimated sample distribution without making any assumptions about the distribution of variables or the sampling distribution. This is achieved by repeatedly resampling the same-sized sample randomly with replacement and projecting, as shown in Fig 1(Bollen and Robert 1990; Bollen and Stine 1992; Davison and David 1997; DiCiccio and Bradley 1996; Efron 1979; Fritz and David 2007; Hayes and Schaarkow 2013; Shrout and Niall 2002; Tibshirani and Bradley 1993).

Two methods have been suggested for verifying the mediating effect with bootstrapping. First, one is to determine whether zero is included in the confidence interval of the re-extracted sample distribution, and the second is to identify the effect with the significant probability of total indirect effect through the decomposition of testing mediating method. The specific methods that calculate the confidence interval can be categorized into Percentile and Bias-corrected.

The method to examine the mediating effect by utilizing the bootstrapping method has been introduced by several scholars since the 1990s (Bollen and Robert 1990; Hayes 2009; Hayes 2017; MacKinnon et al. 2002; Preacher and Hayes 2004; Preacher and Hayes 2008; Shrout and Niall 2002). Despite the strengths the method has, the reason why the bootstrapping method has not been widely accepted was derived from the difficulty in executing a considerable amount of calculation without using a computer, and there were inevitable limitations in application due to its complexity in programming. However, along with the significant advancement in computer development and the simplified procedures in using bootstrapping through the various statistic packages, the utilization ratio of the method is getting higher in different academic fields. Therefore, in this paper, the statistical significance of the fuzzy mediation model was explained using bootstrap.



**Fig. 1** Distribution estimation using bootstrap

2.3.1 Percentile bootstrap

When the magnitude of the influence of the independent variable on the parameter is set as a, and the magnitude of the effect of the parameter on the dependent variable, controlling the influence of the independent variable, is set as b, the indirect effect can be defined as ab. Among the reasoning methods that do not require assumptions on the sampling distribution of ab that refers to the magnitude of the indirect effect, there is a typical method that test the indirect effect by utilizing the confidence interval of a bootstrap. One of the procedures to set the confidence interval (95%) by the percentile bootstrap method is as follows (Shrout and Niall 2002). All procedures are automatically carried out in PROCESS macro, a computer program developed by Hayes.

Reinforcement is extracted from the original sample with sample size extracted from the population, and a bootstrap sample with the same size as the original sample is extracted.

1. Using the bootstrap sample obtained in step 1, estimate the statistics of indirect effects in the resampling.
2. Repeat steps 1 and 2 times to generate bootstrap samples and estimate and store indirect effects using them.
3. Sort the indirect effect estimates from lowest to highest.
4. In the case of using a 95% confidence interval, the lower limit is defined as the statistic value corresponding to the 0.5th (100-95)th percentile of the distribution of the previously obtained statistic value. The upper bound is defined as the statistic corresponding to the [100-0.5 (100-95)]th percentile from the distribution of statistics arranged in ascending order. The lower and upper bound values are determined as the endpoints of the 95% confidence interval.

If 0 is not included in this 95% confidence interval, the indirect effect is said to be statistically significant.

**3. Fuzzy mediation and moderated-mediation analysis**

In this section, referring to the basic concepts in (Hayes 2017), we introduce the definition of fuzzy numbers by Zadeh (Zadeh 1965), and simple fuzzy mediation models with mediators and fuzzy moderated-mediation model introduced by Yoon (Yoon 2020).

**3.1 Fuzzy number**

The definition of a fuzzy number in real numbers, which implies when normalized and convex, is a fuzzy set. An element of a fuzzy set is one that accepts a real value between -1 and 1 as a measure of belonging according to a function known as the membership function. There are no standard rules since the membership function's form might be described in terms of either objective or subjective possibilities. As a result, the -fuzzy numbers parametric class of fuzzy numbers is used in this particular situation. A fuzzy number A is referred to be an LR fuzzy number if it meets the conditions listed below.

(1)

where *L* and *R* are reference functions called left and right shape functions of and have the following properties: , are left-continuous and decreasing function with . And ‘m’ means the mode of the -fuzzy number A. ‘l’ and ‘r’ are greater than 0 and mean the width of the left and right sides. We abbreviate the LR-fuzzy number as . And -fuzzy number, one of the triangular numbers, has the following two operations.

= (, (2)

(3)

.

**3.2 Simple fuzzy mediation model**

Through simple regression analysis, the Baron and Kenny's Simple Mediation Model's mediation analysis method analyzes the causal relationship step-by-step through. A statistical technique called mediation analysis explores ideas on how a causal antecedent variable () influences an outcome variable (). We suggest the following derivation of the three regression equations of Baron and Kenny: regression between independent variables and dependent variables, regression between independent variables and mediators, and regression between mediators and dependent variables.

(4)

(5)

(6)

The regression constants and are significant in this model. Here, X's estimated "direct effect" on Y is represented by the number "", while X's estimated "indirect effect" on Y through M is represented by the number " " and the number " ", which is the product of the two. Additionally, X's direct and indirect effects are added together to get , which is known as the "total effect" and equal to . It demonstrates that the direct effect () is less significant than the total effect (.

It makes more sense to describe ambiguous concepts as variables with fuzzy numbers than crisp numbers, such as "more," "less," and "happy." The following is the suggested fuzzy mediation model.

(7)

(8)

(9)

In the model as above, is the total effect, is the indirect effect, and is the direct effect. Note that it is easily checked that

**텍스트, 시계, 손목시계이(가) 표시된 사진

자동 생성된 설명**

**Fig. 2** The simple fuzzy mediation model

**3.3 Fuzzy mediation model for multiple mediators**

It is often advisable to utilize a basic fuzzy mediation for various mediators rather than a simple fuzzy mediation model since the real world is much more complicated and multi-causal.

The following is a simple fuzzy mediation model with parameters .

(10)

(11)

(12)

where

In the model as above, is the total effect, is the indirect effect through and on , and is the direct effect. In other words, there are k indirect effects.



**Fig. 3** The fuzzy mediation model with multiple mediators

**3.4 Fuzzy moderated-mediation model**

**3.4.1 Moderated-mediation model**

Moderated-mediation is the mechanism by which the moderating variable (), the fourth variable in the causal relationship, may adjust the indirect effect from the independent variable () to the dependent variable () through the parameter (). The terms "Adjusted mediating effect" and "conditional indirect effect" are presently used interchangeably and have the same meaning in statistics. (Preacher, Rucker & Hayes, 2007)

Understanding the conditional character of the mechanism by which one variable effect another variable and testing hypotheses about these conditional effects are the purpose of conditional process analysis, which combines conditioning analysis with mediation analysis. As a result, if the direct and indirect effects of the independent variable (X), which is important in the moderation analysis, are calculated and the influence of the independent variable () on the parameter (), which is organized by the moderating variable (), then X is not a single number but the effect on is a function of . Rather, the effect of X on M becomes a **function of W**, meaning that X’s impact on M varies depending on the level of the moderator.

**3.4.2 Fuzzy moderated-mediation model**

Let's say is a fuzzy predictor variable, is a fuzzy response variable, is a fuzzy mediator, and is a fuzzy moderator in a causal relationship involving ambiguous variables. The model shown in the Fig. below is a moderated mediation model, in which controls the route from to but does not influence any other pathways.

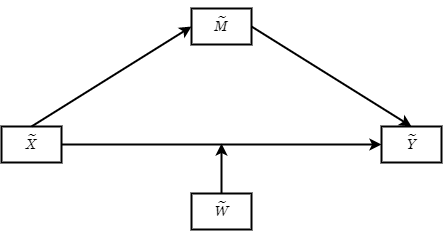
Because controls it, the direct eﬀect of is dependent on . Fig. 1 expresses the following model.

(13)

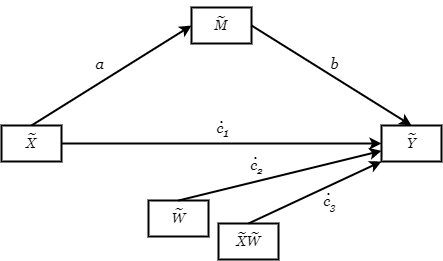
= (14)

. (15)

Here, is defined as a function of = and represents the "fuzzy conditional effect" of in . Within this model, the indirect effect is defined as the product of a and b, identical to the simple fuzzy mediation model. The indirect effect in this model is deemed unconditional, as no moderation is imposed upon the or path. Therefore, in this model, indirectly affects through and also directly affects while depending on .

