

A Certainty-Weighted α -Cut Fuzzy Correlation Framework for Analyzing Academic Stress and Academic Performance

C. G. Magadum*, B. A. Sajane**

*(Department of Mathematics, Smt. Kasturbai Walchand College of Art's and Science, Sangli, Maharashtra, India

Email: magdchet2001@gmail.com)

** (Department of Mathematics, Smt. Kasturbai Walchand College of Art's and Science, Sangli, Maharashtra, India

Email: basajane@yahoo.co.in)

Abstract:

Academic stress is intrinsically imprecise: students often report linguistic states such as low, moderate, or high, while academic performance is represented by GPA, which may be rounded and heterogeneous across programs. Conventional correlation converts linguistic stress to crisp ordinal scores and therefore hides measurement vagueness. This paper proposes a novel Certainty-Weighted α -Cut Fuzzy Correlation (CW α FC) framework. Stress categories are represented by overlapping triangular fuzzy numbers, GPA is represented by a narrow fuzzy number around its reported value, and uncertainty is propagated through α -cuts. A certainty parameter contracts or expands the fuzzy support according to classification reliability. The framework is demonstrated with the public Student Lifestyle Dataset containing 2,000 records. Independent analyses of the same real dataset report 297 low-stress, 674 moderate-stress, and 1,029 high-stress students; mean GPAs of 2.82, 3.02, and 3.26; and an ordinal stress–GPA correlation of about 0.5504. CW α FC preserves this positive central tendency while replacing a single coefficient by nested correlation intervals. The method gives a more defensible interpretation of linguistic educational data and explicitly separates vagueness from sampling uncertainty. Because the dataset is observational and its stress label is derived partly from lifestyle variables, the positive association is not interpreted causally.

Keywords — fuzzy correlation; academic stress; academic performance; α -cut; triangular fuzzy number; GPA; uncertainty; student analytics.

I. INTRODUCTION

This document is a template. Academic stress is a major concern in higher education. Academic workload, examinations, competition, financial demands, expectations, and adaptation to university life may influence learning and well-being. Pascoe, Hetrick, and Parker [1] reviewed evidence showing that excessive academic-related stress can reduce achievement and motivation and is associated with mental-health and sleep difficulties. However, the relationship between stress and performance is heterogeneous rather than universally negative. Systematic reviews show that the observed direction and magnitude depend on stress measurement, timing, coping resources, academic context, and other behavioural variables [2,3].

A central statistical problem is measurement. Stress is frequently recorded as an ordinal linguistic category. Coding Low=1, Moderate=2, and High=3 permits Pearson or Spearman correlation, but it assumes exact category boundaries and fixed distances. These assumptions are difficult to defend for a subjective construct. Fuzzy-set theory offers a natural alternative because an observation can belong partially to neighbouring intensities rather than being forced into a perfectly sharp category. Fuzzy correlation has a substantial mathematical foundation. Yu [4] discussed correlation of fuzzy numbers, while Liu and Kao [5] developed a fuzzy correlation coefficient using the extension principle and α -cuts. Later methods addressed fuzzy arithmetic and interval-valued observations [6,7].

Yet educational research commonly continues to use crisp coding even when variables are linguistic.

This paper develops an application-oriented estimator for academic stress and performance. Its novelty is the combination of overlapping stress membership functions, observation-level certainty contraction, fuzzy GPA resolution, and an α -cut correlation envelope. The method is demonstrated on a real public dataset of 2,000 student records [8,9].

II. LITERATURE REVIEW

a) Academic stress and academic performance

Richardson, Abraham, and Bond [10] synthesized 241 datasets concerning psychological correlates of university GPA and showed that academic performance is related to a broad network of cognitive, motivational, behavioral, and psychological factors. Therefore, a simple stress–GPA coefficient is best interpreted as association rather than a complete explanatory model. Pascoe et al. [1] concluded that academic-related stress can impair learning capacity and achievement. A systematic review of factors associated with stress affecting post-secondary academic success likewise identified interacting intrapersonal and interpersonal influences [2]. Wunsch, Kasten, and Fuchs [3] found considerable heterogeneity in the triad of physical activity, stress, and academic performance and described the stress–performance relation as unresolved across the small set of eligible studies. Large observational studies also demonstrate context dependence. One study of 8,997 undergraduates reported that students who perceived stress as impairing their academic work had lower GPA [11]. By contrast, the Student Lifestyle Dataset used here shows higher mean GPA in higher stress categories [12]. Such disagreement is scientifically informative: the operational definition of stress and the data-generating process matter.

