

تحلیل رگرسیون استوار فازی با داده‌های خروجی و پارامترهای فازی بر پایه رتبه‌بندی مجموعه‌های فازی

الهام رحیمیان^۱، محمدرضا ربیعی^۲، داود شاهسونی^۳

تاریخ دریافت: ۱۳۹۵/۷/۷

تاریخ پذیرش: ۱۳۹۶/۴/۲۴

چکیده:

هنگامی که در مجموعه داده‌ها، مشاهدات دور افتاده وجود دارند روش رگرسیون استوار، جایگزین مناسبی برای رگرسیون معمولی است. همچنین اگر مشاهدات، فازی باشند نیز روش‌های رگرسیون معمول، نمی‌توانند راه‌گشای مدل‌بندی اینگونه از مشاهدات باشند و در این حالت روش رگرسیون فازی، روش جایگزین مناسبی است. برای حالتی که مشاهدات، فازی بوده و در مجموعه داده‌ها، مشاهدات دور افتاده وجود داشته باشند از روش‌های جایگزین استوار فازی استفاده می‌شود. در این مقاله برای حالتی که متغیرهای وابسته و ضرایب رگرسیونی اعداد فازی بوده و مجموعه داده‌ها حاوی مشاهدات دور افتاده است، تحلیل رگرسیون کمترین توان‌های دوم فازی اصلاح شده‌ای مطرح می‌شود. در این روش برای مقایسه مجموعه‌های فازی، باقی‌مانده‌ها رتبه‌بندی می‌شوند. باقی‌مانده‌ها با استفاده از شاخص حضور سراسری برای هر مجموعه فازی (OM^۴) به دست می‌آیند. سپس ماتریس وزن توسط تابع عضویت باقیمانده‌ها تعریف می‌شود و براوردهای کمترین توان‌های دوم فازی موزون با استفاده از ماتریس وزن بدست می‌آیند. برای نشان دادن عملکرد روش پیشنهادی، دو مثال را مطرح و نتایج حاصل از آنها ارائه می‌شود.

واژه‌های کلیدی: رگرسیون استوار، داده دور افتاده، رگرسیون فازی، شاخص OM.

۱ مقدمه

هستند؛ لذا حذف آنها امکان‌پذیر نیست و در این صورت از روش‌های جایگزین رگرسیون استوار استفاده می‌کنیم. در واقع در این موارد، روش‌های استوار به روش LS ترجیح داده می‌شوند. حال اگر متغیرها یا مشاهدات مربوط به آنها دقیق نباشند نیز نمی‌توان از روش‌های کلاسیک استفاده کرد. رگرسیون فازی، یک روش جایگزین مناسب برای این حالت است. بنا بر این هنگامی که در مجموعه داده‌ها، مشاهدات دورافتاده وجود داشته باشند و متغیرها یا مشاهدات مربوط به آنها نادقیق باشند از تحلیل رگرسیون استوار فازی استفاده می‌شود.

اولین پژوهش در مورد تحلیل رگرسیون خطی فازی توسط تاناکا

روش کمترین توان‌های دوم (LS^۵)، روشی معمول است که سال‌هاست برای برآورد پارامترهای رگرسیون به کار می‌رود. اگر مفروضات روش LS برای یک مجموعه داده برقرار باشد، برآوردهای روش LS، به‌عنوان بهترین برآوردها شناخته می‌شوند. اما اگر مشاهدات دور افتاده در مجموعه داده‌ها وجود داشته باشند، روش متعارف، پیش‌بینی‌های خوبی را نتیجه نمی‌دهد. در این روش، مشاهدات دور افتاده از مجموعه داده‌ها حذف می‌شوند و پس از آن روش کلاسیک می‌تواند به کار رود. ولی در برخی تحقیقات، این مشاهدات حاوی اطلاعات مفیدی

^۱ دانشجوی کارشناسی ارشد گروه آمار، دانشگاه صنعتی شاهرود، ایران

^۲ عضو هیئت علمی گروه آمار، دانشگاه صنعتی شاهرود، ایران

^۳ عضو هیئت علمی گروه آمار، دانشگاه صنعتی شاهرود، ایران

^۴ Overall Measurement

^۵ Least-Squares

مقایسه با رتبه‌بندی اعداد فازی یا مجموعه‌های فازی برای اهداف عملی بسیار مهم هستند. زمانی که نظریه مجموعه فازی برای ساختن مدل‌های ریاضی به کار می‌رود با مسائلی مانند مقایسه یا رتبه‌بندی اعداد فازی مواجه می‌شویم. اعداد فازی نظم خطی ندارند و بنا بر این مقایسه آنها ساده نیست. روش‌های متنوعی برای مقایسه و رتبه‌بندی مجموعه‌های فازی پیشنهاد شده‌اند که هر یک دارای عیوب و کاستی‌هایی است. روش‌های ذکر شده از α -برش‌ها استفاده می‌کنند بنا بر این ممکن است از اطلاعات کافی استفاده نکرده و نتایج اربیی حاصل شوند. به‌منظور استفاده از همه اطلاعات ممکن، باید همه α -برش‌ها در نظر گرفته شوند. به دلیل پیچیدگی و اشکالات مسئله، یک روش کلی توسط چانگ و لی [۹] پیشنهاد شد که در آن با استفاده از مفهوم حضور (وجود)، رتبه‌بندی مجموعه‌های فازی، فرمول‌بندی می‌شود. می‌توان این روش را برای همه نوع تابع عضویت مثل محدب، غیر محدب، نرمال، غیرنرمال، پیوسته یا پیوسته تکه‌ای به کار برد. در این مقاله، یک تحلیل رگرسیون استوار فازی (FRR)^{۱۰} پیشنهاد می‌شود که در آن از یک روش رتبه‌بندی فازی برای رتبه دادن به باقی‌مانده‌ها استفاده شده است و تابع عضویت را برای باقی‌مانده‌ها تعریف می‌کند. از آن‌جا که هر باقی‌مانده با توجه به درجه عضویت، روی مدل تأثیر خواهد گذاشت مدل پیشنهادی یک روش FRR خواهد بود. ساختار این مقاله به این صورت تنظیم شده است که در بخش دوم، به بررسی رگرسیون استوار و M-برآوردگرها پرداخته و در ادامه، روش رگرسیون فازی مورد بحث قرار می‌گیرد. در بخش سوم، فاصله دیاموند معرفی شده و پارامترهای فازی مدل بر اساس این فاصله محاسبه می‌شوند. رتبه‌بندی مجموعه‌های فازی و معرفی سطح حضور سراسری^{۱۱} و شاخص OM در بخش چهارم مورد بررسی قرار می‌گیرد. در بخش پنجم به تحلیل رگرسیون استوار فازی بر پایه رتبه‌دهی به مجموعه‌های فازی پرداخته و با استفاده

و همکاران [۹] انجام شد. تعمیمی از این روش، برای هنگامی که اعداد فازی از نوع LR^۶ هستند، توسط باردوسی [۹] توسعه داده شد. دیاموند [۹]، روش کمترین توان‌های دوم فازی (FLS)^۷ را پیشنهاد کرد. میرزایی و ارقامی [۹] مروری بر چند رویکرد در رگرسیون فازی داشته‌اند. کلکین نما و طاهری [۹] رگرسیون فازی بر اساس کمترین قدر مطلق انحرافات را زمانی که داده دور افتاده وجود داشته باشد پیشنهاد کردند. چانگ و لی [۹] برای مواردی که در آنها داده‌های دورافتاده حضور دارند، تعمیمی از روش کمترین توان‌های دوم موزون فازی را پیشنهاد کردند که به هر درجه عضویت وزنی را اختصاص می‌دهد و بر پایه یک اثر متقابل با نظر تصمیم گیرنده است. برای رگرسیون فازی خطی ساده، یانگ و کو [۹] یک الگوریتم تکراری را برای تحلیل کمترین توان‌های دوم موزون فازی (WFLS)^۸ پیشنهاد کردند. این الگوریتم دو مرحله دارد؛ اولین مرحله، انتخاب روش طبقه‌بندی فازی است (که به مشاهدات طبقات، عضویت می‌دهد) و در دومین مرحله، مقادیر عضویت به‌عنوان وزن‌ها استفاده می‌شوند.

یانگ و لین [۹]، تحلیل رگرسیون FLS را برای ورودی فازی و خروجی فازی پیشنهاد کردند. آنها کاربردی از روش خوشه‌بندی را برای تعیین ناهمگنی مشاهدات و داده‌های دور افتاده ارائه کردند. یانگ و لیو [۹] الگوریتم FLS را برای مدل‌های رگرسیون خطی با اثر متقابل پیشنهاد کردند. این الگوریتم برای برآورد مدل‌های FLR^۹ به‌خصوص برای رگرسیون ساده، استوار است. در این الگوریتم، شرایط تعامد اضافه شده تا مسئله بهینه‌سازی را محدود کند. هانگ و یانگ [۹] به‌منظور بررسی مسئله داده‌های دور افتاده، یک شیوه حذف برای روش برنامه ریزی خطی تاناکا پیشنهاد کردند. این روش، قدرت امتحان کردن رفتار میزان تغییرات در تابع هدف مدل رگرسیونی فازی را زمانی که مشاهدات حذف می‌شوند داشت.