****

(a) Conceptual diagram

****

(b) Statistical diagram

**Fig. 4** Moderated-mediation analysis for conditional direct effect

Yoon (Yoon 2020) has so far put up a number of fuzzy mediation models. However, no study has used bootstrapping to evaluate this fuzzy model. The initial proposals for bootstrapping-based analysis of parameters are made in Sect. 4. for simple fuzzy parameter analysis and Sect. 5’s proposals are for analyses of different data types and climatic data.

**4. Bootstrapping for muzzy mediation and moderated-mediation analysis**

In Sect. 4, using the FLSE and FLAD method, we estimated regression coefficients and explored causal relationships among variables. The FLSE approach minimizes the squared L-distance between predicted and observed fuzzy numbers, which parallels classical least squares but is extended to fuzzy domains. On the other hand, FLAD focuses on the absolute deviation, providing robustness against outliers or skewed distributions. Both serve as fuzzy analogues to OLS and LAD, with interpretation linked to minimizing expected fuzzy prediction error under different norms. Furthermore, with objective functions anchored in or -metrics, we extended our analysis to compare the introduced accuracy metrics, FMSE and FMAE, across different optimization techniques.

**4.1 Fuzzy mediation and moderated-mediation analysis based on FLSE**

When using Least Squares Estimation with fuzzy data (FLSE), it is necessary to have a suitable metric in the fuzzy set spaces. A helpful type of metric can be established through the use of support functions. The support function of any compact, convex set can be represented as which is determined by the following formula for all :

(16)

where is the (d-1)-dimensional unit sphere in and represents the scalar product in . It should be noted that for compact and convex sets the support function is uniquely defined. A metric in a fuzzy number set is established through the use of the *-* metric in the space of Lebesgue integrable functions, represented as:

(17)

This leads to the definition of an *-* metric for fuzzy numbers as:

(18)

A fuzzy regression model was previously introduced in the author's studies [24,25] and is expressed as follows:

. (19)

The variables are represented by and for It is assumed that are the fuzzy random errors that account for the fuzziness. It is worth mentioning that all cases can be covered by defining and as follows:

(20)

(21)

where represent the left and right spreads of respectively.

The estimators are obtained by minimizing the following objective function:

(22)

where *q* is the number of the regression model in this fuzzy mediation analysis and *k=1,2,…,q*,.

**4.1.1 Fuzzy mediation and moderated-mediation analysis using FLSE through mathematical formula**

The aforementioned objective function is based on the *-*metric, and the *-* distance can be calculated as (Yoon 2013):

(23)

To minimize the above equation, we obtain the normal equation applying

The normal equation has as its solution, and for each value of , the normal equation can be written as follows:

(24)

To determine the solution vector, we introduce a *triangular fuzzy matrix* *(t.f.m.)* which is expressed as

(25)

and abbreviated as , where is a triangular fuzzy number forandAdditionally, we define a triangular fuzzy vector

. (26)

To minimize the objective function mentioned above, we apply the fuzzy operations, fuzzy numbers and estimators defined in our previous studies (Bollen and Robert 1990;

Fritz and MacKinnon 2007; Kim et al. 2020; Shrout and Niall 2002;). The fuzzy operations are as follows:

(27)

*.* (28)

The following operations are defined for two triangular fuzzy matrices, , , and a crisp matrix :

, (29)

*,* (30)

*,* . (31)

(32)

. (33)

The solutions to the normal equation fuzzy estimators are derived for each by using the above operations and algebraic properties, with

(34)

where

and , for Note that the solution (Hayes and Schaarkow 2013) exists only if . (35)

**4.2 Fuzzy mediation and moderated-mediation analysis based on FLAD**

Least Absolute Deviations (LAD) is a method for measuring error in linear regression models. This method minimizes the sum of the absolute values of the prediction errors for each data point.

While the LSE method, which uses squared errors, is heavily influenced by outliers, LAD, by utilizing absolute errors, is less susceptible to the effects of outliers. Furthermore, LAD provides stable estimates even when the distribution of errors does not follow a normal distribution.

Given these advantages, LAD paves the way for the development of an L\_1-metric for fuzzy numbers, leading us to FLAD (Fuzzy Least Absolute Deviations). FLAD extends LAD's approach to fuzzy data, ensuring more robust and outlier-resistant error measurement and estimation. This progression naturally culminates in the definition of an L\_1-metric for fuzzy numbers as follows:

(36)

where ,

We assume that represent the fuzzy random errors accounting for the inherent fuzziness. It should be noted that a comprehensive representation can be achieved by appropriately defining both and .

Incorporating the *-* metric into a fuzzy regression model, we obtain an objective function that seeks to minimize the cumulative distance between the observed and predicted fuzzy numbers. For the regression model indicated as (36), the objective function using the *-* metric is given by:

, (37)

for , where is the number of the regression model in this fuzzy analysis.

To find the optimal set of coefficients that minimize the objective function (37), computational algorithms such as the GA and HS are employed. These algorithms iteratively adjust the regression coefficients in a manner that reduces the L1 distance between the fuzzy observed and predicted values.

**4.3 Estimation of fuzzy mediation and moderated-mediation analysis based on evolutionary algorithm**

**4.3.1 Genetic algorithm**

Genetic Algorithms (GA) are computational optimization techniques underpinned by the principles of biological evolution. They operationalize strategies for solution identification, harnessing natural processes such as selection, crossover, mutation, and heredity. An initial cohort, termed the 'population', consists of randomly generated solutions, each represented as a 'chromosome'. Subsequently, the performance of each chromosome is assessed via a fitness function, as detailed by Sampson in his review of John H. Holland's work (Sampson 1976).

Chromosomes with superior fitness are earmarked as progenitors for the subsequent generation, and offspring chromosomes are engendered through crossover operations. Mutation operations, executed sporadically on select chromosomes, serve to preserve genetic diversity, and counteract premature convergence to local optima. Through iterative cycles of these processes, GA tend to converge to an optimal or near-optimal solution.

The allure of GA lies in their scalability, versatility, and adeptness in navigating intricate problem landscapes. They have found utility in a myriad of domains, including but not limited to machine learning, economics, medicine, and engineering, as tools for addressing multifaceted optimization challenges.

**4.3.2 Harmony search**

Harmony Search (HS) is a metaheuristic optimization technique inspired by the improvisation of musicians. This algorithm is recognized in the field of computational intelligence for its unique approach to finding solutions resembling musical harmonies.

At the heart of the algorithm lies the 'Harmony Memory' (HM), a repository that can be seen as a collection of solutions symbolizing musical harmonies. During its iterative process, HS either utilizes existing solutions stored in HM, modifies them slightly (pitch adjustment), or introduces new harmonies. This procedure is primarily driven by two probabilistic parameters: the Harmony Memory Considering Rate (HMCR) and the Pitch Adjusting Rate (PAR).

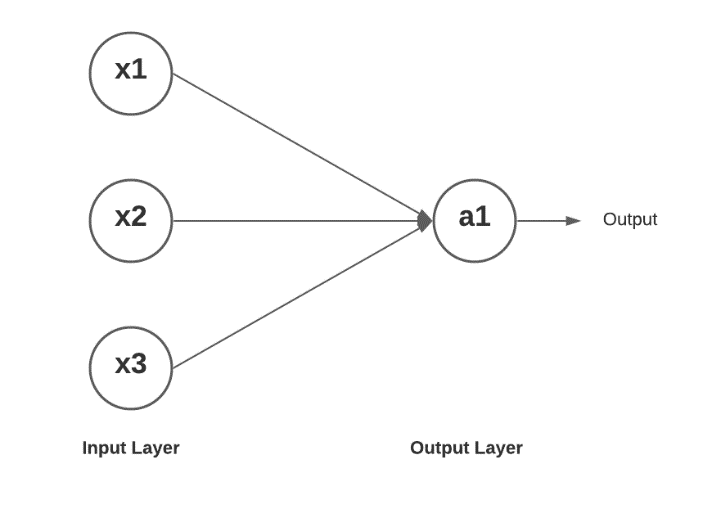
The algorithm emulates the process wherein musicians repeatedly draw from past experiences and attempt new sequences in their pursuit of the perfect harmony. Just as musicians strive to find the right balance in tunes, HS continually innovates and refines to seek the optimal solution, as discussed in the work of Geem, Z. W., Kim, J. H., & Loganathan, G. V. (Geem et al. 2001).

Owing to its adaptability and broad applicability, HS has been effectively deployed in diverse domains, ranging from engineering design optimization to intricate real-world challenges. Especially noteworthy is the algorithm's balanced capability in global exploration and local search, allowing it to efficiently navigate intricate search landscapes, thereby exhibiting high performance in various optimization problems.

