رگرسیون خطی چندگانه کمترین مربعات بر اساس مشاهدات فازی فاصله‌ای مقدار و با استفاده از فاصله علامت‌دار یائو–ویو

فائزه ترکیان – محسن عارفی – محمد قاسم اکبری

گروه آمار، دانشکده علوم، دانشگاه بیرجند

چکیده: در این مقاله، ابتدا فاصله علامت‌دار یائو–ویو برای داده‌های فازی فاصله‌ای مقدار تعمیم و سپس بر اساس آن، پارامترهای مدل رگرسیونی خطی چندگانه برآورد شده‌اند. در ادامه، بر اساس یک اندازه مشابهت بین اعداد فازی فاصله‌ای مقدار، به بررسی نیکویی برازش مدل و حذف داده‌های پرت پرداخته شده است.

واژه‌های کلیدی: مجموعه فازی فاصله‌ای مقدار، رگرسیون خطی، اندازه مشابهت

۱ مقدمه و مفاهیم اوّلیه

تحلیل رگرسیونی یکی از روش‌های آماری برای تحلیل داده‌ها و یافتن ارتباط بین مجموعه‌ای از متغیرها است که در زمینه‌های مختلف کاربرد دارد. تحلیل رگرسیونی در یک محیط فازی به وسیله نویسندگان بسیاری مورد بررسی قرار گرفته است. برای بررسی مدل‌های رگرسیونی در یک محیط فازی می‌توان به تاناکا و همکاران (۱۹۸۲)، کلمینس (۱۹۸۷)، ناتر و وانچ (۲۰۰۲)، عربپور و تاتا (۲۰۰۸)، حسن‌پور و همکاران (۲۰۱۰) اشاره کرد. همچنین در سال‌های اخیر، مدل‌های رگرسیونی در یک محیط فازی فاصله‌ای مقدار(شهودی) نیز مطالعه گردیده است. برای بررسی برخی از این مدل‌ها به ترکیان و همکاران (۱۳۹۰) با مدل رگرسیون خطی ساده با مشاهدات شهودی، و به عارفی و طاهری (۱۳۹۰) با مدل رگرسیون کمترین مربعات و با ورودی و خروجی مبهم (فازی فاصله‌ای مقدار) و ضرایب مبهم مراجعه نمایید.

کار حاضر تعمیم روش معرفی شده در ترکیان و همکاران (۱۳۹۰) بر اساس متر یائو–ویو و بر اساس یک مدل رگرسیونی چندگانه است. همچنین با معرفی یک اندازه‌ی مشابهت، به بررسی نیکویی برازش مدل و یافتن داده‌های پرت پرداخته شده است.

در ادامه برخی مفاهیم اوّلیه مورد نیاز در مقاله را بر اساس آتاناسوف (۱۹۹۹) و گوها و چاکرابورتی (۲۰۱۰) مورد بررسی قرار می‌دهیم.

تعریف ۱ یک مجموعه فازی فاصله‌ای مقدار $\tilde{A}$، از مجموعه مرجع X، به صورت سه‌تایی $\tilde{A} = \{(x, \mu_{\tilde{A}}(x), \nu_{\tilde{A}}(x)) : x \in X\}$ تعریف می‌شود که $\mu_{\tilde{A}}(x) : X \to [0, 1]$ تابع عضویت و $\nu_{\tilde{A}}(x) : X \to [0, 1]$ تابع عدم عضویت x در $\tilde{A}$ هستند و برای هر $x \in X$ در شرط $0 \leq \mu_{\tilde{A}}(x) + \nu_{\tilde{A}}(x) \leq 1$ صدق می کنند.
در یک حالت خاص اگر تابع عدم عضویت به صورت $\nu_{\tilde{A}}(x) = 1 - \mu_{\tilde{A}}(x)$ باشد، به مجموعه‌های فازی کاهش می‌یابد.

تعریف ۲ فرض کنید $\tilde{A}$ یک مجموعه فازی فاصله‌ای مقدار باشد، آنگاه α−برش $\tilde{A}$ به وسیله دو مجموعه زیر تعریف می‌شود:

$$\tilde{A}[\alpha] = \{A_\alpha^\mu, A_\alpha^{1-\nu}\},$$

که در آن $A_\alpha^\mu = \{x : \mu_{\tilde{A}}(x) \geq \alpha\}$ و $A_\alpha^{1-\nu} = \{1 - \nu_{\tilde{A}}(x) \geq \alpha\}$.

تعریف ۳ یک مجموعه فازی فاصله‌ای مقدار $\tilde{A}$ را یک عدد فازی فاصله‌ای مقدار گویند، اگر خواص زیر را دارا باشد:
۱) مقدار m وجود داشته باشد که به ازای آن $\nu_{\tilde{A}}(m) = 0$ و $\mu_{\tilde{A}}(m) = 1$.
۲) α−برش‌های توابع μ و $1 - \nu$ فواصل بسته و کراندار باشند.

تعریف ۴ مجموعه $\tilde{A}$ را یک عدد فازی فاصله‌ای مقدار مثلثی نامند، اگر توابع عضویت و عدم عضویت آن برای هر $s_2 \geq s_4$ و $s_1 \geq s_3$ به صورت زیر باشند (گوها و چاکرابورتی؛۲۰۱۰):

$$\mu_{\tilde{A}}(x) = \begin{cases} 1 - \frac{m-x}{s_1} & m - s_1 \leq x < m, \\ 1 & x = m, \\ 1 - \frac{x-m}{s_2} & m < x < m + s_2, \\ 0 & \text{جاهای دیگر}, \end{cases} \qquad \nu_{\tilde{A}}(x) = \begin{cases} \frac{m-x}{s_3} & m - s_3 \leq x < m, \\ 0 & x = m, \\ \frac{x-m}{s_4} & m < x < m + s_4, \\ 1 & \text{جاهای دیگر}. \end{cases}$$

که به صورت $\tilde{A} = (m; s_1, s_2, s_3, s_4)_T$ نشان داده می‌شود.