^۶ LR type fuzzy numbers^۷ Fuzzy Least Squares^۸ Weighted Fuzzy Least Square^۹ Fuzzy Linear Regression^{۱۰} Fuzzy Robust Regression^{۱۱} Overall existence level

از دو مثال، به کارایی مدل پیشنهادی اشاره خواهد شد.

۲ رگرسیون استوار

۱.۲ روش کمترین توان‌های دوم (LS)

یک مدل رگرسیونی خطی، که $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)'$ بردار مشاهدات متغیر وابسته، $\mathbf{X}_{n \times p}$ ماتریس مشاهدات متغیرهای مستقل، $\epsilon_n \times 1$ بردار خطای تصادفی و $\beta_{p \times 1}$ بردار ضرایب رگرسیونی است. به صورت ماتریسی به شکل

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon$$

$$= \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix} \quad (1)$$

نمایش داده می‌شود. با مینیمم کردن عبارت

$$\sum_{i=1}^n \epsilon_i^2 = \epsilon' \epsilon = (\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta)$$

برآورد LS ضرایب رگرسیونی به صورت

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y} \quad (2)$$

به دست می‌آید. ایراد مهم این روش آن است که به حضور داده‌های دورافتاده بسیار حساس بوده به این مفهوم که اگر داده‌های دورافتاده در مجموعه داده‌ها وجود داشته باشند، منجر به پیش‌بینی‌های اریبی می‌شود. در تحلیل رگرسیون، مشکل حضور داده‌های دورافتاده در مجموعه‌های داده‌ها را به دو شیوه می‌توان برطرف کرد:

الف) کشف داده‌های دور افتاده: تعریف شاخصی برای تشخیص، حذف یا احتمالاً جایگزین کردن داده‌های غیرعادی.
ب) برآورد استوار: تعریف روش‌های برآورد مناسب برای خنثی کردن اثر نامطلوب حضور داده‌های دور افتاده در فرایند برآورد.

در روش متعارف، مشاهدات دورافتاده از مجموعه داده‌ها حذف می‌شوند و پس از آن روش LS می‌تواند به کار رود. اما اگر این داده‌های دور افتاده حاوی اطلاعات مهمی باشند حذف آنها منجر به برآوردهایی می‌شود که کاملاً متفاوت از آنچه باید باشند، هستند. برای برآورد کردن مقادیر پارامتر در چنین شرایطی استفاده از روش‌های استوار به روش LS ترجیح داده می‌شود. یکی از روش‌هایی که تا حدودی نسبت به این مشکل استوارتر عمل می‌کند، روش کمترین توان‌های دوم موزون (WLS)^{۱۲} است.

۲.۲ روش کمترین توان‌های دوم موزون

روش LS معمول، فرض می‌شود که واریانس‌ها خطای مدل هستند. روش WLS، زمانی مورد استفاده قرار می‌گیرد که فرض ثابت بودن واریانس خطا برقرار نباشد. مدل مورد بررسی (؟؟) را با ماتریس واریانس-کوواریانس غیر ثابت

$$\begin{pmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_n^2 \end{pmatrix}$$

در نظر بگیرید. با فرض $W_i = \frac{1}{\sigma_i^2}$, $i = 1, 2, \dots, n$ ، ماتریس قطری W مبین وزن‌هاست که به صورت

$$W = \begin{pmatrix} w_1 & 0 & \dots & 0 \\ 0 & w_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & w_n \end{pmatrix}$$

تعریف می‌شود. پس برآورد کمترین توان‌های دوم موزون عبارت است از

$$\hat{\beta}_{WLS} = \operatorname{argmin}_{\beta} \sum_{i=1}^n w_i \epsilon_i^2 = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1} \mathbf{X}'\mathbf{W}\mathbf{Y} \quad (3)$$

در اصل روش‌های استوار بر پایه M-برآوردگرها^{۱۳} شکل می‌گیرند که در ادامه به شرح آنها خواهیم پرداخت (برای مطالعه بیشتر به [؟] مراجعه کنید).

^{۱۲} Weighted Least Squares

^{۱۳} "M" for "Maximum likelihood-type"

۳.۲ M - برآورد گرها

M- برآورد گرها از نوع برآورد گرهای درست‌نمایی ماکسیم هستند. مدل رگرسیونی (؟؟) را در نظر بگیرید. فرض کنید خطاها (ϵ_i) مستقل و هم توزیع با تابع چگالی احتمال $f(\epsilon)$ باشند. برآورد گر درست‌نمایی ماکسیم β مقداری است که کمیت

$$\prod_{i=1}^n f(\epsilon_i) = \prod_{i=1}^n f(y_i - x_i' \beta)$$

یا به طور معادل $\sum_{i=1}^n \ln f(y_i - x_i' \beta)$ را ماکسیم می‌کند؛ به طوری که x_i' -i-امین سطر X در مدل (؟؟) باشد. در روش کمترین توان‌های دوم، $i = 1, 2, \dots, n$ ، $\epsilon_i \sim N(0, \sigma^2)$ با استفاده از روش درست‌نمایی ماکسیم برای برآورد β ، تابع درست‌نمایی آن عبارت است از:

$$\begin{aligned} L(\beta) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{\epsilon_i^2}{2\sigma^2}\right\} \\ &= \frac{1}{(2\pi\sigma)^{\frac{n}{2}}} \exp\left\{-\frac{\sum_{i=1}^n \epsilon_i^2}{2\sigma^2}\right\} \end{aligned}$$

در نتیجه ماکسیم کردن تابع درست‌نمایی، منجر به مینیم کردن مجموع توان‌های دوم خطاها یعنی

$$\sum_{i=1}^n \epsilon_i^2 = \sum_{i=1}^n (y_i - x_i' \beta)^2 \quad (۴)$$

می‌شود. اگر در رابطه (؟؟) تعریف کنیم $\rho(\epsilon_i) = (y_i - x_i' \beta)^2$ در این صورت این رابطه به صورت

$$\sum_{i=1}^n \rho(\epsilon_i) = \sum_{i=1}^n (y_i - x_i' \beta)^2$$

به دست می‌آید. حال اگر خطاها را استاندارد کنیم برآورد ضرایب رگرسیونی را می‌توان از مینیم کردن $\sum_{i=1}^n \rho(\frac{\epsilon_i}{\sigma})$ به دست آورد. از آنجا که σ معمولاً مجهول است می‌توان از برآورد آن، $\hat{\sigma}$ ، استفاده کرد. شیوه M- برآورد گرها، مینیم کردن توابع مختلف $\rho(z_i)$ است که در آن $z_i = \frac{\epsilon_i}{\sigma}$ و $e_i = y_i - x_i' \hat{\beta}$ باقیمانده i-امین مشاهده است. در حالتی که خطاها به طور نرمال توزیع شده باشند انحرافات استاندارد نمونه، $\hat{\sigma}$ ، اندازه‌های مناسبی برای تغییرپذیری هستند ولی زمانی که در مجموعه داده‌ها مشاهدات دور افتاده داریم به اندازه‌های استوار

دیگری از تغییرپذیری، نیاز است. یکی از اندازه‌های استوار تغییرپذیری، به صورت

$$\hat{\sigma}_{MAD} = d = \frac{\text{median}|e_i - \text{median}(e_i)|}{0.6745}, \quad i = 1, \dots, n \quad (۵)$$

تعریف می‌شود. بنا بر این گوئیم یک برآورد گر استوار است اگر عبارت

$$\sum_{i=1}^n \rho\left(\frac{e_i}{d}\right) \quad (۶)$$

را مینیم کند. پس، باقیمانده‌های استاندارد شده جدید، به صورت $z_i = \frac{e_i}{d}$ تعریف می‌شوند. $\rho(z)$ تابعی از z است. ملاحظه می‌شود که اگر $z^2 = \rho(z)$ با مینیم کردن رابطه (؟؟) به همان مدل کمترین توان‌های دوم معمول می‌رسیم.

روش کمترین توان‌های دوم در به دست آوردن برآورد پارامترها به هر مشاهده، وزن یکسانی می‌دهد. روش‌های استوار قادرند به مشاهدات، وزن‌های نابرابر اختصاص دهند. به طور کلی مشاهداتی که مانده‌های بزرگی را تولید می‌کنند به وسیله برآورد استوار، کم وزن‌ترند. پس می‌توان گفت که انتخاب $\rho(z)$ خاص، در حقیقت انتخاب تابع وزنی است که می‌خواهیم برای رتبه‌بندی مانده‌ها با اندازه‌های مختلف تخصیص دهیم. در ادامه پیشنهاداتی برای به دست آوردن $\rho(z)$ مناسب ارائه شده است.

با گرفتن اولین مشتق جزئی از معادله $\sum \rho(\frac{e_i}{d})$ نسبت به $\hat{\beta}_j$ و با برابر صفر قرار دادن آن به شکل

$$\sum_{i=1}^n x_{ij} \Psi \left[\frac{y_i - \sum_{j=1}^p x_{ij} \beta_j}{d} \right] = 0, \quad j = 1, 2, \dots, n \quad (۷)$$

که $\Psi(z) = d(\rho(z))/d(z)$ ، ضرایب رگرسیونی به دست می‌آیند.

