

رگرسیون خطی فاصله‌ای

علیرضا عرب پور^۱ و زینب محمدی^۲

تاریخ دریافت: ۱۳۹۴/۶/۲۵

تاریخ پذیرش: ۱۳۹۵/۸/۳۰

چکیده:

رگرسیون خطی فاصله‌ای، تعمیمی از رگرسیون معمولی است که از آن برای محاسبه رابطه بین متغیرهای مستقل و متغیر وابسته در یک محیط فازی استفاده می‌شود. هنگامی که پارامترهای مدل رگرسیونی خطی به جای توابع احتمالی با توابع عضویت فازی بیان شوند، این مدل، مدل رگرسیونی خطی فاصله‌ای نامیده می‌شود. در این مقاله نخست روش‌های تحلیل رگرسیون خطی فاصله‌ای را مطرح می‌کنیم و سپس برای بهبود این روش‌ها، روشی پیشنهاد می‌کنیم که ابهام مدل را کاهش می‌دهد. در نهایت با چند مثال عددی، کارایی روش پیشنهادی را بررسی می‌کنیم. محاسبات انجام شده در مثال‌ها با استفاده از بسته آلاباما^۳ به وسیله نرم‌افزار R انجام شده است.

واژه‌های کلیدی: رگرسیون خطی فاصله‌ای، فاصله‌های فازی، شمول کل، تشخیص مدل، برنامه‌ریزی خطی.

۱ مقدمه

سیستم یادگیری گفتار (لیو [۱۴]) است. رگرسیون امکانی، از طریق مدل‌های مختلفی حل می‌شود که مدل‌های اولیه آن فرمول برنامه‌ریزی خطی را به کار می‌برند (لی و تاناکا [۱۳])، تاناکا [۱۷]، تاناکا و واتادا [۲۱]) که تا حدودی ساده است و محاسبات کامپیوتری کمی نیاز دارد؛ اما اکثر برآوردهای پارامترهای مدل‌های رگرسیونی، دقیق هستند. برای برطرف کردن این مشکل، روش‌های دیگری از جمله برنامه‌ریزی درجه دوم (تاناکا و لی [۱۹]) یا ماشین‌های بردار پشتیبان (هائو [۷])، هانگ و کائو [۱۱]) مطرح شدند. این نتایج هنوز رضایت‌بخش نیستند و روش‌ها از نظر محاسباتی پیچیده هستند. طاهری و کلکین نما [۱۶] بر اساس روش قدر مطلق خطا و تعریف یک متر جدید، مدلی در رگرسیون خطی فازی ارائه کردند که در مقایسه با روش‌های قبلی، برازش خوبی روی داده‌های فازی متقارن و نامتقارن انجام می‌دهد.

ساختار مقاله بدین صورت است که در بخش ۱ به‌طور مختصر بعضی از مفاهیم به کار رفته در این مقاله را ذکر می‌کنیم. در ادامه تعدادی از روش‌های تحلیل رگرسیون خطی فاصله‌ای را

مبحث رگرسیون فاصله‌ای از اوایل دهه ۱۹۸۰ میلادی و پس از معرفی نظریه مجموعه‌های فازی توسط لطفی‌زاده [۲۳] مورد توجه قرار گرفت. نخستین بار تاناکا و همکاران [۲۰] در سال ۱۹۸۲ میلادی، رگرسیون فاصله‌ای را مورد بررسی قرار دادند. تاناکا فرض کرد که داده‌ها اعداد فازی مثلثی هستند و با کمینه کردن یک شاخص فازی، ضرایب رگرسیونی را برآورد کرد. اساس کار وی روش‌های برنامه‌ریزی ریاضی بود. برای مدل‌های رگرسیونی فاصله‌ای مفاهیم مختلفی وجود دارد. مفهوم امکانی^۴ را تاناکا و همکاران در سال ۱۹۸۲ میلادی مطرح کردند و بعداً به وسیله تاناکا [۱۷] در سال ۱۹۸۷ میلادی به کار گرفته شد. در سال ۱۹۹۸ میلادی تاناکا و لی [۱۹] بهترین تحقیق را ارائه کردند، به طوری که قابل اجرا در تعدادی از مسائل علمی از جمله کاربرد در کار پژوهشی (چانگ و همکاران [۶])، پیش‌بینی میزان فروش بازار (حشمتی و کندل [۸])، سیستم تشخیص هویت (کانه‌یوشی و همکاران [۱۲])،

^۱ استادیار بخش آمار، دانشگاه شهید باهنر کرمان

^۲ دانش‌آموخته کارشناسی ارشد آمار ریاضی، دانشگاه شهید باهنر کرمان

^۳ Alabama

^۴ possibilistic

فرض کنید $A \in F(X)$ باشد. تکیه گاه A که با $S(A)$ نشان داده می‌شود، به صورت زیر تعریف می‌شود [۱]:

$$S(A) = \{x \in X \mid A(x) > 0\}.$$

برای هر $0 \leq \alpha \leq 1$ ، α -برش A که با $(A)_\alpha$ نشان داده می‌شود، به صورت زیر تعریف می‌شود [۱]:

$$(A)_\alpha = \{x \in X \mid A(x) \geq \alpha\}.$$

فاصله فازی A به کمک از مجموعه عناصری که بین حد بالا و پایین قرار می‌گیرند [۲]:

$$A = [A^-, A^+] = \{x \mid A^- \leq x \leq A^+, x \in \mathbb{R}\}, \quad (1)$$

حال اگر $A^- < A^+$ باشد، فاصله A پهن^۵ است و اگر $A^- = A^+$ باشد، فاصله باریک^۶ (نقطه‌ای) است.

برای این که عدم قطعیت را از طریق فاصله‌ها پردازش کنیم و نشان دهیم به نمایش مؤثری از فاصله‌ها نیاز داریم به طوری که بتواند فاصله‌های دلخواه و مجموعه‌های‌شان را در یک نمودار نشان دهد.

برای ارائه یک تعریف منحصر به فرد از فاصله، به دو پارامتر نیاز داریم. در نتیجه فضای فاصله‌ها دوبعدی است. نمایش‌های زیادی از فاصله وجود دارد که پرکاربردترین آن‌ها نمایش ارائه شده در بالا است. اما مشکل اصلی این نمایش، مربوط به عدم قطعیت (شعاع) است که به صورت مناسب و مستقیم در فاصله نشان داده نمی‌شود. برای حل این مشکل، نمایش دیگری اتخاذ شده است. برای آسان کردن نمایش عدم قطعیت، به مختصات مرکز (M) و شعاع (R) نیاز است (شکل ۲). در این حالت، فاصله به صورت (M, R) نمایش داده می‌شود.

شده است. در نمایش (M, R) واضح است که نادقیق‌ترین عنصر $C(A)$ خود فاصله A است که متناظر با بیشترین مقدار شعاع، یعنی رأس مثلث است. برعکس قاعده مثلث، همه مقادیر دقیق بین کران بالا و پایین را نشان می‌دهد (مجموعه همه فاصله‌های نقطه‌ای). فاصله معمولی^۷ a به کمک از مجموعه

در بخش ۳ ارائه می‌کنیم. در بخش ۴ به معرفی روش پیشنهادی خواهیم پرداختیم. در این بخش، روشی ارائه خواهیم کرد که باعث بهبود روش‌های قبلی است. در این مقاله مجموع انحرافات مثبت و منفی را به ضابطه تابع هدف اضافه می‌کنیم. همچنین در قیده‌های تابع هدف، برای این که علامت \leq را به $=$ تغییر دهیم، انحرافات مثبت را اضافه و انحرافات منفی را از قیدها کم می‌کنیم. از آنجا که انحرافات مثبت و منفی بزرگ‌تر یا مساوی صفر هستند، برای برآورد ضرایب رگرسیونی خطی فاصله‌ای باید تابع هدف را طوری کمینه کنیم یا به عبارتی پارامترهای مدل را طوری برآورد کنیم، که مقدار برآورد شده منطبق بر مقدار مشاهده شده باشد. در پایان نیز برای مقایسه کارایی روش پیشنهادی با سایر روش‌های تحلیل رگرسیون خطی فاصله‌ای چند مثال عددی در بخش ۵ ارائه خواهیم کرد.

۲ تعریف‌های اولیه

فرض کنید X مجموعه‌ای ناتهی باشد. هر زیرمجموعه فازی A از X توسط یک تابع عضویت $\mu_A : X \rightarrow [0, 1]$ مشخص می‌شود که در آن به ازای هر $x \in X$ مقدار $\mu_A(x)$ در بازه $[0, 1]$ است و میزان عضویت X را در A نشان می‌دهد [۱].