تذکر ۱ بر اساس تعریف ۱، α−برش یک عدد فازی فاصله‌ای مقدار مثلثی $\tilde{A} = (m; s_1, s_2, s_3, s_4)_T$ به صورت زیر خلاصه می‌شود:

$$\tilde{A}[\alpha] = \{[A_\mu^L(\alpha), A_\mu^R(\alpha)], [A_{1-\nu}^L(\alpha), A_{1-\nu}^R(\alpha)]\},$$

که در آن

$$\begin{cases} [A_\mu^L(\alpha), A_\mu^R(\alpha)] = [m - s_1(1 - \alpha), m + s_2(1 - \alpha)], \\ [A_{1-\nu}^L(\alpha), A_{1-\nu}^R(\alpha)] = [m - s_3(1 - \alpha), m + s_4(1 - \alpha)]. \end{cases}$$

بر اساس اصل گسترش تعمیم یافته روی اعداد فازی فاصله‌ای مقدار(طاهری و زارعی؛ ۲۰۱۱)، عملگرهای حسابی به صورت زیر تعریف می‌شوند.

قضیه ۷ اگر $\tilde{M} = (m; s_1, s_2, s_3, s_4)_T$ و $\tilde{N} = (n; r_1, r_2, r_3, r_4)_T$ دو عدد فازی فاصله‌ای مقدار مثلثی باشند و $\lambda \in R - \{0\}$، آنگاه داریم:

$$\begin{aligned}\tilde{M} \oplus \tilde{N} &= (m; s_1, s_2, s_3, s_4)_T \oplus (n; r_1, r_2, r_3, r_4)_T \\ &= (m+n; s_1+r_1, s_2+r_2, s_3+r_3, s_4+r_4)_T\end{aligned}$$

$$\lambda \otimes \tilde{M} = \begin{cases}(\lambda m; \lambda s_1, \lambda s_2, \lambda s_3, \lambda s_4)_T & \lambda > 0, \\ (\lambda m; -\lambda s_2, -\lambda s_1, -\lambda s_4, -\lambda s_3)_T & \lambda < 0,\end{cases}$$

تعریف ۵ اگر $\tilde{A}$ و $\tilde{B}$ دو مجموعه فازی فاصله‌ای مقدار باشند، آنگاه

$$(i)\ \tilde{A} \subseteq \tilde{B} \iff \mu_{\tilde{A}}(x) \le \mu_{\tilde{B}}(x)\ ,\ \ \nu_{\tilde{A}}(x) \ge \nu_{\tilde{B}}(x) \quad \forall x \in X,$$
$$(ii)\ \tilde{A} = \tilde{B} \iff \mu_{\tilde{A}}(x) = \mu_{\tilde{B}}(x)\ ,\ \ \nu_{\tilde{A}}(x) = \nu_{\tilde{B}}(x) \quad \forall x \in X.$$

۲ تعمیم فاصله علامت‌دار یائو–ویو بر اساس مشاهدات فازی فاصله‌ای مقدار

تعریف ۶ فرض کنید $\tilde{A}$ و $\tilde{B}$ دو عدد فازی فاصله‌ای مقدار باشند، آنگاه فاصله علامت‌دار یائو–ویو به صورت زیر تعمیم داده می‌شود (ترکیان و همکاران؛۱۳۹۰):

$$d(\tilde{A}, \tilde{B}) = \tfrac{1}{2}\int_0^1 (M_\alpha(\tilde{A}_\mu) - M_\alpha(\tilde{B}_\mu))d\alpha + \tfrac{1}{2}\int_0^1 (M_\alpha(\tilde{A}_{1-\nu}) - M_\alpha(\tilde{B}_{1-\nu}))d\alpha,$$

که در آن

$$M_\alpha(\tilde{A}_\mu) = \frac{A_\mu^L(\alpha)+A_\mu^R(\alpha)}{2}, \qquad M_\alpha(\tilde{A}_{1-\nu}) = \frac{A_{1-\nu}^L(\alpha)+A_{1-\nu}^R(\alpha)}{2},$$
$$M_\alpha(\tilde{B}_\mu) = \frac{B_\mu^L(\alpha)+B_\mu^R(\alpha)}{2}, \qquad M_\alpha(\tilde{B}_{1-\nu}) = \frac{B_{1-\nu}^L(\alpha)+B_{1-\nu}^R(\alpha)}{2}.$$

در یک حالت خاص اگر $\tilde{A} = (a; r_1, r_2, r_3, r_4)_T$ و $\tilde{B} = (b; s_1, s_2, s_3, s_4)_T$ دو عدد فازی فاصله‌ای مقدار مثلثی باشند، داریم:

$$d(\tilde{A}, \tilde{B}) = a - b + \frac{(r_2 - r_1 + r_4 - r_3) - (s_2 - s_1 + s_4 - s_3)}{8}.$$

تذکر ۲ اگر $\tilde{A}$، $\tilde{B}$ و $\tilde{C}$ اعداد فازی فاصله‌ای مقدار باشند، آنگاه فاصله علامت‌دار بیان شده در تعریف ۶، دارای خصوصیات زیر است:

$$\begin{array}{l} i)\ d(\tilde{A},\tilde{B}) = -d(\tilde{B},\tilde{A}),\\ ii)\ d(\tilde{A},\tilde{B}) + d(\tilde{B},\tilde{C}) = d(\tilde{A},\tilde{C}),\\ iii)\ \tilde{A}\approx\tilde{B} \Longleftrightarrow \tilde{B}\approx\tilde{A},\\ iv)\ \tilde{A}\approx\tilde{B},\ \tilde{B}\approx\tilde{C} \Longrightarrow \tilde{A}\approx\tilde{C}. \end{array}$$

که در رابطه‌ی فوق داریم:

$$\tilde{A}\approx\tilde{B} \iff d(\tilde{A},0) = d(\tilde{B},0) \iff d(\tilde{A},\tilde{B}) = 0.$$