توابع زیادی از Ψ ، توسط افراد مختلفی تعریف شده است که تابع Ψ همپل از آن جمله است که یکی از توابعی است که بیشتر مورد استفاده قرار می‌گیرد و از همان پیشنهاداتی است که پیش‌تر در مورد آنها بحث شد. این تابع در زیر آمده و نمودارش

^{۱۴} Hampel

تعریف ۲.۳. $(\alpha$ -برش‌ها): مجموعه عناصری را از X (مجموعه مرجع) که درجه عضویت آنها در مجموعه فازی A دست کم به بزرگی α ($0 < \alpha \leq 1$) باشد، α -برش A گوئیم و با $A_\alpha = \{x \in X | A(x) \geq \alpha\}$ یعنی A_α نشان می‌دهیم. توجه به این که حدود پایین و بالای عدد فازی مثالی A به ترتیب $[A_L(\alpha), A_U(\alpha)] = m_A - l_A$ و $m_A + r_A$ هستند، A_α به صورت $[m_A - (1 - \alpha)l_A, m_A + (1 - \alpha)r_A]$ به دست می‌آید.

جمع دو عدد فازی و ضرب یک عدد دقیق در یک عدد فازی، بر اساس مفهوم اصل توسیع تعریف می‌شود. [۴]

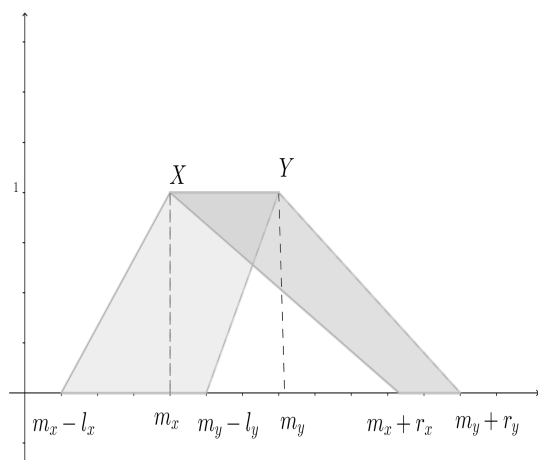
گزاره ۳.۳. اگر $A_1 = (m_1, l_1, r_1)_T$ و $A_2 = (m_2, l_2, r_2)_T$ دو عدد فازی در $T(\mathbb{R})$ باشند و $\lambda \in \mathbb{R}$ آنگاه

$$\lambda A_1 = \begin{cases} (\lambda m_1, \lambda l_1, \lambda r_1)_T & \lambda > 0 \\ (\lambda m_1, -\lambda r_1, -\lambda l_1)_T & \lambda < 0 \end{cases}$$

تعریف ۴.۳. [۴] (فاصله دیاموند): اگر داشته باشیم $X = (m_x, l_x, r_x)$ و $Y = (m_y, l_y, r_y)$ که $X, Y \in T(\mathbb{R})$ آنگاه $d^*(X, Y)$ دیاموند، به صورت

$$\begin{aligned} d^*(X, Y) = & [(m_X - l_X) - (m_Y - l_Y)]^2 \\ & + [(m_X + r_X) - (m_Y + r_Y)]^2 \\ & + [m_X - m_Y]^2 \end{aligned} \quad (9)$$

تعریف شده و در شکل ۲ نشان داده می‌شود.

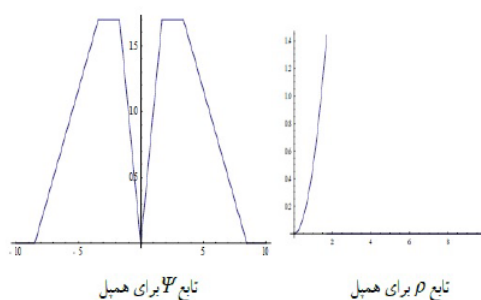


شکل ۲. ناحیه رنگی، بیان کننده فاصله دیاموند برای دو عدد فازی مثالی X و Y

در شکل ۱ نشان داده شده است. - تابع Ψ همپل^{۱۴}

$$\Psi(z) = \begin{cases} |z| & 0 \leq |z| < a \\ a & a \leq |z| \leq b \\ a \left(\frac{c-|z|}{c-b} \right) & b \leq |z| < c \\ 0 & c \leq |z| \end{cases} \quad (8)$$

$a = 1/7$, $b = 3/4$, $c = 8/5$ مقادیر خوبی برای ثابت‌ها هستند.



شکل ۱. تابع ρ و Ψ همپل

۳ تحلیل رگرسیون فازی

در این بخش به بررسی مختصری از اعداد فازی پرداخته، فاصله دیاموند را معرفی کرده و سپس به محاسبه پارامترهای فازی مدل بر اساس این فاصله می‌پردازیم. برای اطلاعات بیشتر به [۴، ۵] و مراجعه کنید.

۱.۳ شرح مختصری از اعداد فازی

تعریف ۱.۳. اگر A یک عدد فازی مثالی با تابع عضویت $A(\cdot)$ باشد به شکل $A = (m_A, l_A, r_A)_T$ نشان داده شده و به صورت

$$A(x) = \begin{cases} \frac{x - (m_A - l_A)}{l_A} & m_A - l_A \leq x \leq m_A \\ \frac{(m_A + r_A) - x}{r_A} & m_A \leq x \leq m_A + r_A \\ 0 & \text{OW} \end{cases}$$

تعریف می‌شود. که در آن $m_A \in \mathbb{R}$ و $l_A, r_A \geq 0$ به ترتیب نشان‌دهنده مقدار مرکز، پهنای چپ و پهنای راست عدد فازی است. مجموعه تمام اعداد فازی مثالی را با $T(\mathbb{R})$ نمایش می‌دهند.

۲.۳ رگرسیون کمترین توان‌های دوم فازی

هدف آن است که بر پایه مشاهداتی به صورت $(Y_i, x_{i1}, x_{i2}, \dots, x_{ip})$ که $(i = 1, \dots, n; j = 1, \dots, p; x_{ij} \in \mathbb{R})$ و $Y_i = (m_{Y_i}, l_{Y_i}, r_{Y_i})_T$ به شکل $(m_{Y_i}, l_{Y_i}, r_{Y_i})_T$ باشند، یک مدل بهینه با ضرایب فازی به صورت

$$Y = A_0 + A_1x_1 + \dots + A_px_p \quad (10)$$

که در آن $A_j = (m_{A_j}, l_{A_j}, r_{A_j})_T; j = 0, 1, \dots, p$ برای توصیف و تحلیل داده‌ها و پیش‌بینی بر پایه آن، به دست آوریم. روش کمترین توان‌های دوم فازی که مشابه روش کمترین توان‌های دوم معمولی است، با تعریف فاصله بین دودعد فازی و مینیم کردن مجموع توان‌های دوم فاصله مشاهدات از مقادیر برآورد شده آنها، معادلاتی مشابه معادلات نرمال در روش کمترین توان‌های دوم معمولی به دست می‌دهد که با حل دستگاه معادلات حاصل، پارامترهای مدل برآورد می‌شوند. یعنی

$$\begin{aligned} M(A_0, A_1, \dots, A_p) \\ = \min \sum_{i=1}^n d^2(A_0 + A_1x_{i1} + \dots + A_px_{ip}, Y_i) \end{aligned} \quad (11)$$

پس برازش مدل، بر پایه تعریف خاصی از فاصله (d) بین دو عدد فازی است که در این مقاله آن را بر پایه فاصله دیاموند تعریف شده در رابطه (۹۹) به دست می‌آوریم. بنا بر این داریم

$$\begin{aligned} d(A_0 + A_1x_{i1} + \dots + A_px_{ip}, Y_i)^2 \\ = (m_{A_0} + m_{A_1}x_{i1} + \dots + m_{A_p}x_{ip} - m_{Y_i})^2 \\ + (m_{A_0} + m_{A_1}x_{i1} + \dots + m_{A_p}x_{ip} \\ - l_{A_0} - l_{A_1}x_{i1} - \dots - l_{A_p}x_{ip} - m_{Y_i} + l_{Y_i})^2 \\ + (m_{A_0} + m_{A_1}x_{i1} + \dots + m_{A_p}x_{ip} \\ + r_{A_0} + r_{A_1}x_{i1} + \dots + r_{A_p}x_{ip} - m_{Y_i} - r_{Y_i})^2 \end{aligned} \quad (12)$$

می‌توان پارامترها را توسط مشتق جزئی نسبت به m_{A_j} و l_{A_j} و r_{A_j} ها ($j = 0, 1, \dots, p$) و برابر صفر قرار دادن آنها یا به عبارت دیگر به وسیله مینیم سازی رابطه (۹۹)، برآورد کرد. در نتیجه FLS، بر حسب برآورد پارامترها به دست می‌آید.

۴ رتبه‌بندی مجموعه‌های فازی

اعداد فازی مثل اعداد واقعی، نظم و ترتیب خطی ندارند و مقایسه آنها ساده نیست. بسیاری از محققان در مورد روش‌های رتبه بندی فازی مطالعاتی انجام داده‌اند. در این قسمت، روش مقایسه اعداد فازی را که توسط چانگ و لی [۹] مطرح شده است، بیان می‌کنیم.

