

یک مدل رگرسیون خطی با داده‌های فازی نوتروسوفیک

زهرا بهدانی*، مجید دره میرکی

گروه ریاضی و آمار، دانشگاه صنعتی خاتم الانبیاء بهبهان، بهبهان، ایران

تاریخ دریافت: ۱۴۰۲/۰۹/۰۶ تاریخ پذیرش: ۱۴۰۳/۰۵/۰۸

نوع مقاله: علمی-پژوهشی

چکیده. تحلیل رگرسیون آماری روشی شناخته شده برای فرمول بندی رابطه بین متغیر پاسخ (خروجی) و برخی از متغیرهای توضیحی (ورودی) با استفاده از مجموعه مشاهدات بر اساس فرض توزیع‌های نرمال است. رگرسیون خطی فازی اساسی‌ترین روش در حوزه مدل‌سازی فازی است که در آن رابطه نامشخص بین متغیرهای هدف و توضیحی تخمین زده می‌شود و به طور مکرر در طیف گسترده‌ای از کاربردهای دنیای واقعی به طور موثر مورد استفاده قرار گرفته است. در این مقاله ما مدل رگرسیون فازی با ضرایب اعداد فازی نوتروسوفیک را مورد بررسی قرار می‌دهیم برای این کار ابتدا تعمیمی از اندازه فاصله دیاموند را برای این اعداد می‌نویسیم و سپس به کمک روش کمترین توان دوم پارامترهای مدل را که اعداد فازی مثلثی نوتروسوفیک هستند را برآورد می‌نماییم و در نهایت با ذکر یک مثال، کاربردی از مدل ارائه شده را بیان می‌کنیم.

2010 Mathematics Subject Classification. 62A86; 62J86

* Corresponding author

E-mails: behdani@bkatu.ac.ir and darehmiraki@bkatu.ac.ir.

عبارات و کلمات کلیدی. رگرسیون فازی، نوتروسوفیک، اندازه فاصله، اعداد مثلثی.

۱. سرآغاز

آمار، علم جمع آوری، تجزیه و تحلیل، ارائه و تفسیر داده ها است. نیازهای دولت به داده‌های سرشماری و همچنین اطلاعات در مورد انواع فعالیت‌های اقتصادی، انگیزه اولیه را برای حوزه آمار فراهم کرد. در حال حاضر نیاز به تبدیل مقادیر زیادی از داده‌های موجود در بسیاری از زمینه‌های کاربردی به اطلاعات مفید، تحولات نظری و عملی در آمار را تحریک کرده است. آمار توصیفی و آمار استنباطی در کنار آمار ناپارامتری از حوزه های اصلی آمار هستند. آمار توصیفی ابزارهایی را برای توصیف یک نمونه فراهم می کند. اصطلاح آمار توصیفی شامل روش های آماری برای توصیف داده ها با استفاده از ویژگی های آماری، نمودارها، نمودارها یا جداول می شود. با شروع از نمونه، اکنون می توان از آمار استنباطی برای بیان جامعه استفاده کرد. آمار ناپارامتری به روشی آماری اشاره دارد که در آن داده‌ها از مدل‌های تجویز شده‌ای که با تعداد کمی از پارامترها تعیین می‌شوند، به دست نمی‌آیند. نمونه‌هایی از این مدل‌ها شامل مدل توزیع نرمال و مدل رگرسیون خطی است. آمار ناپارامتری گاهی اوقات از داده‌هایی استفاده می‌کند که ترتیبی هستند، به این معنی که بر اعداد متکی نیست، بلکه به رتبه‌بندی یا ترتیبی از انواع متکی است.

تحلیل رگرسیون، تحلیل مولفه‌های اصلی و تحلیل کمترین توان دوم جزئی به طور گسترده‌ای برای تجزیه و تحلیل داده‌ها در زمینه‌های مختلف استفاده شده است. مطالعه تحلیل رگرسیون برای مشاهده تأثیر متغیر مستقل (ها) بر متغیر وابسته (ها) به کار می‌رود. تحلیل رگرسیون یک روش آماری قدرتمند برای بررسی رابطه بین یک یا چند متغیر توضیحی و متغیر وابسته است. این روش به طور گسترده برای اهداف توصیفی، برآورد، کنترل و پیش بینی استفاده می‌شود. کاربردهای این روش‌ها را می‌توان در عبدالوهاب و همکاران [۳]، سرویگون و همکاران [۷]، کومار و چونگ [۱۹] و کاراماکوسکا و همکاران [۱۷] مشاهده کرد.

روش حداقل مربعات فرآیند یافتن یک خط رگرسیون یا بهترین خط برازش برای هر مجموعه داده ای است که توسط یک معادله توصیف می شود. این روش مستلزم کاهش مجموع مجذور قسمت های باقیمانده نقاط از منحنی یا خط است و روند نتایج به صورت کمی پیدا می شود. عدم قطعیت مدل ایجاد شده ناشی از تصادفی بودن ذاتی مشاهدات است. در دنیای واقعی، مشاهدات ممکن است به عنوان مقادیر کمی اندازه‌گیری نشوند، اما در اصطلاحات زبانی (فازی)، مانند "حدود ۱۰ سانتی متر"، "تقریباً برابر با ۵۰ تومان"، "کم"، "بالا" و غیره اندازه‌گیری می‌شوند. توابع عضویت برای توصیف چنین اصطلاحات زبانی تعریف شده است

[۲۶]. برای مقابله با این نوع داده‌ها، تحلیل رگرسیون خطی با مدل‌های فازی [۲۳] برای انجام توابع تحلیل رگرسیون آماری توسعه داده شد. رگرسیون فازی با رگرسیون آماری متفاوت است زیرا از توزیع احتمال تبعیت نمی‌کند. انحرافات (خطاهای تخمینی) به عدم دقت مقادیر مشاهده شده، نامشخص بودن ساختار مدل یا هر دو نسبت داده می‌شود. عدم قطعیت مدل به دلیل فازی بودن مشاهدات است که نویز را به پارامترهای فازی مدل وارد می‌کند. بنابراین، یک مرحله مهم در فرآیندهای برازش مدل، تنظیم پارامترهای فازی برای برازش مدل با استفاده از نمونه‌های موجود است.