برای سادگی، زیرمجموعه فازی A از X را با تابع عضویت $A : X \rightarrow [0, 1]$ و درجه عضویت x در A را با $A(x)$ نشان می‌دهیم. همچنین مجموعه همه زیرمجموعه‌های فازی X را با $F(X)$ نشان می‌دهیم. به عبارتی،

$$F(X) = \{A \mid A : X \rightarrow [0, 1]\}.$$

مرکز و شعاع فاصله A به صورت زیر تعریف می‌شود:

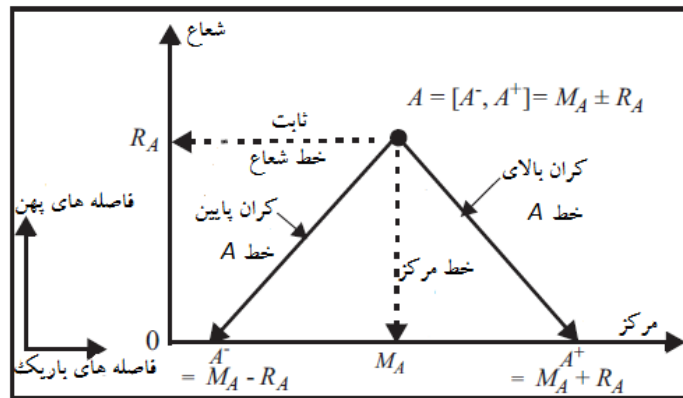
$$M_A = \frac{A^- + A^+}{2}, \quad R_A = \frac{A^+ - A^-}{2}. \quad (2)$$

همه فاصله‌های شمول در فاصله A را با $C(A)$ نشان می‌دهیم. همان‌طور که در شکل ۲ ملاحظه می‌کنید، دامنه تعریف $C(A)$ از طریق خط‌های کران بالا و پایین به صورت مثلث نشان داده

^۵ thick

^۶ thin

^۷ conventional interval



شکل ۱. نمایش مرکز-شعاع A و C(A)

برای کامل کردن تعریف فاصله فازی، دو نوع اطلاعات مورد نیاز است. این اطلاعات، تکیه گاه و مرکز فاصله ها هستند که روی محور افقی تعریف می شوند، اما محور عمودی مربوط به دو سطح متفاوت ۰ و ۱ است (شکل ۲). برای تکمیل تعریف فاصله فازی، دو تابع برای ارتباط تکیه گاه^{۱۱} (S_A) و مرکز^{۱۲} (K_A) به صورت رابطه زیر به کار می روند:

$$(A^-)_\alpha = \inf \{x | \mu_A(x) \geq \alpha; \quad x \geq S_{A^-}\},$$

$$(A^+)_\alpha = \sup \{x | \mu_A(x) \geq \alpha; \quad x \leq S_{A^+}\}. \quad (۵)$$

که $\alpha \in [0, 1]$ است.

هنگامی که روی فاصله فازی A، α -برش ایجاد می کنیم، فاصله معمولی را به دست می آوریم:

$$(A)_\alpha = [(A^-)_\alpha, (A^+)_\alpha]. \quad (۶)$$

زمانی که نمودار توابع، خطی فرض شود، فاصله فازی A از نوع ذوزنقه ای است و از طریق تکیه گاه و مرکز فاصله ها تعریف می شود. تکیه گاه به صورت $S_A = [S_{A^-}, S_{A^+}]$ و مرکز به صورت $K_A = [K_{A^-}, K_{A^+}]$ است (شکل ۲ سمت راست). فاصله معمولی a به صورت $[a^-, a^+]$ نمایش داده می شود و فاصله فازی ذوزنقه ای^{۱۳} A از طریق کران های مرکز و تکیه گاه به صورت زیر تعریف می شود:

عناصری که بین حدود بالا (a^+) و پایین (a^-) به صورت زیر قرار می گیرند:

$$a = \{x | a^- \leq x \leq a^+, x \in \mathbb{R}\}, \quad (۳)$$

به طوری که $a^-, a^+ \in \mathbb{R}$ هستند. باید توجه داشت که فاصله های معمولی را با حروف کوچک مانند a و فاصله های فازی را با حروف بزرگ مانند A نمایش می دهیم.

برای فاصله a، مرکز^{۱۴} آن M(a) و شعاع^{۱۵} آن R(a)، به صورت زیر تعریف می شوند [۴]:

$$M(a) = \frac{(a^- + a^+)}{2}, \quad R(a) = \frac{(a^+ - a^-)}{2}. \quad (۴)$$

فاصله a به عنوان مورد خاص می تواند عدد فازی^{۱۶} در نظر گرفته شود که تابع عضویت آن، مقدار ثابت ۱ را بین a^- و a^+ می گیرد و صفر را هر جای دیگری می گیرد [۴] (شکل ۲، سمت چپ). فاصله فازی A از طریق تابع عضویت خود، یعنی μ_A نمایش داده می شود. برای نشان دادن شکل فاصله فازی باید دو محور در نظر گرفته شود: اولین محور، افقی است که شبیه نمایش فاصله ای، یعنی خط حقیقی \mathbb{R} است. دومین محور، عمودی است که مربوط به تغییرات درجه عضویت در فاصله $[0, 1]$ است (شکل ۲ سمت راست).

^{۱۴} midpoint

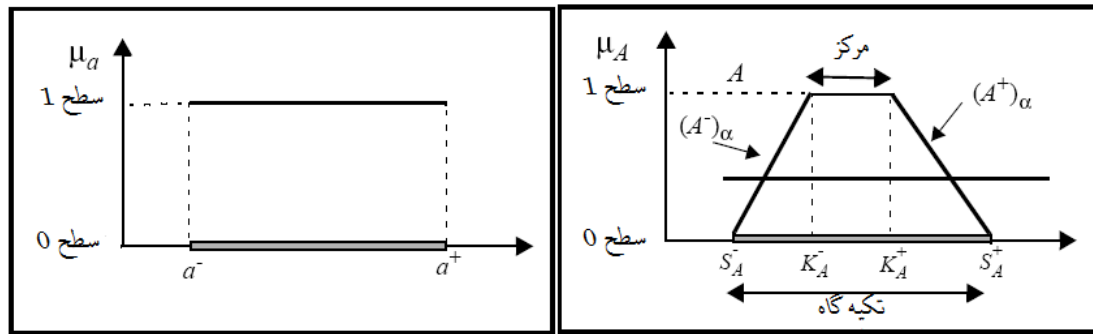
^{۱۵} radius

^{۱۶} fuzzy number

^{۱۱} support

^{۱۲} kernel

^{۱۳} trapezoidal fuzzy interval



شکل ۲. نمایش فاصله فازی و ذوزنقه‌ای

با توجه به تعاریف فوق، در ادامه مدل رگرسیونی خطی فاصله‌ای و انواع مدل‌ها را بیان می‌کنیم.

۳ روش‌های تحلیل رگرسیون خطی فاصله‌ای

در این بخش به‌طور مختصر به معرفی مدل رگرسیونی خطی ساده فاصله‌ای، و تعدادی از روش‌های تحلیل رگرسیون خطی فاصله‌ای از جمله روش تاناکا و همکاران، روش HC^{۱۴}، ضابطه حجم^{۱۵}، ضابطه مجموع^{۱۶} و روش بیسریه^{۱۷} می‌پردازیم.

۱.۳ مدل رگرسیونی خطی فاصله‌ای

مدل رگرسیونی خطی ساده فاصله‌ای به‌صورت زیر تعریف می‌شود [۳]:

$$Y = A. \oplus A_1 \otimes x \oplus E, \quad (13)$$

به‌طوری که Y خروجی (متغیر وابسته فازی)، x ورودی غیر فازی (دقیق) و E خطای فازی مربوط به مدل رگرسیونی است. A و A_1 نیز ضرایب فازی هستند که باید از روی داده‌های مشاهده‌شده برآورد شوند.

دامنه تعریف D متناظر با مدل رگرسیونی خطی فاصله‌ای است. این دامنه تعریف با توجه به داده‌های مشاهده‌شده مشخص می‌شود؛ یعنی:

$$A = (K_A, S_A) = ([K_{A-}, K_{A+}], [S_{A-}, S_{A+}]) \quad (V)$$

یک حالت خاص از فاصله‌های فازی ذوزنقه‌ای، مثالی متقارن است. در این مورد داریم:

$$\begin{aligned} A &= (K_A, S_A) = (K_A, [S_{A-}, S_{A+}]) \\ &= (K_A, [K_A - R(S_A), K_A + R(S_A)]) \quad (8) \end{aligned}$$

که در آن $K_A = M(S_A)$.