یک زیرمجموعه فازی، بر پایه تکیه گاه و تابع عضویت تعریف می‌شود. درجه تعلق یک عنصر معلوم در تکیه گاه به مفهومی نادقیق (فازی)، تابع عضویت این عنصر را تشکیل می‌دهد. بنا بر این برای مقایسه چندین مجموعه فازی، علاقه‌مند به مقایسه‌هایی هستیم که بر پایه درجه تعلق (تابع عضویت) عناصر تکیه گاه یا مکان این عناصر در تکیه گاه باشد.

در نظر گرفتن این خصوصیات برای مقایسه دو مجموعه فازی، گاهی ویژگی‌های نامطلوبی را برای روش رتبه‌بندی فازی به همراه دارد. برای توضیح این مطلب، دو مجموعه فازی A و B را در نظر بگیرید (شکل ۳). اگر a و b به ترتیب بیان کنند عناصر موجود در تکیه گاه برای مجموعه‌های فازی A و B ، و همچنین h بیان کنند درجه عضویت باشد، آنگاه $a(h)$ و $b(h)$ مقادیری از تکیه گاه مجموعه‌های فازی A و B را بر پایه درجه یکسان h از تابع عضویت، نشان می‌دهند. بر این اساس نتیجه می‌گیریم که مجموعه فازی A کوچک‌تر از مجموعه فازی B است؛ ($A < B$).

اما اگر توابع عضویت را بر پایه عنصر یکسانی از تکیه گاه $a (= b)$ مقایسه کنیم، یعنی مقایسه بین $h(a)$ و $h(b)$ ، نتیجه‌ای که به دست می‌آید کاملاً متفاوت خواهد بود ($B < A$). بنا بر این مجموعه‌های فازی باید بر پایه درجه یکسانی از عضویت یا سطح حضور مقایسه شوند نه بر پایه عنصر یکسان در تکیه گاه؛ به عبارتی دیگر نابرابری دو مجموعه فازی، حد اقل در تعدادی عضویت یا سطح حضور، تفاوت عناصر در تکیه گاه‌ها را به دنبال خواهد داشت.

تاناکا و همکاران [۹] به منظور استفاده از همه اطلاعات موجود و با استفاده از سطح $h \in (0, 1]$ ، شاخص مقایسه دو مجموعه فازی A و B را معرفی کردند.

تعریف ۲.۴. دو مجموعه فازی A و B را با توابع عضویت $A(x)$ و $B(y)$ در نظر بگیرید. در سطح $h \in (0, 1]$ شاخص مقایسه این دو مجموعه فازی، به صورت

$$I = \int_0^1 g(\{A^{-1}(h)\})dh - \int_0^1 g(\{B^{-1}(h)\})dh \quad (14)$$

تعریف می شود. که در آن A^{-1} و B^{-1} تصاویر معکوس توابع عضویت A و B در سطوح متفاوتی از h تعریف شده است و نیز تابعی بر روی تصویر توابع عضویت است. بر اساس رابطه (۱۴) می گوییم A از B بزرگ تر است هرگاه $I > 0$.

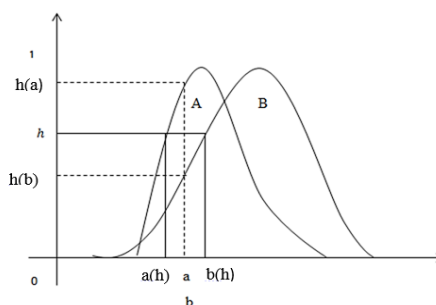
یاجر [۹]، کافمن [۹] و کمپس گونزالس [۹، ۹] روش هایی مشابه روش تاناکا و همکارانش [۲۰] مطرح کردند. توابع g متفاوتی را می توان در رابطه (۱۴) تعریف کرد. یک تعریف مفید و ساده g ، مجموع وزن داده شده بر روی تصویر تابع عضویت به صورت

$$I = \int_0^1 \omega(\sum x(h))dh - \int_0^1 \omega(\sum y(h))dh \quad (15)$$

است. که در آن ω تابع وزنی از h است، و $x(h)$ نشان گر عناصر $\{A^{-1}(h)\}$ ، $y(h)$ نشان گر عناصر $\{B^{-1}(h)\}$ و A و B مجموعه های فازی پیوسته، محدب و نرمال اند.

۱.۴ روش رتبه بندی فازی با مشارکت تصمیم گیرنده

یک جنبه مهم مجموعه های فازی، ذهنی بودن آنهاست. این ویژگی هنگام مقایسه مجموعه های فازی بسیار مشابه، کاملاً مشهود است. به عنوان مثالی برای تشریح این مطلب، دو عدد فازی مثلثی $A_1 = (m_{A_1}, l_{A_1}, r_{A_1})_T$ و $A_2 = (m_{A_2}, l_{A_2}, r_{A_2})_T$ را در شکل ۵ در نظر بگیرید؛ که در آن $m_{A_1} = m_{A_2} = m$. هر دوی این اعداد، تعبیر فازی عدد حدوداً m هستند. بسته به این که تصمیم گیرنده به چه میزان ابهام معتقد باشد یکی از اعداد A_1 یا A_2 را برای مسئله ای خاص انتخاب می کند. اگر از نگاه ریاضی



شکل ۳. ناحیه رنگی، بیان کننده فاصله دیاموند برای دو عدد فازی مثلثی مقایسه دو عدد فازی A و B

تعریف ۱.۴. فرض کنید A یک زیرمجموعه فازی با تابع عضویت $A(x)$ باشد $\{A^{-1}(h)\}$ ، زیرمجموعه دلخواهی است که نشان گر تصویر معکوس تابع عضویت $A(x)$ به ازای $h \in (0, 1]$ است هرگاه

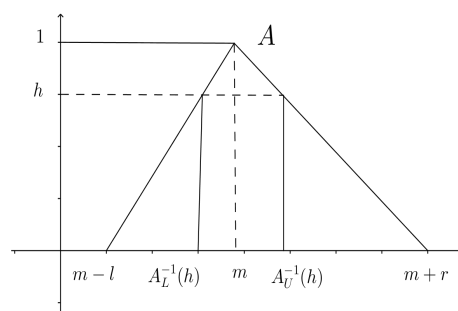
$$\{A^{-1}(h)\} = \{x : A(x) = h\}; x \in \mathbb{R}$$

بر اساس مطالب فوق، در سطوح متفاوتی از h ، روش های مختلفی برای مقایسه دو به دوی مجموعه های فازی صورت گرفته است (به [۹، ۹، ۹، ۹، ۹، ۹، ۹] مراجعه کنید).

برای مثال برای یک عدد فازی مثلثی $A = (m, l, r)_T$ داریم (شکل ۴) (۱۳)

$$\begin{aligned} A_L^{-1}(h) &= \{x : A(x) = h, m-l \leq x \leq m\} \\ &= \{m - (1-h)l\} \end{aligned}$$

$$\begin{aligned} A_U^{-1}(h) &= \{x : A(x) = h, m \leq x \leq m+r\} \\ &= \{m + (1-h)r\} \end{aligned} \quad (13)$$



شکل ۴. نمایش تصویر معکوس ضابطه چپ و راست عدد فازی مثلثی در سطح حضور h

به صورت‌های

$$\omega(h) = \frac{h}{\int_0^1 h dh} \quad \omega(h) = \frac{h(1-h)}{\int_0^1 h(1-h)dh}$$

داده شده است. اندازه‌های وزنی ذهنی χ_1 و χ_2 را معمولاً در ساده‌ترین حالت طوری انتخاب می‌کنند که $\chi_1 + \chi_2 = 1$ باشد (۱۹)

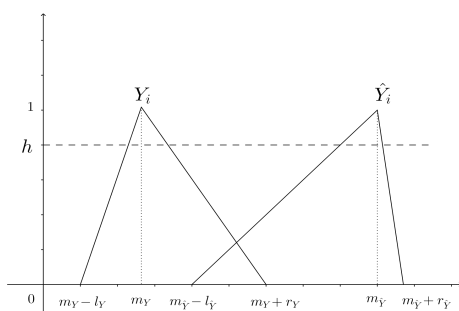
$$\chi_1 = \chi_1(h), \chi_2 = \chi_2(h), \chi_1, \chi_2 \in [0, 1], \chi_1 + \chi_2 = 1$$

قضیه ۴.۴. اگر $\hat{Y}_i = (m_{\hat{Y}_i}, l_{\hat{Y}_i}, r_{\hat{Y}_i})$ و $Y_i = (m_{Y_i}, l_{Y_i}, r_{Y_i})$ در شرایط (؟؟) صدق کنند، باقی‌مانده مدل، یعنی E_i به صورت

$$\begin{aligned} E_i &= d(\hat{Y}_i, Y_i) \\ &= \chi_1 \left((m_{\hat{Y}_i} - m_{Y_i}) - \frac{1}{3}(l_{\hat{Y}_i} - l_{Y_i}) \right) \\ &\quad + \chi_2 \left((m_{\hat{Y}_i} - m_{Y_i}) + \frac{1}{3}(r_{\hat{Y}_i} - r_{Y_i}) \right) \quad (20) \end{aligned}$$

به دست می‌آید.

