رگرسیون خطی فازی و کاربردهای آن در پژوهش‌های علوم اجتماعی

ابراهیم خدایی *

(تاریخ دریافت: ۱۳۸۸/۶/۳۰ ، تاریخ تصویب: ۱۳۸۸/۱۲/۱۸)

چکیده

در این مقاله، تئوری رگرسیون کلاسیک و رگرسیون فازی مقایسه شده‌اند و مثالی کاربردی از کاربرد رگرسیون فازی در علوم اجتماعی عرضه می‌شود. رگرسیون کلاسیک برپایهٔ فرض دقیق‌بودن متغیرهای مورد مطالعه و مشاهدات مرتبط با آنها استوار بوده و در نهایت، روابط بین متغیرها را نیز دقیقاً مشخص می‌کند. در مدل‌سازی مسائل اجتماعی، عموما با مشاهدات نادقیق یا روابط مبهم روبروییم. بنابراین استفاده از روش‌های برازش توابع که به تبیین ساختار مبهم داده‌ها و روابط بین آنها می‌پردازند، ضروری است. در بررسی حاضر، رگرسیون خطی فازی معدل کتبی دیپلم روی نمرات ریاضی، فیزیک و جنسیت داوطلبان آزمون سراسری سال ۱۳۸۷ در قالب مثالی کاربردی با فرض فازی‌بودن ضرایب و غیر فازی بودن داده‌ها بررسی می‌شود. برای آزمون رگرسیون فازی از شاخص اطمینان و مقدار ابهام مدل استفاده خواهد شد.

واژگان کلیدی: رگرسیون خطی کلاسیک، مجموعهٔ فازی، رگرسیون فازی خطی، آزمون سراسری.

* استادیار آمار سازمان سنجش آموزش کشور، khodaie@sanjesh.org

مقدمه

پیش از گسترش علوم، خصوصاً علوم انسانی، بررسی و تحلیل جوامع انسانی عموماً در فلسفه صورت می‌گرفت. در این دوره، مسائل اجتماعات انسانی با دیدگاه‌هایی فلسفی که اساسا ذهنی و انتزاعی بودند تفسیر می‌شدند و اندیشمندان به دنبال بررسی تجربی روابط و پدیده‌های انسانی و اجتماعی نبودند. روایی این‌گونه حکم‌ها و نظرات نیز با معیارهای ذهنی و عقلی ارزشیابی می‌شد، نه با ملاک‌های عینی و تجربی. توسعهٔ دیدگاه‌های علمی در اوایل قرن بیستم و ظهور اثبات‌گرایی دانشمندان حوزهٔ علوم انسانی را به پژوهش‌های عینی و تجربی رهنمون ساخت و صاحب‌نظران را به عرضهٔ معرفتی جدید از واقعیت‌های اجتماعی و تحقیق درخصوص مسائل جامعه با روش عینی هدایت کرد. در این میان، نظریه‌پردازی و روش‌شناسی در علوم اجتماعی به استفاده از علم آمار به مثابه ابزاری دقیق برای محک تجربی پدیده‌ها گرایش پیدا کرد. با نگاهی جزئی‌تر می‌توان جامعه‌شناسی را به دو حوزهٔ نظری و تجربی تقسیم کرد. در بحث‌های نظری، با اندیشه‌ها و نظریه‌ها و مفاهیم انتزاعی و روابط بین آنها سرو کار داریم، اما در حوزهٔ تجربی با جهان تجربی و داده‌ها روبروییم و در هر دو حوزه، برای نتیجه‌گیری و تولید اندیشه و حل مسائل به آمار و ریاضیات به مثابه یک ابزاری مناسب برای تلخیص و استنباط داده‌ها نیاز است. یکی از ابزارهای بسیار مهم در تحلیل روابط بین متغیرها رگرسیون است. کاربردهای روش‌های رگرسیونی به تفصیل در حوزهٔ علوم اجتماعی به بحث و بررسی گذاشته شده است و درخصوص لزوم این ابزار برای تحقیق‌های علوم اجتماعی اجماع نظر وجود دارد. با توجه به این‌که به‌کارگیری شیوه‌ها و ابزارهای نظریهٔ مجموعه‌های فازی در گسترش و تعمیق روش‌های آماری چندی است مورد توجه محققان حوزه‌های علوم، به‌خصوص علوم اجتماعی و جامعه‌شناسی، قرار گرفته است، هدف این مقاله نگاهی اجمالی به روش رگرسیون فازی از نظر آموزش علاقه‌مندان و محققان حوزهٔ علوم اجتماعی است. با این دیدگاه، در این مقاله کلیات روش‌ها، اهداف و فرض های لازم در رگرسیون کلاسیک و نظریهٔ مجموعه‌ها، آمار، تئوری و آزمون‌های رگرسیون فازی به بحث و بررسی گذاشته می‌شود. سپس، رابطهٔ بین نمرات، معدل و جنسیت داوطلبان در آزمون سراسری سال ۱۳۸۷ از منظر رگرسیون کلاسیک و فازی تجزیه و تحلیل و نتایج به‌دست‌آمده با همدیگر مقایسه می‌شوند.