در سال ۱۹۶۵، نظریه مجموعه‌های فازی برای بررسی داده‌های مبهم و نادقیق توسط لطفی زاده [۲۶] ارائه شد. پس از لطفی زاده مجموعه‌های فازی به ویژه اعداد فازی به طور گسترده مورد مطالعه قرار گرفته و در زمینه‌های مختلفی مانند تصمیم‌گیری، تشخیص الگو، تئوری بازی‌ها و غیره به کاربرد برده شدند. در مجموعه‌های فازی، درجه عضویت عناصر در یک مجموعه یک مقدار منحصر به فرد است، اما این مقادیر منفرد نمی‌توانند اطلاعات اضافی را ارائه دهند، زیرا در عمل، اطلاعات مربوط به عناصر مربوط به یک مفهوم فازی ممکن است ناقص باشد. تئوری مجموعه‌های فازی قادر به مقابله با کمبود دانش نیست. آتاناسوف [۲] نظریه مجموعه‌های فازی شهودی را با بسط مجموعه‌های فازی لطفی زاده با استفاده از درجه عدم عضویت ارائه کرد تا با وجود ابهام و تردید ناشی از دانش یا اطلاعات نادرست مقابله کند. به دلیل محدودیت « ≥ 0 درجه عضویت + درجه غیرعضویت ≥ 1 »، مجموعه‌های فازی شهودی توسط ساماندراخ در سال ۱۹۹۵ به مجموعه‌های نوتروسوفیک تعمیم داده شد که می‌توانند اطلاعات ناقص را مدیریت کنند [۲۵، ۱۱].

خمر و همکاران [۱۸] تابع هدف را در قالب توابع زیان مختلف و زیر میانگین‌ها معرفی کرد. آنها تفاوت بین α برش خطاها را برای تخمین ضرایب مدل رگرسیون فازی خطی مینیمم کردند. مدل آنها بر روی برخی از برنامه‌های شبیه‌سازی شده و واقعی پیاده‌سازی شده است. جیانگ و همکاران [۱۵] مسئله بهینه‌سازی چند هدفه برای رگرسیون فازی خطی با ورودی‌های دقیق پیشنهاد کردند. مدل آنها شامل سه هدف: کمینه کردن میانگین مطلق، درصد خطای مدل‌سازی، به حداقل رساندن اثر نقاط پرت و به حداقل رساندن فازی بودن خروجی‌های فازی بود. برتری مدل پیشنهادی آنها نسبت به روش‌های رگرسیون فازی تاناکا، رگرسیون فازی پیترز، رگرسیون کمترین توان دوم فازی و رگرسیون فازی احتمالی در کاربرد طراحی یک چایی ساز تایید شده است. چن و نین [۱۰] مسائل برنامه‌ریزی ریاضی را بر اساس کمترین توان دوم مطلق برای برآورد مدل رگرسیون فازی شهودی با ضرایب فازی شهودی پیشنهاد نمودند. عملکرد

روش آنها بر اساس اندازه شباهت و فاصله ارزیابی شد. نویسندگان مقاله [۱۳] از مدل کمترین توان دوم برای تخمین پارامترها در FLR استفاده کردند. نویسندگان [۱۴] رگرسیون ناپارامتری را با پاسخ‌های فازی و پیش‌بینی‌کننده‌های دقیق با اجرای یک روش مرکز و دامنه گسترده در یک روش سه مرحله‌ای بررسی کردند. چاجی و همکاران در [۸] آزمون فرض و فاصله اطمینان را برای ضرایب فازی در مدل رگرسیون وزن‌دار فازی با ورودی‌های دقیق و خروجی‌های فازی مورد بحث قرار دادند.

از آنجایی که پس از پیدایش مجموعه‌های فازی توسط زاده محققان نسخه‌های متفاوتی از مجموعه‌های فازی مانند مجموعه‌های فازی شهودی، فازی نوع ۲، فازی دوقطبی و ... را پیشنهاد نمودند، مدل‌های رگرسیونی برای این مجموعه‌ها نیز ارائه شد که در ادامه به برخی از آنها اشاره می‌کنیم. بجستانی و همکاران در [۴] یک مدل رگرسیونی برای داده‌های فازی نوع ۲ ارائه نمودند. پرواتی و همکاران [۲۱] یک مدل رگرسیونی برای اعداد فازی شهودی ارائه کردند. تاناکا و لی [۲۴] مدل‌های رگرسیونی برای داده‌های بازه‌ای را مورد توجه قرار دادند. مدل رگرسیونی برای اعداد فازی دوقطبی توسط بهدانی در [۵] مورد بررسی قرار گرفت. بوکوزولا و همکاران رگرسیون در محیط فازی بازه‌ای را مورد بررسی و تحلیل قرار دادند [۶]. علیرغم تحقیقات فراوانی که در زمینه تحلیل رگرسیون فازی انجام شده است، تا آنجا که نویسندگان می‌دانند، به جز آثار ناگاراگان و برومی [۲۰]، مطالعه‌ای در زمینه تحلیل رگرسیون در محیط فازی نوتروسوفیک صورت نگرفته است. از آنجایی که اعداد نوتروسوفیک می‌توانند درجات عضویت و عدم عضویت و عدم تعین را برای مفاهیم نادقیق نشان دهند، دانش و بازنمایی معنادارتر و کاربردی‌تر دارند. این تعمیم از اعداد فازی معمولی به طور گسترده‌ای مورد مطالعه قرار گرفته و در زمینه‌های مختلفی مانند برنامه‌نویسی منطقی کاربرد دارند.

در ادامه مقاله در بخش بعد به بیان برخی از مقدمات فازی مربوط به داده‌های فازی نوتروسوفیک می‌پردازیم. در بخش سوم مدل رگرسیونی فازی نوتروسوفیک و اندازه فاصله بین دو عدد نوتروسوفیک را معرفی می‌نماییم. در بخش چهارم، با ذکر یک مثال کاربردی توانایی مدل پیشنهادی را بر روی یک مثال واقعی نمایش می‌دهیم و سرانجام در بخش پنجم مقاله جمع بندی و نتیجه‌گیری انجام می‌شود.