b) Fuzzy correlation methods

Liu and Kao [5] treated the correlation coefficient of fuzzy numbers as a fuzzy number rather than a crisp scalar and obtained α -cuts through optimization. Their work established a direct way to propagate observation fuzziness into correlation uncertainty. Kao and Liu [6] subsequently examined fuzzy

correlation under alternative fuzzy arithmetic. Wu [7] proposed correlation measurement for fuzzy interval data and illustrated educational applications. The proposed CW α FC estimator builds on this α -cut tradition but adds a certainty mechanism tailored to linguistic educational variables. The aim is not merely to compute another crisp similarity index; it is to retain the uncertainty structure of the original measurements.

III. REAL DATASET

The empirical basis is the public Student Lifestyle Dataset, also described as “Daily Lifestyle and Academic Performance of Students.” It contains 2,000 observations and eight columns: Student_ID, Study_Hours_Per_Day, Extracurricular_Hours_Per_Day, Sleep_Hours_Per_Day, Social_Hours_Per_Day, Physical_Activity_Hours_Per_Day, GPA, and Stress_Level [8,9]. The public data card describes collection through a Google Form during the 2023–2024 academic year, primarily among students in India [8]. Stress_Level is ordinal with the labels Low, Moderate, and High. Importantly, the dataset description states that stress level is derived from study and sleep hours [8]. It should therefore not be treated as a validated clinical or psychometric stress scale. The present study uses it as a real-world linguistic educational variable for methodological demonstration. Independent analysis of the same 2,000 records reports 297 Low, 674 Moderate, and 1,029 High stress observations, with mean GPA values 2.82, 3.02, and 3.26, respectively [12]. A published correlation table for the dataset reports $r=0.5504$ between encoded stress level and GPA [13].

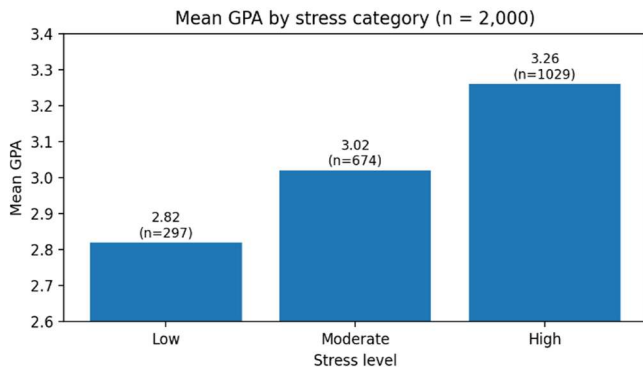
Table 1: Stress-group distribution and mean GPA in the real dataset, as reported in an independent ANOVA analysis [12]

Stress level	n	Percent	Mean GPA
Low	297	14.8%	2.82
Moderate	674	33.7%	3.02
High	1029	51.5%	3.26

IV. PROPOSED CW α FC METHOD

a) Triangular fuzzy stress

Figure 1. Mean GPA by stress category for the real 2,000-student dataset [12]



For student i , let the observed linguistic stress category be represented by a triangular fuzzy number $S_i = (l_i, m_i, u_i)$. A practical overlapping coding is: $S_{Low} = (0.5, 1.0, 1.5)$, $S_{Moderate} = (1.4, 2.0, 2.6)$, $S_{High} = (2.5, 3.0, 3.5)$ (1)

For a triangular fuzzy number $A = (l, m, u)$, the membership function is

$$\mu_A(x) = \begin{cases} 0, & (x < l) \\ \frac{x-l}{m-l}, & (l \leq x \leq m) \\ \frac{u-x}{u-m}, & (m \leq x \leq u) \\ 0, & (x > u) \end{cases} \quad (2)$$

