Journal of Statistical Sciences, Autumn and Winter, 2025
Vol. 19, No. 2, pp 431-448
DOI: 10.61882/jss.19.2.09

Application of Fuzzy Statistics in Experimental Design Models

Kashani, M. [1], Ghasemi, R. [2]
[1]Department of Mathematics and Statistics, Gonbad Kavous University, Gonbad Kavous, Iran.
[2]Department of Statistics, Payam Noor University, Tehran, Iran.

Corresponding author: R. Ghasemi, r.ghasemi62@pnu.ac.ir

Received: 10/4/2025 **Revised:** 18/8/2025 **Accepted and Published Online:** 22/8/2025.

Introduction
In many scientific fields the design of experiments has long been a fundamental tool for analyzing the effects of controlled factors on response variables. Classical approaches, rooted in Fisher's work, typically assume normality of responses and rely on mean-based analyses such as analysis of variance. However, the presence of outliers, arising from measurement errors, invalid data, or unexpected variations, can severely distort results, challenging these assumptions and reducing model reliability. To address these limitations, fuzzy regression has emerged as a powerful alternative. Unlike classical models, it does not require strict distributional assumptions and instead employs fuzzy numbers to capture uncertainty and variability flexibly. This adaptability makes fuzzy regression particularly suitable for handling noisy or imprecise data, where traditional methods fall short. This paper explores the application of fuzzy regression as a robust extension to the design of experiments, demonstrating its potential to improve analysis in the presence of outliers and offering a viable alternative to classical approaches.

Materials and Methods
Two real-world datasets were employed: seedling growth measurements, which capture biological variation in agricultural experiments, and boiling point quality data, representing chemical and industrial contexts. The performance of the classical experimental design was systematically compared with five alternative methods designed to address the influence of outliers: Hu-

ber regression, bi-square regression, mean substitution, ranking, and fuzzy regression. Each method was implemented under identical experimental conditions and assessed using predictive accuracy and goodness-of-fit metrics.

Results and Discussion
The fuzzy regression consistently outperformed both the classical experimental design and the set of robust alternatives across all evaluation metrics. It achieved higher adjusted R^2 values and substantially lower MSPE, indicating superior predictive performance. Its advantage stems from the ability to explicitly manage uncertainty in the data and remain resilient against the distorting effects of anomalous observations. The empirical findings suggest that, in practice, fuzzy regression reduces sensitivity to data contamination while enhancing the interpretability of model outcomes. These results highlight the method's substantial potential for practical applications in domains where noisy or imprecise data are common.

Conclusion
Fuzzy regression is a robust and reliable alternative to classical and robust regression methods. Unlike traditional approaches, it does not rely on strict distributional assumptions, which makes it particularly suitable for real-world datasets often affected by uncertainty and irregularities. The method demonstrated improved predictive accuracy, effective handling of outliers, and stable model performance even under challenging data conditions. In addition to its practical benefits, fuzzy regression also provides a flexible framework for future methodological extensions, such as incorporating qualitative factors through fuzzy coding or integrating with hybrid computational approaches. These features underline its practical relevance and adaptability across a wide range of disciplines, from agricultural sciences and biological research to engineering applications.

Keywords: Design of Experiments, Fuzzy Regression, Outlier Data, Evaluation Criteria

Mathematics Subject Classification (2020): 62A86, 62K05.

مجله علوم آماری، پاییز و زمستان ۱۴۰۴
جلد ۱۹، شماره ۲، ص ۴۳۱ -- ۴۴۸
DOI: 10.61882/jss.19.2.09
مقاله پژوهشی

کاربرد آمار فازی در طرح آزمایش

مجتبی کاشانی[۱]، رضا قاسمی نجف آبادی[۲]

۱گروه ریاضی و آمار، دانشکده علوم پایه و فنی و مهندسی، دانشگاه گنبد کاووس
۲گروه آمار، دانشگاه پیام نور، تهران

نویسنده مسئول : رضا قاسمی نجف آبادی، r.ghasemi62@pnu.ac.ir

تاریخ دریافت: ۱۴۰۴/۱/۲۱ تاریخ بازنگری: ۱۴۰۴/۵/۲۷ تاریخ پذیرش و انتشار: ۱۴۰۴/۵/۳۱

چکیده: در پژوهش‌های آماری، طرح‌های آزمایش، برای بررسی اثر متغیرهای کنترل بر پاسخ‌های خروجی به‌کار می‌روند. این روش‌ها مبتنی بر فرض نرمال‌بودن توزیع داده‌ها بوده و در مواجهه با نقاط دورافتاده با چالش‌های اساسی مواجه می‌شوند. مطالعه حاضر به مقایسه پنج رویکرد مختلف علاوه بر روش کلاسیک طرح آزمایش برای مقابله با این چالش می‌پردازد: روش‌های مقاوم‌سازی شامل هوبر، دو مربعی، میانگین جایگزین، رتبه‌بندی و رگرسیون فازی. با ارائه شواهد تجربی از داده‌های واقعی رشد گیاهچه و کیفیت جوشکاری، نشان داده می‌شود رگرسیون فازی می‌تواند به‌عنوان جایگزینی کارآمد برای روش‌های متداول در شرایط وجود نقاط دورافتاده مورد استفاده قرار گیرد. نتایج حاکی از آن است که رویکرد فازی نه‌تنها از روش کلاسیک طرح آزمایش، بلکه از روش‌های مقاوم‌سازی استاندارد نیز در مواجهه با داده‌های دورافتاده عملکرد بهتری دارد.

واژه‌های کلیدی: طرح آزمایش ها، رگرسیون فازی، داده دورافتاده، روش‌های مقاوم‌سازی.
کد موضوع‌بندی ریاضی (۲۰۲۰): 62A86، 62K05.

۱ مقدمه

طرح آزمایش‌ها به عنوان یکی از روش‌های کلاسیک در تحلیل داده‌ها، از دیرباز مورد توجه محققان بوده است. این روش که ریشه در کارهای فیشر در دهه ۱۹۲۰ دارد، به‌طور گسترده در علوم کشاورزی، صنعت و سایر حوزه‌ها مورد

استفاده قرار گرفته است (فیشر، ۱۹۳۵). تحلیل‌های مرتبط با طرح آزمایش‌ها معمولاً بر اساس فرض نرمال بودن توزیع متغیر خروجی استوار است و از روش‌هایی مانند تحلیل واریانس برای بررسی اثر متغیرهای کنترل استفاده می‌کند (مونتگومری، ۲۰۱۷). با این حال، یکی از چالش‌های اصلی در استفاده از این روش‌ها، وجود نقاط دورافتاده در داده‌ها است. نقاط دورافتاده می‌توانند به دلایل مختلفی از جمله خطاهای اندازه‌گیری، داده‌های نامعتبر یا تغییرات غیرمنتظره در سیستم تحت مطالعه ایجاد شوند. این نقاط می‌توانند تأثیر نامناسبی بر نتایج تحلیل‌های آماری داشته باشند و منجر به برآوردهای نادرست از پارامترهای مدل شوند (بارنت و لوئیس، ۱۹۴۴).