اثبات: موردهای (i)، (iii) و (iv) به سادگی قابل اثبات است. قسمت (ii) به صورت زیر اثبات می‌شود:

$$\begin{aligned} d(\tilde{A},\tilde{B}) &= \tfrac{1}{2}\int_0^1 (M_\alpha(\tilde{A}_\mu) - M_\alpha(\tilde{B}_\mu))d\alpha + \tfrac{1}{2}\int_0^1 (M_\alpha(\tilde{A}_{1-\nu}) - M_\alpha(\tilde{B}_{1-\nu}))d\alpha\\ &= \tfrac{1}{2}[\int_0^1 (M_\alpha(\tilde{A}_\mu) - M_\alpha(\tilde{C}_\mu))d\alpha + \int_0^1 (M_\alpha(\tilde{A}_{1-\nu}) - M_\alpha(\tilde{C}_{1-\nu}))d\alpha]\\ &\quad -\tfrac{1}{2}[(\int_0^1 (M_\alpha(\tilde{B}_\mu) - M_\alpha(\tilde{C}_\mu))d\alpha + \int_0^1 (M_\alpha(\tilde{B}_{1-\nu}) - M_\alpha(\tilde{C}_{1-\nu}))d\alpha]\\ &= d(\tilde{A},\tilde{C}) - d(\tilde{B},\tilde{C}). \end{aligned}$$

۳ رگرسیون‌خطی چندگانه با مشاهدات فازی فاصله‌ای مقدار

مدل رگرسیونی خطی زیر را در نظر بگیرید:

$$\tilde{y_i} = \beta_0 \oplus (\beta_1\otimes\tilde{x}_{i1}) \oplus (\beta_2\otimes\tilde{x}_{i2}) \oplus \dots \oplus (\beta_k\otimes\tilde{x}_{ik}),\quad i = 1,\dots,n$$

که در آن $\tilde{x}_{i1},\tilde{x}_{i2},\dots,\tilde{x}_{ik}$ متغیرهای مستقل، $\tilde{y_i}$ متغیر وابسته(پاسخ) و $\beta_0,\beta_1,\dots,\beta_k$ ضرایب مدل رگرسیونی است. می‌خواهیم به روش کمترین مربعات خطا و با استفاده از تعریف ۶، ضرایب مدل رگرسیونی را برآورد نمائیم. فرض کنید $(\tilde{y}_1,\tilde{x}_{11},\dots,\tilde{x}_{1k})$, $\dots,(\tilde{y}_n,\tilde{x}_{n1},\dots,\tilde{x}_{nk})$، یک نمونه n تایی از جامعه است که در آن $\tilde{y_i}$ و $\tilde{x}_{ij}$ به صورت مشاهدات فازی فاصله‌ای مقدار مثلثی زیر درنظر گرفته شده‌اند:

$$\tilde{y_i} = (y_i; r_{i1}, r_{i2}, r_{i3}, r_{i4})_T,\qquad i = 1,\dots,n$$

$$\tilde{x}_{ij} = (x_{ij}; s_{ij1}, s_{ij2}, s_{ij3}, s_{ij4})_T.\quad j = 1,\dots,k$$

بر اساس تعمیم فاصله علامت‌دار یائو-ویو، مجموع مربعات خطا به صورت زیر به‌دست می‌آید:

$$\begin{aligned} SSE &= \sum_{i=1}^{n} d^2(\tilde{y_i}, \beta_0 \oplus (\beta_1\otimes\tilde{x}_{i1}) \oplus (\beta_2\otimes\tilde{x}_{i2}) \oplus \dots \oplus (\beta_k\otimes\tilde{x}_{ik}))\\ &= \sum_{i=1}^{n}((y_i - \beta_0 - \sum_{j=1}^{k}\beta_j x_{ij}) + \frac{1}{8}[R_i - \sum_{j=1}^{k}\beta_j S_{ij}])^2, \end{aligned}$$

که در آن

$$R_i = r_{i2} - r_{i1} + r_{i4} - r_{i3},\qquad S_{ij} = s_{ij2} - s_{ij1} + s_{ij4} - s_{ij3},$$
$$i = 1,\dots,n,\quad j = 1,\dots,k.$$

به منظور به دست آوردن مقادیر بهینه ضرایب $\beta_k, ..., \beta_1, \beta_0$، مجموع مربعات خطا را مینیمم می‌کنیم. با مشتق گرفتن نسبت به پارامترهای مدل داریم:

$$\frac{\partial SSE}{\partial \beta_0} = \sum_{i=1}^{n}(y_i + \frac{1}{8}R_i) - n\beta_0 - \sum_{i=1}^{n}\sum_{j=1}^{k}\beta_j(x_{ij} + \frac{1}{8}S_{ij}),$$

$$\frac{\partial SSE}{\partial \beta_j} = \sum_{i=1}^{n}(x_{ij} + \frac{1}{8}S_{ij})(y_i + \frac{1}{8}R_i) - \sum_{i=1}^{n}(x_{ij} + \frac{1}{8}S_{ij})(\beta_0 + \sum_{j=1}^{k}\beta_j(x_{ij} + \frac{1}{8}S_{ij})).$$

روابط بالا را می‌توان به صورت ماتریسی زیر نیز خلاصه کرد:

$$(X + \frac{S}{8})'(X + \frac{S}{8})\beta = (X + \frac{S}{8})'(Y + \frac{R}{8}),$$

که در آن

$$X = \begin{bmatrix} 1 & x_{11} & \dots & x_{1k} \\ 1 & x_{21} & \dots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{nk} \end{bmatrix}_{n\times(k+1)}, \qquad S = \begin{bmatrix} 0 & S_{11} & \dots & S_{1k} \\ 0 & S_{21} & \dots & S_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & S_{n1} & \dots & S_{nk} \end{bmatrix}_{n\times(k+1)},$$

$$R = \begin{bmatrix} R_1 \\ R_2 \\ \vdots \\ R_n \end{bmatrix}, \qquad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \qquad Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}.$$

اگر معکوس ماتریس $(X+\frac{S}{8})'(X+\frac{S}{8})$ وجود داشته باشد، آنگاه برآورد ضرایب مدل رگرسیونی به صورت زیر به‌دست می‌آید:

$$\hat{\beta} = \left[(X + \frac{S}{8})'(X + \frac{S}{8})\right]^{-1}\left[(X + \frac{S}{8})'(Y + \frac{R}{8})\right]. \qquad (1)$$

۴ معیار نیکویی برازش مدل

در این بخش، برای ارزیابی نیکویی برازش مدل به تعریف یک اندازه مشابهت بین دو عدد فازی فاصله‌ای مقدار می‌پردازیم.