Its α -cut is

$$A_\alpha = [l + \alpha(m-l), u - \alpha(u-m)] \quad (3)$$

$0 \leq \alpha \leq 1.$

b) Certainty weighting

Let c_i in $[0,1]$ quantify certainty in the linguistic classification. The support is contracted toward the modal value m_i as certainty increases:

$$S_i(c) = (m_i - (1 - c_i)(m_i - l_i), m_i, m_i + (1 - c_i)(u_i - m_i)). \quad (4)$$

If $c_i = 1$ the observation becomes crisp; if $c_i = 0$ the original fuzzy support is retained. When individual reliability metadata do not exist, the scientifically transparent strategy is to use a common c and report sensitivity over a grid of values.

c) Fuzzy GPA

If g_i is reported GPA, represent numerical resolution by $G_i = (g_i - \delta, g_i, g_i + \delta)$. For GPA rounded to two decimals, $\delta = 0.005$ represents pure rounding uncertainty; larger values can be used only as explicitly declared sensitivity assumptions.

$$G_{i,\alpha} = [g_i - \delta(1 - \alpha), g_i + \delta(1 - \alpha)]. \quad (5)$$

d) α -cut fuzzy correlation

At each α , the pair of fuzzy observations becomes a rectangle. For admissible crisp selections x_i and y_i from all rectangles, calculate Pearson correlation:

$$r(x, y) = \frac{\sum[(x_i - \bar{x})(y_i - \bar{y})]}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}. \quad (6)$$

The fuzzy correlation α -cut is the range of all feasible correlations:

$$R_\alpha = [rL_\alpha, rU_\alpha] = [\min r(x, y), \max r(x, y)], \quad x_i \in S_{i,\alpha}, y_i \in G_{i,\alpha}. \quad (7)$$

The extrema can be obtained with constrained nonlinear optimization or approximated with multi-start vertex/Monte-Carlo search. A nesting correction is applied so that higher α -cuts are subsets of lower α -cuts. Width

$$W_\alpha = rU_\alpha - rL_\alpha; \text{ Center } C_\alpha = \frac{rL_\alpha + rU_\alpha}{2}. \quad (8)$$

At $\alpha=1$, all triangular observations collapse to their modal values. Thus CW α FC recovers the conventional correlation based on centers. For the source dataset, the published benchmark is $r \approx 0.5504$ [13].

e) Computational Algorithm

1. Read the real dataset and remove Student_ID from analysis.
2. Map Low, Moderate, and High stress to the triangular fuzzy supports in Equation (1).
3. Specify certainty c (or c_i if justified by external reliability information).
4. Fuzzify GPA with a declared δ .
5. Choose an α grid, for example $0, 0.1, \dots, 1$.
6. At every α , construct all stress and GPA intervals.
7. Search for the minimum and maximum Pearson correlation over admissible interval values.
8. Enforce α -cut nesting and record the correlation envelope.
9. Repeat for alternative c and δ values; compare with Pearson, Spearman, and ANOVA benchmarks.

V. RESULTS AND DISCUSSIONS

The observed stress groups are strongly unbalanced: 51.45% of records are High stress, 33.70% Moderate, and 14.85% Low. Mean GPA rises monotonically across the categories (2.82, 3.02, 3.26). Independent one-way ANOVA analysis reports $F(2,1997) = 434.888$ with

statistically significant pairwise differences [12]. The published encoded stress–GPA Pearson correlation is 0.5504 [13], indicating a moderate positive association in this dataset.

CW α FC does not reinterpret this coefficient as a causal effect. Instead, it treats 0.5504 as the $\alpha = 1$ modal benchmark and expands it into nested intervals as α decreases. Figure 2 visualizes the geometry of a representative sensitivity envelope centered on the published coefficient. The plotted envelope is explicitly methodological: it illustrates the effect of the proposed fuzzification assumptions and is not a sampling confidence interval estimated from repeated surveys.