اثبات. وقتی $\hat{Y}_i$ و Y_i اعداد فازی مثالی باشند فاصله‌ی بین Y_i و $\hat{Y}_i$ با استفاده از معادلات (؟؟)، (؟؟) و (؟؟) از طریق نمایش شکل ۶ محاسبه می‌شود.

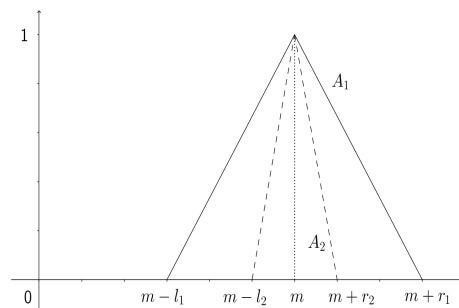


شکل ۶. نمایش تصاویر معکوس Y_i و $\hat{Y}_i$ در سطح حضور h

تحت شرایط (؟؟) و بر اساس روابط (۱۷) و (۱۸)، $OM(Y_i)$ و $OM(\hat{Y}_i)$ را محاسبه می‌کنیم

$$\begin{aligned} OM(Y_i) &= \frac{\int_0^1 h (\chi_1(m_{Y_i} - (1-h)l_{Y_i}) + \chi_2(m_{Y_i} + (1-h)r_{Y_i})) dh}{\int_0^1 h dh} \\ &= \chi_1 \left(\frac{l_{Y_i}}{3} + \frac{m_{Y_i} - l_{Y_i}}{2} \right) + \chi_2 \left(\frac{m_{Y_i} + r_{Y_i}}{2} - \frac{r_{Y_i}}{3} \right) \end{aligned}$$

غیر فازی، به مسئله نگاه کنیم و هیچ‌گونه ابهامی را برای این اعداد در نظر نگیریم، هر دو عدد A_1 و A_2 با هم برابرند در حالی که بنا بر نظر تصمیم گیرنده (از نگاه فازی)، این دو عدد با هم متفاوتند. بنا بر این نیاز به شاخصی برای رتبه‌بندی اعداد فازی است که نظر تصمیم گیرنده را مشارکت دهد.



شکل ۵. مثالی از نمایش عدم اختلاف دو عدد فازی بدون در نظر گرفتن درجه فازی آنها

شاخص زیر به منظور مقایسه دو عدد فازی LR، تعریف می‌شود. برای بررسی در حالت کلی به چانگ و لی [۲] مراجعه کنید.

تعریف ۳.۴. فرض کنید A و B دو عدد فازی LR به ترتیب با توابع عضویت $A(x)$ و $B(y)$ باشند. همچنین فرض کنید $A_L^{-1}(h)$ و $A_U^{-1}(h)$ نیز کران‌های پایین و بالای تصویر تابع عضویت در سطح h بر روی تکیه‌گاه باشند. تفاضل بین B و A به صورت

$$d(A, B) = \int_0^1 g_A(\{A^{-1}(h)\}) dh - \int_0^1 g_B(\{B^{-1}(h)\}) dh \quad (16)$$

$$= OM(A) - OM(B)$$

تعریف می‌شود. که در آن

$$(17)$$

$$g_A(\{A^{-1}(h)\}) = \omega(h) (\chi_1(h)A_L^{-1}(h) + \chi_2(h)A_U^{-1}(h))$$

و

$$(18)$$

$$g_A(\{B^{-1}(h)\}) = \omega(h) (\chi_1(h)B_L^{-1}(h) + \chi_2(h)B_U^{-1}(h))$$

و اندازه‌های وزن $\omega(h)$ و $\chi_1(h)$ و $\chi_2(h)$ می‌بایست به‌طور ذهنی توسط تصمیم گیرنده معین شوند. پیشنهادهایی برای $\omega(h)$

نظر می‌گیریم. پس از آن که مدل رگرسیونی به دست آمد، باقی‌مانده‌ها را از رابطه (۲۲) محاسبه کرده و با استفاده از تابع عضویت باقی‌مانده‌ها، ماتریس وزن را تعریف می‌کنیم و در نهایت، برآوردهای WFLS به وسیله ماتریس وزن ساخته می‌شوند.

برای رگرسیون کلاسیک، برآوردهای WLS در بخش ۲.۲ توضیح داده شد. ما در این مقاله نیز برای برآورد پارامترهای فازی، از این روش استفاده می‌کنیم؛ به این صورت که برای مراکز و پهنای Y_i وزن W را در نظر می‌گیریم؛ که همان ماتریس وزنی است و از طریق محاسبه تابع عضویت باقی‌مانده‌ها به دست می‌آید. بنا بر این مقدار برآورد مراکز و پهنای پارامترها به ترتیب به صورت بسته $\hat{A}_{m_j} = (X'WX)^{-1}X'WY_{m_j}$ و $\hat{A}_{s_j} = (X'WX)^{-1}X'WY_{s_j}$ به دست می‌آیند؛ $(j = 1, \dots, p)$ که p تعداد پارامترهاست. بنا بر این برآورد مدلی به دست می‌آید که تحت تأثیر داده‌های دورافتاده قرار نمی‌گیرد. برای کسب اطلاعات بیشتر در زمینه WFLS به [۲] مراجعه کنید.

۱.۵ مدل رگرسیونی استوار فازی پیشنهاد شده

در تحلیل رگرسیون چندگانه، وقتی x_i دقیق و Y_i یک عدد فازی مثلثی متقارن باشد و در مجموعه داده‌ها، مشاهده دور افتاده باشیم، برای مدل‌های رگرسیون برآورد شده، گام‌های الگوریتم روش FRR به صورت زیر پیشنهاد می‌شود؛ $\hat{A}$ برآورد پارامترهای رگرسیون، k تعداد تکرارها و $\epsilon > 0$ عدد بسیار کوچکی است که توسط تصمیم گیرنده تعیین می‌شود. (این جا مقدار $\epsilon = 0.05$ در نظر گرفته شد)

گام ۱: وقتی x_i دقیق و $Y_i = (m_{Y_i}, l_{Y_i}, r_{Y_i})_T$ یک عدد فازی مثلثی متقارن باشد، از رابطه (۲۲)، برآورد اولیه FLS را می‌یابیم.

گام ۲: مقادیر برآورد مرکز عدد فازی $(m_{\hat{Y}_i})$ و باقیمانده‌ها (E_i) محاسبه می‌شوند. باقیمانده‌ها را از معادله (۲۳)، توسط شاخص OM به دست می‌آوریم.

گام ۳: میانه بر حسب قدر مطلق مقدار باقیمانده‌ها تعیین می‌شود. فاصله‌ها را از دستور زیر محاسبه می‌کنیم

$$\begin{aligned} OM(\hat{Y}_i) &= \frac{\int_0^1 h \left(\chi_1(m_{\hat{Y}_i} - (1-h)l_{\hat{Y}_i}) + \chi_2(m_{\hat{Y}_i} + (1-h)r_{\hat{Y}_i}) \right) dh}{\int_0^1 h dh} \\ &= \chi_1 \left(\frac{l_{\hat{Y}_i}}{3} + \frac{m_{\hat{Y}_i} - l_{\hat{Y}_i}}{2} \right) + \chi_2 \left(\frac{m_{\hat{Y}_i} + r_{\hat{Y}_i}}{2} - \frac{r_{\hat{Y}_i}}{3} \right) \end{aligned}$$

طبق (۲۲)، E_i عبارت است از

$$\begin{aligned} E_i &= d(\hat{Y}_i, Y_i) = OM(\hat{Y}_i) - OM(Y_i) \\ &= \chi_1 \left((m_{\hat{Y}_i} - m_{Y_i}) - \frac{1}{3}(l_{\hat{Y}_i} - l_{Y_i}) \right) \\ &\quad + \chi_2 \left((m_{\hat{Y}_i} - m_{Y_i}) + \frac{1}{3}(r_{\hat{Y}_i} - r_{Y_i}) \right) \end{aligned} \quad (21)$$

□

نتیجه ۵.۴. تحت شرایط (۲۲)، اگر $Y_i = (m_{Y_i}, l_{Y_i}, r_{Y_i})$ و $\hat{Y}_i = (m_{\hat{Y}_i}, l_{\hat{Y}_i}, r_{\hat{Y}_i})$ اعداد فازی مثلثی متقارن باشند و همچنین $l_{\hat{Y}_i} = r_{\hat{Y}_i} = s_{\hat{Y}_i}$ و $l_{Y_i} = r_{Y_i} = s_{Y_i}$ با استفاده از معادلات (۲۲)، (۲۳) و (۲۴) به صورت

$$\begin{aligned} E_i &= d(\hat{Y}_i, Y_i) \\ &= (\chi_1 + \chi_2)(m_{\hat{Y}_i} - m_{Y_i}) + \frac{1}{3}(\chi_2 - \chi_1)(s_{\hat{Y}_i} - s_{Y_i}) \end{aligned} \quad (22)$$

محاسبه می‌شود.