GA and HS are garnering attention as metaheuristic algorithms designed to overcome the complexities and uncertainties inherent in fuzzy data analysis. These methods offer profound insights and effective solutions for fuzzy data through their capabilities in optimizing across expansive search spaces, maintaining a balance between exploration and exploitation, efficiently integrating with fuzzy systems, and adeptly handling the uncertainties of the data.

**4.4 Estimation of fuzzy mediation and moderated-mediation analysis based on neural-networks**

Deep learning is primarily designed to discern complex patterns in data and forecast outcomes based on given inputs. Among traditional statistical methodologies, bootstrapping stands out, particularly in scenarios with limited datasets. Bootstrapping creates multiple samples by randomly resampling with replacement from the provided data, facilitating diverse estimations and hypothesis testing. When integrating deep learning with bootstrapping, valuable insights regarding the parameters of neural-networks (N-N) and their stability can be obtained. Essentially, bootstrapping assists in assessing the uncertainty of predictions rendered by deep learning models, enhancing the interpretability and reliability of the model.



**Fig. 5** Perceptron – neural network (N-N)

In this study, we propose a method that employs an N-N structure for the estimation of coefficients, emphasizing simplicity and transparency. The proposed structure aligns closely with statistical methods for coefficient estimation and offers more intuitive and interpretable results compared to more complex models. The focus of this structure is on capturing the essential linear relationships within the data rather than the more intricate nonlinear interactions. The complexity of modern datasets often renders traditional statistical analysis insufficient. While N-Ns are adept at uncovering complex patterns within data, relying solely on these networks can lead to issues with model reliability and interpretability. Integrating N-Ns with the methodological robustness of bootstrapping can substantially mitigate these issues. Furthermore, this research aims to utilize a variety of optimization techniques grounded in gradient descent.

**4.5 Model performance measurement metrics**

**FRMSE (Fuzzy Root Mean Squared Error)**

FRMSE (Fuzzy Root Mean Squared Error) is an extension of the conventional RMSE (Root Mean Squared Error) that incorporates fuzzy set theory. RMSE measures the differences between predicted values by a model and observed values. By using the square root of the average squared differences, RMSE gauges how accurately a model predicts observed values.

Incorporating fuzzy logic, FRMSE factors in the uncertainty associated with predictions. Fuzzy logic handles data that has a degree of membership within sets rather than absolute values. In FRMSE, this logic measures the uncertainty in both predictions and observations, offering a more detailed understanding of model performance in scenarios with unclear data. Unlike RMSE's clear numerical value, FRMSE provides a range representing possible error, depending on the fuzziness.

(38)

**FMAE (Fuzzy Mean Absolute Error)**

FMAE (Fuzzy Mean Absolute Error) is an error metric combining traditional MAE with fuzzy set theory. By integrating fuzzy logic, FMAE weighs errors based on their degree of uncertainty. This approach is beneficial for data with outliers or imprecise measurements. Yet, its computation demands well-defined fuzzy membership functions and potentially fuzzy rules to capture data's inherent fuzziness.

(39)

Compared to traditional metrics, these fuzzy-based ones are more robust to outliers and better handle data uncertainty. Using fuzzy theory is vital in datasets with inherent uncertainties or ambiguities.

**4.6 Statistical inferences of fuzzy mediation model and moderated-mediation analysis**

FLSE is derived exclusively through mathematical formulation.

**4.6.1 Interval estimation using Sobel method**

The assumption that Fuzzy least squares estimators follow a normal distribution asymptotically is well-known from previous research. When the population variance is unknown, the t-distribution is applied, and when the sample size is large, it is assumed that the data follows a normal distribution, and the (1-α)100% confidence interval for the total effect and direct effect can be represented using the z-value.

*CI for the total effect :*

(40)

*CI for the direct effect*

(41)

The standard error, *se* of the direct effect and total effect is defined as follows.

(42)

(43)

For the sample mean, the following property can be applied.

*,* where is a fuzzy random variable.

For the indirect effect, the confidence interval can be inferred through the "Sobel test" or "delta method" or “product of coefficients method”. The indirect effect ab is an estimated value of based on the sample. The significance level of the indirect effect is determined by dividing the estimate of the mediating effect by the standard error of the estimate and using it as the test statistic to judge the normal distribution ().

(44)

where .(second order standard error estimator)

means the standard errors of a and b, respectively.

*CI for the indirect effect*

(45) where .

In the case of the Sobel test, the normality of the sample distribution of the indirect effect is assumed, and the standard normal distribution is used for derivation. Such assumptions are reasonable for large samples, but not for small samples. Assumptions that vary depending on the situation generally produce low test power. Nevertheless, many experts in various fields still use the Sobel test.

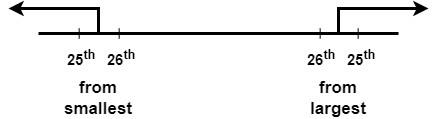
**4.6.2 Proposed interval estimation using bootstrapping**

In this Section, we propose using the bootstrap and confidence intervals for statistical inference on the model proposed in Sect. 3.

Bootstrap was first proposed by Efron (Efron 1979) to overcome the limitations of mediating effect test methods that assume normality. Bootstrap is a non-parametric and empirical resampling method that does not require assumptions about the sampling distribution of the mediating effect estimate. In the previous section, we examined the classical approach to constructing confidence intervals. We propose a method of estimating confidence intervals by using the sample percentiles of the bootstrap distribution among various bootstrap methods.

A resample of size is taken from the original sample, and the estimate is calculated for the resample. This is repeated several times ( times) to determine the dispersion of the resample estimates. A bootstrap sample , , , is obtained by resampling with replacement from the given probability distribution , *, ,,*. The desired statistic can be obtained for this bootstrap sample, and this is used to estimate the bootstrap distribution of the estimated statistic . To calculate an accurate estimate of the confidence interval, at least 1000 bootstrap samples are appropriate, and a larger value of is better.

The percentile bootstrap method *PB* constructs the *(1−α)* × 100% confidence interval for , obtained by listing the estimated statistics {, , } through bootstrap samples in ascending order ≤ ≤ · · · ≤ as [ , ]. Here, BL (Bootstrap Lower bound) is the average of the percentile of and the next number, and BU(Bootstrap Upper bound) is the average of the percentile of and the previous number.



**Fig. 6** Confidence interval estimation using bootstrap when

As depicted in Figure 6, we propose a 95% bootstrap confidence interval through 5000 iterations. Therefore, α=0.05, BL is the average of the 2.5th and 2.6th percentiles, which becomes the average of the estimated statistics at 125th and 126th, and BU is the average of the estimated statistics at 4,874th and 4,875th.

The following algorithm outlines the implementation steps for conducting fuzzy mediation and moderated-mediation analysis using FLAD (Fuzzy Least Absolute Deviations) and bootstrapping. The methodology incorporates fuzzification of variables, iterative resampling, coefficient estimation using metaheuristic optimization (GA/HS), and percentile-based confidence interval computation. The proposed framework is scalable, interpretable, and applicable in settings with uncertain or linguistically imprecise data.

|  |
| --- |
| **Algorithm*: General Procedure for Fuzzy Mediation Analysis using FLAD and Bootstrapping*** |

*Input:*

* Fuzzified data set with variables
* Number of bootstrap iterations
* Optimization algorithm: Genetic Algorithm (GA) or Harmony Search (HS)
* Error metric: *L1* fuzzy distance (FLAD objective)

*Step1] Preprocessing*

* Fuzzify input variables using triangular (or LR-type) fuzzy numbers.
* Define membership functions for each variable.
* Initialize population (if using evolutionary algorithm).

*Step2] Bootstrap Sampling*

* For to :
* Randomly sample (with replacement) a new fuzzy dataset from original data.

*Step3] FLAD Estimation per Bootstrap Sample*

* For each sampled dataset :
* Define fuzzy mediation model equations:
* Define the FLAD objective function:
* Minimize the sum of L1 fuzzy distances between observed and predicted fuzzy numbers.
* Use GA or HS to estimate coefficients .

*Step4] Store Bootstrap Estimates*

* Store the estimated coefficients for each iteration .
* Compute indirect effect as

*Step5] Confidence Interval Estimation*

* After all iterations:

Sort the list of bootstrap estimates for the indirect effect.

* Compute percentile-based confidence interval (e.g., and ).