رگرسیون خطی

نظریهٔ رگرسیون را اولین بار فرانسیس گالتون در ۱۸۸۵ مطرح کرد. معنی کلمهٔ رگرسیون بازگشت است و گالتون نشان داد که میانگین قد فرزندان افراد بلندقد و کوتاه‌قد به میانگین جامعهٔ اصلی بازگشت پیدا می‌کند. بعد از عرضهٔ تئوری رگرسیون، مثل تئوری احتمال که با پیش‌بینی برد و باخت قماربازان شکل گرفت، محققان حوزه‌های گوناگون علوم دریافتند که کاربردهای رگرسیون بسیار فراتر از بحث اولیهٔ آن بوده است، به‌طوری که در حال حاضر اگر بخواهیم تعریفی برای این تئوری به‌دست دهیم، می‌توان به شکل خلاصه گفت که تحلیل رگرسیون روشی آماری است که تغییرات متغیر وابسته از طریق متغیر یا متغیرهای مستقل تبیین و پیش‌بینی می‌شود. روش‌های رگرسیونی از نظر تئوری به سه نوع رگرسیون خطی، غیر خطی و ناپارامتری تقسیم می‌شوند. در این مقاله، فقط رگرسیون خطی مورد بحث و بررسی قرار خواهد گرفت. رگرسیون خطی را برحسب تعداد متغیرهای وابسته و مستقل می‌توان در قالب سه نوع تقسیم‌بندی زیر لحاظ کرد:

الف) رگرسیون سادهٔ خطی: در این حالت، یک متغیر وابسته و یک متغیر مستقل وجود دارد.

ب) رگرسیون چندگانه: در این حالت، یک متغیر وابسته و بیش از یک متغیر مستقل فرض می‌شود.

ج) رگرسیون چندمتغیره: در این حالت، چند متغیر وابسته و چند متغیر مستقل برای تجزیه و تحلیل فرض می‌شوند.

در رگرسیون خطی ساده، اگر فرض کنیم دو متغیر y_i ، x_i و (y_i, x_i) , $i=1,2,\ldots,n$ وجود داشته باشد، رابطه به صورت زیر است:

$$y_i = \beta_\circ + \beta_1 x_i + \varepsilon_i \qquad \varepsilon_i = 1,2,\ldots,n \quad (۱)$$

در رگرسیون چندگانه، بیش از یک متغیر مستقل وجود دارد، به‌طوری‌که اگر فرض کنیم دو متغیر مستقل z و x و متغیر وابسته y موجود باشد، رابطهٔ مورد نظر به صورت زیر خواهد بود:

$$y_i = \beta_\circ + \beta_1 x_i + \beta_2 z + \varepsilon_i$$

در رابطهٔ (۱) $\beta_\circ$ و β_1 را پارامترهای مدل نامیده و هدف برآورد $\beta_\circ$ و β_1 است. برای برآورد پارامترهای فوق روش‌های متفاوتی وجود دارد که می‌توان به روش‌های حداقل مربعات وحداکثردرست‌نمایی اشاره کرد. در روش حداقل مربعات، $\beta_\circ$ و β_1 طوری برآورد می‌شوند که مقدار باقیمانده یا خطا در رابطهٔ (۱)، یعنی $\varepsilon_i = y_i - \hat{\beta}_\circ - \hat{\beta}_1 x_i$ در رابطه $\sum_{i=1}^{n} \varepsilon_i^2$، حداقل مقدار خود را داشته باشد.

با استفاده از مشتق‌گیری و سایر روش‌های ریاضی، برآورد پارامترها به صورت زیر خواهد بود:

$$\hat{\beta}_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \qquad (۲)$$

$$\hat{\beta}_\circ = \bar{y} - \hat{\beta}_1 \bar{x} \qquad (۳)$$

که در رابطهٔ (۲) و (۳)، $\bar{y}$ و $\bar{x}$ میانگین‌های نمونهٔ x و y اند.

در ساختن مدل رگرسیونی (۱) فرض می‌شود که ε_i ها مستقل و دارای توزیع نرمال $N(\circ, \sigma^2)$ اند (σ^2 نامعلوم است). در رگرسیون کلاسیک، پس از پیداکردن معادلهٔ رگرسیونی از آزمون‌های کلاسیک آماری از قبیل آزمون F ، آزمون t به‌منظور اعتباربخشی به مدل استفاده می‌شود و علاوه بر این، همان‌طوری‌که در مقدمه ذکر شد، هدف تئوری رگرسیون پیش‌بینی و تبین واریانس متغیر وابسته به‌وسیلهٔ متغیرهای مستقل است و آماره‌ای که برای این امر استفاده می‌شود عبارت است از ضریب تبیین:

$$r^2 = \frac{\text{واریانس معلوم}}{\text{کل واریانس}} = \frac{\frac{\sum (\hat{y} - \bar{y})^2}{n-2}}{\frac{\sum (y - \bar{y})^2}{n-2}} \qquad (۴)$$

r^2 یا توان دوم ضریب همبستگی نشان‌دهندهٔ واریانس مشترک x و y است یا به عبارت دیگر، مقدار r^2 نشان‌دهندهٔ میزان تغییرات متغیر وابسته است که به‌وسیلهٔ مدل رگرسیونی تبیین می‌شود. همچنین، r^2 را گاهی معیار کاهش نسبی خطا نیز می نامند.