از سوی دیگر در رگرسیون فازی (تاناکا و همکاران، ۱۹۸۲)، به‌جای استفاده از مقادیر قطعی، از اعداد فازی برای مدل‌سازی رابطه بین متغیرها استفاده می‌شود. این روش به‌ویژه در مواجهه با داده‌های نادقیق یا دارای تغییرات زیاد، انعطاف‌پذیری بیشتری ارائه می‌دهد (راس، ۲۰۰۵). تولید متغیرهای فازی عمدتاً ناشی از دو منبع اصلی عدم قطعیت است: منبع اول ابهام ذاتی در تعیین مقادیر دقیق متغیرها که می‌تواند ناشی از محدودیت‌های اندازه‌گیری یا ماهیت کیفی متغیر باشد و منبع دوم تغییرپذیری سیستماتیک در ساختارهای به ظاهر ثابت (عدم قطعیت ذاتی) که منجر به پراکندگی مقادیر حول یک حالت مطلوب می‌شود. این دو ویژگی به‌ویژه در سیستم‌های پیچیده‌ای که مرزهای دقیقی بین حالت‌های مختلف وجود ندارد بیشتر مشهود بوده و توجیه‌کننده کاربرد روش‌های فازی در چنین شرایطی است. رگرسیون فازی به دلیل تنوع و انعطاف پذیری آن هم به لحاظ نوع متغیر و هم عدم الزام به نرمال بودن پاسخ ها بسیار مورد توجه قرار گرفته است. این مدل‌ها با استفاده از کمینه کردن مترهای مختلف فازی توانایی خود را در پوشش حداکثری مقابله با مشکلات متعدد در رگرسیون را نشان داده‌اند. در شن و همکاران (۲۰۱۰)؛ زنگ و همکاران (۲۰۱۷) بهینه کردن مدل رگرسیونی رقم می‌خورد. ملکوموا و شاتسکیخ (۲۰۱۷)؛ اکبری و حسامیان (۲۰۱۹) تلاش های خوبی برای مقابله با مشکل هم خطی انجام داده‌اند. اثر نقاط دورافتاده، ناهمواری در مشاهدات خروجی نیز توسط چاچی و حسامیان (۱۳۹۲)؛ خمر و همکاران (۲۰۲۰)؛ کاشانی و همکاران (۲۰۲۱) بررسی شده و برای بهبود آن و استواری مدل پیشنهاداتی ارائه گردیده است. همچنین قبادی و جوانمرد (۱۳۸۵) به بررسی نحوه به‌کارگیری طراحی آزمایش‌ها در محیط فازی پرداختند و روش‌های تحلیل یک طرح آزمایش با متغیر پاسخ فازی در حضور متغیرهای مستقل غیر فازی را مورد بررسی قرار داند. در این مقاله با ارائه یک مدل رگرسیونی فازی، کارایی این روش در بهبود تحلیل‌های طرح آزمایش‌ها در حضور نقاط دورافتاده مورد بررسی قرار می‌گیرد. در ادامه، در بخش ۲ مبانی آمار فازی مرور می‌شود. بخش ۳ به معرفی مفاهیم و مقدمات رگرسیون فازی اختصاص دارد. در بخش ۴، طرح آزمایش فازی برای نمونه‌های مکرر مطرح شده و همچنین چهار روش استوار برای مقابله با داده‌های پرت در طرح آزمایش مرور می‌شود. بخش ۵ شامل مثال‌های تجربی است که کارایی طرح آزمایش فازی را نشان می‌دهند و نتایج حاصل با سایر روش‌ها مقایسه می‌شوند. در پایان نیز یک آزمون فرض فازی برای بررسی تأثیر متغیرها در مدل پیشنهادی ارائه می‌گردد.

۲ آمار فازی

در بسیاری از علوم مانند روانشناسی، شیمی و حتی ریاضی صفات یا متغیرهایی با مفاهیم نادقیق وجود دارد. نظریه فازی، به عنوان تعمیمی از نظریه مجموعه‌ها، پوشش مناسبی برای این موارد است. در این مسیر، از آنجا که در بررسی توزیع، ارتباط و میزان اثرگذاری یا اثرپذیری متغیرهای مختلف از علم آمار بهره می‌بریم، وجود شاخه‌ای به نام آمار فازی نیز اجتناب‌ناپذیر خواهد بود. در نظریه مجموعه‌های فازی، برای هر مقدار $\alpha \in [۰, ۱]$، α–برش یک مجموعه فازی $\tilde{A}$ روی مجموعه مرجع X به صورت $\tilde{A}_\alpha = \{x \in X \mid \mu_{\tilde{A}}(x) \geq \alpha\}$ تعریف می‌شود که در آن $\mu_{\tilde{A}}(x)$ تابع عضویت مجموعه فازی است. به طور خاص، هنگامی که $\alpha = ۰$، $\tilde{A}_۰ = X$ خواهد بود، یعنی کل مجموعه مرجع را شامل می‌شود، و برای $\alpha = ۱$، $\tilde{A}_۱$ که به آن هسته مجموعه فازی گفته می‌شود، به دست می‌آید.

یک مجموعه فازی $\tilde{A} \in \mathbb{R}$ عدد فازی نامیده می‌شود اگر اولاً نرمال باشد، یعنی نقطه‌ای مانند $x_۰ \in \mathbb{R}$ وجود داشته باشد که $\mu_{\tilde{A}}(x_۰) = ۱$؛ ثانیاً محدب باشد، یعنی برای هر $\alpha \in (۰, ۱]$، $\tilde{A}_\alpha$ یک بازه بسته باشد؛ و ثالثاً تابع عضویت آن در $(-\infty, x_۰)$ غیر نزولی و در $(x_۰, +\infty)$ غیر صعودی باشد. این ویژگی‌ها باعث می‌شود که هر α–برش یک عدد فازی، بازه‌ای بسته و محدب باشد که نمایش مناسبی از عدم قطعیت در مقادیر عددی ارائه می‌دهد (راس، ۲۰۰۵).

تعریف ۱ (عدد فازی LR). یک عدد فازی $\tilde{M} = (m, a, b)_{LR}$ به صورت کامل با تابع عضویت

$$\mu_{\tilde{M}}(x) = \begin{cases} L\left(\frac{m-x}{a}\right) & x \leq m \text{ و } a > ۰ \\ R\left(\frac{x-m}{b}\right) & x > m \text{ و } b > ۰ \\ ۰ & \text{سایر موارد} \end{cases}$$

تعریف می‌شود، که در آن m نقطهٔ مد با درجهٔ عضویت ۱ و a و b به‌ترتیب پهنای چپ و راست عدد فازی هستند. همچنین L و R توابع مرجع نزولی‌اند که شرایط $L(۰) = R(۰) = ۱$، $L(۱) = R(۱) = ۰$ و پیوسته و نزولی بودن را ارضا می‌کنند.

حالت خاص - عدد فازی مثلثی: وقتی $L(z) = R(z) = \max(۰, ۱ - z)$ باشد، عدد فازی LR به عدد فازی مثلثی T به صورت $\tilde{M} = (m, a, b)_T$ تبدیل می‌شود که در آن تابع عضویت برابر است با

$$\mu_{\tilde{M}}(x) = \max\left(۰, \min\left(\frac{x - m + a}{a}, \frac{m + b - x}{b}\right)\right).$$

- **ضرب در اسکالر:** برای $\theta \in \mathbb{R}$ و $\tilde{M} = (m, a, b)_{LR}$

$$\theta\tilde{M} = \begin{cases} (\theta m, \theta a, \theta b)_{LR} & \theta \geq ۰ \\ (\theta m, |\theta| b, |\theta| a)_{LR} & \theta < ۰ \end{cases}$$

- **جمع و تفریق فازی:** برای $\tilde{M}_۱ = (m_۱, a_۱, b_۱)_{LR}$ و $\tilde{M}_۲ = (m_۲, a_۲, b_۲)_{LR}$

$$\tilde{M}_۱ \oplus \tilde{M}_۲ = (m_۱ + m_۲, a_۱ + a_۲, b_۱ + b_۲)_{LR}$$
$$\tilde{M}_۱ \ominus \tilde{M}_۲ = (m_۱ - m_۲, a_۱ + a_۲, b_۱ + b_۲)_{LR}$$