قضیه ۸ فرض کنید $IVFS(X)$ مجموعه تمام مجموعه‌های فازی فاصله‌ای مقدار از مجموعه مرجع X باشد. نگاشت $S : IVFS(X)\times IVFS(X) \longrightarrow [0, 1]$ را یک اندازه مشابهت نامند، اگر شرایط زیر را دارا باشد:

$$\begin{aligned} &i)\ S(\tilde{A}, \tilde{B}) \in [0, 1], \\ &ii)\ S(\tilde{A}, \tilde{B}) = 1 \Longleftrightarrow \tilde{A} = \tilde{B}, \\ &iii)\ S(\tilde{A}, \tilde{B}) = S(\tilde{B}, \tilde{A}), \\ &iv)\ \tilde{A} \subset \tilde{B} \subset \tilde{C} \Longrightarrow S(\tilde{A}, \tilde{C}) \leq \min(S(\tilde{A}, \tilde{B}), S(\tilde{B}, \tilde{C})). \end{aligned}$$

تعریف ۷ اگر $\tilde{A}=(a_0,a_1,a_2,a_3,a_4)_T$ و $\tilde{B}=(b_0,b_1,b_2,b_3,b_4)_T$ دو عدد فازی فاصله‌ای مقدار مثلثی باشند، اندازه مشابهت بین آنها به‌صورت‌زیر تعریف‌می‌شود:

$$S(\tilde{A},\tilde{B})=\frac{1}{1+d(\tilde{A},\tilde{B})}=\frac{1}{1+\frac{1}{5}\sum_{i=0}^{4}\mid a_i-b_i\mid^p}\ ,\qquad p\geq 1$$

که در آن $d(\tilde{A},\tilde{B})=\frac{1}{5}\sum_{i=0}^{4}\mid a_i-b_i\mid^p$ یک متر بین دو عدد فازی فاصله‌ای مقدار مثلثی می‌باشد.

لم ۱ اندازه تعریف شده در تعریف ۷، بر اساس قضیه۸، یک اندازه مشابهت است.

اثبات :

$$i)\ 0<\mid a_i-b_i\mid^p<\infty\quad \forall i=0,...,4\implies 0<\frac{1}{1+\frac{1}{5}\sum_{i=0}^{4}\mid a_i-b_i\mid^p}<1$$

$$ii)\ S(\tilde{A},\tilde{B})=1\iff 1+\tfrac{1}{5}\sum_{i=0}^{4}\mid a_i-b_i\mid^p=1\iff \tfrac{1}{5}\sum_{i=0}^{4}\mid a_i-b_i\mid^p=0$$

$$\iff \mid a_i-b_i\mid^p=0\quad i=0,...,4\iff \tilde{A}=\tilde{B}$$

$$iii)\ \mid a_i-b_i\mid^p=\mid b_i-a_i\mid^p\implies S(\tilde{A},\tilde{B})=S(\tilde{B},\tilde{A})$$

$$iv)\ \tilde{A}\subset\tilde{B}\subset\tilde{C}\implies a_0=b_0=c_0\ ,\ a_i\leq b_i\leq c_i\quad i=1,...,4$$

$$\implies \mid a_i-b_i\mid^p\leq\mid a_i-c_i\mid^p\implies 1+\tfrac{1}{5}\sum_{i=0}^{4}\mid a_i-b_i\mid^p\leq 1+\tfrac{1}{5}\sum_{i=0}^{4}\mid a_i-c_i\mid^p$$

$$\implies \frac{1}{1+\frac{1}{5}\sum_{i=0}^{4}\mid a_i-c_i\mid^p}\leq\frac{1}{1+\frac{1}{5}\sum_{i=0}^{4}\mid a_i-b_i\mid^p}\implies S(\tilde{A},\tilde{C})\leq S(\tilde{A},\tilde{B})$$

به طور مشابه $S(\tilde{A},\tilde{C})\leq S(\tilde{B},\tilde{C})$. بنابراین اثبات کامل است.

برای ارزیابی نیکویی برازش یک مدل رگرسیون فازی فاصله‌ای مقدار می‌توان میزان مشابهت هر مشاهده $\tilde{y}_i$ را از مقدار برآورد آن محاسبه کرد. در این مقاله، فرض می‌شود که بزرگ بودن این اعداد نشانگر برازش خوب مدل و بر عکس کوچک بودن این اعداد نشان دهنده برازش ضعیف مدل به مشاهدات است. همچنین برای بررسی دو مدل رگرسیونی می‌توان از میانگین اندازه مشابهت به صورت $MS=\frac{1}{n}\sum_{i=1}^{n}S(\tilde{y}_i,\hat{\tilde{y}}_i)$ نیز استفاده نمود و مدلی را انتخاب کرد که بیشترین مقدار MS را دارا باشد.

۵ شناسایی داده‌های دورافتاده (پرت)

در بین مشاهدات، داده‌ای که تفاوت بسیاری با بقیه مشاهدات داشته باشد و خط رگرسیونی را تحت تأثیر خود قرار دهد به عنوان داده پرت درنظر گرفته می‌شود. بر اساس اندازه مشابهت تعریف شده، می‌توان داده‌های دورافتاده را شناسایی کرد. داده‌ای که کمترین میزان مشابهت $S(\tilde{y}_i,\hat{\tilde{y}}_i)$ را داشته باشد، به عنوان داده پرت درنظر گرفته

می‌شود. اینکه مقدار S چقدر کوچک باشد، تا داده‌ی متناظر با آن را داده پرت تلقی کنیم، به نظر کاربر بستگی دارد.