۲. مقدمات فازی

در این بخش به معرفی مفاهیم اساسی اعداد فازی می‌پردازیم که در ادامه مورد استفاده قرار خواهند گرفت. علاوه بر این، برخی از عملیات حسابی و یک متریک فاصله بین اعداد فازی

ارائه شده است. فرض کنیم \tilde{A} یک مجموعه فازی با تابع عضویت $\mu_{\tilde{A}(x)}$ باشد. مجموعه فازی \tilde{A} یک عدد فازی نامیده می شود اگر در شرایط تحدب، نرمال بودن و مرزی صدق کند. عدد فازی \tilde{A} یک عدد فازی LR گفته می شود اگر تابع عضویت آن به صورت زیر باشد.

$$(۱.۲) \quad \mu_{\tilde{A}} = \begin{cases} L\left(\frac{m-x}{s_1}\right), & x \leq m \\ R\left(\frac{x-m}{s_2}\right), & m < x. \end{cases}$$

که $L(x)$ و $R(x)$ توابع نزولی اکید از \mathbb{R}^+ به $[0, 1]$ هستند و همچنین $L(0) = R(0) = 1$ است. s_1 و s_2 به ترتیب پهنای چپ و راست عدد فازی هستند. ما یک عدد فازی را با $\tilde{A} = (m, s_1, s_2)$ نمایش می دهیم. در حالت خاص که عدد فازی مثلثی باشد داریم $L(x) = R(x) = \max(1 - x, 0)$. یک عدد فازی مثلثی نوتروسوفیک از سه تابع عضویت $\mu(x)$ ، عدم عضویت $\nu(x)$ و عدم قطعیت $\tau(x)$ که به صورت زیر تعریف می شوند تشکیل شده است:

$$\mu_{\tilde{A}} = \begin{cases} L\left(\frac{m-x}{s_1}\right), & x \leq m \\ R\left(\frac{x-m}{s_2}\right), & m < x. \end{cases}$$

$$\nu_{\tilde{A}} = \begin{cases} 1 - L\left(\frac{m-x}{s'_1}\right), & x \leq m \\ 1 - R\left(\frac{x-m}{s'_2}\right), & m < x. \end{cases}$$

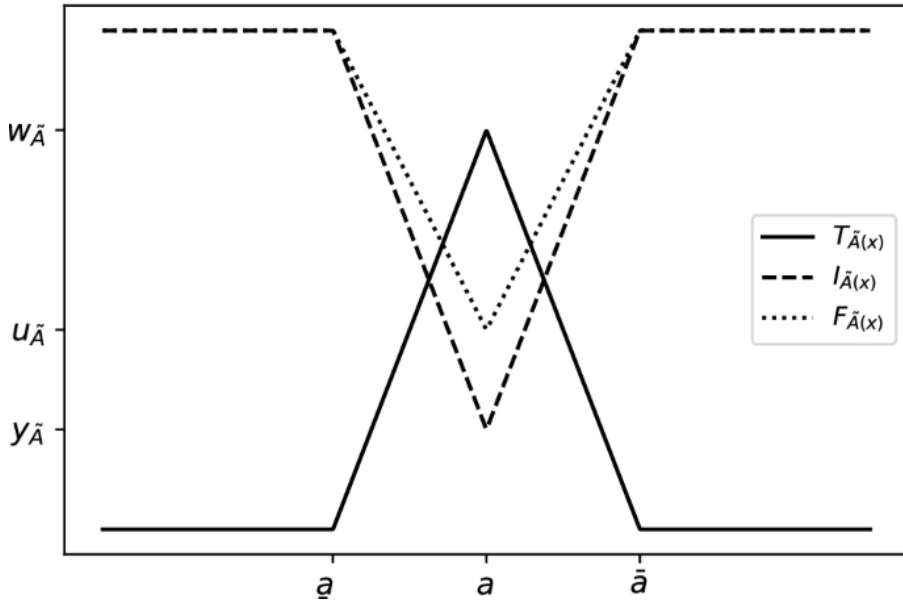
$$\tau_{\tilde{A}} = \begin{cases} 1 - L\left(\frac{m-x}{s''_1}\right), & x \leq m \\ 1 - R\left(\frac{x-m}{s''_2}\right), & m < x. \end{cases}$$

شکل ۱ یک عدد فازی مثلثی نوتروسوفیک را نمایش می دهد. عملیات جبری روی اعداد فازی نوتروسوفیک به صورت زیر تعریف می شود:

تعریف ۱.۲. فرض کنیم $\tilde{A} = (a, s_1, s_2, s'_1, s'_2, s''_1, s''_2)$ و $\tilde{B} = (b, r_1, r_2, r'_1, r'_2, r''_1, r''_2)$ دو عدد فازی نوتروسوفیک و λ یک عدد حقیقی دلخواه باشد در این صورت داریم:

$$\tilde{A} + \tilde{B} = (a+b, s_1+r_1, s_2+r_2, s'_1+r'_1, s'_2+r'_2, s''_1+r''_1, s''_2+r''_2) \quad (۱)$$

$$\tilde{A} - \tilde{B} = (a-b, s_1-r_2, s_2-r_1, s'_1-r'_2, s'_2-r'_1, s''_1-r''_2, s''_2-r''_1) \quad (۲)$$



شکل ۱: یک عدد فازی نوتروسوفیک

(۳)

$$\lambda \tilde{A} = \begin{cases} (\lambda a, \lambda s_1, \lambda s_2, \lambda s'_1, \lambda s'_2, \lambda s''_1, \lambda s''_2) & \lambda \geq 0 \\ (-\lambda a, -\lambda s_1, -\lambda s_2, -\lambda s'_1, -\lambda s'_2, -\lambda s''_1, -\lambda s''_2) & \lambda < 0 \end{cases}$$