مجموعه‌ها و آمار فازی

ریاضیات یک فرامجموعه از منطق بولی است که بر مفهوم درستی نسبی دلالت می‌کند. منطق کلاسیک هر چیزی را بر اساس یک سیستم دوتایی نشان می‌دهد (درست یا غلط، ۰ یا ۱، سفید یا سیاه)، ولی منطق فازی درستی هر چیزی را با یک عدد که مقدار آن بین صفر و یک است نشان می‌دهد. مثلاً اگر رنگ سیاه را با عدد صفر و رنگ سفید را با عدد ۱ نشان دهیم، آنگاه رنگ خاکستری عددی نزدیک به صفر خواهد بود. در ۱۹۶۵، دکتر لطفی‌زاده نظریهٔ سیستم‌های فازی را معرفی کرد. منطق فازی، معتقد است که ابهام در ماهیت علم است. لطفی‌زاده معتقد است که باید به دنبال ساختن مدل‌هایی بود که ابهام را به منزلهٔ بخشی از سیستم برخلاف دیگران که معتقدند باید تقریب‌ها را دقیق‌تر کرد تا بهره‌وری افزایش یابد مدل کند. در منطق ارسطویی، یک دسته‌بندی درست و نادرست وجود دارد. همهٔ گزاره‌ها درست یا نادرست‌اند. اما در منطق فازی، گزاره‌ها مقداری درست و مقداری نادرست‌اند. مثلاًگزارهٔ "هوا سرد است" یک گزارهٔ منطقی فازی است که درستی آن بسته به شرایط جغرافیایی و فرد قضاوت‌کننده گاهی کم و گاهی زیاد است.

قبل از بحث دربارهٔ مجموعه‌های فازی بهتر است نگاهی اجمالی به نظریهٔ مجموعه‌های کلاسیک بیاندازیم که با تعریف مجموعه شروع می‌شود. گردایهٔ معینی از اشیاء را مجموعه می‌نامند. اشیاء این گردایه اعضا یا عناصر مجموعه نامیده می‌شوند. مثلاًاعداد طبیعی کمتر یا مساوی ۸ یک مجموعه را تشکیل می‌دهند. اگر این مجموعه را با A نشان دهیم، خواهیم داشت:

$$A:\{1,2,\ldots,8\}$$

مجموعهٔ A را می‌توان به صورت‌های $A=\{x\in N\mid x\le 8\}$ یا $A=\{x\in X\mid P(x)\}$ نیز نشان داد که P ویژگی x را نشان می دهد. برای متعلق بودن عناصر یک مجموعه می‌توان از توابع نشانگر استفاده کرد. مثلاً برای مجموعه فوق‌الذکر تابع نشانگر به صورت زیر تعریف می‌شود:

$$I_A(x)=\begin{cases}1 & x\in A\\ 0 & x\notin A\end{cases}$$

چنان‌که ملاحظه می‌شود، دامنهٔ تابع نشانگر مجموعهٔ مرجع و برد آن مجموعهٔ دو عضوی $\{0,1\}$ است. یعنی $I_A(x):X\to\{0,1\}$. واضح است که هر مجموعه یک تابع نشانگر دارد و بالعکس.

در مثال بالا، ویژگی مجموعهٔ A دقیقاً تعریف شده است و عضویت یک عنصر در مجموعه کاملاً مشخص است، اما در زندگی روزمره با مفاهیم و استدلال‌هایی سروکار داریم که نادقیق‌اند، از قبیل قد بلند، سردی هوا، مسافت طولانی، فشار خون بالا. نظریهٔ مجموعه‌های فازی نظریه‌ای برای مدل‌بندی و تجزیه‌وتحلیل مفاهیم نادقیق فوق‌الذکر است. در نظریهٔ مجموعه‌های کلاسیک، برد تابع نشانگر مجموعه $\{0,1\}$ است، اما در منطق فازی برد مجموعه به جای مجموعه $\{0,1\}$ بازهٔ $[0,1]$ در نظر گرفته می‌شود. به عبارت دیگر، در این حالت به جای تعلق قطعی، یا عدم تعلق قطعی میزان عضویت در بازهٔ $[0,1]$ تعریف می‌شود و هر قدر میزان عضویت به عدد یک نزدیک‌تر باشد، یعنی تعلق بیشتر عضو به مجموعه، و هر قدر به صفر نزدیک‌تر باشد، یعنی عدم تعلق به مجموعه.

در این تئوری، به جای مجموعه نشانگر از تابع عضویت $\mu_A(x): X \to [0,1]$ استفاده می‌شود، به عبارت دیگر، $\mu_A(x)$ میزان عضویت x در مجموعه فازی A را نشان می‌دهد. با توجه به این‌که در تئوری مجموعه‌ها خصوصیات ریاضی گوناگونی از اجتماع، اشتراک متمم، مکمل و افراز وجود دارد، طبیعتاً این خصوصیات در تئوری مجموعه‌های فازی نیز قابل تعریف‌اند.[۲]

با توجه به گسترش کاربردی منطق فازی در علوم گوناگون از قبیل فنی‌مهندسی، علوم انسانی و علوم پزشکی مجال بحث و بررسی درخصوص این منطق در این مقاله فراهم نیست و فقط نکتهٔ مهمی که باید در اینجا بدان اشاره شود این است که آیا با توجه به حوزه مورد بحث، یعنی احتمالات، اشتراکی بین تئوری منطق فازی و علم آمار وجود دارد یا خیر؟ یا آیا بین منطق فازی و علم آمار مرزبندی دقیقی وجود دارد یا خیر؟ در پاسخ می‌توان گفت نظریهٔ آمار و نظریهٔ مجموعه‌های فازی هر دو برای مطالعهٔ الگوها و سیستم‌هایی شامل عدم قطعیت و پویا وضع شده‌اند، اما با این تفاوت که نظریهٔ آمار برای مطالعهٔ الگوهای مبتنی بر عدم قطعیت آماری (منسوب به پیشامدهای تصادفی) و نظریهٔ مجموعه‌های فازی برای مطالعهٔ الگوهای مبتنی بر عدم قطعیت امکانی (ناشی از ابهام و نادقیق‌بودن) گسترش یافته‌اند.