عملگرهای فوق هنگامی دقیق هستند که توابع L و R یکسان باشند. همچنین برای حالت‌های کلی‌تر، محاسبات بر اساس اصل گسترش[1] انجام می‌شود

$$\mu_{\tilde{M}_۱ \oplus \tilde{M}_۲}(z) = \sup_{z=x+y} \min(\mu_{\tilde{M}_۱}(x), \mu_{\tilde{M}_۲}(y)).$$

۳ رگرسیون فازی

وقتی یکی از متغیرهای ورودی یا خروجی مدل دارای ذات مبهم و فازی باشند، می‌توان رگرسیون فازی را به جای رگرسیون کلاسیک متصور شد. به‌علاوه ممکن است تمامی شرایط بکارگیری رگرسیون کلاسیک، خصوصاً شرایط توزیعی، در مسئله برقرار نباشد. رگرسیون فازی به‌عنوان یک روش جایگزین، به‌ویژه در شرایطی که داده‌ها دارای عدم قطعیت یا نقاط دورافتاده هستند، بسیار مفید است (یانگ و همکاران، ۲۰۱۳؛ حسامیان و اکبری، ۲۰۲۰). مدل‌های رگرسیونی فازی بر اساس اینکه ورودی، خروجی یا ضرایب فازی باشند به چند دسته مختلف تقسیم می‌شوند (کاچپشیک و فدریزی، ۱۹۹۲). در رگرسیون فازی، رویکردهای مختلفی برای کم کردن خطای مدل برآورد شده وجود دارد. یکی از این رویکردها، مدل‌های امکانی است. در این دیدگاه، کار بر اساس حل یک مدل برنامه‌ریزی خطی یا غیرخطی و بر مبنای تابع هدفی که تحت کمینه یا بیشینه کردن ضرایب مدل برآورد شده تولید می‌شود، انجام می‌گیرد. این روش به‌ویژه در شرایطی که داده‌ها دارای عدم قطعیت هستند، بسیار کارآمد است (زنگ و همکاران، ۲۰۱۷). در این رویکرد با توجه به اثر گذاری نقاط دورافتاده در مقادیر خروجی، استفاده از مترهایی که کمتر تحت تاثیر آنها قرار می گیرند راه گشا خواهد بود.

مدل رگرسیون کلاسیک، معمولا بر اساس یک سری پیش‌فرض‌های توزیعی و در قالب شاخص‌هایی مانند متوسط خطای پیشگویی یا متوسط توان دوم خطای پیشگویی برازش داده می‌شود. اما در رگرسیون فازی، شرایط

[1] Extension Principle

توزیعی خاصی مانند رگرسیون کلاسیک وجود ندارد و معمولاً به دنبال کمینه کردن فاصله بین پاسخ‌های برآورد شده و پاسخ‌های مشاهده شده هستیم. از اینرو، نیازمند استفاده از مترهای اندازه‌گیری فازی برای محاسبه فواصل بین اعداد فازی و دستیابی به مدلی با کمترین فاصله هستیم (تاناکا و همکاران، ۱۹۸۲؛ راس، ۲۰۰۵). فرض کنید $\tilde{B} = (m_b, \ell_b, r_b)_{LR}$ و $\tilde{A} = (m_a, \ell_a, r_a)_{LR}$ دو عدد فازی LR باشند. برای اندازه‌گیری فاصله بین اعداد فازی از متری به صورت

$$D_z(\tilde{A}, \tilde{B}) = \frac{1}{3}\big(|m_A - m_B| + |\ell_A - \ell_B| + |r_A - r_B|\big).$$

استفاده می‌شود (زنگ و همکاران، ۲۰۱۷). این متر به دلیل سادگی و دقت بالا در محاسبه فاصله بین اعداد فازی، همچنین کنترل و کم کردن تاثیر نقاط دورافتاده، انتخاب مناسبی برای تحلیل داده‌های فازی با حضور نقاط دورافتاده است.

اگر $\tilde{\boldsymbol{y}}$ و $\hat{\tilde{\boldsymbol{y}}}$ به ترتیب متغیرهای مشاهده شده و برآورد شده باشند، دو شاخص مهم برای سنجش خطای مدل خطای میانگین مطلق پیش‌بینی و خطای میانگین مربعات پیش‌بینی هستند که به صورت

$$\text{MPE}(\tilde{\boldsymbol{y}}, \hat{\tilde{\boldsymbol{y}}}) = \frac{1}{n}\sum_{i=1}^{n} D_z(\tilde{y}_i - \hat{\tilde{y}}_i)$$
$$\text{MSPE}(\tilde{\boldsymbol{y}}, \hat{\tilde{\boldsymbol{y}}}) = \frac{1}{n}\sum_{i=1}^{n} D_z^2(\tilde{y}_i - \hat{\tilde{y}}_i)$$

تعریف می‌شوند. علاوه بر این، از ضریب تعیین $R^2 = 1 - \frac{\sum(\tilde{y}_i - \hat{\tilde{y}}_i)^2}{\sum(\tilde{y}_i - \bar{\tilde{y}}_i)^2}$ و ضریب تعیین تصحیح شده $R^2_{\text{adj}} = 1 - (1 - R^2)\frac{n-1}{n-p-1}$ نیز برای ارزیابی مدل استفاده می‌شود، که در آن p تعداد متغیرهای مستقل در مدل است و $\bar{\tilde{y}}_i$ بر اساس میانگین مراکز، پهناهای چپ و پهناهای راست بدست می‌آید.

۴ طرح آزمایش فازی با نمونه‌های مکرر

طرح آزمایش‌های با نمونه‌های مکرر به ما امکان می‌دهند با تغییر دادن عوامل (متغیرهای ورودی)، تاثیر آن‌ها را بر متغیر پاسخ بررسی کنیم. این روش به دلیل تکرار آزمایش‌ها روی یک یا چند عامل، قدرت آماری بالایی دارد. یعنی می‌تواند اثر عوامل را دقیق‌تر اندازه‌گیری کند و با کنترل تغییرات، به نتایج قابل اعتمادتری برسد. به‌جای اینکه فقط میانگین نتایج را در نظر بگیریم، می‌توانیم از یک ساختار فازی استفاده کنیم. این کار به ما کمک می‌کند تا در شرایطی که داده‌ها به دلیل وجود نقاط دورافتاده یا تغییرات زیاد، پراکنده هستند، تحلیل دقیق‌تری داشته باشیم. استفاده از رویکرد فازی، همه مقادیر به‌دست‌آمده را در نظر می‌گیرد و فقط به میانگین اکتفا نمی‌کند.

بررسی تغییرات میانگین پاسخ‌ها در روش طرح آزمایش، مبنای تحلیل‌ها و استنباط‌های آماری است. با این

حال، وجود نقاط دورافتاده می‌تواند باعث افزایش خطا در برآوردها و استنباط‌های نادرست شود، زیرا میانگین به شدت تحت تاثیر این نقاط قرار می‌گیرد. همچنین، این نقاط ممکن است پیش‌فرض‌های آماری مانند نرمال بودن توزیع داده‌ها را نقض کنند. برای مقابله با این مشکل، روش‌های آماری مختلفی برای کاهش تاثیر داده‌های دورافتاده وجود دارد. در این پژوهش، ما چهار روش را برای کاهش اثر نقاط دورافتاده بررسی و با رویکرد اصلی این مقاله، یعنی تحلیل فازی، مقایسه می‌کنیم. در ادامه، ابتدا این چهار روش و سپس ایده اصلی پژوهش معرفی خواهند شد.