تعریف ۲.۲. فرض کنیم $\tilde{A} = (a, s_1, s_2, s'_1, s'_2, s''_1, s''_2)$ یک عدد فازی نوتروسوفیک LR باشد در این صورت α برش این عدد به صورت زیر تعریف می شوند:

$$\begin{aligned} \tilde{A}_{\mu} &= [a - s_1 L^{-1}(\alpha), a - s_2 R^{-1}(\alpha)], \\ \tilde{A}_{1-\nu} &= [a - s'_1 L^{-1}(\alpha), a - s'_2 R^{-1}(\alpha)], \\ \tilde{A}_{1-\tau} &= [a - s''_1 L^{-1}(\alpha), a - s''_2 R^{-1}(\alpha)], \end{aligned}$$

جدول ۱: تفاوت میان نگرش‌های مختلف آمار: قطعی، فازی، فازی شهودی و فازی نوتروسوفیک

۱	آمار قطعی	آمار کلاسیک برای تجزیه و تحلیل داده‌ها زمانی که همه مشاهدات و پارامترهایی که در نمونه یا جامعه وجود دارند دقیق و قطعی باشند استفاده می‌شود.
۲	آمار فازی	آمار فازی برای تجزیه و تحلیل داده‌های دارای مشاهدات (پارامترهای) غیردقیق، نامطمئن و فازی استفاده می‌شود. آمار بر اساس آمار فازی است و معیار عدم قطعیت را در نظر نمی‌گیرد.
۳	آمار فازی شهودی	فازی شهودی (IF) بسط منطق فازی کلاسیک است. عضویت و عدم عضویت را متعلق به بازه واحد واقعی در نظر می‌گیرد. بنابراین آمارهای مبتنی بر IF بسط آمار فازی هستند.
۴	آمار فازی نوتروسوفیک	آمار نوتروسوفیک (NS) بر اساس ایده منطق نوتروسوفیک است. منطق نوتروسوفیک بسط منطق فازی است و معیار عدم تعین را در نظر می‌گیرد. بنابراین، NS گسترش آمار کلاسیک است که به تجزیه و تحلیل در شرایط عدم قطعیت می‌پردازد.

۳. رگرسیون فازی

در برخی از سیستم‌های واقعی، به ویژه سیستم‌های محیطی، داده‌های موجود لزوماً به عنوان مقادیر دقیق جمع‌آوری (گزارش) نمی‌شوند. در چنین شرایطی، ما نیاز به توسعه رویکردهای مناسب جدید برای مقابله با داده‌های غیر دقیق داریم. در این مقاله، مدل‌سازی رگرسیونی جدیدی معرفی شده است که می‌تواند در مواردی که مقادیر متغیر (های) مستقل و همچنین مقادیر متغیر وابسته (پاسخ) به صورت اعداد فازی نوتروسوفیک مشاهده (گزارش) می‌شوند، مورد استفاده قرار گیرد. لازم به ذکر است که تا جایی که نویسندگان اطلاعات دارند، تنها چند اثر با موضوع مدل‌سازی رگرسیون در محیط‌های نوتروسوفیک وجود دارد. ناگاراچان و برومی [۲۰] یک مدل رگرسیون فازی را برای مجموعه‌های فازی نوتروسوفیک بررسی کردند. هیچ معیار نیکویی برازشی در کار آنها وجود ندارد، اما برای ارزیابی مدل‌های به‌دست‌آمده، ما در اینجا یک شاخص نیکویی برازش برای داده‌های فازی نوتروسوفیک معرفی می‌نماییم. مدل‌های رگرسیون فازی بر اساس شرایط متغیرهای مستقل و وابسته به سه دسته طبقه‌بندی شود:

(۱) داده‌های ورودی غیر فازی اما داده‌های خروجی فازی هستند.

(۲) داده‌های ورودی و خروجی هر دو غیر فازی هستند.

(۳) داده‌های ورودی و خروجی هر دو فازی هستند.

ما در این مقاله از مدل رگرسیون فازی برای دسته اول استفاده خواهیم کرد. مدل رگرسیون فازی که در آن داده‌های ورودی غیر فازی و داده‌های خروجی فازی هستند به صورت زیر بیان می‌شود:

$$(۱.۳) \quad Y_i = A_0 + A_1 x_{i1} + A_2 x_{i2} + \dots + A_p x_{ip} + \epsilon_i, \quad i = 1, 2, \dots, n$$

که x_{ij} ها متغیرهای مستقل و اعداد حقیقی مثبت و A_j اعداد مثلثی نوتروسوفیک می‌باشند. در این مدل Y_i متغیر وابسته عدد فازی نوتروسوفیک $(y_i, r_{i1}, r_{i2}, r'_{i1}, r'_{i2}, r''_{i1}, r''_{i2})$ است و ϵ خطای تصادفی است.

در اینجا ما $A_j = (a_j, s_{j1}, s_{j2}, s'_{j1}, s'_{j2}, s''_{j1}, s''_{j2})$ را اعدادی فازی مثلثی نوتروسوفیک اختیار می‌نماییم در این صورت با توجه به مطالب بخش قبل داریم:

$$(۲.۲) \quad \sum_{j=1}^p A_j x_{ij} = \left(\sum_{j=1}^p a_j, \sum_{j=1}^p s_{j1} x_{ij}, \sum_{j=1}^p s_{j2} x_{ij}, \sum_{j=1}^p s'_{j1} x_{ij}, \sum_{j=1}^p s'_{j2} x_{ij}, \sum_{j=1}^p s''_{j1} x_{ij}, \sum_{j=1}^p s''_{j2} x_{ij} \right)$$

برای برآورد ضرایب رگرسیون در ۱.۳ از روش کمترین توان دوم خطا استفاده می‌کنیم. برای این منظور لازم است که ابزاری برای اندازه‌گیری فاصله دو اعداد فازی را معرفی کنیم. برای این منظور تعمیمی از فاصله دیاموند [۱۲] را ارائه می‌دهیم.