برای فاصله‌های معمولی a و b ، رابطه شمول a در b به‌صورت زیر تعریف می‌شود [۵]:

$$\begin{aligned} a \subseteq b &\Leftrightarrow \begin{cases} b^- \leq a^- \\ a^+ \leq b^+ \end{cases} \Leftrightarrow \begin{cases} M(b) - R(b) \leq M(a) - R(a) \\ M(a) + R(a) \leq M(b) + R(b) \end{cases} \\ &\Leftrightarrow \begin{cases} M(b) - M(a) \leq R(b) - R(a) \\ M(a) - M(b) \leq R(b) - R(a) \end{cases} \quad (9) \end{aligned}$$

از رابطه فوق نتیجه می‌شود:

$$a \subseteq b \Leftrightarrow |M(b) - M(a)| \leq R(b) - R(a). \quad (10)$$

فرض کنید $\alpha \in \mathbb{R}$ یک اسکالر باشد. آن‌گاه:

$$\alpha \in b \Leftrightarrow |M(b) - \alpha| \leq R(b). \quad (11)$$

برای دو فاصله فازی A و B ، رابطه شمول A در B به‌صورت زیر تعریف می‌شود [۵] (شکل):

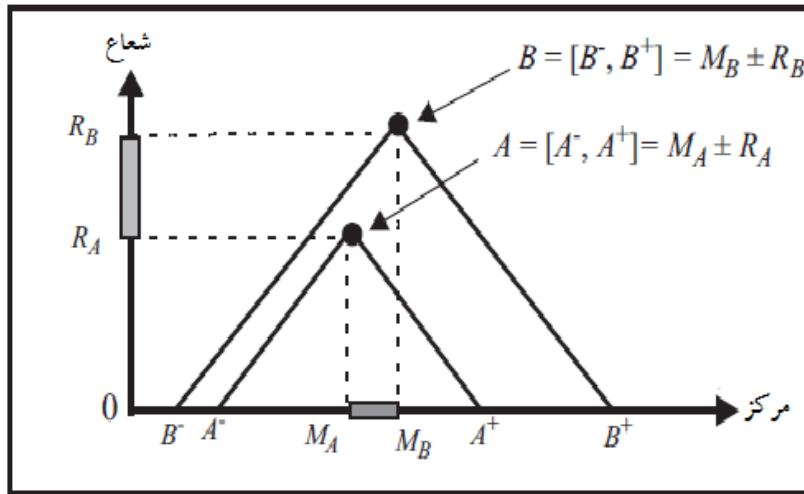
$$A \subseteq B \Leftrightarrow \forall x : \mu_A(x) \leq \mu_B(x). \quad (12)$$

^{۱۴} Hladik, Cerny

^{۱۵} volume

^{۱۶} sum

^{۱۷} bisserier



شکل ۳. شمول فاصله \$A\$ در \$B\$

دو نکته را باید در نظر بگیریم:

۱. تعریف تابع هدف، همراه با قیدهایی برای بهینه‌سازی مسئله،

۲. داشتن ضابطه تشخیص یا نیکویی برازش.

تاناکا و همکاران در سال‌های ۱۹۸۲ و ۱۹۸۹ مدل رگرسیونی خطی فاصله‌ای را مطرح کردند که در آن پارامترهای \$A_1\$ و \$A_1\$، فاصله‌های فازی مثلثی متقارن هستند و ورودی‌های مدل دارای مقادیری دقیق‌اند و خروجی مدل نیز فاصله فازی مثلثی متقارن است [۲۰].

وقتی زامین داده مشاهده شده را در نظر می‌گیرید، خروجی‌اش، فاصله فازی مثلثی متقارن است:

$$Y_j = (K_{Y_j}, S_{Y_j}).$$

خروجی پیش‌بینی متناظر با \$Y_j\$، یعنی \$\hat{Y}_j\$ نیز فاصله فازی مثلثی متقارن به صورت زیر است:

$$\begin{aligned} \hat{Y}_j &= (K_{\hat{Y}_j}, S_{\hat{Y}_j}) = (K_{\hat{Y}_j}, [S_{\hat{Y}_j}^-, S_{\hat{Y}_j}^+]) \\ &= (K_{\hat{Y}_j}, [K_{\hat{Y}_j} - R(\hat{Y}_j(x)), K_{\hat{Y}_j} + R(\hat{Y}_j(x))]) \end{aligned} \quad (15)$$

۱.۱.۳ گرایش مدل

مدل \$\hat{Y}(x) = A_1 \oplus A_1 \otimes x\$ را در نظر بگیرید. در این مدل تغییرات خروجی همانند مقدار ورودی است. بنا بر این غیرممکن

$$D^- = \min_{j=1, \dots, n} x_j, \quad D^+ = \max_{j=1, \dots, n} x_j$$

بنا بر این \$D\$ به کمک از فاصله کراندار حقیقی که ورودی \$x\$ به آن تعلق دارد:

$$D = [D^-, D^+] \subset \mathbb{R}.$$

یک نمونه \$n\$ تایی از مجموعه داده‌های مشاهده شده که روی فاصله \$D\$ تعریف می‌شوند، در نظر بگیرید. زامین نمونه به صورت زوج \$(x_j, y_j)\$، \$j = 1, \dots, n\$ نمایش داده می‌شود. \$x_j\$ها ورودی‌های دقیق و \$y_j\$ها خروجی فازی متناظر هستند. همانند هر روش رگرسیونی، هدف رگرسیون فاصله‌ای این است که باید رابطه تابعی پیش‌بینی \$\hat{Y}\$ بین ورودی‌های \$x\$ و خروجی‌های \$Y\$ را مشخص کند:

$$\hat{Y} = A_1 \oplus A_1 \otimes x, \quad (14)$$

به طوری که \$\hat{Y}\$ خروجی پیش‌بینی شده است.

یکی از اهداف مسئله رگرسیون فاصله‌ای، برآورد پارامترهای فاصله‌ای \$A_1\$ و \$A_1\$ است. برای این منظور، روش رگرسیون فاصله‌ای امکانی، اتخاذ شده است که هدف آن برآورد پارامترهای فازی است به طوری که داده‌های مشاهده شده در پیش‌بینی شده، قرار گیرند (شکل).

در خصوص مسئله رگرسیون خطی فاصله‌ای، روش‌های زیادی معرفی شده است. به طور کلی بدون توجه به روش انتخابی،

آسان‌تر کردن اجرای مدل فازی دارد، نمی‌تواند شمول کل را حفظ کند. در واقع زمانی که تکیه‌گاه‌ها شمول هستند، شمول کل فاصله‌های فازی برقرار است اگر و تنها اگر مقادیر مرکزها (مشاهده‌شده و برآوردشده) منطبق باشند، که عملاً رخ دادن چنین حالتی امکان‌پذیر نیست.

۲.۳ روش HC

هدف تحلیل رگرسیون فاصله‌ای، یافتن پارامترهای فاصله‌ای برای مدل رگرسیونی است که در آن داده‌ها دقیق یا به‌صورت فاصله‌ای هستند. روش‌های گوناگونی برای این مسئله مطرح شده است؛ اما ویژگی بسیاری از آنها این است که بعضی از پارامترهای رگرسیونی محاسبه‌شده، دقیق هستند. این ویژگی در کاربردهای گوناگون ناخوشایند است. برای برطرف کردن این مشکل، هادیک و کرنی [۹] در سال ۲۰۱۰ میلادی روش HC را مطرح کردند که انگیزه آن تحلیل تحمل در سیستم خطی است. مدل رگرسیونی خطی چندگانه فاصله‌ای به‌صورت زیر نمایش داده می‌شود:

$$y = A_0 + A_1 X_{*1} + \dots + A_k X_{*k} = AX$$

یا

$$y_j = A_0 + A_1 X_{j1} + \dots + A_k X_{jk} \\ = AX_{j*} \quad j = 1, 2, \dots, n, \quad (18)$$

که در آن X ماتریس ورودی $((n \times (k+1)))$ ، X_{j*} نشان‌دهنده زامین سطر ماتریس X ، y بردار خروجی n -بعدی، A بردار فاصله‌ای از پارامترهای مدل رگرسیونی است. اولین ستون ماتریس X که شامل یک است، معرف جمله ثابت است. ضرایب فاصله‌ای A_i ، $i = 0, 1, \dots, k$ ، به‌صورت $A_i = [a_i - c_i, a_i + c_i]$ نشان داده می‌شوند. سمت راست رابطه (۱۸) یک فاصله به‌صورت زیر است [۱۱، ۱۳، ۱۹]:

$$y_j = [a^t X_j - c^t |X_j|, a^t X_j + c^t |X_j|];$$

زیرا

$$y_j = (a_0, c_0) + (a_1, c_1)x_{j1} + \dots + (a_k, c_k)x_{jk} \\ = (a_0 + a_1 x_{j1} + \dots + a_k x_{jk}, \\ c_0 + c_1 |x_{j1}| + \dots + c_k |x_{jk}|) \\ = (a^t X_j, c^t |X_j|),$$

است که برای ورودی‌های مثبت (منفی)، کاهش (افزایش) پهنا داشته باشیم.