الف- روش استوار هوبر[1]: در این روش، به‌جای برآورد معمول با کمترین مربعات، از رویکردی استفاده می‌شود که خطاهای بزرگ را با وزن کمتری در نظر می‌گیرد. تابع وزن‌دهی هوبر به‌صورت

$$w_H(e) = \begin{cases} ۱ & |e| \le k \\ \frac{k}{|e|} & |e| > k \end{cases}$$

است، که در آن e نمایانگر خطای مدل و k آستانه‌ای برای کنترل حساسیت به داده‌های دورافتاده است. مقدار پیشنهادی برای این آستانه معمولاً $k = ۱٫۳۴۵$ در نظر گرفته می‌شود. این روش در واقع بین رویکرد کمترین مربعات و کمترین قدرمطلق خطاها پل می‌زند و برآوردهایی با تعادل بین دقت و مقاومت ارائه می‌دهد (لامبرت-لاکروا و زوالد، ۲۰۱۱).

ب- روش استوار دو مربعی[2]: این روش نیز مشابه روش هوبر از رویکرد وزن‌دهی استفاده می‌کند، اما تابع وزن‌دهی آن شکلی نرم‌تر دارد و برای خطاهای بزرگ، وزن صفر اختصاص می‌دهد. تابع وزن‌دهی دو مربعی به صورت

$$w_H(e) = \begin{cases} \left(۱ - \left(\frac{e}{k}\right)^۲\right)^۲ & |e| \le k \\ ۰ & |e| > k \end{cases}$$

تعریف می شود. در این رابطه نیز e نشان‌دهنده خطا و k آستانه‌ای برای برش مقادیر دورافتاده است. مقدار معمول برای این آستانه $k = ۴٫۶۸۵$ پیشنهاد می‌شود. در این روش، مشاهداتی که دارای خطای بسیار زیاد باشند، هیچ تأثیری در برآورد نهایی نخواهند داشت (بتی و همکاران، ۲۰۲۳).

ج- روش میانگین جایگزین: در این روش، هنگامی که داده‌ای به‌عنوان نقطه دورافتاده شناسایی شود، به‌جای حذف کامل آن، مقدار آن با میانگین سایر مشاهدات معتبر همان بخش یا گروه جایگزین می‌شود. این رویکرد به‌ویژه زمانی مفید است که حذف داده‌ها به‌دلیل حجم نمونه کم ممکن نباشد و حفظ ساختار طرح ضروری باشد (لیتل و روبین، ۲۰۱۹).

د- روش رتبه‌بندی: در این روش، به‌جای استفاده از مقدار واقعی پاسخ‌ها، رتبه آن‌ها در تحلیل به‌کار می‌رود. این روش باعث می‌شود تأثیر نقاط دورافتاده کاهش یابد و تحلیل بر اساس موقعیت نسبی داده‌ها انجام شود، نه

[1] Huber robust method
[2] Bi-square robust method

مقادیر مطلق آن‌ها (صالح و همکاران، ۲۰۲۲).

۵- روش رگرسیون فازی: فرض کنید برای هر ترکیب از متغیرهای ورودی، مقادیر ممکن برای متغیر خروجی متناظر با آن‌ها در تکرارهای مختلف به صورت $\boldsymbol{y}^* = (y_1^*, \ldots, y_w^*)$ تعریف گردد. در این صورت، عدد فازی LR به فرم $\tilde{y} = (y_c, \ell, r)_{LR}$ تعریف می‌شود، که در آن y_c مرکز عدد فازی و مقادیر مثبت ℓ و r به ترتیب پهناهای چپ و راست عدد فازی هستند. بر اساس تکرارهای به‌دست‌آمده، y_c را میانگین مقادیر تکرار شده در نظر می‌گیریم. بنابراین، پهنای چپ و پهنای راست عدد فازی به ترتیب به صورت $\ell = y_c - y_{\min} - \delta$ و $r = y_{\max} - y_c + \delta$ است، که در آن $y_{\min}$ کمترین و $y_{\max}$ بیشترین مقدار به‌دست‌آمده برای متغیر خروجی در تمام تکرارها است و δ یک عدد مثبت برای ساخت ابتدا و انتهای عدد فازی است.

باید توجه داشت که، چون کم‌ترین و بیش‌ترین مقدار تولید شده در خروجی‌ها جزئی از نمونه‌های ممکن هستند، باید در عدد فازی ما دارای اعتبار و درجه عضویت غیرصفر باشند. بنابراین، مقدار عددی مثبتی مانند δ به پهناهای چپ و راست اضافه می‌شوند تا این مقادیر در نقاط ابتدایی و انتهایی با درجه اعتبار صفر قرار نگیرند. میزان افزایش ایجاد شده برای پهناها می‌تواند به تعداد نمونه‌های دریافتی یا درصد سهم کم‌ترین و بیش‌ترین خروجی از کل داده‌های تکرار شده باشد. همچنین، این عدد می‌تواند توسط یک متخصص نیز اعلام شود. در مثال‌های کاربردی، با توجه به نظر کارشناس، ۱۰ درصد از مقدار عددی مرکز را در نظر گرفته و سپس میانگین این اعداد را به عنوان δ در ایجاد پهنای اعداد فازی قرار داده‌ایم. اکنون با فرض اینکه $\tilde{y}_i = (y_i; \ell_i, r_i)_{LR}$، i امین پاسخ فازی باشد، مدل رگرسیونی فازی به صورت $\tilde{y}_i = \tilde{\beta}_0 + \sum_{j=1}^{p} \tilde{\beta}_j X_{ij}$ تعریف می شود که در آن $\tilde{\beta}_j$ $(j = 1, \ldots, p)$ ضرایب مدل رگرسیونی فازی هستند و $\tilde{\beta}_0$ عرض از مبدا مدل است. سپس برای رسیدن به برآوردهایی با بهترین عملکرد، تابع هدف برای کمینه‌سازی تفاوت بین مقادیر واقعی و پیش‌بینی‌شده برابر $\min D_z = \sum_{i=1}^{p} |\tilde{\boldsymbol{y}}_i - \hat{\tilde{\boldsymbol{y}}}_i|$ است، که در آن D_z متر فازی (زنگ و همکاران، ۲۰۱۷) است که مجموع قدر مطلق تفاوت بین مقادیر واقعی و پیش‌بینی‌شده را اندازه‌گیری می‌کند. بنابراین با فرض $\tilde{\beta}_j = (c_j, a_j, b_j)_{LR}$ به عنوان ضرایب فازی مدل، تابع هدف کمینه به صورت

$$\min f(c_j, a_j, b_j) = \sum_{i=1}^{n} \left(|y_i - \sum_{j=1}^{p} c_j X_j| + |\ell_i - \sum_{j=1}^{p} a_j X_j| + |r_i - \sum_{j=1}^{p} b_j X_j| \right)$$

تعریف می شود، که در آن c_j، a_j، و b_j به ترتیب مرکز، پهنای چپ و پهنای راست پارامترهای مدل هستند و برای کمینه شدن تابع هدف باید تحت شرایط

$$a_j \geq 0, \quad b_j \geq 0, \quad \sum_{j=1}^{p} a_j X_{ij} \geq 0, \quad \sum_{j=1}^{p} b_j X_{ij} \geq 0.$$

برآورد شوند. در نهایت مدل برآوردشده به فرم $\hat{\tilde{y}}_j = \hat{\tilde{\beta}}_0 + \sum_{j=1}^{p} \hat{\tilde{\beta}}_j X_j$ ارائه می‌شود.