این دو نظریه متناقض یکدیگر نبوده و شامل همدیگر هم نیستند، اما می‌توان برای حل یک مسئله از هر دو روش‌های کلاسیک آماری و روش‌های فازی استفاده کرد. در این زمینه مثلاً می‌توان به آزمون فرض‌های فازی رگرسیون فازی اشاره داشت.

رگرسیون خطی فازی

چنان‌که در بخش قبلی ملاحظه شد، رگرسیون کلاسیک فرضیاتی را در زمینهٔ توزیع احتمالی خطاها در نظر می‌گیرد. اگرچه مدل رگرسیون خطی کلاسیک کاربردهای بسیار دارد، اما در بعضی مواقع ساختن مدل با مشکلاتی مواجه است که عبارت‌اند از تعداد کم یا نامناسب بودن مشاهدات، مشکلات تعریف تابع توزیع مناسب، ابهام در رابطهٔ بین متغیرهای وابسته و مستقل، ابهام در وقوع یا درجهٔ وقوع رویدادها، بی‌دقتی و خطا. مثلاً به‌کاربردن تحلیل رگرسیون آماری ممکن است باعث نتیجه‌گیری اشتباه شود. برای حل این مسئله می‌توان از روش‌های دیگر از قبیل رگرسیون استوار و رگرسیون خطی فازی استفاده کرد که در اینجا رگرسیون خطی فازی مورد بحث قرار می‌گیرد. رگرسیون فازی در حالت کلی به سه نوع تقسیم می‌شود:

الف) رگرسیون فازی در حالتی که روابط بین متغیرها (ضرایب مدل رگرسیونی) فازی فرض شوند.

ب) رگرسیون فازی در حالتی که مشاهدات در متغیر وابسته و متغیرهای مستقل نادقیق و فازی باشند.

ج) رگرسیون فازی در حالتی که هم روابط بین متغیرها و هم مشاهدات فازی در نظر گرفته شوند.

رگرسیون خطی با ضرایب فازی را اولین بار تاناکا و همکارانش در ۱۹۸۲ معرفی کردند. بعد از اولین مقالهٔ ایشان، مقالات متعددی در خصوص تئوری فوق‌الذکر به‌دست ایشان و افراد دیگری با تکیه بر گسترش تئوری و مثال‌های کاربردی منتشر شد که حشمتی و کاندل (۱۹۸۵) برای پیش‌بینی هوا از آن جمله است.

در رگرسیون خطی با ضرایب فازی، فرض می‌شود که مشاهدات و متغیرها دقیق و ابهام در مدل و ضرایب رگرسیون است. فرض کنیم y متغیر وابسته و $X_1, X_2, \ldots, X_p$ متغیرهای مستقل و تعداد مشاهدات n باشد، صورت کلی مدل رگرسیون خطی فازی به شکل زیر خواهد بود:

$$\tilde{Y} = f(x, \tilde{A}) = \tilde{A}_0 + \tilde{A}_1 x_1 + \tilde{A}_2 x_2 + \cdots + \tilde{A}_p x_p \qquad (۵)$$

هدف برآورد پارامترهای مدل (۵) یعنی $\tilde{A}_0, \tilde{A}_1, \ldots, \tilde{A}_p$ است، به قسمتی که مدل (۵) بهترین برازش را برای داده‌ها به‌دست دهد. برای پیدا کردن پارامترهای فوق از تابع عضویت مثلثی زیر استفاده خواهیم کرد. توجه شود که می‌توان از توابع عضویت دیگر، از قبیل نرمال، استفاده کرد،اما در این مقاله فقط تابع عضویت مثلثی مورد بحث و بررسی قرار می‌گیرد. تابع عضویت مثلثی به صورت زیر تعریف می‌شود:

---

[۲]. نک منابع گوناگون فارسی، به ویژه عادل آذر (۱۳۸۶)

$$\widetilde{A}_{(x)}=\begin{cases}1-\dfrac{a-x}{s^L} & a-s^L\le x\le a\\ 1-\dfrac{x-a}{s^R} & a<x\le a+s^R\end{cases} \qquad (۶)$$

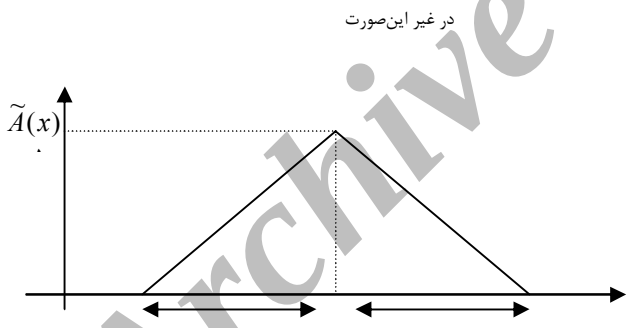
که a مقدار میانه، s^L کران پایین و s^R کران بالای $\widetilde{A}$ را نشان می‌دهند، به قسمیت که $\widetilde{A}=(a,s^L,s^R)_T$. اگر $s^L=s^R=s$ باشد، در این صورت $\widetilde{A}$ عدد مثلثی متقارن و در غیر این‌صورت نامتقارن خواهد بود. در صورت متقارن بودن، $\widetilde{A}$ را با $(a,s)_T$ نشان خواهیم داد. به عبارت دیگر، در حالت $\widetilde{A}=(a,s)_T$ ، تابع عضویت $\widetilde{A}(x)$ به صورت زیر خواهد بود:

$$\widetilde{A}_{(x)}=\begin{cases}1-\dfrac{a-x}{s} & a-s\le x\le a\\ 1-\dfrac{x-a}{s} & a<x\le a+s\end{cases} \qquad (۷)$$