با توجه به این که A_0 و A_1 فاصله‌های فازی مثلثی هستند و x ورودی دقیق، $\hat{Y}(x)$ نیز فاصله فازی مثلثی متقارن است. مرکز $M(\hat{Y}(x))$ و پهنا $R(\hat{Y}(x))$ به‌صورت زیر هستند:

$$M(\hat{Y}(x)) = K_{\hat{Y}_j} \\ M(A_0) + M(A_1) \cdot x = K_{A_0} + K_{A_1} \cdot x \\ R(\hat{Y}(x)) = R(A_0) + R(A_1) \cdot x \quad (16)$$

همان‌طور که x روی دامنه D تغییر می‌کند، تغییرات معادله (۱۶) نیز باید مطابق علامت x تحلیل شود. با توجه به معادله (۱۶)، تغییرات $M(\hat{Y}(x))$ به علامت $K(A_1)$ بستگی دارد و برای هر مقدار ورودی x می‌تواند افزایش یا کاهش داشته باشد. همچنین تغییرات $R(\hat{Y}(x))$ نیز به علامت ورودی بستگی دارد. چون $R(A_1)$ همیشه مثبت است، اگر x مثبت (منفی) باشد، شعاع خروجی افزایش (کاهش) خواهد یافت.

در این زمینه، امکان و نوع تغییرات از مرکز خروجی متناسب با علامت $K(A_1)$ وجود دارد، همچنان که تغییرات شعاع خروجی نیز به علامت ورودی x محدود است.

۲.۱.۳ تعریف محدودیت (قید)

برای α -برش معین، فاصله فازی تبدیل به فاصله معمولی می‌شود؛ یعنی نمایش فاصله فازی، تعمیمی از فاصله معمولی است. تاناکا اصل α -برش را برای تعریف محدودیت‌های (قید) شمول در مسئله بهینه‌سازی به کار برد (داده‌های مشاهده‌شده در پیش‌بینی‌شده شمول هستند).

در این مورد، برای مجموعه داده‌های مشاهده‌شده، پارامترهای مدل فازی، یعنی A_0 و A_1 ، طوری برآورد می‌شوند که برای هر α -برش، همه داده‌های مشاهده‌شده زیرمجموعه (شمول) پیش‌بینی‌شده قرار گیرند. به عبارتی:

$$[Y_j]_\alpha \subseteq [\hat{Y}_j]_\alpha \Leftrightarrow \quad (17)$$

$$|M([\hat{Y}_j]_\alpha) - M([Y_j]_\alpha)| \leq R([\hat{Y}_j]_\alpha) - R([Y_j]_\alpha)$$

بعد از این که روش بهینه‌سازی انجام شد، فرض می‌شود که پارامترهای حاصل که برای یک α -برش معین محاسبه شده‌اند، برای همه $\alpha \in [0, 1]$ معتبر باشند. اگرچه این رویکرد (نمایش α -برش) مزایایی از جمله کاهش دادن پیچیدگی محاسبات و

که در آن

۱. اگر $\forall x \in D, x - shift \geq 0$ باشد، یعنی $shift \leq$

$\inf(D)$ است؛ پس خروجی مدل، افزایش پهنا روی D دارد.

۲. اگر $\forall x \in D, x - shift < 0$ باشد، یعنی $shift >$

$\sup(D)$ است؛ پس خروجی مدل، کاهش پهنا روی D دارد.

برای سادگی، مقدار $shift = \inf(D)$ را زمانی که خروجی مدل افزایش پهنا دارد، انتخاب می‌کنیم و برعکس، برای کاهش پهنا مقدار $shift = \sup(D)$ در نظر می‌گیریم.

۱.۳.۳ تعریف قید (محدودیت)

با استفاده از مدل دوزنقه‌ای، مسئله شمول همه ورودی‌های مشاهده شده در مقدار پیش‌بینی، در هر سطح α حفظ می‌شود. چون پارامترهای فازی، دوزنقه‌ای هستند، خروجی مدل $\hat{Y}(x)$ نیز دوزنقه‌ای است. می‌دانیم که خروجی مشاهده شده Y_j ، فازی مثلی متقارن است و خروجی مدل \hat{Y}_j نیز فرض شده که دوزنقه‌ای باشد. در این صورت، مسئله شمول تبدیل به شمول مراکز و تکیه‌گاه‌ها می‌شود. بنا بر این برای تعمیم روش فاصله‌ای تاناکا و حل مسئله شمول، دو قید شمول در روش تشخیص باید مد نظر قرار بگیرند: شمول تکیه‌گاه:

$$[Y_j]_{\alpha=0} \subseteq [\hat{Y}_j]_{\alpha=0},$$

و شمول مرکز:

$$[Y_j]_{\alpha=1} \subseteq [\hat{Y}_j]_{\alpha=1}. \quad (24)$$

چون شکل فاصله‌های فازی، دوزنقه‌ای فرض شده است، اگر رابطه (۲۴) برقرار باشد، در این صورت شمول برای هر سطح $\alpha \in [0, 1]$ تضمین می‌شود؛ یعنی

$$\forall \alpha \in [0, 1] : [Y_j]_{\alpha} \subseteq [\hat{Y}_j]_{\alpha}. \quad (25)$$

قیدهای شمول نیز به صورت زیر نوشته می‌شوند:

• قید شمول مرکز ($\alpha = 1$):

$$\begin{aligned} [Y_j]_{\alpha=1} &\subseteq [\hat{Y}_j]_{\alpha=1} \\ \Leftrightarrow K_{Y_j} &\in [K_{\hat{Y}_j}^-, K_{\hat{Y}_j}^+] \\ \Leftrightarrow |M(K_{\hat{Y}_j}) - K_{Y_j}| &\leq R(K_{\hat{Y}_j}). \end{aligned} \quad (26)$$

$$a = (a_1, \dots, a_k)^t, \quad c = (c_1, \dots, c_k)^t,$$

$$|X_j| = (1, |x_{j1}|, \dots, |x_{jk}|)^t$$

و مرکز و شعاع (پهنا) y_j به ترتیب $a^t X_j$ و $c^t |X_j|$ هستند.

c را به صورت $c = \delta \cdot c^\Delta$ در نظر می‌گیریم، به طوری که c^Δ داده شده، میزان تحمل نامنفی است. در حالت خاص، $c^\Delta = |a|$ به صورت $c^\Delta = |a|$ است. برای حل مسئله باید کمینه تحمل را به گونه‌ای پیدا کرد که رابطه (۱۸) با $A = [a - \delta \cdot c^\Delta, a + \delta \cdot c^\Delta]$ صدق کند؛ یعنی:

$$\forall j \in \{1, \dots, p\}$$

$$\exists a' \in [a - \delta \cdot c^\Delta, a + \delta \cdot c^\Delta] : y_j = X_{j*} a'. \quad (19)$$

اگر $|X_{j*}| c^\Delta > 0$ باشد، δ^* میزان کمینه تحمل است و از رابطه زیر به دست می‌آید:

$$\delta^* := \max_{j: |X_{j*}| c^\Delta > 0} \frac{|y_j - X_{j*} a|}{|X_{j*}| c^\Delta}. \quad (20)$$

۳.۳ روش ضابطه حجم

در این روش، فرض می‌شود که پارامترهای فاصله‌ای فازی A و A_1 اعداد دوزنقه‌ای باشند. در این صورت، خروجی مدل، یعنی \hat{Y}_j نیز دوزنقه‌ای است:

$$\hat{Y}_j = (K_{\hat{Y}_j}, S_{\hat{Y}_j}) = \left([K_{\hat{Y}_j}^-, K_{\hat{Y}_j}^+], [S_{\hat{Y}_j}^-, S_{\hat{Y}_j}^+] \right). \quad (21)$$

در این روش، خروجی مدل می‌تواند هر نوع تغییرات پهنا برای هر علامت x با ایجاد یک انتقال^{۱۸} روی ورودی مدل اصلی داشته باشد. بر اساس این روش، مدل خطی فاصله‌ای که روی دامنه D تعریف شده است، به مدل زیر تبدیل می‌شود:

$$Y = A \oplus A_1 \otimes (x - shift) = A \oplus A_1 \otimes w, \quad (22)$$

که در آن $w = x - shift$ در نظر گرفته می‌شود.