۵ مثال‌های کاربردی

برای نمایش کاربرد و کارایی ایده این پژوهش، از یک مثال واقعی استفاده شده است. این بخش شامل دو مرحله است. ابتدا، داده‌هایی که توزیع نرمال دارند، با استفاده از روش‌های معمول در طرح آزمایش‌ها و همچنین مدل رگرسیون فازی تحلیل می‌شوند. سپس، با اضافه کردن نقاط دورافتاده به داده‌ها، تحلیل‌ها در هر دو روش معمولی و فازی تکرار شده و نتایج مقایسه می‌شوند.

۵.۱ داده‌های رشد گیاه

داده‌های رشد گیاه (آریایی‌فر و همکاران، ۱۳۹۷) ارائه شده در جدول ۵ شامل ۶۰ نمونه و تکرار ۴ تایی در هر بخش بر اساس دو متغیر ورودی با عنوان‌های «زمان ماندگاری در هیپوکلریت سدیم بر حسب دقیقه» (x_1) و «شماره هفته» (x_2) به همراه یک متغیر خروجی با عنوان «رشد گیاهچه بر حسب سانتی متر» (y) هستند که بر اساس تعاریف ارائه شده در ایده تحقیق، به صورت فازی شده در جدول ۲ نمایش داده شده‌اند. براساس مدل طرح آزمایش و طبق نتایج

جدول ۱. داده‌های واقعی رشد گیاهچه

بخش اول				بخش دوم				بخش سوم			
ردیف	x_1	x_2	y	ردیف	x_1	x_2	y	ردیف	x_1	x_2	y
۱	۵	۱	۳	۲۱	۱۰	۳	۳	۴۱	۲۰	۲	۳
۲	۵	۱	۴٫۵	۲۲	۱۰	۳	۳٫۵	۴۲	۲۰	۲	۳
۳	۵	۱	۳٫۵	۲۳	۱۰	۳	۳	۴۳	۲۰	۲	۳٫۵
۴	۵	۱	۴٫۵	۲۴	۱۰	۳	۳	۴۴	۲۰	۲	۴٫۵
۵	۵	۲	۳٫۵	۲۵	۱۵	۱	۲٫۵	۴۵	۲۰	۳	۴٫۵
۶	۵	۲	۴٫۵	۲۶	۱۵	۱	۳	۴۶	۲۰	۳	۳٫۵
۷	۵	۲	۳	۲۷	۱۵	۱	۴٫۵	۴۷	۲۰	۳	۳
۸	۵	۲	۲٫۵	۲۸	۱۵	۱	۳	۴۸	۲۰	۳	۴
۹	۵	۳	۳	۲۹	۱۵	۲	۳	۴۹	۲۵	۱	۹
۱۰	۵	۳	۴	۳۰	۱۵	۲	۴	۵۰	۲۵	۱	۷
۱۱	۵	۳	۳٫۵	۳۱	۱۵	۲	۲٫۵	۵۱	۲۵	۱	۹
۱۲	۵	۳	۴	۳۲	۱۵	۲	۳	۵۲	۲۵	۱	۸
۱۳	۱۰	۱	۴٫۵	۳۳	۱۵	۳	۴٫۵	۵۳	۲۵	۲	۷٫۵
۱۴	۱۰	۱	۳٫۵	۳۴	۱۵	۳	۳	۵۴	۲۵	۲	۸
۱۵	۱۰	۱	۴٫۵	۳۵	۱۵	۳	۲٫۵	۵۵	۲۵	۲	۷٫۵
۱۶	۱۰	۱	۳٫۵	۳۶	۱۵	۳	۳	۵۶	۲۵	۲	۸
۱۷	۱۰	۲	۳	۳۷	۲۰	۱	۳	۵۷	۲۵	۳	۸٫۵
۱۸	۱۰	۲	۴٫۵	۳۸	۲۰	۱	۳	۵۸	۲۵	۳	۹
۱۹	۱۰	۲	۳٫۵	۳۹	۲۰	۱	۳	۵۹	۲۵	۳	۹
۲۰	۱۰	۲	۳	۴۰	۲۰	۱	۳	۶۰	۲۵	۳	۸٫۵

حاصل از آن (جدول ۳)، عدد ثابت و میزان زمان ماندگاری گیاهچه در هیپوکلریدریک دارای اثر معنی‌دار هستند، در حالی که اثر متغیر شماره هفته معنی‌دار نیست. در این مسیر و بر اساس مدل‌سازی رگرسیون فازی و با استفاده

جدول ۲. فازی‌سازی داده‌های اصلی

ردیف	x_1	x_2	$\tilde{y}_i = (y_i; \ell_i, r_i)_T$
۱	۵	۱	$(۳٫۸۷۵, ۱٫۳۷۵, ۱٫۱۲۵)_T$
۲	۵	۲	$(۳٫۳۷۵, ۱٫۳۷۵, ۱٫۶۲۵)_T$
۳	۵	۳	$(۳٫۶۲۵, ۱٫۱۲۵, ۰٫۸۷۵)_T$
۴	۱۰	۱	$(۴٫۰۰, ۱٫۰۰, ۱٫۰۰)_T$
۵	۱۰	۲	$(۳٫۵۰, ۱٫۰۰, ۱٫۵۰)_T$
۶	۱۰	۳	$(۳٫۳۷۵, ۰٫۸۷۵, ۱٫۱۲۵)_T$
۷	۱۵	۱	$(۳٫۲۵۰, ۱٫۲۵۰, ۱٫۷۵۰)_T$
۸	۱۵	۲	$(۳٫۱۲۵, ۱٫۱۲۵, ۱٫۳۷۵)_T$
۹	۱۵	۳	$(۳٫۲۵۰, ۱٫۲۵۰, ۱٫۷۵۰)_T$
۱۰	۲۰	۱	$(۳٫۰۰, ۰٫۵۰, ۰٫۵۰)_T$
۱۱	۲۰	۲	$(۳٫۵۰, ۱٫۰۰, ۱٫۵۰)_T$
۱۲	۲۰	۳	$(۳٫۷۵۰, ۰٫۷۵۰, ۱٫۲۵۰)_T$
۱۳	۲۵	۱	$(۸٫۲۵۰, ۱٫۷۵۰, ۱٫۲۵۰)_T$
۱۴	۲۵	۲	$(۷٫۷۵۰, ۰٫۷۵۰, ۰٫۷۵۰)_T$
۱۵	۲۵	۳	$(۸٫۷۵۰, ۰٫۷۵۰, ۰٫۷۵۰)_T$

جدول ۳. آزمون اثرات متغیرها

منابع تغییر	مجموع مربعات	درجه آزادی	میانگین مربعات	آماره F	سطح معنی‌داری	ضریب اثر اتا
مدل تصحیح شده	۴۷۵٫۲۲۵	۱۴	۱۰۵٫۱۶	۲۸۶٫۳۷	۰٫۰۰۰	۰٫۹۲۱
عدد ثابت	۸۳۷٫۱۱۷	۱	۸۳۷٫۱۱۷۴	۲۷۱۹٫۸۸۱	۰٫۰۰۰	۰٫۹۸۴
هفته‌ها	۰٫۹۷۵	۲	۰٫۴۸۷	۱٫۱۲۹	۰٫۳۳۲	۰٫۰۴۸
هیپوکلریدریک	۲۲۰٫۸۹۲	۴	۵۵٫۲۲۳	۱۲۷٫۸۴۷	۰٫۰۰۰	۰٫۹۱۹
اثر متقابل	۳٫۶۰۸	۸	۰٫۴۵۱	۱٫۰۴۴	۰٫۴۱۸	۰٫۱۵۷
خطا	۱۹٫۴۳۸	۴۵	۰٫۴۳۲			
کل	۱۴۱۹٫۷۵۰	۶۰				
تصحیح شده کل	۲۴۴٫۹۱۳	۵۹				

از نرم‌افزار $Matematica$، برآورد مدل رگرسیونی با استفاده از متر D_z به صورت

$$\hat{\tilde{\boldsymbol{y}}} = (۳٫۷۲۳, ۱٫۵, ۱٫۵)_T \oplus (-۰٫۱۹۵, ۰, ۰)_T \cdot x_1 \oplus (۰٫۰۳۳, ۰, ۰)_T \cdot x_2,$$

حاصل شده است. جدول ۴ بیانگر عملکرد مدل رگرسیون فازی برآورد شده و طرح آزمایش ها است.