۶ مثال کاربردی

داده های جدول ۲، مربوط به یک فرآیند رنگرزی در صنعت نساجی هستند(توانایی و همکاران؛ ۲۰۰۴). متغیر مستقل اول غلظت رنگ مورد استفاده برحسب گرم در لیتر، متغیر مستقل دوم زمان پروسه بر حسب دقیقه و متغیر پاسخ میزان جذب رنگ توسط پارچه بر حسب k/s است. به دلیل عدم نتیجه‌گیری با دقت کافی در اندازه‌گیری، مشاهدات به‌صورت اعداد فازی فاصله‌ای مقدار مثلثی در نظر گرفته شده‌اند. فرض کنید بتوان مدل رگرسیونی زیر را بین این داده‌ها درنظر گرفت:

$$\tilde{y}_i = \beta_۰ \oplus (\beta_۱ \otimes \tilde{x}_{i۱}) \oplus (\beta_۲ \otimes \tilde{x}_{i۲}). \quad i = ۱, ..., ۲۰$$

برای برآورد ضرایب مدل رگرسیونی فوق، ماتریس‌های X و S و بردارهای Y و R به صورت زیر به‌دست می‌آیند:

$$X = \begin{bmatrix} ۱ & ۰/۷۵ & ۳۶ \\ ۱ & ۱/۵۰ & ۳۶ \\ \vdots & \ddots & \vdots \\ ۱ & ۶ & ۴۸ \end{bmatrix}, \quad S = \begin{bmatrix} ۰ & ۰/۱۸۷۵ & ۳ \\ ۰ & ۰/۳۷۵۰ & ۳ \\ \vdots & \ddots & \vdots \\ ۰ & ۱/۵۰۰۰ & ۱۲ \end{bmatrix}, \quad Y = \begin{bmatrix} ۰/۵۵۷۵ \\ ۰/۹۲۶۶ \\ \vdots \\ ۲/۸۲۵۱ \end{bmatrix}, \quad R = \begin{bmatrix} ۰/۱۳۹۴ \\ ۰/۲۳۱۷ \\ \vdots \\ ۰/۷۰۶۳ \end{bmatrix}.$$

بنابراین بر اساس رابطه (۱) ضرایب به صورت زیر برآورد می‌شوند:

$$\hat{\beta} = \left[(X + \tfrac{S}{۸})'(X + \tfrac{S}{۸})\right]^{-۱} \left[(X + \tfrac{S}{۸})'(Y + \tfrac{R}{۸})\right] = \begin{bmatrix} ۰/۱۴۳۳ \\ ۰/۳۹۴۷ \\ ۰/۰۰۵۳ \end{bmatrix}.$$

پس، مدل بهینه برابر می‌شود با

$$\tilde{y} = ۰/۱۴۳۳ \oplus (۰/۳۹۴۷ \otimes \tilde{x}_۱) \oplus (۰/۰۰۵۳ \otimes \tilde{x}_۲).$$

اکنون فرض کنید مقدار غلظت رنگ مورد استفاده «حدوداً ۳/۱۵» گرم در لیتر و مدت زمان پروسه «حدوداً ۳۰» دقیقه باشد، که به ترتیب با $(۳/۱۵; ۰/۳۲, ۰/۱۶, ۰/۶۳, ۱/۵۶)_T$ و $(۳۰/۰۰; ۳/۰۰, ۱/۵۰, ۶/۰۰, ۱۵/۰۰)_T$ مشخص شده باشند، آنگاه مقدار پیش‌بینی جذب رنگ توسط پارچه به صورت زیر به‌دست می‌آید:

$$\begin{aligned} \tilde{y} &= ۰/۱۴۳۳ \oplus (۰/۳۹۴۷ \otimes (۳/۱۵; ۰/۳۲, ۰/۱۶, ۰/۶۳, ۱/۵۶)_T) \\ &\quad \oplus (۰/۰۰۵۳ \otimes (۳۰/۰۰; ۳/۰۰, ۱/۵۰, ۶/۰۰, ۱۵/۰۰)_T) \\ &= (۱/۵۵; ۰/۱۴۲۲, ۰/۰۷۱۱, ۰/۲۸۰۵, ۰/۶۹۵۲)_T. \end{aligned}$$

بنابراین، اگر غلظت رنگ «حدوداً ۳/۱۵» و مدت زمان پروسه «حدوداً ۳۰» دقیقه

باشد، در این مسأله، با امکان ۱ میزان جذب رنگ برابر **۱/۵۵** و با امکانی بین ۰/۴۶ تا **۰/۹۳** میزان جذب رنگ برابر **۱/۴۲** می‌باشد.
برای بررسی نیکویی برازش مدل، مقدار اندازه مشابهت $S(\tilde{y}_i, \hat{\tilde{y}}_i)$ را برای همه داده‌ها