تابع عضویت ضریب فازی i ام که یک عدد فازی مثلثی فرض می‌شود به صورت زیر است که در شکل (۱) به تصویر کشیده شده است.

$$\widetilde{A}(x)=\begin{cases}1-\dfrac{|a_i-x|}{s_i} & a_i-s_i\le x\le a_i+s_i\\ \circ & \text{در غیر این‌صورت}\end{cases}$$

شکل (۱)



تابع عضویت عدد فازی $\widetilde{A}$ تابع دو پارامتر a,s است، به طوری که پارامتر a، پارامتر میانه و پارامتر s پارامتر گسترهٔ عددی فازی است که نشان‌دهندهٔ میزان فازی‌بودن عدد است. در واقع، عدد فازی $\widetilde{A}$ با دو پارامتر a,s نشان‌دهندهٔ تقریباً a است.

پارامترهای فازی $\widetilde{A}=\{\widetilde{A}_\circ,\widetilde{A}_1,\ldots,\widetilde{A}_p\}$ می‌توانند به صورت بردار $\widetilde{A}=\{a,s\}$ نشان داده شوند، به‌قسمی‌که داریم:

$$a=(a_\circ,a_1,\ldots,a_p)$$
$$s=(s_\circ,s_1,\ldots,s_p)$$

بنابراین، خروجی رگرسیون می‌تواند به صورت ذیر نشان داده شود:

$$\widetilde{y}=(a_\circ,s_\circ)+(a_1,s_1)x_1+(a_2,s_2)x_2+\cdots,(a_p,s_p)x_p \qquad (۸)$$

در نتیجه، تابع عضویت متغیر خروجی رگرسیون به صورت ذیر بدست می‌آید:

$$\mu_{\widetilde{y}}(y)=\begin{cases}\max\left(\min\left\{\widetilde{A}(x)\right\}\right) & \{x\mid y=f(x,a)=\phi\\ \circ & \text{در غیر اینصورت}\end{cases} \qquad (۹)$$

با جایگزینی رابطهٔ (۹) در رابطهٔ (۸) خواهیم داشت:

$$\mu_{\tilde{y}}(y)=\begin{cases}1-\dfrac{|\,y-\sum_{i=1}^{n}a_i x_i\,|}{\sum_{i=1}^{n}s_i\,|\,x_i\,|} & x_i\neq 0\\ 1 & x_i=0,\ y=0\\ 0 & x_i=0,\ y\neq 0\end{cases} \qquad (۱۰)$$

مثال: فرض کنید یک مدل رگرسیون فازی به شکل زیر باشد:

$$\tilde{Y}=\tilde{A}_1x_1+\tilde{A}_2x_2$$

پارامترهای مدل اعداد فازی مثلثی فرض شده و با پارامترهای a و s به شرح زیر تعریف می‌شوند:

$$\tilde{A}_1=(4,\ 2) \quad , \quad \tilde{A}_2=(5\ ,3)$$

تابع رگرسیون به صورت زیر است:

$$\tilde{Y}=(4,2)x_1+(5,3)x_2$$

در نتیجه، خروجی مدل $\tilde{Y}$ نیز یک عدد فازی مثلثی خواهد بود که تابع عضویت آن به صورت زیر به‌دست می‌آید:

$$\mu_{\tilde{y}}(y)=\begin{cases}1-\dfrac{|\,y-4x_1-5x_2\,|}{2\,|\,x_1\,|+3\,|\,x_2\,|} & x_1,x_2\neq 0\\ 1 & x_1=x_2=y=0\\ 0 & x_1=x_2=0,y\neq 0\end{cases}$$

در این صورت، پارامترهای a_y و s_y برای خروجی مدل رگرسیون فازی به شرح زیر است:

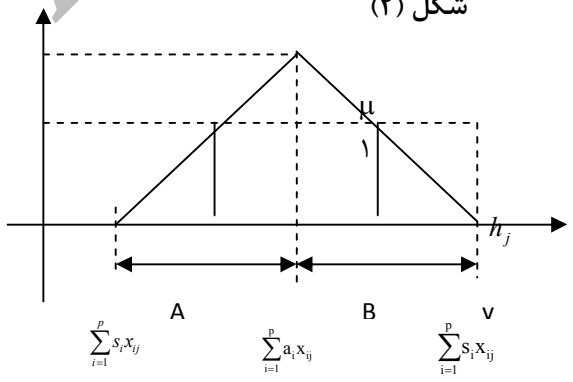
$$a_y=4x_1+5x_2$$
$$s_y=2\,|\,x_1\,|+3\,|\,x_2\,|$$

حالت داده‌های غیر فازی در رگرسیون می‌تواند تبدیل به یک مدل برنامه‌ریزی خطی شود. در این حالت، هدف مدل رگرسیون تعیین مقادیر بهینهٔ پارامترهای $\tilde{A}$ است، به قسمی که مجموعهٔ فازی خروجی مدل رگرسیون شامل (y_i) دارای درجهٔ عضویت بزرگ‌تر یا مساوی h باشد. یعنی:

$$\mu_{\tilde{y}_j}(y_j)\geq h_j \qquad j=1,2,\ldots,n \qquad (۱۱)$$

که n تعداد مشاهدات است و درجهٔ h را کاربر تعیین می‌کند. با افزایش h، میزان فازی‌بودن خروجی کاهش می‌یابد. شکل (۲) تابع عضویت خروجی مدل رگرسیون را نشان می‌دهد:

شکل (۲)



معادلهٔ (۱۱) نشان‌دهندهٔ این است که متغیر وابسته فازی می‌بایست بین دو مقدار A و B قرار گیرد (شکل ۲). در این مدل رگرسیونی، هدف یافتن ضرایب فازی است که گسترهٔ خروجی فازی را برای همهٔ مقادیر مجموعهٔ داده‌ها حداقل کند:

$$O = \min\left\{\sum_{i=1}^{p}\sum_{j=1}^{n} s_i x_{ij}\right\} \qquad (۱۲)$$

تابع هدف مدل در مواجهه با دو محدودیت زیر حداقل می‌شود:

$$y_j \geq \sum_{i=1}^{p} a_i x_{ij} - (1-h)\sum_{i=1}^{p} s_i x_{ij} \qquad (۱۳)$$

$$y_j \leq \sum_{i=1}^{p} a_i x_{ij} + (1-h)\sum_{i=1}^{p} s_i x_{ij} \qquad (۱۴)$$

از آنجا که برای هر مشاهده دو نوع محدودیت وجود دارد، تعداد کل محدودیت‌ها ۲n خواهد بود.

مثال: فرض کنید برای یک مدل رگرسیون با دو متغیر مستقل مقادیر زیر را داریم:

x_1	x_2	y
3	5	12
2	4	8
1	6	11

داده‌های مشاهده‌شده غیر فازی‌اند، اما فرض بر این است که ساختار مدل فازی است. با فرض این‌که پارامترهای مدل (a, s) مثلثی بوده پارامترهای مدل با حل مسئلهٔ برنامه‌ریزی خطی زیر به‌دست می‌آیند.

تابع هدف:

$$O = \min\{s_1(3+2+1) + s_2(5+4+6)\}$$

محدودیت‌ها:

$$12 \geq 3a_1 + 5a_2 - (1-h)(3s_1 + 5s_2)$$
$$12 \leq 3a_1 + 5a_2 + (1-h)(3s_1 + 5s_2)$$
$$8 \geq 2a_1 + 4a_2 - (1-h)(2s_1 + 4s_2)$$
$$8 \leq 2a_1 + 4a_2 + (1-h)(2s_1 + 4s_2)$$
$$11 \geq a_1 + 6a_2 - (1-h)(1c_1 + 6s_2)$$
$$11 \leq a_1 + 6a_2 + (1-h)(1c_1 + 6s_2)$$

با حل مسئلهٔ برنامه‌ریزی خطی فوق فرض می‌کنیم جواب‌های زیر به‌دست می‌آید (با فرض h=0.5)

$$a_1 = 1.04 \;,\; a_2 = 1.62 \;,\; s_1 = 0.53 \;,\; s_2 = 00$$

و همچنین، مقدار مینیمم شدهٔ تابع هدف یعنی O برابر 3.23 و

$$\widetilde{A}_1 = (1.04, 0.53) \;,\; \widetilde{A}_2 = (1.62, 0.00)$$

در این حالت، علاقه‌مندیم به‌ازای مقادیر متغیرهای مستقل، یعنی x_1, x_2 مقدار، y را پیش‌بینی کنیم، در این صورت،

$$x_2 = 6 \;,\; x_1 = 1$$

$$\widetilde{Y}^a(z) = 1.04 \times 1 + 1.62 \times 6 = 10.76$$
$$\widetilde{Y}^s(x) = 0.53 \times 1 + 0.00 \times 6 = 0.53$$

مقدار مشاهده‌شده، یعنی ۱۱، قابل مقایسه است. با تغییر h مقادیر a_iها تغییری پیدا نمی‌کنند، اما مقادیر s_i ها و o تغییر پیدا می‌کنند.

ارزیابی مدل برازش‌شده

مدل‌های رگرسیونی فازی با شاخص‌ها و آماره‌های گوناگونی ارزیابی می‌شوند. یکی از شاخص‌های ارزیابی رگرسیون فازی شاخص اطمینان (Index of Confidence) است که با فرمول زیر بیان می‌شود:

$$IC = 1 - \frac{SSE}{SST} \qquad (۱۵)$$

که SSE و SST عبارت‌اند از:

$$SSE = 2\sum_{i=1}^{n}\left(Y_i - \tilde{Y}_i^{a}\right)^2$$

$$SST = \sum_{i=1}^{n}\left(Y_i - \tilde{Y}_i^{L}\right)^2 + \sum_{i=1}^{n}\left(\tilde{Y}_i^{R} - Y_i\right)^2$$

معمولاً در انتخاب مدل hهای گوناگونی را در نظرگرفته و مدلی را انتخاب می‌کنیم که دارای IC کوچک باشد. البته میزان ابهام مدل را نیز باید در نظر گرفت.