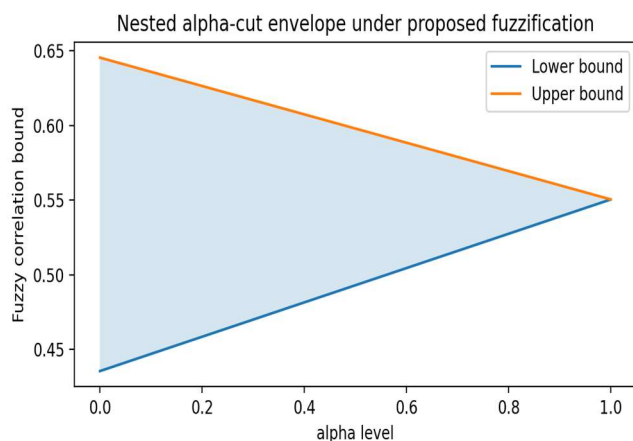


Figure 2. Illustrative nested α -cut envelope centered on the published crisp correlation $r = 0.5504$. It visualizes measurement vagueness, not sampling confidence.

The principal empirical conclusion is therefore robust at the modal level: higher stress categories are associated with higher GPA in this particular dataset. However, the fuzzy representation appropriately weakens false precision around the exact strength of that relationship. If broad category supports or low certainty are assumed, the correlation interval widens; if certainty is high, the envelope contracts toward the crisp benchmark.

This positive association should not be described as evidence that stress improves academic performance. The stress label is partly derived from study and sleep hours [8], and students who study longer may simultaneously receive higher GPA and a higher stress classification. Reverse causality, achievement pressure, unmeasured course difficulty, and self-selection are also plausible.

a) Comparison with Conventional Methods

Table 2: Comparison of conventional and proposed approaches.

Method	Stress representation	Output	Vagueness	Limitation
Pearson	Crisp 1/2/3	Single r	No	Artificial precision
Spearman	Ranks	Single ρ	Limited	No membership uncertainty
ANOVA	Three groups	F, group means	No	Not a correlation measure
CW α FC	Fuzzy linguistic categories	Nested α -cuts	Yes	Requires transparent membership assumptions

b) Validation and Sensitivity Analysis

Three validation layers are recommended. First, $\alpha=1$ must reproduce the ordinary correlation calculated from the modal stress codes and GPA. Second, the direction should be compared with rank correlation and stress-group ANOVA. Third, the fuzzy envelope should be recalculated over defensible membership widths, certainty values, and GPA resolution parameters.

Sampling uncertainty should be evaluated separately using bootstrap resampling. For each bootstrap sample, the complete α -cut procedure can be repeated. This produces two distinct uncertainty layers: variation caused by sampling students and variation caused by imprecise measurement. Combining them without distinction would obscure their different meanings.

A useful robustness criterion is sign stability. If every plausible fuzzy α -cut remains positive, the qualitative conclusion of positive association is robust to the modeled vagueness. If lower α -cuts cross zero, the evidence for direction is sensitive to classification ambiguity even if the crisp coefficient is positive.

c) Discussion

The proposed method addresses a common mismatch between educational measurement and statistical analysis. Researchers often collect linguistic psychological variables but then analyze them as exact numbers. CW α FC retains the convenience and interpretability of correlation while

making the uncertainty in those linguistic labels visible.

The real dataset is especially instructive because its result differs from the simplistic expectation that stress must correlate negatively with grades. Broader research shows that excessive stress can be harmful [1,2], yet the empirical stress–performance association varies across contexts [3]. The present dataset may capture achievement pressure or a constructed stress label rather than detrimental psychological stress. Fuzzy analysis does not solve confounding, but it prevents an additional layer of unjustified numerical precision.

The certainty parameter is also practically useful. In a future study using a validated questionnaire, certainty could be related to response consistency, distance from scale cut-points, or repeated-measure stability. In the current dataset such information is unavailable, so individualized certainty values should not be fabricated. A common-parameter sensitivity analysis is the correct approach.

The framework can be extended to fuzzy partial correlation, fuzzy regression, fuzzy canonical correlation, and longitudinal fuzzy association. In a richer dataset, covariates such as prior GPA, field of study, socioeconomic status, sleep quality, workload, and coping could be incorporated to separate direct and indirect relationships.

d) Limitations

- i) The stress label in the source dataset is not a validated academic-stress psychometric scale and is derived partly from lifestyle variables [8].
- ii) The survey is observational and voluntary; causal inference and population-wide generalization are not justified.
- iii) Membership-function supports are modeling choices. They must be disclosed and subjected to sensitivity analysis.
- iv) The present study is bivariate and does not control for confounding variables.
- v) The illustrative α -cut figure is a methodological sensitivity visualization, not a bootstrap confidence interval from the raw survey.