نتیجه ۶.۴. تحت شرایط (۲۴) اگر Y_i و $\hat{Y}_i$ اعداد فازی مثلثی متقارن و $\chi_1 = \chi_2 = \chi = \frac{1}{3}$ باشند، داریم

$$E_i = d(Y_i, \hat{Y}_i) = m_{\hat{Y}_i} - m_{Y_i} \quad (23)$$

(برای جزئیات بیشتر به [۲] مراجعه کنید.)

۵ تحلیل رگرسیون استوار فازی بر پایه رتبه‌دهی به مجموعه‌های فازی

در این مقاله، یک مدل رگرسیونی چند گانه را در حالتی که X عددی دقیق و ضرایب رگرسیونی و Y فازی هستند در

تعریف می‌کنیم.

(۲۴)

گزاره ۳.۵. [۲۱] فاصله d که در تعریف بالا بیان شد یک متر روی $T(\mathbb{R}) \times T(\mathbb{R})$ است. یعنی برای A, B, C داریم:

$$۱. d(A, B) \geq ۰, d(A, A) = ۰$$

$$۲. d(A, B) = d(B, A)$$

$$۳. d(A, C) \leq d(A, B) + d(B, C)$$

$$D_i = ||abs(E_i) - median(abs(E_i))||, \quad i = ۱, ۲, \dots, n$$

که $||\cdot||$ فاصله اقلیدسی است.

گام ۴: تابع عضویت باقیمانده‌ها را به شکل

$$\mu(E) = \begin{cases} ۱ & abs(E) \leq a \\ \frac{b-abs(E)}{b-a} & a < abs(E) < b \\ ۰ & OW \end{cases} \quad (۲۵)$$

تعریف می‌کنیم. $a = median(D_i)$ و $b = max(D_i) + d$ که

در آن d ، پارامتر استوار تغییرپذیری معرفی شده در (۲۴) است.

گام ۵: مقادیر عضویت را با استفاده از گام ۴ محاسبه کرده و

ماتریس وزن را تشکیل می‌دهیم. ماتریس وزن، ماتریسی قطری

است که عناصر روی قطر آن، درجه عضویت‌ها هستند. پس

از آن، برآوردهای WFLS را از طریق ماتریس وزن به دست

می‌آوریم.

گام ۶: اگر $|\hat{A}^{k+1} - \hat{A}^k| < \epsilon$ ، الگوریتم را متوقف می‌کنیم.

در غیر این صورت به گام ۲ می‌رویم.

۶ مثال‌های عددی

در این قسمت، برای نشان دادن کاربرد روش پیشنهادی، از داده‌های شبیه‌سازی شده و واقعی استفاده و نتایج حاصل تحلیل می‌شود.

۱.۶ داده‌های شبیه‌سازی شده

مسئله، شامل ۳ متغیر تبیینی و ۳۰ مشاهده از داده‌های شبیه‌سازی شده با توزیع نرمال است به‌طوری که

$$X_1 \sim N(\mu = ۲۰, \sigma = ۳)$$

$$X_2 \sim N(\mu = ۵۰, \sigma = ۱۲)$$

$$X_3 \sim N(\mu = ۳۲, \sigma = ۱۳)$$

هستند. ۱۵-امین مشاهده در متغیرهای وابسته با $۲۰ + y_{15}$ جایگزین شده و به‌عنوان یک داده دور افتاده در نظر گرفته می‌شود. مجموعه داده‌ها در جدول ۱ نشان داده شده است.

برای این مجموعه داده، از طریق مدل رگرسیونی فازی FLS که پیش‌تر بیان شد، ضرایب رگرسیونی برآورد می‌شوند. این مقادیر به‌عنوان مقادیر اولیه وارد الگوریتم پیشنهادی شده و تحلیل FRR با توجه به مقادیر متغیر وابسته و مرکز (y_i) و پهنای که به‌صورت $s_i = \frac{y_i}{\lambda}$ ($i = ۱, ۲, \dots, n$) تعریف می‌شوند، محاسبه می‌گردند. باقیمانده‌ها و عناصر روی قطر ماتریس وزن، در جدول ۱ نشان داده شده‌اند.

۲.۵ شاخص نیکویی برازش

یکی از شاخص‌های ارزیابی نیکویی برازش مدل رگرسیونی آماری، معیار $MPE^{۱۵}$ است که در قالب یک تعریف در زیر آورده شده است.

تعریف ۱.۵. تحت مفروضات مدل رگرسیونی (۲۴) اگر

$\hat{Y}_i$ مقدار پیشگویی Y_i باشد آنگاه میانگین خطای پیشگویی

(MPE) برابر است با

$$MPE = \frac{1}{n} \sum_{i=1}^n d(Y_i, \hat{Y}_i)$$

که در این مقاله از d (فاصله) حسن پور [۲۱] استفاده کرده‌ایم که در ادامه به معرفی آن خواهیم پرداخت.

تعریف ۲.۵. حسن پور و همکاران [۲۱]، فاصله بین دو عدد فازی

را به‌صورت زیر بیان کردند.

اگر $A = (m_A, l_A, r_A)_T$ و $B = (m_B, l_B, r_B)_T$ دو عدد

فازی در $T(\mathbb{R})$ باشند، $d: T(\mathbb{R}) \times T(\mathbb{R}) \rightarrow \mathbb{R}$ را به‌صورت

$$d(A, B) = |m_A - m_B| + |l_A - l_B| + |r_A - r_B| \quad (۲۶)$$

^{۱۵} Mean Predicted Error

جدول ۱. مجموعه داده‌ها، وزن‌ها (W_i) و مقایسه مدل FLS و مدل پیشنهادی با توجه به شاخص MPE برای داده‌های شبیه‌سازی