مثال کاربردی: بررسی رابطه بین معدل کتبی دیپلم و نمرات ریاضی و فیزیک و جنسیت داوطلبان

از ۱۳۵۷ و بعد از انقلاب اسلامی، تغییراتی اساسی در سیستم پزیرش دانشجو در کشور رخ داده است. هدف از این تغییرات و اصلاحات ایجاد شانس برابر برای همهٔ داوطلبان و تسهیل ورود متقاضیان به دوره‌های آموزش عالی در کنار توجه به افزایش ظرفیت‌های این دوره ها بوده است. ظرفیت‌های مطلوب آموزش عالی طی ده سال اخیر تقریباً ثابت مانده است و افزایش ظرفیت‌ها بیشتر در دوره‌های پیام نور و غیرانتفاعی اتفاق افتاده است که از نظر داوطلبان دوره‌هایی چندان مطلوب نبوده‌اند. از طرف دیگر به‌رغم افزایش ظرفیت‌های دوره‌های مورد اشاره، تعداد متقاضیان ورود به آموزش عالی کاهش نیافته است. از جمله تعداد شرکت‌کنندگان در آزمون سراسری ۱۳۷۹ برابر ۱.۳۳۹.۱۳۴ و در ۱۳۸۷ برابر ۱.۵۰۶.۰۰۰، امّا ظرفیت‌های دورهٔ روزانهٔ دانشگاه‌ها طی ده سال گذشته تقریبا ثابت و به‌طور متوسط حدود ۸۰۰۰۰ نفر بوده است. بنابراین، داوطلبان برای ورود به دانشگاه‌ها و موسسات آموزش عالی به‌منظور ورود به مقاطع بالاتر، مخصوصاً دوره‌های روزانه، در رقابت بسیار فشرده‌ای درگیرند. رقابت فشردهٔ داوطلبان ورود به دانشگاه‌ها و مراکز آموزش عالی در یک آزمون چهار ساعته باعث ایجاد حساسیت، استرس زیاد داوطلبان و مشکلات اجتماعی متعددی در کشور شده است. این امر مراجع قانون‌گذاری، از قبیل مجلس شورای اسلامی و شورای انقلاب فرهنگی، را واداشته است که قوانین و ضوابط متعددی را به تصویب برسانند. آخرین قانون مربوط به آبان ۱۳۸۶ است. در این قانون مجلس شورای اسلامی مصوب کرد که از ۱۳۹۰، سوابق تحصیلی استاندارد دانش‌آموزان جایگزین آزمون سراسری شود. صرف‌نظر از امکان اجرای این قانون، در این بخش علاوه بر بحث و بررسی رگرسیون کلاسیک و فازی تلاش بر این است که به این سوال پاسخ داده شود:

آیا بین سوابق تحصیلی نهایی موجود دانش‌آموزان (معدل کتبی دیپلم یا سال سوم متوسطه) و نمرات آزمون سراسری رابطه‌ای وجود دارد یا خیر؟

در این بررسی، اطلاعات همهٔ شرکت‌کنندگان گروه آزمایشی علوم ریاضی و فنی در سهمیهٔ منطقهٔ یک در آزمون سراسری ۱۳۸۷، شامل نمرات هفت درس امتحانی (ادبیات فارسی، عربی، فرهنگ و معارف اسلامی، زبان خارجی، ریاضیات، فیزیک و مکانیک و شیمی)، معدل کتبی دیپلم و جنس داوطلبان به منزلهٔ جامعهٔ مورد بررسی تجزیه و تحلیل می‌شود. جامعه آماری عبارت است از اطلاعات همهٔ شرکت‌کنندگان آزمون سراسری سال۱۳۸۷ در گروه آزمایشی علوم ریاضی و فنی در منطقهٔ یک به تعداد ۱۰۳۹۳۶ نفر که از بین این افراد ۴۳۴۳ نفر دادهٔ گمشده و بقیه، در منطقهٔ یک حدود ۹۹.۵۹۳ نفر برای تجزیه و تحلیل استفاده شده است. در ادامه، با توجه به این‌که بحث و بررسی رگرسیون خطی فازی هدف اصلی این مقاله است، نمونه‌ای به حجم ۷۴ را با روش نمونه‌گیری تصادفی ساده و بدون جایگذاری با خطای ۸ درصد استخراج و با دو روش رگرسیون کلاسیک و فازی بررسی و

تجزیه و تحلیل می‌کنیم. نرم افزار مورد استفاده برای رگرسیون کلاسیک SPSS و برای رگرسیون فازی Lingo11.0 و Excel است. این بخش به سه قسمت زیر تقسیم می‌شود:

الف) رگرسیون کلاسیک چندگانهٔ معدل روی نمرات و جنس در جامعهٔ آماری

ب) رگرسیون کلاسیک چندگانهٔ معدل روی نمرات و جنس در نمونهٔ آماری

ج) رگرسیون فازی چندگانهٔ معدل روی نمرات و جنس در نمونهٔ آماری

الف) رگرسیون کلاسیک چندگانهٔ معدل روی نمرات و جنس در جامعهٔ آماری

پس از اجرای رگرسیون چندگانهٔ گام‌به‌گام معدل کتبی دیپلم (متغیر وابسته) روی نمرات هفت درس امتحانی (ادبیات فارسی، عربی، فرهنگ و معارف اسلامی، زبان خارجی، ریاضیات، فیزیک و مکانیک و شیمی) و جنس همهٔ داوطلبان منطقه یک گروه آزمایشی علوم ریاضی و فنی (جامعه آماری شامل ۹۹.۵۹۳ نفر، ۴۴.۵ درصد زن و ۵۵.۵ درصد مرد) مدل زیر به‌دست می‌آید.