پهنای خروجی مدل به صورت زیر داده شده است:

$$\forall x \in D : R(\hat{Y}(x)) = R(A \cdot) + R(A_1) \cdot |w|. \quad (23)$$

طبق رابطه (۲۳) و با مقدار انتقال مناسب، خروجی مدل می‌تواند هر نوع تغییر پهنا روی D داشته باشد. در واقع،

• قید شمول تکیه گاه $(\alpha = 0)$:

که در آن $\Delta = \text{sign}(M(D))$ است.

حجم محدود به خروجی مدل روی کل دامنه D به صورت زیر است:

$$V = \int_{w_{\min}}^{w_{\max}} A(w) dw. \quad (32)$$

با جایگذاری روابط (۳۰) و (۳۱) در رابطه (۳۲) نتیجه می شود:

$$V = R(K_{A_0}) + R(S_{A_0}) + (R(K_{A_1}) + R(S_{A_1})) \cdot M(D) \cdot \Delta \quad (33)$$

توجه داشته باشید که برای سهولت، از این پس رابطه (۳۳) را ضابطه حجم می نامیم.

در این روش، ضابطه حجم، مستقل از داده های یادگیری است. بنا بر این بهینه سازی نه تنها در نقاط یادگیری، بلکه روی دامنه تعریف مدل، مستقل از توزیع داده های یادگیری انجام می شود. این ویژگی، کارایی ضابطه را افزایش می دهد.

۵.۳ روش ضابطه مجموع

تاناکا و همکاران [۱۸] در سال ۱۹۸۲ میلادی ضابطه ای مطرح کردند که مجموع پهناهای پارامترهای مدل فازی را کمینه می کند:

$$J = R(A_0) + R(A_1). \quad (34)$$

اما بهینه سازی این رابطه باعث می شود که اغلب پارامترهای مدل، اعدادی دقیق باشند و خروجی مدل خیلی عریض شود. به همین دلیل، تاناکا و همکاران [۱۸] ضابطه دیگری مطرح کردند، که ابهام مدل را کمینه می کند.

در روش جدید تاناکا، خروجی و پارامترهای مدل، اعداد ذوزنقه ای هستند. لذا تابع هدف نیز تغییر می کند:

$$S = M \cdot (R(S_{A_0}) + R(K_{A_0})) + (R(S_{A_1}) + R(K_{A_1})) \cdot \sum_{j=1}^M |x_j|. \quad (35)$$

برای سهولت، از این پس، رابطه (۳۵) را ضابطه مجموع می نامیم. بنا بر این روش برآورد ضرایب فازی ذوزنقه ای به کمک از کمینه کردن ضابطه مجموع تحت قیدهای (۲۶)-(۲۹).

$$[Y_j]_{\alpha=0} \subseteq [\hat{Y}_j]_{\alpha=0}.$$

$$\Leftrightarrow [K_{Y_j} - R_{Y_j}, K_{Y_j} + R_{Y_j}] \subseteq [S_{\hat{Y}_j}^-, S_{\hat{Y}_j}^+] \\ \Leftrightarrow \left| M(S_{\hat{Y}_j}) - K_{Y_j} \right| \leq R(S_{\hat{Y}_j}) - R_{Y_j}. \quad (27)$$

• قید شمول مرکز در تکیه گاه:

برای به دست آوردن فاصله فازی، قید شمول دیگری باید برقرار باشد؛ یعنی شمول مرکز در تکیه گاه:

$$[K_{\hat{Y}_j}^-, K_{\hat{Y}_j}^+] \subseteq [S_{\hat{Y}_j}^-, S_{\hat{Y}_j}^+] \\ \Leftrightarrow \left| M(K_{\hat{Y}_j}) - M(S_{\hat{Y}_j}) \right| \\ \leq R(S_{\hat{Y}_j}) - R(K_{\hat{Y}_j}). \quad (28)$$

• قید شعاع مثبت:

واضح است که شعاع فاصله های فازی تشخیصی باید مثبت یا صفر باشد؛ یعنی:

$$R(K_{A_i}) \geq 0, \quad R(S_{A_i}) \geq 0, \quad i = 0, 1. \quad (29)$$

در مجموع، روش تشخیص باید تحت قیدهای معرفی شده در روابط (۲۶)-(۲۹) انجام شود.

۴.۳ ضابطه حجم

هدف این بخش، آن است که ضابطه ای مستقل از داده های یادگیری تعریف کنیم. با فرض این که مدل تشخیص باید روی کل دامنه D به کار رود، سنجیده تر است که یک مدل با حداقل فازی را ترجیح دهیم؛ یعنی یک مدل که ابهام آن کمتر باشد. چون بعد عمودی باید در نظر گرفته شود، در حقیقت ابهام کل مدل به کمک از حجم محدود به خروجی مدل روی دامنه D . مساحت خروجی از طریق فاصله فازی ذوزنقه ای به صورت زیر بیان می شود [۲۲]:

$$A(w) = \frac{K_{\hat{Y}}^+ + S_{\hat{Y}}^+}{2} - \frac{K_{\hat{Y}}^- + S_{\hat{Y}}^-}{2}, \quad (30)$$

به طوری که

$$\forall w \in D : \begin{cases} K_{\hat{Y}}^- = K_{A_0}^- + (M(K_{A_1}) - R(K_{A_1})) \cdot \Delta \cdot w \\ K_{\hat{Y}}^+ = K_{A_0}^+ + (M(K_{A_1}) + R(K_{A_1})) \cdot \Delta \cdot w \\ S_{\hat{Y}}^- = S_{A_0}^- + (M(S_{A_1}) - R(S_{A_1})) \cdot \Delta \cdot w \\ S_{\hat{Y}}^+ = S_{A_0}^+ + (M(S_{A_1}) + R(S_{A_1})) \cdot \Delta \cdot w \end{cases} \quad (31)$$

۶.۳ روش بیسریه

قید دوم هم به کمک از:

$$R_{A_1} \geq 0, \quad M_{A_1} \geq R_{A_1} \geq 0. \quad (40)$$

۲.۶.۳ معرفی ضابطه

تابع هدف باید یک مدل با ابهام کمتر روی دامنه تعریف خود مشخص کند. محاسبه ابهام کل روی کل دامنه Δ صورت می گیرد، که به صورت زیر بیان می شود:

$$J = 2R_{A_1} \cdot R_D^+ + 2R_{A_1} \cdot R_D^- \cdot M_D + \frac{2}{3}M_{A_1} \cdot R_D^+ \quad (41)$$

بنا بر این روش تشخیص با کمینه کردن ضابطه (۴۱) تحت قیود (۳۹) و (۴۰) انجام می شود.

۴ روش پیشنهادی

در بخش قبل با تعدادی از روش های تحلیل رگرسیون خطی فاصله ای آشنا شدیم. در این بخش روشی ارائه می کنیم که باعث بهبود ضابطه حجم، ضابطه مجموع و روش بیسریه می شود.

در این روش، مجموع انحرافات مثبت و منفی را به ضابطه حجم، ضابطه مجموع و تابع هدف روش بیسریه اضافه می کنیم که باعث انعطاف بیشتر مدل می شود. به عبارتی با اضافه کردن مجموع انحرافات به مدل، به نوعی قدر مطلق اختلاف بین مقدار مشاهده شده متغیر پاسخ و مقدار برآورد شده را کمینه می کنیم. همچنین در قیدهای (۲۶)، (۲۷)، (۲۸) و (۳۹) برای این که علامت \leq را به $=$ تغییر دهیم، انحرافات مثبت (d_j^+) را به قیدها اضافه و انحرافات منفی (d_j^-) را از قیدها کم می کنیم به طوری که مقدار برآورد شده بر مقدار مشاهده شده منطبق شود. به همین دلیل انتظار داریم که در این مدل ها، مقدار خطا کاهش و ضریب تعیین مدل افزایش یابد.

۱.۴ بهبود ضابطه حجم با استفاده از روش

پیشنهادی

در این روش، ضابطه حجم به ضابطه زیر تغییر می یابد. بنا بر این ضابطه حجم پیشنهادی برابر است با:

$$R(K_{A_1}) + R(S_{A_1}) + (R(K_{A_1}) + R(S_{A_1})) \cdot M(D) \cdot \Delta + \sum_{j=1}^m (d_j^+ + d_j^-) \cdot \quad (42)$$

که در آن m تعداد قیدها است.