جدول ۱: داده های مربوط به فرآیند رنگرزی

i	$(x_{i1}; s_{i11}, s_{i12}, s_{i13}, s_{i14})_T$	$(x_{i2}; s_{i21}, s_{i22}, s_{i23}, s_{i24})_T$	$(y_i; r_{i1}, r_{i2}, r_{i3}, r_{i4})_T$
۱	(۰/۷۵, ۰/۰۸, ۰/۰۳۸, ۰/۱۵, ۰/۳۸)$_T$	(۱۲/۰۰, ۱/۲۰, ۰/۶۰, ۲/۴۰, ۶/۰۰)$_T$	(۰/۵۵۷۵, ۰/۰۶, ۰/۰۳, ۰/۱۱, ۰/۲۸)$_T$
۲	(۱/۵۰, ۰/۱۵, ۰/۰۸, ۰/۳۰, ۰/۷۵)$_T$	(۱۲/۰۰, ۱/۲۰, ۰/۶۰, ۲/۴۰, ۶/۰۰)$_T$	(۰/۹۲۶۶, ۰/۱۰, ۰/۰۵, ۰/۱۹, ۰/۴۵)$_T$
۳	(۳/۰۰, ۰/۳۰, ۰/۱۵, ۰/۶۰, ۱/۵۰)$_T$	(۱۲/۰۰, ۱/۲۰, ۰/۶۰, ۲/۴۰, ۶/۰۰)$_T$	(۱/۵۶۷۳, ۰/۱۶, ۰/۰۸, ۰/۳۱, ۰/۷۸)$_T$
۴	(۴/۵۰, ۰/۴۵, ۰/۲۳, ۰/۹۰, ۲/۲۵)$_T$	(۱۲/۰۰, ۱/۲۰, ۰/۶۰, ۲/۴۰, ۶/۰۰)$_T$	(۲/۰۱۶۳, ۰/۲۰, ۰/۱۰, ۰/۴۰, ۱/۰۱)$_T$
۵	(۶/۰۰, ۰/۶۰, ۰/۳۰, ۱/۲۰, ۳/۰۰)$_T$	(۱۲/۰۰, ۱/۲۰, ۰/۶۰, ۲/۴۰, ۶/۰۰)$_T$	(۲/۴۲۹۳, ۰/۲۰, ۰/۱۲, ۰/۴۹, ۱/۲۱)$_T$
۶	(۰/۷۵, ۰/۰۸, ۰/۰۴, ۰/۱۵, ۰/۳۸)$_T$	(۲۴/۰۰, ۲/۴۰, ۱/۲۰, ۴/۸۰, ۱۲/۰۰)$_T$	(۰/۶۳۶۵, ۰/۲۴, ۰/۰۳, ۰/۱۳, ۰/۳۲)$_T$
۷	(۱/۵۰, ۰/۱۵, ۰/۰۸, ۰/۳۰, ۰/۷۵)$_T$	(۲۴/۰۰, ۲/۴۰, ۱/۲۰, ۴/۸۰, ۱۲/۰۰)$_T$	(۱/۰۳۷۸, ۰/۰۶, ۰/۰۵, ۰/۲۱, ۰/۵۲)$_T$
۸	(۳/۰۰, ۰/۳۰, ۰/۱۵, ۰/۶۰, ۱/۵۰)$_T$	(۲۴/۰۰, ۲/۴۰, ۱/۲۰, ۴/۸۰, ۱۲/۰۰)$_T$	(۰/۵۷۰۴, ۰/۱۰, ۰/۰۳, ۰/۱۱, ۰/۲۹)$_T$
۹	(۴/۵۰, ۰/۴۵, ۰/۲۳, ۰/۹۰, ۲/۲۵)$_T$	(۲۴/۰۰, ۲/۴۰, ۱/۲۰, ۴/۸۰, ۱۲/۰۰)$_T$	(۲/۰۸۶۴, ۰/۰۶, ۰/۱۰, ۰/۴۲, ۱/۰۴)$_T$
۱۰	(۶/۰۰, ۰/۶۰, ۰/۳۰, ۱/۲۰, ۳/۰۰)$_T$	(۲۴/۰۰, ۲/۴۰, ۱/۲۰, ۴/۸۰, ۱۲/۰۰)$_T$	(۲/۶۵۰۲, ۰/۲۱, ۰/۱۳, ۰/۵۳, ۱/۳۳)$_T$
۱۱	(۰/۷۵, ۰/۰۸, ۰/۰۴, ۰/۱۵, ۰/۳۸)$_T$	(۳۶/۰۰, ۳/۶۰, ۱/۸۰, ۷/۲۰, ۱۸/۰۰)$_T$	(۰/۵۴۵۳, ۰/۰۵, ۰/۰۳, ۰/۱۱, ۰/۲۷)$_T$
۱۲	(۱/۵۰, ۰/۱۵, ۰/۰۸, ۰/۳۰, ۰/۷۵)$_T$	(۳۶/۰۰, ۳/۶۰, ۱/۸۰, ۷/۲۰, ۱۸/۰۰)$_T$	(۰/۹۸۱۸, ۰/۱۰, ۰/۰۵, ۰/۲۰, ۰/۴۹)$_T$
۱۳	(۳/۰۰, ۰/۳۰, ۰/۱۵, ۰/۶۰, ۱/۵۰)$_T$	(۳۶/۰۰, ۳/۶۰, ۱/۸۰, ۷/۲۰, ۱۸/۰۰)$_T$	(۱/۶۰۳۲, ۰/۱۶, ۰/۰۸, ۰/۳۲, ۰/۸۰)$_T$
۱۴	(۴/۵۰, ۰/۴۵, ۰/۲۳, ۰/۹۰, ۲/۲۵)$_T$	(۳۶/۰۰, ۳/۶۰, ۱/۸۰, ۷/۲۰, ۱۸/۰۰)$_T$	(۲/۱۸۲۷, ۰/۲۲, ۰/۱۱, ۰/۴۴, ۱/۰۹)$_T$
۱۵	(۶/۰۰, ۰/۶۰, ۰/۳۰, ۱/۲۰, ۳/۰۰)$_T$	(۳۶/۰۰, ۳/۶۰, ۱/۸۰, ۷/۲۰, ۱۸/۰۰)$_T$	(۲/۸۶۸۶, ۰/۲۹, ۰/۱۴, ۰/۵۷, ۱/۴۳)$_T$
۱۶	(۰/۷۵, ۰/۰۸, ۰/۰۴, ۰/۱۵, ۰/۳۸)$_T$	(۴۸/۰۰, ۴/۸۰, ۲/۴۰, ۹/۶۰, ۲۴/۰۰)$_T$	(۰/۵۸۸۸, ۰/۰۶, ۰/۰۳, ۰/۱۲, ۰/۲۹)$_T$
۱۷	(۱/۵۰, ۰/۱۵, ۰/۰۸, ۰/۳۰, ۰/۷۵)$_T$	(۴۸/۰۰, ۴/۸۰, ۲/۴۰, ۹/۶۰, ۲۴/۰۰)$_T$	(۱/۰۰۸۴, ۰/۱۰, ۰/۰۵, ۰/۲۰, ۰/۵۰)$_T$
۱۸	(۳/۰۰, ۰/۳۰, ۰/۱۵, ۰/۶۰, ۱/۵۰)$_T$	(۴۸/۰۰, ۴/۸۰, ۲/۴۰, ۹/۶۰, ۲۴/۰۰)$_T$	(۱/۶۱۶۶, ۰/۱۶, ۰/۰۸, ۰/۳۲, ۰/۸۰)$_T$
۱۹	(۴/۵۰, ۰/۴۵, ۰/۲۳, ۰/۹۰, ۲/۲۵)$_T$	(۴۸/۰۰, ۴/۸۰, ۲/۴۰, ۹/۶۰, ۲۴/۰۰)$_T$	(۲/۱۱۵۷, ۰/۲۱, ۰/۱۱, ۰/۴۲, ۱/۰۶)$_T$
۲۰	(۶/۰۰, ۰/۶۰, ۰/۳۰, ۱/۲۰, ۳/۰۰)$_T$	(۴۸/۰۰, ۴/۸۰, ۲/۴۰, ۹/۶۰, ۲۴/۰۰)$_T$	(۲/۸۲۵۱, ۰/۲۸, ۰/۱۴, ۰/۵۷, ۱/۴۱)$_T$