*Step6] Output*

* Estimated coefficients:
* Confidence intervals for:

Total effect, Direct effect, Indirect effect

**텍스트, 스크린샷, 폰트, 번호이(가) 표시된 사진

자동 생성된 설명**

**5 Data Analysis: Bootstrapping Fuzzy Mediation and Moderated-Mediation Analysis using FLSE through mathematical formula**

**5.1 Fuzzy mediation Analysis for Team Performance Data**

The effect of dysfunctional behavior on team performance has been proposed by many authors (Cole et al. 2008; Diefendorff et al. 2007; Felps et al. 2006; Robinson et al. 1998; Duffy et al. 2006) Furthermore, in (Yoon 2020), a fuzzy mediation analysis is proposed. The variable “Dysfunctional Behavior” () illustrates how frequently team members act to weaken the work of others or to interrupt change and innovation. The variable “Team performance” () represents the supervisor's evaluation of the team's efficiency and ability to complete tasks in a timely manner. The variable “Negative tone” () indicates how frequently team members experience feelings of anger and disgust at work. As in the previous paper, all the variables are fuzzified with spread 0.05. The model is described in Figure 6. The result of the fuzzy mediation model is as follows:

Table 2 indicates the parameter estimates of the proposed model. Here, Indirect effect is the product of the coefficient of and .

* Indirect effect:

And the total effect is the sum of direct effect and indirect effect.

* Total effect:

텍스트이(가) 표시된 사진

자동 생성된 설명

텍스트, 라인, 폰트, 도표이(가) 표시된 사진

자동 생성된 설명

**Fig. 7** The fuzzy mediation analysis of the team performance data

**Table 2** Effects of the Dysfunctional behavior on team performance

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Method | | Effect | | | | |
| Total effect | Direct effect | | Indirect  effect | |
| FMA | 0.122 | | | 0.453 | | -0.331 |

*5.1.1 Statistical Inference on the Total, Direct and Indirect Effect*

The effects of the Dysfunctional behavior on team performance are shown in Table 2. value is approximately the same as value when the degree of freedom is large enough.

i) For 95% confidence interval for the total effect in FMA,

95% CI for ,

where

ii) For 95% confidence interval for the direct effect in FMA,

95% CI for ,

where

iii) For 95% confidence interval for the indirect effect in FMA,

95% CI for ,

where

*5.1.2 Bootstrap Confidence Interval on the Total, Direct and Indirect Effect*

The confidence intervals for the effects in the previous section were calculated under the assumption that the sample distributions of the effects follow a normal distribution. However, the result of previous research shows that the sampling distribution is skewed and does not follow the normal distribution. For this reason, we calculated bootstrap confidence intervals, which does not assume a specific sample distribution. The bootstrap confidence interval was estimated using 5000 bootstrap samples.

i) For 95% bootstrap confidence interval for the total effect

In 95% bootstrap confidence interval, the Bootstrap Lower bound (BL) is the average of the 2.5th percentile number and the next number, and the Bootstrap Upper bound (BU) is the average of the 97.5th percentile number and the previous number.

95% Bootstrap CI for

The Bootstrap Sample Distribution is shown in Figure 7 (a). Looking at the bootstrap sample distribution of the estimate of the total effect and Table 3, it is skewed to the right.

|  |  |
| --- | --- |
|  |  |
|  | |

**Fig. 8** The Bootstrap Sample Distribution of the direct effect (a), indirect effect (b) and total effect (c) of Team data

ii) For 95% bootstrap confidence interval for the direct effect

95% Bootstrap CI for

Like the previous sample, the bootstrap sample distribution of the estimate of the direct effect shown in Figure 7 (b) is skewed to the right.

iii) For 95% bootstrap confidence interval for the indirect effect

95% Bootstrap CI for

Like the previous sample, the bootstrap sample distribution of the estimate of the direct effect shown in Figure 7 (c) is skewed to the right.

**Table 3** Skewness about total, direct and indirect effect of team data

|  |  |  |
| --- | --- | --- |
| Skewness | |  |
| Total effect | Direct effect | Indirect effect1 |
| -0.392 | -0.386 | -0.546 |

**Table 4** Confidence interval about total, direct and indirect effect of team data

|  |  |  |  |
| --- | --- | --- | --- |
| Effect | Method | 95% CI  BL | 95% CI  BU |
| Total | CMA  Bootstrap  FMA  Bootstrap in FMA | -0.258  -0.444  -0.042  -0.401 | 0.479  0.510  0.284  0.508 |
| Direct | CMA  Bootstrap  FMA  Bootstrap in FMA | 0.080  -0.028  0.290  0.020 | 0.803  0.782  0.616  0.782 |
| Indirect | CMA  Bootstrap  FMA  Bootstrap in FMA | -0.568  -0.652  -0.500  -0.650 | -0.094  -0.090  -0.162  -0.094 |

Table 3 shows that all skewness values are negative, indicating that the bootstrap sample distributions are skewed to the right. As a result, we cannot assume that the sample distributions of the effects are normally distributed. Therefore, it is advisable to use the bootstrap confidence interval method in FMA, as the Sobel test assumes a normal distribution. By using the bootstrap method in FMA, we can obtain more accurate confidence intervals. However, due to the small number of team data samples and the right-skewed bootstrap samples, we can see that the confidence interval width is larger in Bootstrap in FMA than in FMA from the Table 4. Additionally, the lower boundaries of confidence intervals are more negative and the upper boundaries of confidence intervals are more positive on the vertical line in Bootstrap in FMA than in FMA. The fuzzy mediation model showed that dysfunctional behavior negatively affects team performance, with emotional tone as a significant mediator.  
**From a managerial perspective,** this suggests that leaders should not only monitor disruptive behaviors but also address emotional climates within teams.  
Implementing early detection systems for negative sentiment or emotional tone could serve as a preventive strategy to maintain high performance.

**5.2 Fuzzy mediation Analysis for Adolescent Hate Speech Data**

The exposure of adolescents to hate speech was examined in a study by (Kim et al. 2020; Kansok et al. 2022). For “Negative degree of hate speech” (), participants were asked to rate how they felt and acted when subjected to hate speech or discriminatory behavior. (1: strongly disagree, 5: strongly agree). The average score of 10 questions was used, with higher scores indicating greater negative emotions and more contracted behavior in response to hate speech or discriminatory behavior. For “Necessity of Response to hate speech” (), participants were asked to rate how much measures they thought were needed for hate/discrimination expressions. (1: not necessary at all, 5: very necessary). The average score of 7 questions was used, with higher scores indicating a greater necessity to respond to hate speech. For “Negative effect on social phenomena'() participants were asked to rate how much they thought the problem of hate speech would expand to society and have a negative impact. (1: strongly disagree, 5: strongly agree). The average score of 9 questions was used, with higher scores indicating a greater negative impact of hate speech on society. For “Seriousness of discrimination against minority groups” () participants were asked to rate how serious the problem of hate speech and discrimination against minority groups (1: not serious at all; 5: very serious). The average score of 11 questions was used, with higher scores indicating a greater seriousness of the problem. All variables in the Adolescent Hate Speech data were fuzzified with a spread of 1. The model is described in Figure 10, and the result of the fuzzy dual mediation model is as follows.

Table 5 indicates the parameter estimates of the proposed model. Here, Indirect effect of the first mediator variable is the product of the coefficient of and .

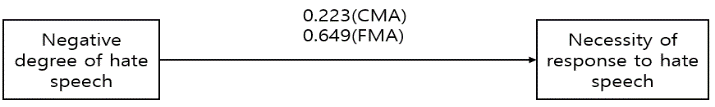
* Indirect effect:

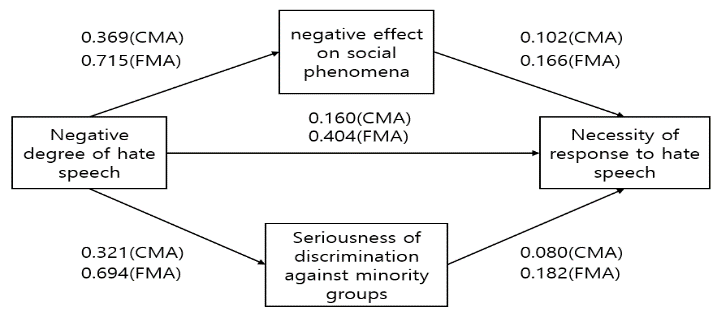
Additionally, Indirect effect of the second mediator variable is the product of the coefficient of and .

* Indirect effect:

And the total effect is the sum of direct effect and specific indirect effects.