تعریف ۱.۳. فرض کنیم $\tilde{A} = (a, s_1, s_2, s'_1, s'_2, s''_1, s''_2)$ و $\tilde{B} = (b, r_1, r_2, r'_1, r'_2, r''_1, r''_2)$ دو عدد فازی نوتروسوفیک باشند. فاصله بین این دو عدد به صورت زیر تعریف می‌شود:

$$(۳.۳) \quad d^2(\tilde{A}, \tilde{B}) = (a - b)^2 + (s_1 - r_1)^2 + (s'_1 - r'_1)^2 + (s''_1 - r''_1)^2 + (s_2 - r_2)^2 + (s'_2 - r'_2)^2 + (s''_2 - r''_2)^2.$$

برای برآورد پارامترها، تابع کمترین توان دوم خطا به صورت زیر نوشته می‌شود:

$$\begin{aligned}
 S(A_0, A_1, \dots, A_p) &= \sum_{i=1}^n d^2(\tilde{y}_i, \sum_{j=0}^p A_j x_{ij}) \\
 &= \sum_{i=1}^n (y_i - \sum_{j=0}^p a_j x_{ij})^2 + \sum_{i=1}^n (s_{i1} - \sum_{j=0}^p r_{1j} x_{ij})^2 \\
 &\quad + \sum_{i=1}^n (s'_{i1} - \sum_{j=0}^p r'_{1j} x_{ij})^2 + \sum_{i=1}^n (s''_{i1} - \sum_{j=0}^p r''_{1j} x_{ij})^2 \\
 &\quad + \sum_{i=1}^n (s_{i2} - \sum_{j=0}^p r_{2j} x_{ij})^2 + \sum_{i=1}^n (s'_{i2} - \sum_{j=0}^p r'_{2j} x_{ij})^2 \\
 &\quad + \sum_{i=1}^n (s''_{i2} - \sum_{j=0}^p r''_{2j} x_{ij})^2.
 \end{aligned}
 \tag{۴.۳}$$

تابع S نسبت به A_0, A_1, \dots, A_p مینیمم می‌شود (در این رابطه $x_{i0} = 1$ است). برآوردگر کمترین توان دوم A_0, A_1, \dots, A_p بایستی در رابطه زیر صدق کند:

$$\frac{\partial S}{\partial A_j} \Big|_{\hat{A}_0, \hat{A}_1, \dots, \hat{A}_p} = 0
 \tag{۵.۳}$$

در مدل‌های رگرسیون چندگانه بهتر است آنها را با نماد ماتریسی نوشت. زیرا این امکان را فراهم می‌سازد که داده‌ها و نتایج فشرده درآیند. لذا اکنون یک صورت ماتریسی از مدل ارائه می‌دهیم.

$$\tilde{y} = XA + \epsilon$$

که در آن

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{bmatrix} \quad R_1 = \begin{bmatrix} r_{11} \\ r_{21} \\ \cdot \\ \cdot \\ \cdot \\ r_{n1} \end{bmatrix} \quad R_2 = \begin{bmatrix} r_{12} \\ r_{22} \\ \cdot \\ \cdot \\ \cdot \\ r_{n2} \end{bmatrix}$$

$$R_{\lambda}^* = \begin{bmatrix} r'_{\lambda\lambda} \\ r'_{\lambda\gamma} \\ \cdot \\ \cdot \\ r'_{n\lambda} \end{bmatrix} \quad R_{\gamma}^* = \begin{bmatrix} r'_{\lambda\gamma} \\ r'_{\gamma\gamma} \\ \cdot \\ \cdot \\ r'_{n\gamma} \end{bmatrix} \quad R_{\lambda}^{**} = \begin{bmatrix} r''_{\lambda\lambda} \\ r''_{\lambda\gamma} \\ \cdot \\ \cdot \\ r''_{n\lambda} \end{bmatrix} \quad R_{\gamma}^{**} = \begin{bmatrix} r''_{\lambda\gamma} \\ r''_{\gamma\gamma} \\ \cdot \\ \cdot \\ r''_{n\gamma} \end{bmatrix}$$

$$X = \begin{bmatrix} \lambda & x_{\lambda\lambda} & x_{\lambda\gamma} & \cdots & x_{\lambda m} \\ \lambda & x_{\gamma\lambda} & x_{\gamma\gamma} & \cdots & x_{\gamma m} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \lambda & x_{n\lambda} & x_{n\gamma} & \cdots & x_{nm} \end{bmatrix}$$

$$A = \begin{bmatrix} a_{\cdot} \\ a_{\lambda} \\ \cdot \\ \cdot \\ \cdot \\ a_p \end{bmatrix} \quad S_{\lambda} = \begin{bmatrix} s_{\cdot\lambda} \\ s_{\lambda\lambda} \\ \cdot \\ \cdot \\ \cdot \\ s_{p\lambda} \end{bmatrix} \quad S_{\gamma} = \begin{bmatrix} s_{\cdot\gamma} \\ s_{\lambda\gamma} \\ \cdot \\ \cdot \\ \cdot \\ s_{p\gamma} \end{bmatrix}$$

$$S_{\lambda}^* = \begin{bmatrix} s'_{\cdot\lambda} \\ s'_{\lambda\lambda} \\ \cdot \\ \cdot \\ \cdot \\ s'_{p\lambda} \end{bmatrix} \quad S_{\gamma}^* = \begin{bmatrix} s'_{\cdot\gamma} \\ s'_{\lambda\gamma} \\ \cdot \\ \cdot \\ \cdot \\ s'_{p\gamma} \end{bmatrix} \quad S_{\lambda}^{**} = \begin{bmatrix} s''_{\cdot\lambda} \\ s''_{\lambda\lambda} \\ \cdot \\ \cdot \\ \cdot \\ s''_{p\lambda} \end{bmatrix} \quad S_{\gamma}^{**} = \begin{bmatrix} s''_{\cdot\gamma} \\ s''_{\lambda\gamma} \\ \cdot \\ \cdot \\ \cdot \\ s''_{p\gamma} \end{bmatrix}$$

با استفاده از این نمادها تابع کمترین توان دوم خطا به صورت زیر بازنویسی می‌شود:

$$\begin{aligned}
 S(A_0, A_1, \dots, A_p) &= (Y - AX)'(Y - AX) + (R_1 - S_1 X)'(R_1 - S_1 X) \\
 &\quad + (R_2 - S_2 X)'(R_2 - S_2 X) + (R_1^* - S_1^* X)'(R_1^* - S_1^* X) \\
 &\quad + (R_2^* - S_2^* X)'(R_2^* - S_2^* X) + (R_1^{**} - S_1^{**} X)'(R_1^{**} - S_1^{**} X) \\
 &\quad + (R_2^{**} - S_2^{**} X)'(R_2^{**} - S_2^{**} X).
 \end{aligned}
 \tag{۶.۳}$$