شده										
d_{Prop}	d_{FLS}	$\hat{y}_{Prop}$	$\hat{y}_{FLS}$	W_{Prop}	W_{FLS}	(y_i, s_i)	x_3	x_2	x_1	مشاهده
۳/۱۸	۴/۹۶	(۱۲۱/۶۷, ۱۵/۲۱)	(۱۲۳/۱۰, ۱۵/۳۹)	۰/۹۷	۱	(۱۱۹/۱۳, ۱۴/۸۹)	۳۰/۴۸	۶۶/۵۷	۲۴/۰۲	۱
۱/۵۸	۴/۸۹	(۱۲۱/۳۹, ۱۵/۱۷)	(۱۲۴/۰۴, ۱۵/۵۱)	۱	۱	(۱۲۰/۱۳, ۱۵/۰۲)	۳۴/۱۹	۶۵/۸۴	۲۱/۱۶	۲
۳/۶۸	۳/۱۰	(۸۷/۸۳, ۱۰/۹۸)	(۸۸/۲۹, ۱۱/۰۴)	۰/۹۶	۱	(۹۰/۷۷, ۱۱/۳۵)	۲۵/۴۸	۳۹/۰۹	۲۱/۱۸	۳
۵/۴۱	۸/۰۷	(۶۲/۶۴, ۷/۸۳)	(۶۴/۷۷, ۸/۱۰)	۰/۹۲	۱	(۵۸/۳۲, ۷/۲۹)	۲۲/۸۳	۲۲/۳۳	۱۴/۸۸	۴
۰/۱۸	۳/۸۹	(۱۳۰/۳۷, ۱۶/۳۰)	(۱۳۳/۶۳, ۱۶/۷۰)	۱	۱	(۱۳۰/۵۲, ۱۶/۳۱)	۳۸/۶۱	۷۱/۴۶	۲۰/۶۸	۵
۲/۲۶	۴/۰۵	(۱۰۴/۵۶, ۱۳/۰۷)	(۱۰۵/۹۹, ۱۳/۲۵)	۰/۹۹	۱	(۱۰۲/۷۶, ۱۲/۸۴)	۲۶/۵۳	۵۴/۶۹	۲۲/۰۶	۶
۶/۴۳	۳/۰۵	(۱۰۴/۰۳, ۱۳/۰۰)	(۱۰۶/۷۳, ۱۳/۳۴)	۰/۹۰	۱	(۱۰۹/۱۷, ۱۳/۶۵)	۳۴/۹۸	۵۰/۲۴	۱۸/۰۹	۷
۰/۶۹	۴/۸۰	(۸۳/۴۴, ۱۰/۴۳)	(۸۶/۷۳, ۱۰/۸۴)	۱	۱	(۸۲/۸۹, ۱۰/۳۶)	۱۹/۵۳	۴۵/۱۳	۱۶/۹۹	۸
۵/۳۷	۵/۹۷	(۸۳/۱۸, ۱۰/۴۰)	(۸۳/۶۶, ۱۰/۴۶)	۰/۹۲	۱	(۷۸/۸۸, ۹/۸۶)	۳۰/۱۰	۳۱/۵۸	۱۹/۴۴	۹
۳/۰۹	۷/۲۵	(۱۱۱/۲۷, ۱۳/۹۱)	(۱۱۴/۶۰, ۱۴/۳۳)	۰/۹۶	۱	(۱۰۸/۸۰, ۱۳/۶۰)	۴۱/۶۸	۵۲/۶۶	۱۶/۸۴	۱۰
۴/۵۷	۳/۵۶	(۷۶/۲۷, ۹/۵۳)	(۷۷/۰۷, ۹/۶۳)	۰/۹۴	۱	(۷۹/۹۲, ۹/۹۹)	۲۰/۴۲	۳۳/۵۱	۱۹/۷۹	۱۱
۶/۶۲	۶/۰۷	(۹۶/۰۹, ۱۲/۰۱)	(۹۶/۵۳, ۱۲/۰۷)	۰/۹۰	۱	(۱۰۱/۳۹, ۱۲/۶۷)	۳۳/۸۱	۳۹/۹۳	۲۰/۸۴	۱۲
۵/۲۹	۵/۷۱	(۱۰۲/۲۲, ۱۲/۷۸)	(۱۰۱/۸۹, ۱۲/۷۴)	۰/۹۳	۱	(۱۰۶/۴۵, ۱۳/۳۱)	۲۸/۹۳	۴۷/۵۰	۲۴/۱۲	۱۳
۴/۳۲	۷/۳۶	(۱۱۱/۳۱, ۱۳/۹۱)	(۱۱۳/۷۵, ۱۴/۲۲)	۰/۹۵	۱	(۱۰۷/۸۶, ۱۳/۴۸)	۳۱/۰۲	۵۹/۰۷	۲۰/۵۴	۱۴
۴۹/۲۴	۴۵/۲۳	(۱۰۱/۱۹, ۱۲/۶۵)	(۱۰۴/۴۰, ۱۳/۰۵)	۰	۱	(۱۴۰/۵۸, ۱۷/۵۷)	۲۷/۳۴	۵۴/۵۱	۱۸/۳۷	۱۵
۴/۲۹	۲/۸۰	(۶۷/۵۸, ۸/۴۵)	(۶۶/۳۹, ۸/۳۰)	۰/۹۵	۱	(۶۴/۱۵, ۸/۰۲)	۴/۹۹	۳۳/۸۶	۲۴/۹۰	۱۶
۱/۸۵	۱/۰۴	(۱۲۰/۸۳, ۱۵/۱۰)	(۱۲۳/۱۴, ۱۵/۳۹)	۱	۱	(۱۲۲/۳۱, ۱۵/۲۹)	۳۰/۱۳	۶۷/۷۸	۲۲/۴۸	۱۷
۰/۹۲	۲/۰۹	(۱۲۱/۲۲, ۱۵/۱۵)	(۱۲۲/۱۵, ۱۵/۲۷)	۱	۱	(۱۲۰/۴۸, ۱۵/۰۶)	۵۰/۱۱	۵۰/۳۹	۲۰/۶۹	۱۸
۱/۸۷	۵/۱۵	(۱۳۴/۵۰, ۱۶/۸۱)	(۱۳۷/۱۲, ۱۷/۱۴)	۱	۱	(۱۳۳/۰۰, ۱۶/۶۳)	۴۰/۴۷	۷۲/۴۵	۲۲/۰۱	۱۹
۲/۸۷	۴/۶۴	(۸۳/۰۲, ۱۰/۳۸)	(۸۴/۴۴, ۱۰/۵۶)	۰/۹۸	۱	(۸۰/۷۲, ۱۰/۰۹)	۲۷/۱۰	۳۵/۴۹	۱۸/۴۸	۲۰
۲/۵۲	۲/۵۴	(۸۸/۸۱, ۱۱/۱۰)	(۸۸/۷۹, ۱۱/۱۰)	۰/۹۹	۱	(۹۰/۸۳, ۱۱/۳۵)	۲۳/۴۰	۴۰/۶۱	۲۲/۵۷	۲۱
۸/۴۱	۷/۸۱	(۹۸/۲۲, ۱۲/۲۸)	(۹۸/۷۱, ۱۲/۳۴)	۰/۸۶	۱	(۱۰۴/۹۵, ۱۳/۱۲)	۳۵/۲۴	۴۰/۷۹	۲۰/۸۱	۲۲
۰/۱۷	۱/۰۲	(۹۹/۱۵, ۱۲/۳۹)	(۱۰۰/۱۱, ۱۲/۵۱)	۱	۱	(۹۹/۲۹, ۱۲/۴۱)	۲۷/۰۱	۴۸/۷۱	۲۱/۸۷	۲۳
۲/۰۱	۵/۱۸	(۸۲/۰۷, ۱۰/۲۶)	(۸۴/۶۱, ۱۰/۵۸)	۱	۱	(۸۰/۴۶, ۱۰/۰۶)	۲۵/۱۳	۳۸/۲۸	۱۶/۸۶	۲۴
۱/۴۴	۳/۳۶	(۹۷/۷۱, ۱۲/۲۱)	(۹۶/۱۸, ۱۲/۰۲)	۱	۱	(۹۸/۸۶, ۱۲/۳۶)	۳۲/۷۲	۳۸/۴۳	۲۴/۶۱	۲۵
۲/۶۳	۰/۱۷	(۹۲/۷۰, ۱۱/۵۹)	(۹۰/۷۳, ۱۱/۳۴)	۰/۹۸	۱	(۹۰/۵۹, ۱۱/۳۲)	۴۸/۳۰	۲۱/۴۵	۲۱/۳۰	۲۶
۴/۴۴	۱/۶۱	(۵۶/۴۷, ۷/۰۶)	(۶۱/۳۱, ۷/۶۶)	۰/۹۴	۱	(۶۰/۰۲, ۷/۵۰)	—/۷۶	۳۹/۹۴	۱۴/۲۵	۲۷
۴/۷۱	۳/۲۴	(۱۱۴/۶۵, ۱۴/۳۳)	(۱۱۵/۸۳, ۱۴/۴۸)	۰/۹۴	۱	(۱۱۸/۴۲, ۱۴/۸۰)	۳۹/۶۰	۵۳/۰۹	۲۱/۴۱	۲۸
۶/۱۳	۶/۳۵	(۹۲/۷۴, ۱۱/۵۹)	(۹۲/۹۱, ۱۱/۶۱)	۰/۹۱	۱	(۸۷/۸۳, ۱۰/۹۸)	۱۸/۹۰	۴۷/۷۹	۲۳/۸۲	۲۹
۱/۹۸	۲/۳۶	(۱۱۴/۷۶, ۱۴/۳۴)	(۱۱۵/۰۶, ۱۴/۳۸)	۱	۱	(۱۱۳/۱۷, ۱۴/۱۵)	۴۴/۲۸	۴۷/۹۹	۲۱/۹۲	۳۰
۴/۹۳۸	۵/۵۷۷	MPE								

جدول ۲. برآورد مدل رگرسیونی داده‌های شبیه‌سازی شده

روش	برآورد رگرسیونی
FLS	$\hat{y} = (12/1214, 1/5152) + (0/5819, 0/0726)x_1$ $+ (1/041, 0/13014)x_2 + (0/9084, 0/1136)x_3$
پیشنهادی	$\hat{y} = (3/9303, 0/4913) + (1/0771, 0/1347)x_1$ $+ (0/94907, 0/1186)x_2 + (0/9425, 0/1177)x_3$

جدول ۳. مجموعه داده‌ها، وزن‌ها (W_i) و مقایسه مدل FLS و مدل پیشنهادی با توجه به شاخص MPE برای داده‌های مثال

سیمان پورتلند						(y_i, s_i)	x_3	x_2	x_1	مشاهده
d_{Prop}	d_{FLS}	$\hat{Y}_{Prop}$	$\hat{Y}_{FLS}$	W_{Prop}	W_{FLS}					
0/23	1/42	(78/72, 6/89)	(79/85, 6/87)	1	1	(78/5, 6/9)	6	26	7	1
2/42	3/87	(72/26, 6/21)	(70/72, 6/26)	0/93	1	(74/3, 6/4)	15	29	1	2
1/62	1/89	(105/53, 9/60)	(105/79, 9/60)	1	1	(104/3, 9/4)	8	56	11	3
1/58	1/13	(88/97, 7/91)	(88/47, 7/93)	1	1	(87/6, 7/8)	8	31	11	4
0/54	2/38	(95/94, 8/65)	(97/87, 8/60)	1	1	(95/5, 8/6)	6	52	7	5
4/89	5/07	(105/03, 9/54)	(104/81, 9/56)	0/73	1	(109/2, 9/9)	9	55	11	6
1/21	0/89	(103/71, 9/40)	(102/11, 9/45)	1	1	(102/7, 9/3)	17	71	3	7
2/67	3/13	(74/74, 6/42)	(70/08, 6/55)	1	1	(72/5, 6/2)	22	31	1	8
2/75	4/96	(90/96, 8/10)	(88/61, 8/17)	0/90	1	(93/3, 8/3)	18	54	2	9
0/49	0/87	(115/45, 10/58)	(115/05, 10/61)	1	1	(115/9, 10/6)	4	47	21	10
13/59	8/47	(80/86, 7/04)	(76/03, 7/18)	0	1	(68, 7/4)	23	40	1	11
1/62	1/48	(112/32, 10/28)	(112/43, 10/29)	1	1	(113/3, 10/6)	9	66	11	12
3/11	3/94	(111/82, 10/25)	(112/67, 10/23)	0/90	1	(109/4, 9/9)	8	68	10	13
2/82	3/03	MPE								