* Total effect:



****

**Fig. 9** The fuzzy mediation analysis of the Adolescent Hate speech data

**Table 5** Effects of the Negative degree of hate speech on Necessity of response to hate speech

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Method | | Effect | | | | | | |
| Total effect | Direct effect | | Indirect effect1 | | Indirect effect2 | |
| FMA | 0.649 | | | 0.404 | | 0.119 | | 0.126 |

*5.2.1 Statistical Inference on specific indirect effects*

Here, standard error of the total effect in FMA is,

Additionally, standard error of the mediator variables is,

i) For 95% confidence interval for the indirect effect of the first mediator variable in FMA,

95% CI for ,

where

ii) For 95% confidence interval for the indirect effect of the first mediator variable in FMA,

95% CI for ,

where

*5.2.2 Bootstrap Confidence Interval on specific indirect effects*

i) For 95% bootstrap confidence interval for the indirect effect of the first mediator variable

95% Bootstrap CI for

The Bootstrap Sample Distribution is shown in Figure 9 (a). Looking at the bootstrap sample distribution of the estimate of the indirect effect of the first mediator variable and Table 5, it is slightly skewed to the right.

|  |  |
| --- | --- |
|  |  |

**Fig. 10** The Bootstrap Sample Distribution of the indirect effect of the first mediator variable (a) and indirect effect of the second mediator variable (b) of Adolescent Hate speech data

ii) For 95% bootstrap confidence interval for the indirect effect of the second mediator variable

95% Bootstrap CI for

Like the previous sample, the bootstrap sample distribution of the indirect effect of the second mediator variable shown in Figure 9 (b) is slightly skewed to the right.

**Table 6** Skewness about specific indirect effects of Adolescent Hate speech data

|  |  |
| --- | --- |
| Skewness | |
| Indirect effect1 | Indirect effect2 |
| -0.056 | -0.032 |

**Table 7** Confidence intervals about specific indirect effects of Adolescent Hate speech data

|  |  |  |  |
| --- | --- | --- | --- |
| Effect | Method | 95% CI  BL | 95% CI  BU |
| Indirect1 | CMA  Bootstrap  FMA  Bootstrap in FMA | 0.0275  0.0259  0.0840  0.0981 | 0.0475  0.0495  0.1543  0.1394 |
| Indirect2 | CMA  Bootstrap  FMA  Bootstrap in FMA | 0.0165  0.0155  0.0946  0.1069 | 0.0350  0.0354  0.1574  0.1454 |

Table 7 shows all of the specific indirect effects were found to be positively significant. As a result of data analysis, you can see the difference of confidence intervals between CMA and FMA because direct effect is greater than crisp data. Furthermore, since the values of skewness of the indirect effects were negative, the bootstrap sample distributions were slightly skewed to the right. Therefore, based on the team data, it is recommended to use the bootstrap method because the distribution is not symmetrical. The analysis revealed that exposure to hate speech significantly affects perceived social harm through emotional response.  
**For public policy managers or educators,** this indicates that interventions should focus not only on suppressing hate content but also on mediating emotional responses and promoting coping mechanisms.

**5.3 Fuzzy moderated-mediation Analysis for Solar Power Data**

Research on solar power generation using weather data has been proposed by many authors (Sun et al. 2020; Chuluunsaikhan et al. 2021). The Solar dataset contains weather status and daily power production data from solar panels, which were combined by Rami Mashkouk. The data used in this study ranges from 2016 to 2019. The variable "Temp" () represents the daily average temperature. The variable "Day Power" () represents the amount of power generated by solar panels in a day. "Humidity" () indicates the daily average humidity, and "Sky Cover" () represents the amount of cloud cover at 3 p.m. This variable is divided into eight categories and is coded from 1 to 8, with higher values indicating greater cloud cover or fog. As temperature and humidity are continuous variables, expressing them as crisp data may result in a loss of information. Therefore, we fuzzified these variables with a spread defined as half the difference between the values of the day and the next day. "Sky Cover" is a linguistic representation of the sky state at 3 p.m. by the observer, and we deemed it reasonable to fuzzify this variable with Spread 1 instead of representing it with crisp data. We normalized all variables except for "Sky Cover" to a range of 0 to 1. The model is described in Figure 13, and the results are presented below.

텍스트이(가) 표시된 사진

자동 생성된 설명

**텍스트, 폰트, 도표, 라인이(가) 표시된 사진

자동 생성된 설명**

**Fig. 11** The fuzzy mediation analysis of the Solar data

Here, Indirect effect is the product of the coefficient of and .

Indirect effect:

Additionally, interaction coefficient is .

*5.3.1 Statistical Inference on the indirect effect and interaction coefficient*.

i) For 95% confidence interval for the indirect effect in FMA,

95% CI for

ii) For 95% confidence interval for the interaction coefficientin FMA,

95% CI for ,

Where = = 0.0442

*5.3.2 Bootstrap Confidence Interval on Indirect effect and interaction coefficient*

i) For 95% bootstrap confidence interval for the indirect effect

95% Bootstrap CI for

The Bootstrap Sample Distribution is shown in Figure 11 (a). Looking at the bootstrap sample distribution of the estimate of the indirect effect and Table 7, it is slightly skewed to the left.

|  |  |
| --- | --- |
|  |  |

**Fig. 12** The Bootstrap Sample Distribution of the indirect effect (a) and of interaction coefficient (b) for the Solar data

ii) For 95% bootstrap confidence interval for interaction coefficient

95% Bootstrap CI for

Looking at the bootstrap sample distribution of the estimate of interaction coefficient shown in Figure 11 (a) and Table 7, it is slightly skewed to the right.

**Table 8** Skewness of indirect effect and interaction coefficient

|  |  |
| --- | --- |
| Skewness | |
| Indirect effect | Interaction coefficient |
| 0.016 | -0.004 |

**Table 9** Confidence intervals about Indirect effect and interaction coefficient

|  |  |  |  |
| --- | --- | --- | --- |
| Effect | Method | 95% CI  BL | 95% CI  BU |
| Indirect | CMA  Bootstrap  FMA  Bootstrap in FMA | 0.318  0.312  0.299  0.272 | 0.400  0.408  0.329  0.356 |
| Interaction  coefficient | CMA  Bootstrap  FMA  Bootstrap in FMA | 0.0031  0.0027  0.0354  0.0426 | 0.0648  0.0651  0.0746  0.0672 |

Table 9 shows the indirect effect is found to be positively significant. This indicates that an increase in temperature compensates for a decrease in day power due to humidity. Also, interaction coefficient is positively significant, suggesting that an increase in temperature compensates for a decrease in day power caused by sky cover. The use of fuzzy data has provided more accurate analysis results compared to the use of crisp data. Additionally, the direct effect is found to be larger in FMA than in CMA, indicating that temperature has a greater impact on day power in FMA. Furthermore, the coefficients of other variables have slightly decreased when compared to the crisp data, implying that the use of crisp data can result in slightly understated or exaggerated results. Finally, it is recommended to use the bootstrap method due to the asymmetric distribution of the bootstrap sample.

**6 Data Analysis: Comparison of Performances based on various optimization methods**

Previously, we conducted a causal analysis to find out the influence of variables on the dependent variable. The regression coefficient, estimated using the previous FLSE method through a closed-form mathematical solution (CS) (Yoon and Choi 2013) (Yoon and Choi 2013), is considered accurate. Nevertheless, we would like to do a further causal analysis to compare accuracies using various algorithms in this section. We conducted analysis using Team Performance data and Solar data used from Section 5. We applied Neural-Networks to estimate the regression coefficient and we utilized several optimization methods that based on gradient descent. At this point, we used FMSE, also called L2 Loss as the cost function. This method can alter the estimates of regression coefficients by adjusting hyperparameters like epochs and learning rates. We trained the regression model for 800 epochs with a learning rate of , applying optimization algorithms. We also utilized FLAD, also called L1 Loss, as the cost function. However, methods that based on gradient descent can only be used for differentiable functions. For this reason, we used both GA and HS to address both constrained and unconstrained optimization problems. Furthermore, we used bootstrapping, a commonly used data sampling technique, to increase data when there is little data. The bootstrap method is particularly advantageous when the dataset size is limited. Acquiring new data demands both time and resources. So, we generated 5000 bootstrap samples for the regression coefficient using the bootstrap method, and then estimated regression coefficients by calculating the mean of bootstrap samples. The results for each dataset are as follows.