در تشخیص مدل رگرسیونی دو مسئله مطرح است: اولین آن تشخیص ساختار مدل مناسب برای مجموعه داده ها و دومین آن برآورد مدل داده شده است. بیسریه و همکاران [۴] در سال ۲۰۰۹ میلادی فرض کردند ساختار مدل، خطی باشد. در این صورت، مسئله رگرسیون تبدیل به مسئله برآورد پارامترهای مدل می شود. آن ها روش خود را به صورت زیر مطرح کردند:

یک نمونه n تایی از داده های مشاهده شده را در نظر بگیرید. Z امین نمونه به صورت جفت (X_j, Y_j) نمایش داده می شود که در آن X_j ورودی فاصله ای و Y_j خروجی فاصله ای متناظر است. فرض می شود که همه فاصله های ورودی مشاهده شده، در فاصله D قرار بگیرند. فاصله $D = [D^-, D^+]$ به صورت زیر داده شده است:

$$D^- = \min(X_j^-), \quad D^+ = \max(X_j^+). \quad (36)$$

هدف رگرسیون فاصله ای باید رابطه تابع پیش بینی بین ورودی X و خروجی Y را مشخص کند. دامنه تعریف مدل تشخیصی به کمک از همه فاصله هایی که در D قرار می گیرند؛ یعنی

$$\Delta = c(D).$$

بیسریه برای اجرای روش پیشنهادی خود دو نکته را مد نظر قرار داد:

- قیود شمول فاصله ای، برای بهینه سازی مسئله معرفی شوند.
- ضابطه تشخیص، بهینه سازی شود.

۱.۶.۳ قیود بهینه سازی

قیود شمول به صورت زیر داده شده اند:

$$Y_j \subseteq \hat{Y}_j \Leftrightarrow |M_{\hat{Y}_j} - M_{Y_j}| \leq R_{\hat{Y}_j} - R_{Y_j}, \quad (37)$$

به طوری که

$$M_{\hat{Y}_j} = M_{A_1} + M_{A_1 \otimes X_j}, \quad R_{\hat{Y}_j} = R_{A_1} + R_{A_1 \otimes X_j}. \quad (38)$$

با جایگذاری رابطه (۳۸) در (۳۷)، قید شمول به صورت زیر نوشته می شود:

$$|M_{A_1} + M_{A_1} M_{X_j} + R_{A_1} R_{X_j} - M_{Y_j}| \leq R_{A_1} + M_{A_1} R_{X_j} + R_{A_1} M_{X_j} - R_{Y_j}. \quad (39)$$

۳.۴ بهبود روش بیسریه با استفاده از روش قیدهای (۲۶)، (۲۷) و (۲۸) نیز به صورت زیر تغییر می‌یابند:

پیشنهادی

در این روش، تابع هدف به ضابطه زیر تغییر می‌یابد، بنا بر این تابع هدف پیشنهادی برابر است با:

$$J = 2R_{A_1} \cdot R_D^y + 2R_{A_1} \cdot R_D^y \cdot M_D + \frac{2}{3}M_{A_1} \cdot R_D^y + \sum_{j=1}^m (d_j^+ + d_j^-) \cdot \quad (47)$$

قید (۳۹) نیز به صورت زیر تغییر می‌یابد:

$$|M_{A_1} + M_{A_1}M_{X_j} + R_{A_1}R_{X_j} - M_{Y_j} + d_j^+ - d_j^-| = R_{A_1} + M_{A_1}R_{X_j} + R_{A_1}M_{X_j} - R_{Y_j} \quad (48)$$

از آنجا که انحرافات مثبت و منفی، بزرگ‌تر یا مساوی صفر هستند، برای برآورد ضرایب باید تابع هدف (۴۷) را تحت قید (۴۸) کمینه کنیم.

۵ مثال‌های عددی

در این بخش، برای بهبود روش HC و همچنین مقایسه کارایی روش پیشنهادی با روش‌های ضابطه حجم، ضابطه مجموع و روش بیسریه، چند مثال ارائه می‌کنیم.

۱.۵ بهبود روش HC

مدل برآورد قیمت خانه را در نظر بگیرید ([۱۳] و [۱۹]):

$$y = A_0 + A_1x_1 + A_2x_2 + A_3x_3,$$

که در آن x_1 کیفیت مصالح، x_2 مساحت طبقه اول (m^2)، x_3 مساحت طبقه دوم (m^2)، y نیز قیمت فروش است (جدول ۱). با استفاده از داده‌های جدول ۱ می‌خواهیم مدل فوق را برازش دهیم. روش HC به صورت زیر انجام می‌شود، در ابتدا a را از روش کمترین مربعات معمولی (OLS) برآورد می‌کنیم:

$$a = (X^T X)^{-1} X^T y = (-239/2748, 264/8359, 7/4659, 6/7569)^T$$

سپس قرار می‌دهیم $|a| = c^A$. بنا بر این $\delta^* = 0.0482$ را به دست می‌آوریم. از این مقدار δ^* متوجه می‌شویم که برای همه a_i ها برای $i = 1, \dots, 4$ ، $\delta^* a_i$ ها ابهام تا ۴.۸۲ درصد تحمل دارند.

$$\begin{aligned} [Y_j]_{\alpha=1} &\subseteq [\hat{Y}_j]_{\alpha=1} \\ &\Leftrightarrow |M(K_{\hat{Y}_j}) - K_{Y_j} + d_j^+ - d_j^-| \\ &= R(K_{\hat{Y}_j}). \end{aligned} \quad (43)$$

$$\begin{aligned} [Y_j]_{\alpha=0} &\subseteq [\hat{Y}_j]_{\alpha=0} \\ &\Leftrightarrow |M(S_{\hat{Y}_j}) - K_{Y_j} + d_j^+ - d_j^-| \\ &= R(S_{\hat{Y}_j}) - R_{Y_j}. \end{aligned} \quad (44)$$

$$\begin{aligned} [K_{\hat{Y}_j}^-, K_{\hat{Y}_j}^+] &\subseteq [S_{\hat{Y}_j}^-, S_{\hat{Y}_j}^+] \\ &\Leftrightarrow |M(K_{\hat{Y}_j}) - M(S_{\hat{Y}_j}) + d_j^+ - d_j^-| \\ &= R(S_{\hat{Y}_j}) - R(K_{\hat{Y}_j}). \end{aligned} \quad (45)$$

از آنجا که انحرافات مثبت و منفی، بزرگ‌تر یا مساوی صفر هستند، برای برآورد ضرایب رگرسیونی خطی فاصله‌ای باید ضابطه حجم پیشنهادی را تحت قیدهای (۴۳)، (۴۴) و (۴۵) کمینه کنیم. به عبارتی پارامترهای مدل را طوری برآورد می‌کنیم که مقدار برآورد شده منطبق بر مقدار مشاهده شده باشد.

۲.۴ بهبود ضابطه مجموع با استفاده از روش

پیشنهادی

در این روش، ضابطه مجموع به ضابطه زیر تغییر می‌یابد. بنا بر این ضابطه مجموع پیشنهادی برابر است با:

$$M \cdot (R(S_{A_1}) + R(K_{A_1})) + (R(S_{A_1}) + R(K_{A_1})) \cdot \sum_{j=1}^m |x_j| + \sum_{j=1}^m (d_j^+ + d_j^-) \cdot \quad (46)$$

از آنجا که انحرافات مثبت و منفی، بزرگ‌تر یا مساوی صفر هستند، برای برآورد ضرایب رگرسیونی خطی فاصله‌ای باید ضابطه مجموع پیشنهادی را تحت قیدهای (۴۳)، (۴۴) و (۴۵) کمینه کنیم.

جدول ۱. داده‌های قیمت خانه

x_3	x_2	x_1	y	j
۳۶/۴۳	۳۸/۰۹	۱	۶۰۶	۱
۲۶/۵۰	۶۲/۱۰	۱	۷۱۰	۲
۴۴/۷۱	۶۳/۷۶	۱	۸۰۸	۳
۳۸/۰۹	۷۴/۵۲	۱	۸۲۶	۴
۴۱/۱۰	۷۵/۳۸	۱	۸۶۵	۵
۲۶/۴۹	۵۲/۹۹	۲	۵۲۸	۶
۲۶/۴۹	۶۲/۹۳	۲	۹۱۷	۷
۳۳/۱۲	۷۲/۰۴	۲	۱۰۳۱	۸
۴۳/۰۶	۷۶/۱۲	۲	۱۰۹۲	۹
۴۲/۶۴	۹۰/۲۶	۲	۱۲۰۳	۱۰
۳۱/۳۳	۸۵/۷۰	۳	۱۳۹۴	۱۱
۲۷/۶۴	۹۵/۲۷	۳	۱۴۲۰	۱۲
۲۷/۶۴	۱۰۵/۹۸	۳	۱۶۰۱	۱۳
۶۶/۸۱	۷۹/۲۵	۳	۱۶۳۲	۱۴
۳۲/۲۵	۱۲۰/۵۰	۳	۱۶۹۹	۱۵

۱. نمونه‌ای به اندازه ۱ از توزیع یکنواخت^{۱۹} روی $(0, 0.5)$

تولید می‌کنیم و در d قرار می‌دهیم.