جدول ۴. برآورد مدل رگرسیونی داده‌های مثال سیمان پورتلند

روش	برآورد رگرسیونی
FLS	$\hat{y} = (53/53, 3/796) + (1/43, 0/170)x_1$ $+ (0/69, 0/066)x_2 + (-0/28, 0/023)x_3$
پیشنهادی	$\hat{y} = (48/9385, 3/9244) + (1/6536, 0/1636)x_1$ $+ (0/6622, 0/067)x_2 + (0/1642, 0/0105)x_3$

۲.۶ داده‌های سیمان پورتلند

در این مثال اثر ترکیب سیمان پورتلند روی گرمای آزاد شده در حین سفت شدن آن مورد بررسی قرار می‌گیرد [؟] داده‌ها در جدول ۳ نمایش داده شده‌اند. این مسئله با ۱۳ مشاهده شامل ۳ متغیر تبیینی به صورت x_1 ، مقدار آلومینات تری کلسیم، x_2 ، مقدار سیلیکات تری کلسیم، x_3 ، مقدار آلومینو فریت تترا کلسیم

در جدول ۲ برای روش FLS که توسط فاصله دیاموند محاسبه شد و همین‌طور برای روش پیشنهادی، برآوردهای مدل رگرسیونی آورده شده‌اند. با استفاده از شاخص نیکویی برازش که در بخش؟؟ معرفی شد، مقدار MPE برای روش FLS معادل ۵/۵۸ و برای روش پیشنهادی ۴/۹۴ به دست می‌آید. بنا بر این با توجه به کمتر شدن مقدار MPE روش پیشنهادی به روش FLS ، برتری روش پیشنهادی مشخص می‌شود (جدول ۱).

می باشد. (y_i, s_i) مقدار گرمای آزاد شده در هر گرم سیمان است. ۱۱-آمین مشاهده متغیر وابسته را از $(۸۳/۸, ۷/۴)$ به $(۶۸, ۷/۴)$ تغییر داده و آن را به عنوان یک داده دور افتاده در نظر می گیریم. مجموعه داده ها و وزن های روش پیشنهادی در جدول ۳ نشان داده شده و در جدول ۴، برای روش FLS و روش پیشنهادی، برآوردهای مدل رگرسیونی آورده شده اند. در جدول ۳ وزن ۱۱-آمین مشاهده، در روش پیشنهادی، تقریباً «۰» به دست آمد. وزن هایی که به عنوان نتیجه روش پیشنهادی پیدا می شوند، درجات عضویت هر مشاهده هستند. این عضویت ها، اثرات مشاهدات را روی مدل نشان می دهند. همان طور که در جدول ۳ قابل مشاهده است داده دور افتاده با درجه کوچکی از عضویت، روی مدل تأثیر می گذارد. درجه عضویت سایر مشاهدات، یک یا نزدیک به یک هستند که اثر مهمی روی برآورد مدل رگرسیونی دارند. در این مثال هم کمتر بودن مقدار MPE روش پیشنهادی که $۲/۸۲$ به دست آمد به مقدار MPE روش FLS که معادل $۳/۰۳$ است، نشان دهنده برتری روش پیشنهادی به روش FLS است (جدول ۳).

برآوردهای خوبی از روش های معمول LS حاصل نخواهند شد و باید از روش های جایگزین استوار فازی استفاده کرد. در این پژوهش، برای محاسبه برآورد مدل رگرسیونی خطی چندگانه، زمانی که مشاهدات متغیر وابسته، فازی و در مجموعه داده ها، مشاهده دور افتاده داشته باشیم، از تحلیل رگرسیون استوار فازی استفاده شده و یک روش تکراری پیشنهاد شد. در این روش، باقی مانده ها با استفاده از شاخص OM چانگ و لی [۹] محاسبه شده سپس با استفاده از تابعی بسیار شبیه به تابع استوار معرفی شده در انتهای بخش؟؟، برای باقی مانده ها تابع عضویت در نظر گرفته شد.

الگوریتم پیشنهاد شده در این پژوهش تا رسیدن به یک دقت مطلوب برای پارامترها ادامه پیدا می کند. در دو مثال مطرح شده در این مقاله، وزن های به دست آمده از روش رگرسیون استوار پیشنهادی و رگرسیون حداقل توان های دوم با هم مقایسه شده اند (جدول ۱ و ۳). همان طور که در این جداول مشاهده می شود، داده های دور افتاده در روش پیشنهادی بوضوح از وزن های کمتری نسبت به روش FLS برخوردارند و همان طور که در مثال ها بیان شد شاخص نیکویی برازش MPE برتری روش پیشنهادی را نسبت به روش FLS مشخص می کند.

۷ نتیجه گیری

هرگاه در بین مجموعه داده ها، مشاهده دور افتاده وجود داشته باشد و متغیرها یا مشاهدات مربوط به آنها دقیق نباشند،

مراجع

- [۱] کلکین نما، م. و طاهری، س. م. رگرسیون فازی بر اساس کمترین قدر مطلق انحرافات، اندیشه آماری، سال ۱۲ شماره ۱ (شماره پیاپی ۲۳) (شماره ویژه دومین کارگاه آمار و احتمال فازی)، صص ۶۷-۵۹، ۱۳۸۶.
- [۲] میرزایی، ی. و همکاران. رگرسیون فازی: مروری بر چند رویکرد. اندیشه آماری، سال ۱۲ شماره ۱ (شماره پیاپی ۲۳) (شماره ویژه دومین کارگاه آمار و احتمال فازی)، صص ۴۷-۳۵، ۱۳۸۶.
- [3] Adamo, J. M. (1980). Fuzzy decision trees. *Fuzzy sets and systems*, 4(3), 207-219.
- [4] Asai, H. and Tanaka, H. (1982). Linear regression analysis with fuzzy model, *IEEE Trans. Systems, Man and Cybernetics*, 12, 903-907.
- [5] Bárdossy, A. (1990). Note on fuzzy regression, *Fuzzy Sets and Systems*, 37(1), 65-75.

- [6] Buckley, J. J. and Chanas, S. (1989). A fast method of ranking alternatives using fuzzy numbers, *Fuzzy sets and systems*, **30(3)**, 337-338.
- [7] Campos, L. M. and González, A. (1989). A subjective approach for ranking fuzzy numbers, *Fuzzy sets and systems*, **29(2)**, 145-153.
- [8] Chang, P. T. and Lee, E. S. (1994). Ranking of fuzzy sets based on the concept of existence, *Computers Mathematics with Applications*, **27(9-10)**, 1-21.
- [9] Chang, P. T. and Lee, E. S. (1996). A generalized fuzzy weighted least-squares regression, *Fuzzy sets and systems*, **82(3)**, 289-298.
- [10] Diamond, P. (1988). Fuzzy least squares. *Information Sciences*, **46(3)**, 141-157.
- [11] Draper, N. R. and Smith, H. (2014). *Applied regression analysis*, John Wiley Sons.
- [12] Dubois, D. and Prade, H. (1980). *Fuzzy sets and systems: Theory and Applications*, Academic Press.
- [13] González, A. (1990). A study of the ranking function approach through mean values, *Fuzzy sets and systems*, **35(1)**, 29-41.
- [14] Hasanpour, H., Maleki, H. R. and Yaghoubi, M. A. (2010). Fuzzy linear regression model with crisp coefficients: a goal programming approach, *Iranian Journal of Fuzzy Systems*, **7(2)**, 19-39.
- [15] Hung, W. L. and Yang, M. S. (2006). An omission approach for detecting outliers in fuzzy regression models, *Fuzzy sets and systems*, **157(23)**, 3109-3122.
- [16] Kaufmann, A. (1986). Hybrid Data-Variou Associations Between Fuzzy Subsets and Random Variables, *In Fuzzy Sets Theory and Applications*, 171-211.
- [17] Kula, K. Ş. and Apaidin, A. (2008). Fuzzy robust regression analysis based on the ranking of fuzzy sets. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, **16(5)**, 663-681.
- [18] Maeder, R. E. (1991). *Programming in mathematica*, Addison-Wesley Longman Publishing Co., Inc.
- [19] Nanda, S. (1989). On sequences of fuzzy numbers, *Fuzzy sets and systems*, **33(1)**, 123-126.
- [20] Tanaka, H. (1984). A formulation of fuzzy linear programming based on comparison of fuzzy numbers, *Control and cybernetics*, **13(3)**, 185-194.
- [21] Xu, R. and Le, C. (2001). Multidimensional least-squares fitting with a fuzzy model, *Fuzzy Sets and Systems*, **119(2)**, 215-223.

- [22] Yager, R. (1979). Ranking fuzzy subsets over the unit interval. *In Decision and Control including the 17th Symposium on Adaptive Processes*, 1435-1437.
- [23] Yang, M. S. and Ko, C. H. (1997). On cluster-wise fuzzy regression analysis, *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, **27(1)**, 1-13.
- [24] Yang, M. S. and Lin, T. S. (2002). Fuzzy least-squares linear regression analysis for fuzzy input-output data, *Fuzzy sets and systems*, **126(3)**, 389-399.
- [25] Yang, M. S. and Liu, H. H. (2003). Fuzzy least-squares algorithms for interactive fuzzy linear regression models, *Fuzzy sets and systems*, **135(2)**, 305-316.
- [26] Zadeh, L. A. (1978). Fuzzy sets as a basis for a theory of possibility, *Fuzzy Sets and Systems*, **1(3-28)**, 61-72.