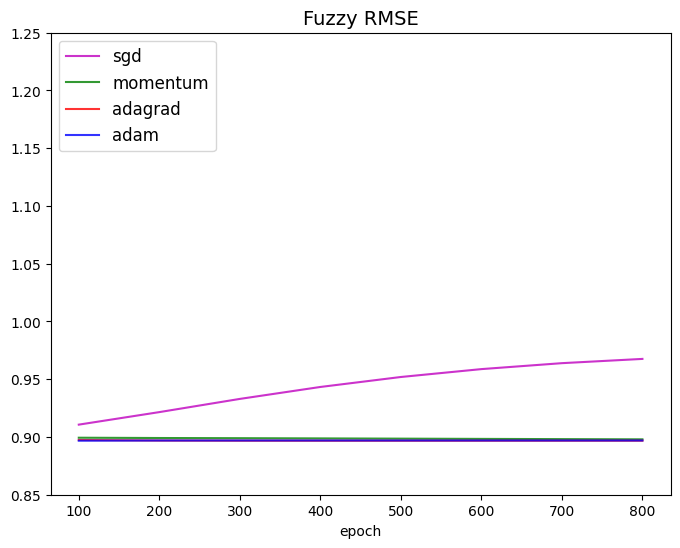
**6.1 Team Data**

In Table 10, differences in FRMSE and FMAE can be observed among the closed-form mathematical solution (CS) (Yoon and Choi 2013), Neural-Networks using four optimization methods based on gradient descent, GA and HS. While examining FRMSE, it was found that in some methods, the lowest value was . In the case of FMAE, the lowest value, , was observed for the SGD. This discrepancy arises from the inherent structural distinctions among the optimization methods. The closed-form mathematical solution (CS) was calculated by utilizing a closed-form formula. In other words, the solution obtained through this is the true solution. On the contrary, optimization method is a technique that updates the model to minimize the cost function. Thus, in the basic regression model, the closed-form mathematical solution

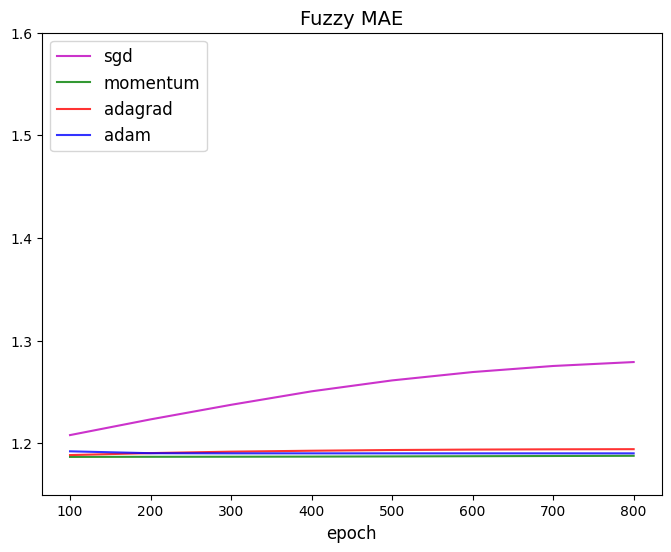
**Table 10** FRMSE, FMAE to optimization methods using Team date

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Measurements | FLSE |  | |  |  | |  |  |  | FLAD | |
| Closed-form Solution (CS) | Evolutionary Algorithms | | | Neural Networks | | | | | Evolutionary Algorithms | |
| GA | HS | | SGD | MOMENTUM | | ADAGRAD | ADAM | GA | HS |
| FRMSE  FRMSE (bootstrap)  FMAE  FMAE (bootstrap) | 0.8968 | 0.8968 | 0.8968 | | 0.8976 | 0.8968 | | 0.8974 | 0.8968 | 0.8976 | 0.8978 |
| 0.8968 | 0.8970 | 0.8968 | | 0.9676 | 0.8980 | | 0.8968 | 0.8970 | 0.8977 | 0.8978 |
| 1.1932 | 1.1919 | 1.1931 | | 1.1885 | 1.1921 | | 1.2001 | 1.1921 | 1.1932 | 1.1941 |
| 1.1914 | 1.1903 | 1.1931 | | 1.2792 | 1.1879 | | 1.1944 | 1.1903 | 1.1925 | 1.1941 |

(CS) method yields the lowest values for FRMSE and FMAE. However, the Team data model is a mediation model and is a composite model, so the results may be different. Furthermore, the results of the analysis using the bootstrap method exhibited higher or lower values compared to those obtained using the original data. While examining FRMSE (bootstrap), it was observed that in some methods, the lowest value was . In the case of FMAE (bootstrap), the lowest value, , was observed for the MOMENTUM. You can see that the results using bootstrap are slightly different from using the original data. This discrepancy arises because, unlike the process of finding optimal solution in the original data, the bootstrap method utilizes optimized results derived from bootstrap samples based on the original data. Additionally, you can see variation of FRMSE and FMAE using bootstrap for neural network models according to epoch in Figure 12, 13. In general, the figure decreases as epoch increases, but because mediation model is a composite model, there is a slight difference in the results. Consequently, more important thing than analytical figures is that the bootstrap method mitigates parameter estimation instability with limited data. Therefore, we recommend the use of the bootstrap method for these reasons.



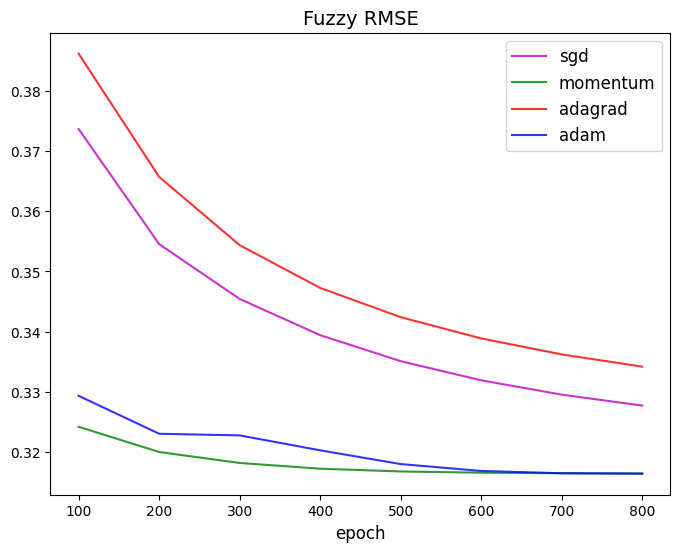
**Fig. 13** Variation of FRMSE of Team data using bootstrap for neural network models according to epoch



**Fig. 14** Variation of FMAE of Team data using bootstrap for neural network models according to epoch

**6.2 Solar Power Data**

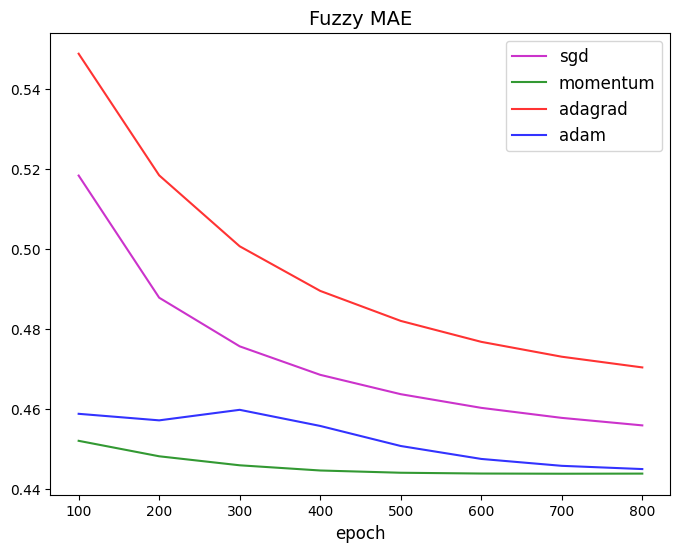
In Table 11, it was found that in some methods, the lowest value of FRMSE was . In the case of FMAE, the lowest value, , was observed for the GA(L1). As mentioned in Team data, the Solar data model is a moderated-mediation model and is a composite model, so the results may be different. In the case of FRMSE (bootstrap), it was found that the lowest value , was obtained in ADAM. In the case of FMAE (bootstrap), the lowest value, , was observed for the GA(L1). As mentioned before, the results obtained using bootstrap are slightly different from those obtained using the original data. Additionally, variation of FRMSE and FMAE using bootstrap for neural network models according to epoch described in Figure 14, 15. You can see that the performance of the ADAM and MOMENTUM methods was better than other methods.