محاسبه می‌کنیم. بر اساس اطلاعات جدول ۲، میزان مشابهت هر مقدار متغیر پاسخ و مقدار برآوردشده آن زیاد است پس می‌توان نتیجه گرفت که مدل برازش شده، یک مدل مناسب است که البته در بین مشاهدات داده شماره ۸ کمترین میزان مشابهت را دارد و به نظر می‌رسد برازش خوبی نداشته باشد انتخاب این داده به عنوان داده پرت به نظر کاربر بستگی دارد. برای بررسی بیشتر این داده را از بین مشاهدات حذف می‌کنیم و مدل را دوباره برازش می‌دهیم و معیار مشابهت را برای آنها محاسبه می‌کنیم. از مقایسه مقادیر میانگین اندازه مشابهت (MS) برای مدل اولیه و مدل جدید، در می‌یابیم که میانگین اندازه مشابهت از ۰/۹۴۵۱ در مدل اولیه به مقدار **۰/۹۴۹۴** در مدل جدید افزایش یافته است بنابراین در مقایسه دو مدل اولیه و جدید، مدل جدید را که MS بیشتری دارد انتخاب می‌کنیم.

بحث و نتیجه‌گیری

در این مقاله، رویکرد جدیدی برای برآورد ضرایب مدل رگرسیونی خطی چندگانه با مشاهدات فازی فاصله‌ای مقدار به روش کمترین مربعات و با استفاده از تعمیم فاصله علامت‌دار یائو–ویو ارائه گردید. همچنین بر اساس یک معیار مشابهت جدید به بررسی آزمون نیکویی برازش و یافتن داده‌های پرت پرداخته شد.