جدول ۴. مقایسه معیارهای ارزیابی مدل فازی و غیرفازی

مدل	$\mathbf{R}^2$	$\mathbf{R}^2_{\text{adj}}$	**MSPE**	**MPE**
فازی	۰٫۸۷۳	۰٫۸۵۱	۰٫۵۸۱	۰٫۶۵۴
طرح آزمایش	۰٫۹۲۱	۰٫۸۹۵	۰٫۴۳۲	-

مقایسه نتایج حاصل از روش‌ها، بیانگر آن است که تفاوت ایجاد شده در $MSPE$ و R^2 به ترتیب ۰٫۱۴۹ و ۷ درصد است. در واقع، عملکرد روش طرح آزمایش بهتر از مدل رگرسیون فازی است. اما در اینجا ذکر دو نکته قابل تأمل است. اول اینکه تعداد نمونه‌ها در روش طرح آزمایش ۴ برابر نمونه‌ها در مدل‌سازی رگرسیون فازی است.

نکته دوم این که در داده‌های اصلی، نقطه یا نقاط دورافتاده وجود ندارد، بنابراین تغییرات زیادی بین مقادیر مختلف در نمونه‌ها و میانگین داده‌ها ایجاد نمی‌شود.

۵.۲ داده‌های رشد گیاهچه با نقاط دورافتاده

در این بخش، در داده‌های اصلی، نمونه شماره ۵۰ را از ۷ به ۳ و نمونه شماره ۶۰ را از ۸٫۵ به ۳ تغییر داده‌ایم. نتیجه حاصل از روش طرح آزمایش و مدل رگرسیون فازی به ترتیب در جداول ۵ و ۶ ارائه شده است. همانطور که در

جدول ۵. آزمون اثرات متغیرها با وجود نقاط دورافتاده

منابع تغییر	مجموع مربعات	درجه آزادی	میانگین مربعات	آماره F	سطح معنی‌داری	ضریب اثر اتا
مدل تصحیح شده	۱۵۹٫۸۵۸	۱۴	۱۱٫۴۱۸	۷٫۰۰۳	۰٫۰۰۰	۰٫۶۸۵
عدد ثابت	۱۰۷۵٫۲۶۷	۱	۱۰۷۵٫۲۶۷	۶۵۹٫۴۴۸	۰٫۰۰۰	۰٫۹۳۶
هفته‌ها	۰٫۱۰۸	۲	۰٫۰۵۴	۰٫۰۳۳	۰٫۹۶۷	۰٫۰۰۱
هیپوکلریدریک	۱۵۷٫۰۶۷	۴	۳۹٫۲۶۷	۲۴٫۰۸۲	۰٫۰۰۰	۰٫۶۸۲
اثر متقابل	۲٫۶۸۳	۸	۰٫۳۳۵	۰٫۲۰۶	۰٫۹۸۸	۰٫۰۳۵
خطا	۷۳٫۳۷۵	۴۵	۱٫۶۱۳			
کل	۱۳۰۸٫۵۰۰	۶۰				
تصحیح شده کل	۲۳۳٫۲۳۳	۵۹				

جدول ۵ ملاحظه می‌شود، به‌واسطه وجود نقاط دورافتاده، میانگین مربعات خطای مدل برابر با ۱٫۶۳۱ حاصل شده است و R^2 و $\mathbf{R}^2_{\text{adj}}$ به ترتیب برابر با ۰٫۶۸۵ و ۰٫۵۸۸ به‌دست آمده‌اند. حال مدل رگرسیون فازی برای داده‌های جدید به صورت

$$\hat{\tilde{\boldsymbol{y}}} = (۱٫۸۳۵, ۰٫۹۸۶, ۰٫۹۸۶)_T \oplus (۰٫۴۹۳, ۰٫۰۰۷, ۰٫۰۰۳)_T \cdot x_۱ \oplus (۰٫۰۶۱, ۰, ۰٫۰۱۸)_T \cdot x_۲.$$

است. همچنین، جدول ۶ بیانگر عملکرد مدل رگرسیون فازی برآورد شده و سایر روش ها است.

جدول ۶. نتایج معیارها برای روش‌های مختلف

روش	$\mathbf{R}^2$	$\mathbf{R}^2_{\text{adj}}$	MSPE	MPE
فازی	۰٫۷۸۶	۰٫۷۶۳	۰٫۵۹۴	۰٫۷۰۸
طرح آزمایش	۰٫۶۸۵	۰٫۵۸۸	۱٫۶۳۱	–
هوبر	۰٫۷۲۵	۰٫۶۷۸	۱٫۰۰۶	–
دو مربعی	۰٫۵۱۰	۰٫۶۲۶	۳٫۱۶۳	–
رتبه‌بندی	۰٫۶۲۰	۰٫۵۰۲	۳٫۵۴۷	–
میانگین جایگزین	۰٫۸۰۵	۰٫۷۵۵	۰٫۷۹۲	–

نتایج مقایسه عملکرد روش‌های مختلف در شرایط وجود داده‌های دورافتاده نشان می‌دهد که روش فازی با دارا بودن بالاترین مقدار R^2_{adj} (۰٫۷۶۳) و کمترین خطای مربعات پیش‌بینی ($MSPE = ۰٫۵۹۴$)، به‌عنوان کارآمدترین روش شناخته می‌شود. این برجستگی احتمالاً ناشی از توانایی ذاتی سیستم‌های فازی در مدیریت عدم قطعیت‌ها و

پردازش داده‌های دورافتاده است. در مقابل، اگرچه روش میانگین جایگزین با مقدار $R^2 = ۰/۸۰۵$ بالاترین ضریب تعیین را نشان می‌دهد، اما با توجه به R^2_{adj} پایین‌تر (۰/۷۵۵) و خطای پیش‌بینی بیشتر ($MSPE = ۰/۷۹۲$)، نتوانسته جایگاه روش فازی را به چالش بکشد. روش هوبر نیز با مقادیر $R^2 = ۰/۷۲۵$ و $MSPE = ۱/۰۰۶$ عملکردی متوسط ارائه داده که حاکی از مقاومت نسبی آن در برابر داده‌های دورافتاده است. از سوی دیگر، روش‌های کلاسیک مانند طرح آزمایش ($R^2_{adj} = ۰/۵۸۸$، $MSPE = ۱/۶۳۱$) و به‌ویژه روش رتبه‌بندی ($R^2_{adj} = ۰/۵۰۲$، $MSPE = ۳/۵۴۷$) به‌وضوح تحت تأثیر منفی داده‌های دورافتاده قرارگرفته‌اند. این یافته‌ها به روشنی تأیید می‌کنند که در محیط‌های دارای نویز و داده‌های دورافتاده، روش‌های مقاوم مانند فازی و هوبر از کارایی بالاتری نسبت به روش‌های مرسوم برخوردارند، با این حال روش فازی به دلیل قابلیت‌های منحصر به فرد در مدل‌سازی سیستم‌های غیرقطعی، به‌عنوان گزینه بهینه در چنین شرایطی شناخته می‌شود.