۲. این نمونه را جایگزین δ^* می‌کنیم، برای محاسبه SSR و

SST که در بالا شرح داده شده است.

۳. مراحل (۱) و (۲) را ۱۰۰ بار تکرار می‌کنیم.

۴. بیشترین مقدار R^2 را فرا می‌خوانیم.

۵. مقدار d مربوط به بیشترین R^2 را فرا می‌خوانیم.

۶. برای دستیابی به d مناسب‌تر، مراحل (۱) تا (۴) را ۱۰۰۰ بار تکرار می‌کنیم.

۷. میانگین d ها را محاسبه می‌کنیم، که مقدار 0.004 حاصل می‌شود.

۸. این میانگین را جایگزین δ^* می‌کنیم.

۹. مقدار R^2 را برای خط پایینی رگرسیون به دست می‌آوریم، که $R^2 = 0.99$ حاصل می‌شود.

بنا بر این مدل به صورت زیر است:

$$y = [-250/817, -227/732]x_1 + [252/060, 277/612]x_2 + [7/105, 7/826]x_3 + [6/431, 7/083]x_3.$$

همان گونه که ملاحظه می‌کنید هیچ کدام از ضرایب صفر نیستند.

این یعنی روش HC برای برآورد پارامترها مناسب است. حال با

استفاده از نرم‌افزار R ، ضریب تعیین R^2 را برای کران‌های پایین

و بالای \hat{y} محاسبه می‌کنیم. برای خط پایینی رگرسیون، نخست

بردار $V = a - \delta^* \cdot |a|$ را به دست می‌آوریم، سپس SSR و SST

را محاسبه می‌کنیم. به این ترتیب با توجه به رابطه $R^2 = \frac{SSR}{SST}$ ،

مقدار ضریب تعیین 0.93 به دست آمد. به طور مشابه، برای خط

بالایی رگرسیون نیز R^2 را به دست می‌آوریم؛ با این تفاوت که

بردار $V = a + \delta^* \cdot |a|$ است. مقدار $R^2 = 1$ حاصل شد.

حال می‌خواهیم ضریب تعیین R^2 را با استفاده از شبیه‌سازی

به دست آوریم. الگوریتم زیر، مراحل انجام کار را نشان می‌دهد.

الگوریتم تعیین δ^* و بهبود ضریب تعیین R^2 با استفاده از

شبیه‌سازی:

^{۱۹}Uniform

$$d_2 = \sum_{j=1}^n d^2(\hat{Y}_j, Y_j) \\ = \sum_{j=1}^n (S_{\hat{Y}_j}^- - S_{Y_j}^-)^2 + (K_{\hat{Y}_j}^- - K_{Y_j}^-)^2 \\ + (S_{\hat{Y}_j}^+ - S_{Y_j}^+)^2 + (K_{\hat{Y}_j}^+ - K_{Y_j}^+)^2. \quad (50)$$

- مقادیر ضابطه مجموع و حجم: میزان فازی بودن (ابهام) مدل را مشخص می‌کنند.

با استفاده از روش ضابطه مجموع و روش پیشنهادی، ضرایب را برآورد کرده‌ایم. نتایج حاصل را در جدول ۴ ملاحظه می‌کنید. با توجه به جدول ۴ ملاحظه می‌کنید که مقدار فاصله دیاموند که همان میزان خطای مدل است، از ۷۱/۰۵ به ۲۸/۸۶ کاهش یافته است. همچنین مقدار مجموع (مجموع پهنای فاصله‌های پیش‌بینی‌شده) از ۳۷/۰۸ به ۲۰/۶۲ و مقدار حجم (میزان ابهام مدل) از ۳/۲۸ به ۱/۸ کاهش یافته است و این یعنی روش پیشنهادی باعث بهبود روش ضابطه مجموع شده است. با توجه به جدول ۵ ملاحظه می‌کنید که مقدار فاصله دیاموند، مقدار مجموع و مقدار حجم در روش پیشنهادی کمتر از روش ضابطه حجم شده است. بنا بر این روش پیشنهادی باعث بهبود روش ضابطه حجم شده است.

۳.۵ مقایسه کارایی روش پیشنهادی با روش

بیسریه

روش پیشنهادی و روش بیسریه را روی داده‌های به کار گرفته‌شده در [۱۰] و [۱۵] اعمال می‌کنیم. داده‌ها در جدول ۶ آورده شده‌اند. همچنین نتایج حاصل را در جدول ۷ ملاحظه می‌کنید. همان‌گونه که مشاهده می‌کنید که مقدار ابهام مدل در روش پیشنهادی بسیار کمتر از روش بیسریه است. بنا بر این روش پیشنهادی باعث بهبود روش بیسریه شده است.

۶ نتیجه‌گیری

اعتبار روش‌های رگرسیون کلاسیک بستگی به برقراری فرضیاتی دارد که مدل‌سازی و استنباط‌های مربوط بر پایه آن فرضیات پایه‌گذاری می‌شوند. از جمله این فرضیات می‌توان به فرض نرمال بودن خطاها، در دسترس بودن مشاهدات دقیق و

۱۰. مقدار R^2 را برای خط بالایی رگرسیون به دست می‌آوریم که $R^2 = 1$ حاصل می‌شود.

حال، R^2 حاصل شده از روش HC را با روش شبیه‌سازی مقایسه می‌کنیم (جدول ۲ را ببینید). با توجه به جدول ۲ ملاحظه می‌کنید که R^2 خط پایینی رگرسیون روش شبیه‌سازی بهتر از روش HC است؛ زیرا مقدار آن به یک نزدیک‌تر است. R^2 خط بالایی رگرسیون در هر دو روش برابر با ۱ است. بنا بر این با استفاده از روش شبیه‌سازی، δ^* پیدا کردیم که روش HC را بهبود دادیم.

۲.۵ مقایسه کارایی روش پیشنهادی با روش ضابطه مجموع و ضابطه حجم

داده‌های جدول ۳ را که شامل ۸ مقدار مربوط به یک متغیر مستقل و مقادیر متغیر وابسته نظیرشان که اعداد فازی مثلثی متقارن هستند، در نظر بگیرید [۱۹] و [۱۰]. شاخص‌های کارایی استفاده شده عبارت‌اند از:

- فاصله دیاموند^{۲۰}: برای تعیین برآزش مدل به داده‌ها مورد استفاده قرار می‌گیرد و به صورت زیر برای فاصله‌های فازی مثلثی متقارن (رابطه (۴۹)) و دوزنقه‌ای (رابطه (۵۰)) تعریف شده است:

$$d_1 = \sum_{j=1}^n d^2(\hat{Y}_j, Y_j) \\ = \sum_{j=1}^n (S_{\hat{Y}_j}^- - S_{Y_j}^-)^2 + (K_{\hat{Y}_j}^- - K_{Y_j}^-)^2 \\ + (S_{\hat{Y}_j}^+ - S_{Y_j}^+)^2, \quad (49)$$

که در آن

$$S_{\hat{Y}_j} = S_{A.} + S_{A_1} \cdot x_j \\ \Rightarrow S_{\hat{Y}_j}^- = S_{A.}^- + S_{A_1}^- \cdot x_j, \\ S_{\hat{Y}_j}^+ = S_{A.}^+ + S_{A_1}^+ \cdot x_j \\ K_{\hat{Y}_j} = K_{A.} + K_{A_1} \cdot x_j \\ \Rightarrow K_{\hat{Y}_j}^- = K_{A.}^- + K_{A_1}^- \cdot x_j, \\ K_{\hat{Y}_j}^+ = k_{A.}^+ + K_{A_1}^+ \cdot x_j$$

^{۲۰} diamond distance

جدول ۲. مقایسه R^2 ی روش HC با روش شبیه سازی

شبیه سازی روش	HC روش	
۰/۹۹	۰/۹۳	رگرسیون پایینی خط R^2
۱	۱	رگرسیون بالایی خط R^2

جدول ۳. مجموعه داده های مشاهده شده

y_j	x_j	j
(۲/۲۵, [۱/۵, ۳])	۰/۱	۱
(۲/۸۷۵, [۲, ۳/۷۵])	۰/۲	۲
(۲/۵, [۱/۵, ۳/۵])	۰/۳	۳
(۴/۲۵, [۲/۵, ۶])	۰/۴	۴
(۴/۰, [۲/۵, ۵/۵])	۰/۵	۵
(۵/۲۵, [۴, ۶/۵])	۰/۶	۶
(۷/۵, [۵/۵, ۹/۵])	۰/۷	۷
(۸/۵, [۷, ۱۰])	۰/۸	۸

وجود داده های کافی اشاره کرد. هنگامی که برقراری فرضیات رگرسیون کلاسیک با مشکل مواجه شود، می توان از رگرسیون خطی فاصله ای به عنوان جایگزینی برای رگرسیون کلاسیک استفاده کرد. در این مقاله روش های مختلف تحلیل رگرسیون خطی فاصله ای را مطرح کردیم. حال سؤالی که مطرح است وجود داده های کافی اشاره کرد. هنگامی که برقراری فرضیات رگرسیون کلاسیک با مشکل مواجه شود، می توان از رگرسیون خطی فاصله ای به عنوان جایگزینی برای رگرسیون کلاسیک استفاده کرد. در این مقاله روش های مختلف تحلیل رگرسیون خطی فاصله ای را مطرح کردیم. حال سؤالی که مطرح است این است که کدام روش را انتخاب کنیم؟ واضح است باید معقول ترین و کارآمدترین روش را که در عین حال، پیچیدگی کمتری داشته باشد، انتخاب کرد. با اضافه کردن انحرافات مثبت و منفی در قیدها و تابع هدف، میزان خطای برآورد پارامترها در رگرسیون خطی فاصله ای به طور قابل توجهی کاهش می یابد.