با مشتق گیری از عبارت بالا بر حسب پارامترهای مدل، برآورد پارامترهای مجهول به صورت زیر برآورد می‌شود:

$$\begin{aligned}
 \hat{A} &= (X'X)^{-1} X'Y & \hat{S}_1 &= (X'X)^{-1} X'R_1 \\
 \hat{S}_2 &= (X'X)^{-1} X'R_2 & \hat{S}_1^* &= (X'X)^{-1} X'R_1^* \\
 \hat{S}_1^{**} &= (X'X)^{-1} X'R_1^{**} & \hat{S}_2^{**} &= (X'X)^{-1} X'R_2^{**}
 \end{aligned}$$

ملاحظه ۲.۳. توجه به این نکته ضروری است که برای برآورد پارامترها، مقادیر متغیرهای مستقل مثبت در نظر گرفته شده است. مطمئناً اگر هر یک از مقادیر منفی باشند در محاسبات و برآوردها تغییرات جزئی با توجه به تعریف ۱.۲ باید صورت گیرد.

۴. اندازه‌گیری کیفیت برازش

بعد از برازش الگوی رگرسیون، قبل از آنکه از مدل برای برازش و برآورد استفاده شود لازم است که کیفیت آن ارزیابی شود. کیفیت برازش الگو را می‌توان با ابزارهای زیادی ارزیابی نمود. از جمله می‌توان میانگین توان دوم خطا و ضریب تعیین را نام برد. میانگین توان دوم خطا در واقع خطای پیشگویی را برآورد می‌کند و برآوردی از واریانس خطاها و متغیر پاسخ نیز است. ضریب تعیین که در رگرسیون ساده برابر با توان دوم ضریب همبستگی بین متغیر پاسخ و مقادیر برازش شده است، را می‌توان به عنوان نسبت کل تغییرات در متغیر پاسخ که توسط متغیرهای پیشگوی بیان می‌شود تفسیر کرد. می‌توانیم ضریب تعیین (R^2) را به صورت زیر بنویسیم:

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}
 \tag{۱.۴}$$

که در آن $SST = \sum_{i=1}^n (y_i - \bar{y})^2$ کل مجموع انحرافات در y است. همچنین $SSR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$ و $SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2$ در ادامه برای محاسبه ضریب تعیین برای

مدل رگرسیونی که متغیر پاسخ و ضرایب مدل اعداد فازی مثلثی نوتروسوفیک از تعریف ۱.۳ استفاده خواهیم کرد. $SST = \sum_{i=1}^n d^2(\tilde{y}_i, \bar{y})$ همچنین $SSR = \sum_{i=1}^n d^2(\hat{y}_i - \bar{y})^2$ و $SSE = \sum_{i=1}^n d^2(\tilde{y}_i - \hat{y}_i)^2$

۵. مثال عددی

در این بخش برای نشان دادن کارایی مدل پیشنهادی مثالی با داده‌های فازی نوتروسوفیک حل می‌نماییم.

مثال ۱.۵. به منظور نشان دادن کاربرد مدل رگرسیونی فازی ارائه شده، مثالی از [۱۶] را در نظر می‌گیریم که اثر ترکیب سیمان پرتلند را بر حرارت ایجاد شده در طول سخت شدن تحلیل می‌کند. مجموعه داده در جدول ۳ ارائه شده است. در این مجموعه داده، خروجی فازی مثلثی متقارن \tilde{Y} بیانگر گرمای حاصل از کالری در هر گرم سیمان است. ورودی x_2 ، x_1 و x_3 به ترتیب میزان تری کلسیم آلومینات، مقدار تری کلسیم سیلیکات و مقدار تترا کلسیم آلومینوفریت را نشان می‌دهند. البته در مقاله اصلی داده‌ها فازی مثلثی معمولی بوده‌اند و ما در اینجا آنها را تبدیل به فازی مثلثی نوتروسوفیک نموده‌ایم. برآورد پارامترهای مدل به روش

جدول ۲: داده‌های مثال ۱.۵

\tilde{Y}	x_3	x_2	x_1	ردیف
(۷۸/۵, ۶/۵, ۶/۲, ۵, ۲/۳, ۴/۳, ۳/۶)	۶	۲۶	۷	۱
(۷۴/۳, ۶/۴, ۶/۶, ۵/۹, ۳/۴, ۵/۲, ۲/۹)	۱۵	۲۹	۱	۲
(۱۰۴/۳, ۹/۴, ۹, ۹/۵, ۲/۵, ۶/۳, ۴/۷)	۶	۲۶	۷	۳
(۸۷/۶, ۷/۸, ۸, ۸/۱, ۶/۵, ۷, ۷/۲)	۸	۳۱	۱۱	۴
(۸۷/۶, ۷/۸, ۷, ۸, ۴/۵, ۶/۵, ۷/۲)	۶	۵۲	۷	۵
(۱۰۹/۲, ۹/۹, ۹/۲, ۹/۵, ۸/۷, ۷/۴, ۶/۹)	۹	۵۵	۱۱	۶
(۱۰۲/۷, ۹/۳, ۹/۷, ۹, ۴/۶, ۷/۵, ۶/۳)	۱۷	۷۱	۳	۷
(۷۲/۵, ۶/۲, ۶, ۶/۶, ۴/۹, ۶/۵, ۶/۳)	۲۲	۳۱	۱	۸
(۹۳/۱, ۸/۳, ۹, ۸/۷, ۶/۷, ۸/۳, ۴/۸)	۱۸	۵۴	۲	۹
(۱۱۵/۹, ۱۰/۶, ۱۰, ۱۱, ۹/۸, ۸/۷, ۷/۵)	۴	۴۷	۲۱	۱۰
(۸۳/۸, ۷/۴, ۷, ۷/۶, ۵, ۶/۴, ۷/۶)	۲۳	۴۰	۱	۱۱
(۱۱۳/۳, ۱۰/۶, ۱۰, ۱۰/۹, ۹/۸, ۸/۷, ۷/۹)	۹	۶۶	۱۱	۱۲
(۱۰۹/۴, ۹/۹, ۱۱, ۸, ۷, ۷/۶, ۴/۷)	۸	۶۸	۱۰	۱۳