۵.۳ داده‌های کیفیت جوشکاری

در این بخش، کیفیت جوشکاری به‌عنوان متغیر پاسخ در نظر گرفته شده است. سه متغیر کنترل‌شده شامل دمای دستگاه، زاویه جوشکاری، و مدت زمان جوشکاری هرکدام در سه سطح به‌عنوان عوامل مؤثر بر متغیر پاسخ ایفای نقش می‌کنند. با توجه به ماهیت کیفی و فازی پاسخ (کیفیت جوش)، و همچنین وجود داده‌های دورافتاده در مشاهدات که با استفاده از معیار فاصله کوک[1] شناسایی شده‌اند، از روش‌های مختلف برای تحلیل داده‌ها و مقایسه عملکرد آن‌ها استفاده شد. نتایج حاصل از به‌کارگیری روش تحلیل فازی، روش متداول طرح آزمایش و سه روش مقاوم شامل هوبر، رتبه‌بندی و میانگین جایگزین در جدول ۷ ارائه شده است. همان‌طور که ملاحظه می‌شود، روش فازی بالاترین مقدار R^2 و کمترین مقدار MSPE را به خود اختصاص داده است و از این رو در شرایطی که متغیر پاسخ ماهیت فازی دارد و داده‌های دورافتاده در مجموعه داده‌ها وجود دارند، عملکرد بهتری نسبت به سایر روش‌ها نشان می‌دهد. پس از آن، روش دو مربعی، میانگین جایگزین و روش رتبه‌بندی نیز عملکرد قابل قبولی دارند. در مقابل، روش معمول طرح آزمایش در مواجهه با داده‌های دورافتاده، ضعیف‌ترین نتایج را ارائه داده است.

جدول ۷. نتایج معیارها برای روش‌های مختلف

روش	R^2	R^2_{adj}	MSPE	MPE
فازی	۰/۸۲۶	۰/۷۶۸	۶۵/۱۲۳	۸/۸۷۵
طرح آزمایش	۰/۷۶۱	۰/۶۹۱	۹۷/۸۵۳	–
هوبر	۰/۷۸۳	۰/۷۲۴	۸۵/۵۷۸	–
دو مربعی	۰/۸۰۵	۰/۷۳۸	۷۶/۷۶۵	–
رتبه‌بندی	۰/۷۹۲	۰/۷۳۵	۷۹/۶۸۷	–
میانگین جایگزین	۰/۸۳۶	۰/۷۶۴	۷۵/۷۲۸	–

[1] Cook's Distance

۶ آزمون فرض فازی

در این بخش، با الهام از ایده مطرح‌شده در (لی و همکاران، ۲۰۱۵)، یک آزمون فرض فازی برای بررسی تأثیر متغیرها در مدل پیشنهادی انجام خواهیم داد. هدف این کار، نشان دادن هم‌خوانی و تطابق نتایج به‌دست‌آمده از مدل فازی با روش‌های سنتی طرح آزمایش در تعیین متغیرهای مؤثر است. با فرض $\tilde{\theta}_{j\alpha} = [\tilde{\theta}^L_{j\alpha}, \tilde{\theta}^U_{j\alpha}]$، به عنوان بازه ایجاد شده برای پارامتر در یک α–برش ($0 \le \alpha \le 1$) از jامین پارامتر فازی، آزمون فرض فازی

$$\begin{cases} H_0: & \tilde{\theta}_{j\alpha} \text{ تقریباً صفر است} \\ H_1: & \tilde{\theta}_{j\alpha} \text{ دور از صفر است} \end{cases}$$

را در نظر بگیرید. بر اساس روش افرون و تیبشیرانی (۱۹۹۴) تعداد $\dot{B}$ نمونه بوت‌استرپی تولید و ضرایب مدل آن‌ها که با $\hat{\tilde{\theta}}^{(\dot{b})}_j$، $\dot{b} = 1, \ldots, \dot{B}$، نمایش داده می‌شود، برآورد می‌کنیم. آماره آزمون به صورت $T_\alpha = \sqrt{D^2\left(\frac{\hat{\tilde{\theta}}_{j\alpha}}{S_{\hat{\tilde{\theta}}_{j\alpha}}}\right)}$ تعریف می شود، که در آن فاصله اقلیدسی برابر

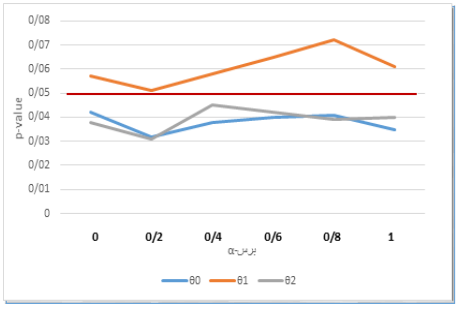
$$D^2\left(\frac{\hat{\tilde{\theta}}_{j\alpha}}{S_{\hat{\tilde{\theta}}_{j\alpha}}}\right) = \left(\frac{1}{2}(\hat{\tilde{\theta}}^L_{j\alpha} + \hat{\tilde{\theta}}^U_{j\alpha})\right)^2 + \left(\frac{1}{2}|\hat{\tilde{\theta}}^L_{j\alpha} - \hat{\tilde{\theta}}^U_{j\alpha}|\right)^2$$

است. همچنین $S_{\hat{\tilde{\theta}}_{j\alpha}}$ نماد انحراف استاندارد پارامتر jام فازی در برش αام و $\bar{x}_j$ به عنوان میانگین مقادیر jامین متغیر ورودی هستند:

$$S_{\hat{\tilde{\theta}}_{j\alpha}} = \sqrt{\frac{\sum_{i=1}^{n} D^2(\hat{\tilde{y}}_{i\alpha}, \tilde{y}_{i\alpha})}{(n-p-1)\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2}}$$

اکنون p–مقدار آزمون برابر $p-value = \frac{1}{\dot{B}}\sum_{\dot{b}=1}^{\dot{B}} I(T_\alpha^{(\dot{b})} \ge T_\alpha)$ است، که در آن $I(\cdot)$ بیانگر تابع مشخصه است. شکل ۱ تغییرات p–مقدار را برای پارامترهای (θ_0، θ_1، θ_2). بررسی این نمودارها اطلاعات مهمی در مورد معنی‌داری آماری هر پارامتر ارائه می‌دهد. در شکل ۱–الف، p–مقدار برای پارامتر θ_1 (منحنی آبی) به طور قابل توجهی بالاتر از سطح ۰٫۰۵ است. این موضوع بیانگر آن است که θ_1 تأثیر معنی‌داری در مدل ندارد و نمی‌توان آن را با اطمینان در نظر گرفت. در مقابل، p–مقدار برای پارامترهای θ_0 (منحنی زرد) و θ_2 (منحنی سیاه با خط‌چین ضخیم) در اغلب نقاط کمتر از۰٫۰۵ بوده، که نشان‌دهنده تاثیر معنی‌دار این پارامترها در مدل است. در شکل ۱–ب، مقادیر p–مقدار در برخی نقاط دستخوش نوسان شده‌اند. مشاهده می‌شود که در برخی برش‌های α، p–مقدار برای θ_1 افزایش یافته و در نزدیکی ۰٫۰۵ قرار گرفته است، اما همچنان بالاتر از این مقدار باقی مانده که بیانگر عدم معنی‌داری آماری آن است. از سوی دیگر، p–مقدار برای θ_0 و θ_2 نیز نوساناتی را تجربه کرده‌اند اما همچنان عمدتاً در محدوده معنی‌دار قرار دارند. این نوسانات می‌تواند به دلیل وجود نقاط دورافتاده در داده‌ها باشد که روی محاسبات آزمون اثر

(الف)



(ب)

شکل ۱. رفتار مقادیر سطح معنی داری برای پارامترهای (θ_0، θ_1، θ_2)، الف: بدون داده دورافتاده و ب: با حضور داده‌های دورافتاده

گذاشته است. به طور کلی، این نتایج نشان می‌دهند که در مدل مورد بررسی، پارامترهای θ_0 و θ_2 نقش معنی‌داری دارند، در حالی که θ_1 تأثیر چشمگیری بر خروجی ندارد و ممکن است حذف آن موجب بهبود مدل شود.