جدول ۴. مقایسه روش ضابطه مجموع با روش ضابطه مجموع پیشنهادی

پیشنهادی مجموع ضابطه روش	مجموع ضابطه روش	
([۰/۵۴ و ۰/۵۵] و [-۱/۴۵ و ۱/۴۷])	([-۱/۹۶ و ۱/۹۲] و [-۱/۳۶ و -۰/۴۶])	A_1
([۹/۹۳ و ۹/۹۳] و [۹/۸۲ و ۱۱/۳۴])	([۸/۹۳ و ۸/۹۳] و [۸/۹۳ و ۱۰/۸۳])	A_1
۲۸/۸۶	۷۱/۰۵	دیاموند فاصله
۲۰/۶۲	۳۷/۰۸	مجموع
۱/۸	۳/۲۸	حجم

جدول ۵. مقایسه روش ضابطه حجم با روش ضابطه حجم پیشنهادی

روش ضابطه حجم پیشنهادی	روش ضابطه حجم	
$([۰/۹۱], [۰/۳۵] و [۲/۵])$	$([۰/۹۲], [۰/۳۶] و [۰/۲۵])$	A_0
$([۵/۸۳], [۷/۹۲] و [۷/۵])$	$([۵/۸۳], [۷/۹۳] و [۷/۵])$	A_1
۹۱/۰۲	۹۱/۲۹	دیاموند فاصله
۳۸/۲۲	۳۸/۴۲	مجموع
۳/۱۲	۳/۱۵	حجم

جدول ۶. داده‌های ورودی، مشاهده شده و پیش‌بینی شده

X_j ورودی‌های	Y_j خروجی‌های مشاهده شده	\hat{Y}_j خروجی‌های پیش‌بینی شده
$[۱/۵ و ۲/۵]$	$[۳/۵ و ۴/۵]$	$[۳/۵ و ۶/۵]$
$[۳ و ۴]$	$[۵ و ۶]$	$[۴/۱۶ و ۷/۲۵]$
$[۴/۵ و ۶/۵]$	$[۶/۵ و ۸/۵]$	$[۴/۸۲ و ۸/۵]$
$[۶/۵ و ۷/۵]$	$[۶ و ۷]$	$[۵/۶۹ و ۹]$
$[۸, ۹]$	$[۸, ۹]$	$[۶/۳۵, ۹/۷۵]$
$[۹/۵ و ۱۱/۵]$	$[۷ و ۹]$	$[۷ و ۱۱]$
$[۱۰/۵ و ۱۱/۵]$	$[۱۰ و ۱۱]$	$[۷/۴۵ و ۱۱]$
$[۱۲ و ۱۳]$	$[۹ و ۱۰]$	$[۸/۱۱ و ۱۱/۷۵]$

جدول ۷. مقایسه روش بیسریه با روش پیشنهادی

روش بیسریه	روش پیشنهادی بیسریه	
$[۲/۸۴۴ و ۵/۲۵]$	$[۶/۵ و ۴۸۴]$	A_0
$[۰/۴۳۸ و ۰/۵]$	$[-۰/۰۰۱ و ۰/۰۰۱]$	A_1
۱۵۳/۹۵	۶۶/۲۰	مدل ابهام

مراجع

- [۱] حسین یزدی، م.، حسن پور، ح.، و ماشین چی، م. (۱۳۸۶). آشنایی با حساب اعداد فازی، فرهنگ و اندیشه ریاضی، ۳۸، ۴۹-۶۲.
- [2] Bissierier, A., Boukezzoula, R., and Galichet, S., (2009). An interval approach for fuzzy linear regression with imprecise data, *Laboratoire d'Informatique, Systèmes, Traitement de l'Information et de la Connaissance (LISTIC), Malaga, Spain*, BP 80439 74941.
- [3] Bissierier, A., Boukezzoula, R., and Galichet, S. (2010). A revisited approach to linear fuzzy regression using trapezoidal fuzzy intervals, *Information Sciences*, **180**, 3653-3673.
- [4] Bissierier, A., Boukezzoula, R., and Galichet, S. (2009). Linear fuzzy regression using trapezoidal fuzzy intervals, *Information Sciences*, **4 (1)**, 59-72, .
- [5] Boukezzoula, R., Galichet, S., and Foulloy, L., (2007). MIN and MAX operators for fuzzy intervals and their potential use in aggregation operators, *IEEE Trans. on Fuzzy Systems*, **15(6)**, 1135-1144.
- [6] Chang, P. T., Lee, E. S., and Konz, S. A. (1996). Applying fuzzy linear regression to VDT legibility, *Fuzzy Sets and Systems*, **80 (2)**, 197-204.
- [7] Hao, P. Y. (2009). Interval regression analysis using support vector networks, *Fuzzy Sets and Systems*, **160 (17)**, 2466-2485.
- [8] Heshmaty, B. and Kandel, A. (1985). Fuzzy linear regression and its applications to forecasting in uncertain environment. *Fuzzy Sets and Systems*, **15 (2)**, 159-191.
- [9] Hladik, M. and Cerny, M. (2010). New approach to interval linear regression, *24th Mini EURO Conference: Continuous Optimization and Information-Based Technologies in the Financial Sector, Izmir, TURKEY*, 167-171.
- [10] Hojati, M., Bector, C. R., and Smimou, K. (2005). A simple method for computation of fuzzy linear regression, *European Journal of Operational Research*, **166**, 172-184.
- [11] Huang, C. H. and Kao, H. Y. (2009). Interval regression analysis with soft-margin reduced support vector machine. *Lecture Notes in Computer Science*, 826-835.
- [12] Kaneyoshi, M., Tanaka, H., Kamei, M., and Furuta, H. (1990). New system identification technique using fuzzy regression analysis, *Proceedings of the First International Symposium on Uncertainty Modeling and Analysis, Baltimore*, 528-533.
- [13] Lee, H. and Tanaka, H. (1999). Upper and lower approximation models in interval regression using regression quantile techniques, *European Journal of Operational Research*, **116(3)**, 653-666.
- [14] Liu, P. (2009). Study on a speech learning approach based on interval support vector regression, *in Proceedings of 4th International Conference on Computer Science and Education, Nanning*, 1009-1012.

- [15] Sakawa, M. and Yano, H. (1992). Multiobjective fuzzy linear regression analysis for fuzzy input-output data, *Fuzzy Sets and Systems*, **47**, 173-181.
- [16] Taheri S. M. and Kelkinnama M. (2012). Fuzzy linear regression based on least absolute deviation, *Iranian Journal of Fuzzy Systems*, **9** (1), 121-140.
- [17] Tanaka, H. (1987). Fuzzy data analysis by possibilistic linear models, *Fuzzy Sets and Systems*, **24** (3), 363–375.
- [18] Tanaka, H., Hayashi, I., and Watada, J. (1989). Possibilistic linear regression analysis for fuzzy data, *European Journal of Operational Research*, **40**, 389-396.
- [19] Tanaka, H. and Lee, H. (1998). Interval regression analysis by quadratic programming approach, *IEEE Transactions on Fuzzy Systems*, **6** (4), 473–481.
- [20] Tanaka, H., Uejima, S., and Asai, K. (1982). Linear regression analysis with fuzzy model, *IEEE Systems Trans. Systems Man. Cybernet.*, **2**, 903-907.
- [21] Tanaka, H. and Watada, J. (1988). Possibilistic linear systems and their application to the linear regression model. *Fuzzy Sets and Systems*, **27** (3), 275–289.
- [22] Yager, R. (2008). Using trapezoids for representing granular objects: applications to learning and OWA aggregation, *Information Sciences*, **178**, 363–380.
- [23] Zadeh, L. A. (1965). Fuzzy sets, *Information and Control*, **8**, 338-353.