کمترین توان دوم خطا برابر است با:

$$\hat{A}_1 = (۶۲/۶۴۳۴; ۴/۹۳۶۸, ۴/۶۰۲۴, ۳/۶۲۷۰, -۳/۶۲۲۴, ۲/۱۲۲۷, ۰/۹۳۸۱)$$

$$\hat{A}_2 = (۱/۴۵۷۲; ۰/۱۵۴۶, ۰/۱۱۶۴, ۰/۲۳۸۵, ۰/۵۰۶۴, ۰/۱۹۵۵, ۰/۲۷۳۲)$$

$$\hat{A}_3 = (۰/۵۰۴۳; ۰/۰۵۳۹۵, ۰/۰۶۵۴۱, ۰/۰۴۳۲, ۰/۰۵۵۴, ۰/۰۴۲۸, ۰/۰۱۸۳۷)$$

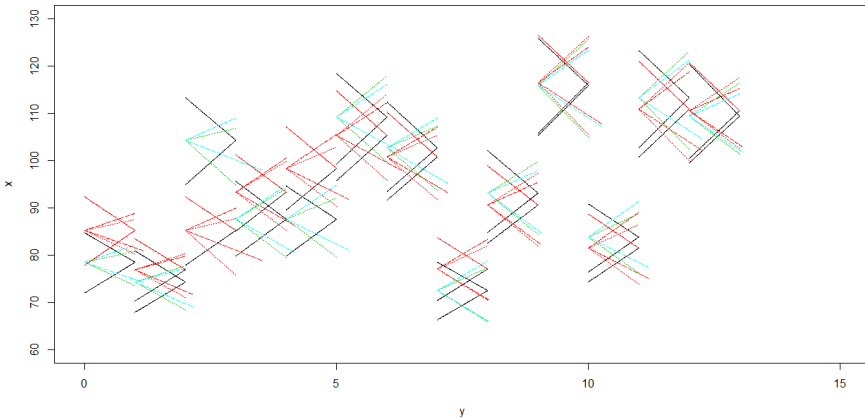
$$\hat{A}_4 = (-۰/۱۲۱, -۰/۰۰۴۱, -۰/۰۰۶۳, ۰/۰۸۴۴, ۰/۲۸۲۷, ۰/۱۲۶۴, ۰/۱۹۲۳)$$

مقادیر برازش شده براساس این مدل در جدول ۱.۵ ثبت شده اند. شکل ۲ مقادیر مشاهده شده و مقادیر برازش شده را با استفاده از برآوردهای بدست آمده برای ضرایب مدل نشان می‌دهد. در این شکل مقادیر برازش شده با رنگ قرمز مشخص شده‌اند.

جدول ۳: مقادیر برازش شده برای داده‌های مثال ۱.۵

ردیف	\hat{Y}
۱	(۸۵/۲۳; ۷/۴, ۷/۰۸, ۶/۹۲, ۳/۰۵, ۵/۳۶, ۴/۴۸)
۲	(۷۶/۹۱; ۶/۶, ۶/۵۲, ۶/۳۸, ۲/۷۳, ۵/۴۶, ۴/۶۳)
۳	(۸۵/۲۳; ۷/۳, ۷/۰۸, ۶/۹۲, ۳/۰۶, ۵/۳۶, ۴/۴۸)
۴	(۹۳/۳۴; ۸/۲۸, ۷/۸۶, ۸/۲۶, ۵/۹۳, ۶/۶۱, ۶/۰۵)
۵	(۹۸/۳۴; ۸/۸, ۸/۷۸, ۸/۰۵, ۴/۵, ۶/۴۸, ۴/۹۶)
۶	(۱۰۵/۳۲; ۹/۵۷, ۹/۴۲, ۹/۳۸, ۷/۵۴, ۷/۷۷, ۶/۶۸)
۷	(۱۰۰/۷۶; ۹/۱۶, ۹/۴۸, ۸/۸۴, ۶/۶۴, ۷/۹, ۶/۳۳)
۸	(۷۷/۰۷; ۶/۶۷, ۶/۶۱, ۷/۰۶, ۴/۸۲, ۶/۴۳, ۶/۰۱)
۹	(۹۰/۶۱; ۸/۰۹, ۸/۲۵, ۷/۹۵, ۵/۴۷, ۷/۱, ۵/۹۴)
۱۰	(۱۱۶/۴۶; ۱۰/۷, ۱۰/۱, ۱۱, ۱۰/۷۵, ۸/۷۵, ۸/۳۱)
۱۱	(۸۱/۴۹; ۷/۱۶, ۷/۱۹, ۷/۵۳, ۵/۶, ۶/۹۴, ۶/۳۷)
۱۲	(۱۱۰/۸۷; ۱۰/۱۶, ۱۰/۱۴, ۹/۸۶, ۸/۱۵, ۸/۲۴, ۶/۸۹)
۱۳	(۱۱۰/۵۴; ۱۰/۱۲, ۱۰/۱۶, ۹/۶۲, ۷/۴۷, ۸, ۶/۴۶)

برای بررسی نیکویی برازش مدل معیارهای کیفیت برازش را برای این داده‌ها محاسبه کردیم. برای این منظور ابتدا میانگین مقادیر مشاهده شده برای متغیر پاسخ را به دست می‌آوریم که برابر است با $\bar{y} = (۹۴/۷۸; ۸/۴۷, ۸/۳۶, ۸/۲۹, ۵/۸۲, ۶/۹۵, ۵/۹۷)$ که با توجه به آن مقدار $SST = ۳۰۰۱/۳۹$ و همچنین $SSE = ۶۸۹/۵۸$ است. لذا برای این مدل ضریب



شکل ۲: مقادیر مشاهده شده و مقادیر برازش شده نوتروسوفیک (رنگ قرمز) برای داده های جدول ۱.۵

تعیین $R^2 = 0.77$ است. این مقدار بیان می کند که ۷۷ درصد از تغییرات متغیرها توسط مدل پیشگویی می شود.