۷ بحث و نتیجه‌گیری

طرح آزمایش کلاسیک به دلیل حساسیت به مفروضاتی مانند نرمال بودن داده‌ها و نقاط دورافتاده، ممکن است نتایج دقیقی ارائه ندهد. برای حل این مشکل، این پژوهش روش‌های آماری مقاوم و همچنین رگرسیون فازی را بررسی کرد. نتایج نشان داد که رگرسیون فازی، به دلیل توانایی ذاتی خود در مدیریت عدم قطعیت و کاهش اثر نقاط دورافتاده، عملکرد بهتری نسبت به روش‌های کلاسیک و مقاوم داشت. اگرچه روش کلاسیک برای داده‌های عادی فاقد دورافتاده برازش بهتری نشان داد، اما خطای پیش‌بینی مدل فازی کمتر بود. رگرسیون فازی برای داده‌های دارای نقاط دورافتاده در تمامی شاخص‌ها، عملکردی برتر از سایر روش‌ها داشت. این پژوهش نشان می‌دهد که رگرسیون فازی ابزاری قدرتمند برای تحلیل داده‌های دارای نقاط دورافتاده است و می‌تواند به عنوان جایگزینی مطمئن برای روش‌های کلاسیک در نظر گرفته شود. توسعه این مدل برای متغیرهای کیفی نیز می‌تواند زمینه تحقیقات آینده را فراهم کند.

تقدیر و تشکر

نویسندگان از نظرات و پیشنهادهای سازنده داوران، سردبیر، هیئت تحریریه و ویراستار محترم مجله که در ارتقا و ارائه بهتر مقاله موثر بودند کمال تشکر را دارند.

مراجع

آریایی‌فر، س. (۱۳۹۷)، تأثیر تیمار هیپوکلریت سدیم بر شاخص‌های طولی گیاهچه. هفتمین کنفرانس ملی مرتع و مرتعداری ایران، ۱۸-۱۹ اردیبهشت ۱۳۹۷، تهران، ایران.

چاچی، ج. و حسامیان، غ. (۱۳۹۲)، مدل‌بندی داده‌های فازی با رگرسیون اسپلاین تطبیقی چندگانه، مجله علوم آماری، ۸(۱)، ۱-۱۸.

قبادی، ش.، جوانمرد، م. (۱۳۸۵)، طراحی و تحلیل آزمایش‌ها در محیط فازی، اولین کنفرانس ملی نگهداری و تعمیرات.

Akbari, M. G., and Hesamian, G. (2019), Elastic Net Oriented to a Fuzzy Semiparametric Regression Model with Fuzzy Explanatory Variables and Fuzzy Responses, *IEEE Transactions on Fuzzy Systems*, **27**(12), 2433-2442.

Barnett, V., and Lewis, T. (1994), Outliers in Statistical Data, John Wiley and Sons.

Bhatti, S. H., Khan, F. W., Irfan, M., and Raza, M. A. (2023), An Effective Approach Towards Efficient Estimation of General Linear Model in Case of Heteroscedastic Errors, *Communications in Statistics-Simulation and Computation*.

Efron, B., and Tibshirani, R. J. (1994), An Introduction to the Bootstrap, *Chapman and Hall/CRC*.

Fisher, R. A. (1935), The Design of Experiments, Oliver and Boyd.

Farnoosh, R., Ghasemian, J., and Solaymanifard, O. (2012), Integrating Ridge-Type Regularization in Fuzzy Nonlinear Regression, *Computational and Applied Mathematics*, **31**, 323-338.

Hesamian, G., and Akbari, M. G. (2020), A Robust Multiple Regression Model Based on Fuzzy Random Variables, *Journal of Computational and Applied Mathematics*, p.113270.

Kacprzyk, J., and Fedrizzi, M.(1992), *Fuzzy Regression Analysis* (2nd ed.), Berlin: Springer-Verlag.

Khammar, A. H., Arefi, M., and Akbari, M. G. (2020), A robust least squares fuzzy regression model based on kernel function, *Iranian Journal of Fuzzy Systems*, **17**, 105-119.

Kashani, M., Arashi, M., and Rabiei, M. R. (2021), Resampling in Fuzzy Regression via Jackknife-after-Bootstrap(JB), *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, **29**(4), 517–535.

Lambert-Lacroix, S., and Zwald, L. (2011), Robust Regression Through the Huber's Criterion and Adaptive Lasso Penalty, *Electronic Journal of Statistics*, **5**, 1015–1053.

Lee, W. J., Jung, H. Y., Yoon, J. H., and et al. (2015), The Statistical Inferences of Fuzzy Regression Based on Bootstrap Techniques, *Soft Computing*, **19**, 883–890. https://doi.org/10.1007/s00500-014-1415-5

Little, R. J. A., and Rubin, D. B. (2019), *Statistical Analysis with Missing Data*, (3rd ed.), Wiley.

Melkumova, L., and Shatskikh, S. (2017), Comparing Ridge and Lasso Estimators for Data Analysis, *Procedia Engineering*, **201**, 746-755.

Montgomery, D. C. (2017), Design and Analysis of Experiments, John Wiley and Sons.

Ross, T. J. (2010), Fuzzy Logic with Engineering Applications, John Wiley and Sons.

Saleh, A. K. M. E., Arashi, M., Saleh, R. A., and Norouzirad, M. (2022), *Rank-Based Methods for Shrinkage and Selection: With Application to Machine Learning*, John Wiley and Sons.

Shen, S. L., Mei, C. L., and Cui, J. L. (2010), A Fuzzy Varying Coefficient Model and its Estimation, *Computers and Mathematics with Applications*, **60**, 1696-1705.

Tanaka, H., Uejima, S., and Asai, K. (1982), Linear Regression Analysis with Fuzzy Model, *IEEE Transactions on Systems, Man, and Cybernetics*, **12**(6), 903-907.

Taheri, S. M., and Chachi, J. (2020), A Robust Variable-Spread Fuzzy Regression Model, In *Recent Developments and the New Direction in Soft Computing Foundations and Applications*(pp. 309-320), Springer, Cham.

Yang, Z., Yin, Y., and Chen, Y. (2013), Robust Fuzzy Varying Coefficient Regression Analysis with Crisp Inputs and Gaussian Fuzzy Output, *Journal of Computing Science and Engineering*, **7**, 263-271.

Zeng, W., Feng, Q., and Li, J. (2017), Fuzzy Least Absolute Linear Regression, *Applied Soft Computing*, **52**, 1009-1019.