**Fig. 15** Variation of FRMSE of Solar data using bootstrap for neural network models according to epoch

**Table 11** FRMSE, FMAE to optimization methods using Solar data

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Measurements | FLSE |  | |  |  | |  |  |  | FLAD | |
| Closed-form Solution (CS) | Evolutionary Algorithms | | | Neural Networks | | | | | Evolutionary Algorithms | |
| GA | HS | | SGD | MOMENTUM | | ADAGRAD | ADAM | GA | HS |
| FRMSE  FRMSE (bootstrap)  FMAE  FMAE (bootstrap) | 0.3252 | 0.3164 | 0.3200 | | 0.3277 | 0.3164 | | 0.3341 | 0.3164 | 0.3189 | 0.3220 |
| 0.3195 | 0.3166 | 0.3200 | | 0.3277 | 0.3165 | | 0.3342 | 0.3164 | 0.3186 | 0.3220 |
| 0.4499 | 0.4442 | 0.4490 | | 0.4558 | 0.4437 | | 0.4703 | 0.4449 | 0.4417 | 0.4469 |
| 0.4486 | 0.4442 | 0.4490 | | 0.4558 | 0.4438 | | 0.4703 | 0.4449 | 0.4416 | 0.4469 |

****

**Fig. 16** Variation of FMAE of Solar data using bootstrap for neural network models according to epoch

7. ****Practical Applicability and Engineering Use Cases****

Through the case study based on solar power generation data, it shows how our model can be used to estimate the impact of meteorological uncertainty (temperature, humidity, sky cover) on energy output. This reflects a real-life scenario for companies in renewable energy sectors.

In addition, we describe implementation strategies using Python and publicly available libraries (e.g., scikit-fuzzy, DEAP, and our own scripts), emphasizing that our model is modular, interpretable, and feasible to deploy, especially with modern computing resources.

Engineering Case Example: Solar Power Forecasting

In renewable energy engineering, predicting daily solar power output under uncertain weather conditions is a critical challenge. Traditional models often fail to incorporate ambiguity in linguistic weather descriptions such as “partly cloudy” or “moderately humid.”

Our fuzzy mediation model addresses this gap by treating weather inputs as fuzzy variables and quantifying both direct and indirect effects (e.g., temperature → power, moderated by sky condition).

This allows solar energy companies to:

* Estimate power output more accurately under ambiguous inputs
* Make risk-aware scheduling and load balancing decisions
* Optimize operational efficiency under uncertainty
* Ease of Implementation

The model is implementable using Python-based optimization libraries. For instance:

FLSE can be solved via matrix algebra (NumPy)

FLAD optimization is done using Genetic Algorithm or Harmony Search (DEAP, PyHarmony)

8. Conclusions

This study proposed a novel approach for fuzzy mediation and moderated-mediation analysis by integrating bootstrapping with fuzzy regression models—Fuzzy Least Squares Estimation (FLSE) and Fuzzy Least Absolute Deviations (FLAD). The use of FLSE, with its closed-form solution, provided globally optimal coefficients, while FLAD employed evolutionary algorithms (GA and HS) to overcome the limitations of non-differentiability. We applied these methods to three real-world datasets (team dynamics, adolescent hate speech, and solar energy generation) to empirically validate the framework. Across these cases, the bootstrap-enhanced fuzzy models produced more robust, interpretable, and statistically reliable confidence intervals than traditional methods like Sobel. Particularly in small or skewed samples, our approach significantly improved estimation stability and captured uncertainty more effectively. Future work could explore several directions. First, expanding the analysis to larger datasets and additional domains—such as healthcare, finance, or climate science—could help confirm the model’s applicability. Second, incorporating alternative fuzzy number representations and multi-level or hierarchical fuzzy models could enrich the analytical framework. Third, optimizing computational efficiency by integrating parallel processing or hybrid metaheuristics could mitigate algorithmic complexity. Lastly, deeper integration with deep learning and interpretable AI models could unlock new possibilities for fuzzy mediation in data-rich but ambiguous environments.

**Funding** This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No.2019M3F2A1073179; RS-2024-00351610).

**Data availability**

Team Performance data will be made available on request.

Adolescent Hate speech data is available at <https://www.nypi.re.kr/archive/mps/program/examinDataCode/view?menuId=MENU00226&pageNum=1&titleId=157&schType=0&schText=&firstCategory=&secondCategory=>

Weather data set of Solar data is available at <https://www.kaggle.com/datasets/ramima/weather-dataset-in-antwerp-belgium>

Day power data of Solar data is available at <https://www.kaggle.com/datasets/fvcoppen/solarpanelspower>

**Declarations**

**Conflict of interest** The authors declare no competing interest.

**Ethical approval** Not applicable.

References

Baron R.M., Kenny D.A.: The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. J Pers Soc Psychol. 51(6), 1173-82 (1986)

Biesanz, Jeremy C., Carl F. Falk, and Victoria Savalei.: Assessing mediational models: Testing and interval estimation for indirect effects. Multivariate Behavioral Research 45.4, 661-701 (2010)

Bollen, Kenneth A., and Robert A. Stine.: Bootstrapping goodness-of-fit measures in structural equation models. Sociological Methods & Research 21.2, 205-229 (1992)

Bollen, Kenneth A., and Robert Stine.: Direct and indirect effects: Classical and bootstrap estimates of variability. Sociological methodology, 115-140 (1990)

Chuluunsaikhan, T., Nasridinov, A., Choi, W.S., Choi, D.B., Choi, S.H., Kim, Y.M.: Predicting the Power Output of Solar Panels based on Weather and Air Pollution Features using Machine Learning. Journal of Korea Multimedia Society, 24(2), 222-232 (2021)

Cole, M.S., Walter, F., Bruch, H.: Affective mechanisms linking dysfunctional behavior to performance in work teams: a moderated mediation study. J. Appl. Psychol. 93, 945–958 (2008)

Davison, Anthony Christopher, and David Victor Hinkley.:  Bootstrap methods and their application. No. 1. Cambridge university press (1997).

DiCiccio, Thomas J., and Bradley Efron.: Bootstrap confidence intervals. Statistical science, 11.3 189-212 (1996)

Diefendorff, J.M., Mehta, K.: The relations of motivational traits with workplace deviance. J. Appl. Psychol. 92, 967–977 (2007)

Duffy, M.K., Ganster, D.C., Shaw, J.D., Johnson, J.L., Pagon, M.: The social context of undermining behavior at work. Organ. Behav. Hum. Decis. Process. 101, 105–126 (2006)

Edwards, J. R., & Lambert, L. S.: Methods for integrating moderation and mediation: a general analytical framework using moderated path analysis. Psychological methods, 12(1), 1 (2007).

Efron, B.: “Bootstrap methods: another look at the jackknife annals of statistics 7: 1–26." View Article PubMed/NCBI Google Scholar 24 (1979).

Efron, Bradley, and Robert Tibshirani.: Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Statistical science, 54-75 (1986)

Fairchild, A. J., & MacKinnon, D. P.: A general model for testing mediation and moderation effects. Prevention science, 10, 87-99 (2009).

Falk, C.F., Vogel, T.A., Hammami, S., Miočević, M.: Multilevel mediation analysis in R: A comparison of bootstrap and Bayesian approaches. Behav. Res. Methods 56, 750–764 (2024). https://doi.org/10.3758/s13428-023-02192-5

Felps, W., Mitchell, T.R., Byington, E.: How, when, and why bad apples spoil the barrel: negative group members and dysfunctional groups. In: Staw, B.M. (ed.) Research in Organizational Behavior, vol. 27, pp. 175–222. Elsevier, Amsterdam (2006)

Francis, S., Rangasamy, S., Balasundaram, S., Iyer, L.S., Elangovan, N.: Corporate social responsibility and its impact on organizational attractiveness: Unveiling the mediating role of perceived organizational support. Cogent Bus. Manag. 11, (2024). https://doi.org/10.1080/23311975.2024.2349269

Fritz, Matthew S., and David P. MacKinnon.: Required sample size to detect the mediated effect.. Psychological science 18.3, 233-239 (2007).

Galvao, A.F., Parker, T., Xiao, Z.: Bootstrap inference for panel data quantile regression. J. Bus. Econ. Stat. 42, 628–639 (2024). <https://doi.org/10.1080/07350015.2023.2210189>

Geem, Z. W., Kim, J. H., & Loganathan, G. V.: A new heuristic optimization algorithm: harmony search. simulation, 76(2), 60-68 (2001)

Groeneboom, P., Jongbloed, G.: Confidence intervals in monotone regression. Scand. J. Stat. 51, 1749–1781 (2024). <https://doi.org/10.1111/sjos.12730>

Hayes and Schaarkow.: Partial and Conditional Effects in Moderation and Mediation Analysis: Bootstrap Experiments. Journal of Rationale, Vol. 9, No. 2, pp. 1-24, (2013)

Hayes, A. F., & Scharkow, M.: The relative trustworthiness of inferential tests of the indirect effect in statistical mediation analysis: does method really matter?. Psychological science, 24(10), 1918-1927 (2013).