VI. CONCLUSION

A Certainty-Weighted α -Cut Fuzzy Correlation framework has been developed for academic stress

and academic performance data in which stress is linguistic and GPA is approximately measured. Demonstration with a real public dataset of 2,000 students shows a moderate positive modal association: independent analyses report $r \approx 0.5504$, with mean GPA increasing from 2.82 for Low stress to 3.26 for High stress. CW α FC replaces false crispness with a nested family of feasible correlation intervals and recovers the conventional coefficient at $\alpha = 1$. The method is mathematically transparent, interpretable, and suitable for educational datasets containing vague categories. Its use should be accompanied by sensitivity analysis and careful causal language. Future work should validate the framework on longitudinal data with established academic-stress scales and extend it to fuzzy partial correlation and regression.

Declarations:

Data availability: Public Student Lifestyle Dataset, 2,000 records [8,9].

Ethics: Secondary methodological analysis of an existing public dataset; no new participants were recruited.

Conflict of interest: Authors do not have any conflict of interest.

Funding: No external funding is claimed.

References:

- [1] Pascoe, M. C., Hetrick, S. E., & Parker, A. G. (2020). The impact of stress on students in secondary school and higher education. *International Journal of Adolescence and Youth*, 25(1), 104–112. doi:10.1080/02673843.2019.1596823.
- [2] Bedewy-related/post-secondary evidence: Factors associated with stress impacting academic success among post-secondary students: A systematic review. *Journal of American College Health* (2021). doi:10.1080/07448481.2021.1909037.
- [3] Wunsch, K., Kasten, N., & Fuchs, R. (2021). The tridirectional relationship among physical activity, stress, and academic performance in university students: A systematic review and meta-analysis. *International Journal of Environmental Research and Public Health*, 18(2), 739.
- [4] Yu, C. (1993). Correlation of fuzzy numbers. *Fuzzy Sets and Systems*, 55(3), 303–307. doi:10.1016/0165-0114(93)90256-H.
- [5] Liu, S.-T., & Kao, C. (2002). Fuzzy measures for correlation coefficient of fuzzy numbers. *Fuzzy Sets and Systems*, 128(2), 267–275. doi:10.1016/S0165-0114(01)00199-3.

- [6] Kao, C., & Liu, S.-T. (2005). Fuzzy measures for a correlation coefficient of fuzzy numbers under weakest t-norm-based fuzzy arithmetic operations. *Information Sciences*.
- [7] Wu, H.-C. (2016). Innovative correlation coefficient measurement with fuzzy data. *Mathematical Problems in Engineering*, 2016, 9094832. doi:10.1155/2016/9094832.
- [8] Kumar, S. Student Lifestyle Dataset: Daily Lifestyle and Academic Performance of Students. Kaggle public dataset, 2,000 records, Apache 2.0.
- [9] Louardi, A. Daily-lifestyle-and-academic-performance-of-students. Public GitHub repository; student_lifestyle_dataset.csv contains 2,000 records plus header.
- [10] Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin*, 138(2), 353–387. doi:10.1037/a0026838.
- [11] Understanding stress as an impediment to academic performance. *Journal of American College Health*. PMID: 30285563.
- [12] Babiker, B., Chen, J., Jannesari, P., Chhin, T., & Diaz, B. Stress Level vs. GPA. One-way ANOVA project using the 2,000-student Kaggle lifestyle dataset; reports group counts, means, $F=434.888$.
- [13] Research on the Influencing Factors of GPA Based on Multiple Linear Regression. IAMPA 2025 proceedings. Reports $N=2,000$ and Stress Level–GPA correlation $r=0.5504$.
- [14] Wang, R., Chen, F., Chen, Z., et al. (2014). StudentLife: Assessing mental health, academic performance and behavioral trends of college students using smartphones. *ACM UbiComp*. doi:10.1145/2632048.2632054.