جدول ۲: داده‌های مربوط به فرآیند رنگرزی و اندازه‌های مشابهت مربوطه

i	$(y_i; r_{i۱}, r_{i۲}, r_{i۳}, r_{i۴})_T$	$(\hat{y}_i; s_{i۱}, s_{i۲}, s_{i۳}, s_{i۴})_T$	$S(\tilde{y}_i, \hat{\tilde{y}}_i)$	$S'(\tilde{y}_i, \hat{\tilde{y}}_i)$
۱	(۰/۵۵۷۵, ۰/۰۶, ۰/۰۳, ۰/۱۱, ۰/۲۸)$_T$	(۰/۵۰۲۶, ۰/۰۴, ۰/۰۲, ۰/۰۷, ۰/۱۸)$_T$	۰/۹۵۷۳	۰/۹۵۹۱
۲	(۰/۹۲۶۶, ۰/۱۰, ۰/۰۵, ۰/۱۹, ۰/۴۵)$_T$	(۰/۷۹۸۶, ۰/۰۷, ۰/۰۳, ۰/۱۳, ۰/۳۳)$_T$	۰/۹۳۳۱	۰/۹۴۴۱
۳	(۱/۵۶۷۳, ۰/۱۶, ۰/۰۸, ۰/۳۱, ۰/۷۸)$_T$	(۱/۳۹۰۷, ۰/۱۲, ۰/۰۶, ۰/۲۵, ۰/۶۲)$_T$	۰/۹۱۷۷	۰/۹۲۷۵
۴	(۲/۰۱۶۳, ۰/۲۰, ۰/۱۰, ۰/۴۰, ۱/۰۱)$_T$	(۱/۹۸۲۷, ۰/۱۸, ۰/۰۹, ۰/۳۷, ۰/۹۲)$_T$	۰/۹۶۴۵	۰/۹۵۸۷
۵	(۲/۴۲۹۳, ۰/۲۰, ۰/۱۲, ۰/۴۹, ۱/۲۱)$_T$	(۲/۵۷۴۸, ۰/۲۴, ۰/۱۲, ۰/۴۹, ۱/۲۲)$_T$	۰/۹۷۱۳	۰/۹۵۳۹
۶	(۰/۶۳۶۵, ۰/۲۴, ۰/۰۳, ۰/۱۳, ۰/۳۲)$_T$	(۰/۵۶۶۰, ۰/۰۴, ۰/۰۲, ۰/۰۸, ۰/۲۱)$_T$	۰/۹۵۲۰	۰/۹۵۷۶
۷	(۱/۰۳۷۸, ۰/۰۶, ۰/۰۵, ۰/۲۱, ۰/۵۲)$_T$	(۰/۸۶۲۱, ۰/۰۷, ۰/۰۴, ۰/۱۴, ۰/۳۶)$_T$	۰/۹۱۸۰	۰/۹۲۲۸
۸	(۰/۵۷۰۴, ۰/۱۰, ۰/۰۳, ۰/۱۱, ۰/۲۹)$_T$	(۱/۴۵۴۱, ۰/۱۳, ۰/۰۷, ۰/۲۶, ۰/۶۶)$_T$	۰/۷۶۷۷	----
۹	(۲/۰۸۶۴, ۰/۰۶, ۰/۱۰, ۰/۴۲, ۱/۰۴)$_T$	(۲/۰۴۶۲, ۰/۱۹, ۰/۱۰, ۰/۳۸, ۰/۹۵)$_T$	۰/۹۶۲۳	۰/۹۵۹۴
۱۰	(۲/۶۵۰۲, ۰/۲۱, ۰/۱۳, ۰/۵۳, ۱/۳۳)$_T$	(۲/۶۳۸۲, ۰/۲۵, ۰/۱۲, ۰/۵۰, ۱/۲۵)$_T$	۰/۹۷۲۰	۰/۹۵۸۷
۱۱	(۰/۵۴۵۳, ۰/۰۵, ۰/۰۳, ۰/۱۱, ۰/۲۷)$_T$	(۰/۶۲۹۵, ۰/۰۵, ۰/۰۲, ۰/۱۰, ۰/۲۴)$_T$	۰/۹۷۳۸	۰/۹۵۶۳
۱۲	(۰/۹۸۱۸, ۰/۱۰, ۰/۰۵, ۰/۲۰, ۰/۴۹)$_T$	(۰/۹۲۵۵, ۰/۰۸, ۰/۰۴, ۰/۱۶, ۰/۳۹)$_T$	۰/۹۵۶۸	۰/۹۵۵۶
۱۳	(۱/۶۰۳۲, ۰/۱۶, ۰/۰۸, ۰/۳۲, ۰/۸۰)$_T$	(۱/۵۱۷۶, ۰/۱۴, ۰/۰۷, ۰/۲۷, ۰/۶۸)$_T$	۰/۹۴۷۰	۰/۹۴۴۹
۱۴	(۲/۱۸۲۷, ۰/۲۲, ۰/۱۱, ۰/۴۴, ۱/۰۹)$_T$	(۲/۱۰۹۶, ۰/۲۰, ۰/۱۰, ۰/۳۹, ۰/۹۸)$_T$	۰/۹۵۰۰	۰/۹۴۸۱
۱۵	(۲/۸۶۸۶, ۰/۲۹, ۰/۱۴, ۰/۵۷, ۱/۴۳)$_T$	(۲/۷۰۱۷, ۰/۲۶, ۰/۱۳, ۰/۵۱, ۱/۲۸)$_T$	۰/۹۲۰۷	۰/۹۱۷۰
۱۶	(۰/۵۸۸۸, ۰/۰۶, ۰/۰۳, ۰/۱۲, ۰/۲۹)$_T$	(۰/۶۹۳۰, ۰/۰۵, ۰/۰۳, ۰/۱۱, ۰/۲۷)$_T$	۰/۹۷۳۳	۰/۹۵۶۳
۱۷	(۱/۰۰۸۴, ۰/۱۰, ۰/۰۵, ۰/۲۰, ۰/۵۰)$_T$	(۰/۹۸۹۰, ۰/۰۸, ۰/۰۴, ۰/۱۷, ۰/۴۲)$_T$	۰/۹۶۹۵	۰/۹۵۹۷
۱۸	(۱/۶۱۶۶, ۰/۱۶, ۰/۰۸, ۰/۳۲, ۰/۸۰)$_T$	(۱/۵۸۱۰, ۰/۱۴, ۰/۰۷, ۰/۲۹, ۰/۷۲)$_T$	۰/۹۶۳۹	۰/۹۵۵۳
۱۹	(۲/۱۱۵۷, ۰/۲۱, ۰/۱۱, ۰/۴۲, ۱/۰۶)$_T$	(۲/۱۷۳۱, ۰/۲۰, ۰/۱۰, ۰/۴۱, ۱/۰۲)$_T$	۰/۹۷۴۶	۰/۹۵۷۸
۲۰	(۲/۸۲۵۱, ۰/۲۸, ۰/۱۴, ۰/۵۷, ۱/۴۱)$_T$	(۲/۷۶۵۲, ۰/۲۶, ۰/۱۳, ۰/۵۲, ۱/۳۱)$_T$	۰/۹۵۵۶	۰/۹۴۵۳
میانگین			۰/۹۴۵۱	۰/۹۴۹۴

مراجع

ترکیان، ف.، اکبری، م. ق.، عارفی، م.، (۱۳۹۰)، برآورد ضرایب رگرسیونی بر اساس داده‌های فازی شهودی و با استفاده از فاصله علامت‌دار یائو-ویو، گزارش یازدهمین کنفرانس سیستم‌های فازی ایران، دانشگاه سیستان و بلوچستان، ۲۶-۱۹.

عارفی، م. طاهری، س. م.، (۱۳۹۰)، رگرسیون کمترین مربعات با ورودی و خروجی مبهم و ضرایب مبهم، گزارش یازدهمین کنفرانس سیستم‌های فازی ایران، دانشگاه سیستان و بلوچستان، ۴۵-۳۸.

Arabpour, A. R. and Tata, M. (2008), Estimating the Parameters of a Fuzzy Linear Regression Model. *Iranian Journal of Fuzzy Systems*, **5(2)**, 1-19.

Atanassov, K. (1999), Intuitionistic Fuzzy Sets: Theory and Applications, *Physics-Verlag, Heidelberg, Germany.*

Celmins, A. (1987), Least Squares Model Fitting to Fuzzy Vector Data. *Fuzzy Sets and Systems*, **22**, 260-269.

Guha, D. and Chakraborty, D. (2010), A theoretical Development of Distance Measure of Intuitionistic Fuzzy Numbers. *International Journal of Mathematics and Mathematical Sciences*, (to appear).

Hasanpour, H. Maleki, H.R. and Yaghoubi, M.A. (2010), Fuzzy Linear Regression Model with Crisp Coefficients: A Goal Programing Approach. *Iranian Journal of Fuzzy Systems*, **7(2)**, 19-39.

Näther, W. and Wünsche, A. (2002), Least Squares Fuzzy Regression with Fuzzy Random Variables. *Fuzzy Set and Systems*, **130**, 43-50.

Taheri, S. M. and Zarei, R. (2011), Extension principle of vague sets and its applications, *Advances in Fuzzy Mathematics*, **6(1)**, 17-26.

Tanaka, H., Vegima, S. and Asai, K. (1982), Linear Regression Analysis with Fuzzy Model, *IEEE Trans, Systems Man Cybernet*, **12**, 903-907.

Tavanai, H., Taheri, S. M. and Nasiri, M. (2005), Modelling of Colour Yield in Polyethylene Terephthalate Dyeing with Statistical and Fuzzy Regression, *Iranian Polymer Journal*, **14**, 954-967.