۶. نتیجه گیری

امروزه با گسترش روز افزون کاربرد علوم داده در علوم و صنایع مختلف، مدل های رگرسیونی که از جمله زیر شاخه های این علم می باشد نیز مورد توجه بسیاری از محققین قرار گرفته اند. با توجه به اینکه برخی از داده ها ذاتا از عدم قطعیت برخوردارند لذا مدل های رگرسیونی فازی نیز به موضوعات روز علم داده تبدیل شده اند. در این مقاله ما یک مدل رگرسیونی برای داده های ورودی حقیقی و خروجی فازی نوتروسوفیک پیشنهاد نموده ایم. برای این منظور ابتدا اندازه فاصله دیاموند را برای محاسبه فاصله بین دو عددی فازی نوتروسوفیک بهبود بخشیدیم و سپس به کمک روش کمترین مربعات پارامترهای مدل را برآورد کردیم. در نهایت با ذکر یک مثال کاربردی توانایی مدل پیشنهادی بررسی شده است. نتایج نشان می دهد که روش معرفی شده با دقت مطلوبی پارامترهای مجهول را برآورد می کند.

مراجع

- [1] Ahmed, J. (2021). LR-type fully single-valued neutrosophic linear programming problems. *Neutrosophic Sets and Systems*, 46(1), 416-444.
- [2] Atanassov, K. T. (2012). *On intuitionistic fuzzy sets theory* (Vol. 283). Springer.
- [3] Abdul-Wahab, S. A., Bakheit, C. S., & Al-Alawi, S. M. (2005). Principal component and multiple regression analysis in modelling of ground-level ozone and factors affecting its concentrations. *Environmental Modelling & Software*, 20(10), 1263-1271.
- [4] Bajestani, N. S., Kamyad, A. V., & Zare, A. (2017). A piecewise type-2 fuzzy regression model. *International Journal of Computational Intelligence Systems*, 10(1), 734-744.
- [5] Behdani, Z. The Least Squares Method for Estimating Regression Model Parameters with Bipolar Fuzzy Numbers. Available at SSRN 4717871.
- [6] Boukezzoula, R., & Coquin, D. (2021). Interval-valued fuzzy regression: Philosophical and methodological issues. *Applied Soft Computing*, 103, 107145.
- [7] Cervigón, R., Moreno, J., Reilly, R. B., Perez-Villacastin, J., & Castells, F. (2012). Quantification of anaesthetic effects on atrial fibrillation rate by partial least-squares. *Physiological Measurement*, 33(10), 1757.
- [8] Chachi, J., Akhond, M. R., & Handali, K. (2023). Statistical inference of fuzzy weighted regression based on bootstrap approach. *Fuzzy Systems and its Applications*.
- [9] Chen, L. H., & Nien, S. H. (2020). Approach for establishing intuitionistic fuzzy linear regression models based on weakest t-norm arithmetic. *IEEE Transactions on Fuzzy Systems*, 29(6), 1431-1445.
- [10] Chen, L. H., & Hsueh, C. C. (2009). Fuzzy regression models using the least-squares method based on the concept of distance. *IEEE Transactions on Fuzzy Systems*, 17(6), 1259-1272.
- [11] Darehmiraki, M. (2020). A solution for the neutrosophic linear programming problem with a new ranking function. In *Optimization theory based on Neutrosophic and Plithogenic Sets* (pp. 235-259). Academic Press.
- [12] Diamond, P. (1988). Fuzzy least squares. *Information sciences*, 46(3), 141-157.
- [13] Hose, D., & Hanss, M. (2019). Fuzzy linear least squares for the identification of possibilistic regression models. *Fuzzy Sets and Systems*, 367, 82-95.
- [14] Hesamian, G., Torkian, F., Johannssen, A., & Chukhrova, N. (2024). A fuzzy nonparametric regression model based on an extended center and range method. *Journal of Computational and Applied Mathematics*, 436, 115377.
- [15] Jiang, L., & Liao, H. (2020). Mixed fuzzy least absolute regression analysis with quantitative and probabilistic linguistic information. *Fuzzy Sets and Systems*, 387, 35-48.

- [16] Kong, L. (2022). Fuzzy linear regression model based on adaptive lasso method. *International Journal of Fuzzy Systems*, 24(1), 508-518.
- [17] Karamacoska, D., Barry, R. J., & Steiner, G. Z. (2019). Using principal components analysis to examine resting state EEG in relation to task performance. *Psychophysiology*, 56(5), e13327.
- [18] Khammar, A. H., Arefi, M., & Akbari, M. G. (2020). A robust least squares fuzzy regression model based on kernel function. *Iranian Journal of Fuzzy Systems*, 17(4), 105-119.
- [19] Kumar, S., & Chong, I. (2018). Correlation analysis to identify the effective data in machine learning: Prediction of depressive disorder and emotion states. *International journal of environmental research and public health*, 15(12), 2907.
- [20] Nagarajan, D., Broumi, S., Smarandache, F., & Kavikumar, J. (2021). Analysis of neutrosophic multiple regression. *Neutrosophic Sets and Systems*, 43, 44-53.
- [21] Parvathi, R., Malathi, C., Akram, M., & Atanassov, K. T. (2013). Intuitionistic fuzzy linear regression analysis. *Fuzzy Optimization and Decision Making*, 12, 215-229.
- [22] Smarandache, F. (2014). Introduction to neutrosophic statistics. *Infinite Study*.
- [23] Tanaka, H., & Watada, J. (1988). Possibilistic linear systems and their application to the linear regression model. *Fuzzy sets and systems*, 27(3), 275-289.
- [24] Tanaka, H., & Lee, H. (1998). Interval regression analysis by quadratic programming approach. *IEEE Transactions on Fuzzy Systems*, 6(4), 473-481.
- [25] Wang, H., Smarandache, F., Zhang, Y., & Sunderraman, R. (2010). Single valued neutrosophic sets. *Infinite study*, 12.
- [26] Zadeh, L. A. (1979). Fuzzy sets and information granularity. *Fuzzy sets, fuzzy logic, and fuzzy systems: selected papers*, 433-448.