Hayes, A. F.: An index and test of linear moderated mediation. Multivariate behavioral research, 50(1), 1-22 (2015).

Hayes, A. F.: Introduction to mediation, moderation, and conditional process analysis: A regression-based approach (2017).

Hayes, Andrew F.: Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. Communication monographs 76.4, 408-420 (2009)

He, Y., Song, P.X.K., Xu, G.: Adaptive bootstrap tests for composite null hypotheses in the mediation pathway analysis. J. R. Stat. Soc. Ser. B Stat. Methodol. 86, 411–434 (2024). <https://doi.org/10.1093/jrsssb/qkad129>

James, L. R., & Brett, J. M.: Mediators, moderators, and tests for mediation. Journal of applied psychology, 69(2), 307 (1984).

Judd, Charles M., and David A. Kenny.: Process analysis: Estimating mediation in treatment evaluations. Evaluation review 5.5, 602-619 (1981)

Kansok-Dusche, J., Ballaschk, C., Krause, N., Zeißig, A., Seemann-Herz, L., Wachs, S., & Bilz, L.: A Systematic Review on Hate Speech among Children and Adolescents: Definitions, Prevalence, and Overlap with Related Phenomena. Trauma, Violence, & Abuse, 0(0) (2022)

Kim, D. J., Koo, Y. Y., Yoon, J. H.: Fuzzy mediation analysis for KOSPI-related variables. *Int. J. Fuzzy Log. Intell. Syst.* 21(1), 49–56 (2021)

Kim, Y.H., Yee, Y.J., Jo, A.M., Lim, S.T.: A Study on How Seriously Adolescents are Exposed to Hate Speech and Solutions to Address the Issue. National Youth Policy Institute of Korea (2020).

Lee, H. E.: Review of methods for testing mediating effects in recent HRD research. The Korean Journal of Human Resource Development Quarterly, 16(2), 225-249 (2014).

MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S.: Mediation analysis. Annu. Rev. Psychol. 58, 593-614 (2007)

MacKinnon, D. P.: Introduction to statistical mediation analysis. Routledge (2012).

MacKinnon, David P., et al.: A comparison of methods to test mediation and other intervening variable effects. Psychological methods 7.1, 83-104 (2002)

Morgan-Lopez, A. A., & MacKinnon, D. P.: Demonstration and evaluation of a method for assessing mediated moderation. Behavior research methods, 38(1), 77-87 (2006).

Preacher, Kristopher J., and Andrew F. Hayes.: Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. Behavior research methods 40.3, 879-891 (2008)

Preacher, Kristopher J., and Andrew F. Hayes.: SPSS and SAS procedures for estimating indirect effects in simple mediation models. Behavior research methods, instruments, & computers 36, 717-731 (2004)

Robinson, S.L., O’Leary-Kelly, A.M.: Monkey see, monkey do: the influence of work groups on the antisocial behavior of employees. Acad. Manag. J. 41, 658–672 (1998)

Sampson, J. R.:Adaptation in natural and artificial systems (John H. Holland). (1976)

Shrout, Patrick E., and Niall Bolger.: Mediation in experimental and nonexperimental studies: new procedures and recommendations. Psychological methods, 7.4, 422 (2002)

Sobel, M.E.: Asymptotic Confidence Intervals for Indirect Effects in Structural Equation Models. Sociological Methodology. **13**, 290-312 (1982)

Sun, M., Feng, C., Zhang, J.: Probabilistic solar power forecasting based on weather scenario generation. Applied Energy, 266 (2020)

Tibshirani, Robert J., and Bradley Efron.: An introduction to the bootstrap.  Monographs on statistics and applied probability 57.1 (1993).

Wywiał, J., Szkutnik, T.: Bootstrap tests for unbiasedness of predictors. Commun. Stat. Simul. Comput. (2024). <https://doi.org/10.1080/03610918.2023.2301048>

Yoon, J. H., Choi, S.H., Grzegorzewski, P.: On asymptotic properties of the multiple fuzzy least squares estimator. Soft Methods for Data Science. **456**, 525-532 (2016)

Yoon, J. H., Choi, S.H.: Fuzzy least squares estimation with new fuzzy operations, Advances in Intelligent Systems and Computing. **190**, 193-202 (2013)

Yoon, J. H., Grzegorzewski, P.: On optimal and asymptotic properties of a fuzzy L2 estimator. Mathematics. **8**, 1956 (2020)

Yoon, J. H.: Fuzzy mediation analysis. Int. J. Fuzzy Syst. **22**(1), 338-349 (2020)

Yoon, J. H.: Fuzzy Moderation and Moderated-Mediation Analysis. Int. J. Fuzzy Syst. **22**(1), 338-349 (2020)

Zadeh, L.A.: Fuzzy sets. Information and control. 8, 338-353 (1965)

인간의 얼굴, 사람, 입술, 인물사진이(가) 표시된 사진

AI가 생성한 콘텐츠는 부정확할 수 있습니다.**Da Jeong Kang** received her B.S. degree in Mathematics from Sejong University, South Korea. She is currently participating in the Undergraduate Research Programme at the School of Mathematics and Computing (Computational Science and Engineering), Yonsei University, South Korea. Her research interests include artificial intelligence, electroencephalography (EEG), inverse problems, and neuroimaging.

**인간의 얼굴, 사람, 턱, 목이(가) 표시된 사진

AI가 생성한 콘텐츠는 부정확할 수 있습니다.Hyeon Gu Kang** received B.S. degree in Mathematics and Statistics from Sejong University, South Korea. He is currently an M.S. student in Computer Science and Engineering at Korea University, South Korea. His research interests include natural language processing and continual learning.

**Zong Woo Geem** received B.Eng. from Chung-Ang University, M.Sc. from Johns Hopkins University, and Ph.D. from Korea University. He researched at Virginia Tech, University of Maryland-College Park, and Johns Hopkins University. He is currently a Professor with the College of IT Convergence, Gachon University, South Korea. He has invented a music-inspired optimization algorithm, harmony search, which has been applied to various scientific and engineering problems. His research interests include phenomenon-mimicking algorithms and their sustainable applications to energy, environment, water, and infrastructure. He has served for various journals in editing (Senior Editor for Engineering Optimization and Guest Editor for Swarm and Evolutionary Computation, International Journal of Bio-Inspired Computation, Journal of Applied Mathematics, Applied Sciences, Complexity, and Sustainability).

사람, 인간의 얼굴, 턱, 목이(가) 표시된 사진

자동 생성된 설명

**Geon Hee Lee** received the B.S. degree in electrical engineering from Gachon University, Seongnam, South Korea, in 2021, and the M.S. degree in Energy IT from Gachon University, Seongnam, South Korea, in 2023. He is currently a Ph.D. student at the Gachon University, Seongnam, South Korea. His research interests include optimization and artificial intelligence.

**A person in a suit and tie

AI-generated content may be incorrect.Sung Wook Baik** is a full professor in the Department of Software at Sejong University and the head of the Intelligent Media Laboratory (IMLab). He was recognized as a Highly Cited Researcher by Clarivate Web of Science in 2024. His research interests include image processing, video analytics, multimedia data processing, energy informatics, Internet of Things (IoT) and related areas. He has published over 200 papers in peer-reviewed international journals, including the top journals in the field such as IEEE Transactions on Industrial Informatics, IEEE Internet of Things Journal, IEEE Transactions on Image Processing, ACM Computing Surveys, IEEE Transactions on Big Data, IEEE Transactions on Industrial Electronics, Information Fusion, Pattern Recognition, and others.

**Jin Hee Yoon** received her B.S., M.S., and Ph.D. degrees in Mathematics from Yonsei University, South Korea. She is currently a faculty member in the Department of Mathematics and Statistics at Sejong University, Seoul, South Korea. Her research interests include soft computing, fuzzy data analysis, intelligent systems, and machine learning. She serves as a board member of the Korean Institute of Intelligent Systems (KIIS) and has been working as an associate editor for IEEE Transactions on Fuzzy Systems (TFS), Soft Computing, and International Journal of Fuzzy Logic and Intelligent Systems (IJFIS). Additionally, she is the co-representative of Korea for the International Fuzzy Systems Association (IFSA) and serves as the vice chair of communications and education for the IEEE Fuzzy Systems Technical Committee.