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta	B	Std. Error
1	ضریب ثابت a	۱۲۴۰.۰۱۷	۱.۰۱۲		۱۲۲۵.۶۷۱	۰.۰۰۰
	جنسjnc	۱۳۵.۷۴۵	۱.۳۰۳	۰.۲۴۴	۱۰۴.۲۱۲	۰.۰۰۰
	ریاضیng5	۰.۰۶۷۴	۰.۰۰۸	۰.۰۳۷۰	۸۸.۱۱۱	۰.۰۰۰
	فیزیکng6	۰.۰۴۸۲	۰.۰۰۷	۰.۲۹۸	۷۱.۰۶۵	۰.۰۰۰

چنان‌که ملاحظه می‌شود، سایر متغیرهای مستقل به علت عدم معنی‌داربودن از مدل حذف شده‌اند. آزمون‌های آماری از قبیل F و نرمال‌بودن باقیمانده‌ها با آزمون کلموگرف و اسمیرینف در سطح ۰.۰۵ تایید شده‌اند. ضریب تبیین یا R^2 برابر ۰.۴۵۳ است. به عبارت دیگر، ۴۵.۳ تغییرات متغیر معدل کتبی دیپلم با متغیرهای جنسیت و نمرات ریاضی و فیزیک داوطلبان تبیین می‌شود.

ب) رگرسیون کلاسیک چندگانهٔ معدل روی نمرات و جنس در نمونهٔ آماری

در این مرحله، یک نمونه به حجم ۷۵ به روش نمونه‌گیری تصادفی ساده و بدون جایگذاری با خطای ۸ درصد استخراج می‌شود. پس از اجرای رگرسیون چندگانه گام‌به‌گام معدل کتبی دیپلم (متغیر وابسته) روی نمرات دروس امتحانی ریاضیات، فیزیک و جنس روی نمونهٔ استخراج شده (نمونهٔ آماری شامل ۱۰۰ نفر،۴۴ درصد زن و ۵۶ درصد مرد) مدل زیر به‌دست می‌آید.

Sig.	t	Standardized Coefficients	Unstandardized Coefficients		Model	
Std. Error	B	Beta	Std. Error	B		
۰.۰۰۰	۳۳.۳۰۲		۳۷.۸۲۰	۱۲۵۹.۴۶۴	ضریب ثابت	1
۰.۰۵۲	۱.۸۹۹	۰.۱۸۳	۴۹.۴۳۵	۹۳.۸۵۷	jncجنسیت	
۰.۰۰۴	۳.۰۰۲	۰.۴۳۷	۰.۲۳۳	۰.۶۹۹	ng5ریاضی	
۰.۰۴۰	۱.۸۲۸	۰.۲۶۳	۰.۲۰۲	۰.۳۷۰	ng6فیزیک	

ضریب تبیین یا R^2 برابر ۰.۴۱ است. به عبارت دیگر، ۴۱ تغییرات متغیر معدل کتبی دیپلم با متغیرهای جنسیت و نمرات ریاضی و فیزیک داوطلبان در نمونه استخراج‌شده در رگرسیون تبیین می‌شود. همچنین، آزمون F یعنی معنی‌داری کل رگرسیون برازش شده نیز معنی‌دار است. نکتهٔ مهم عدم نرمال‌بودن باقیمانده‌ها با آزمون کلموگرف و اسمیرینف در سطح 0.05 است که با توجه به کم‌بودن حجم نمونه است و برای حل این مشکل رگرسیون فازی پیشنهاد می‌شود که در بخش بعد به آن خواهیم پرداخت. با توجه به جداول بالا نتایج به‌دست آمده در نمونه‌ای با خطای ۸ درصد قابل تعمیم به جامعه است.

ج) رگرسیون فازی چندگانهٔ معدل روی نمرات و جنس در نمونهٔ آماری

تئوری مجموعه‌ها و رگرسیون خطی فازی در بخش سوم و چهارم عرضه شده‌اند. در این بخش می‌خواهیم رگرسیون خطی فازی را با فرض ابهام در ضرایب درخصوص معدل کتبی دیپلم (متغیر وابسته) روی متغیرهای مستقل نمرات ریاضی، فیزیک و جنسیت داوطلبان به‌دست بیاوریم. تعداد نمونه‌ها برابر ۷۴ و متغیرهای جنس، نمرهٔ ریاضی و فیزیک را به ترتیب با JNC، NG5 و NG6 نشان می‌دهیم. برای پیدا کردن تابع هدف (۱۲) داریم: $\sum \mathrm{JNC} = 27$، $\sum \mathrm{NG5} = 9212$ و $\sum \mathrm{NG6} = 9295$. بنابراین:

$$O = 148 + 27 \times S_1 + 9212 \times S_2 + 9295 \times S_3$$

می‌خواهیم تابع O را با فرض قیدهای معادله (۱۳) و (۱۴) که در مجموع ۱۴۸ قید است حداقل کنیم. پس از تشکیل قیدها و حل معادلهٔ خطی بالا با توجه به hهای گوناگون، مقادیر $s_3, s_2, s_1, s_\circ, a_3, a_2, a_1, a_\circ$ براساس جدول شماره ۱ با استفاده از نرم‌افزار Lingo11.0 چنین به‌دست می‌آید:

جدول۱. نتایج حاصل از برازش توابع معدل، نمرات و جنس

h	$a_\circ$	$s_\circ$	a_1	s_1	a_2	s_2	a_3	s_3
۰.۲	۱۱۸۳	۳۰۶	۲۵۴	۱۵۷	۰.۵۴	۱.۳	۰.۳۳	۰
۰.۴	۱۱۸۳	۴۰۸	۲۵۴	۲۱۰	۰.۵۴	۱.۷	۰.۳۳	۰
۰.۵	۱۱۸۳	۴۸۹	۲۵۴	۲۵۲	۰.۵۴	۲.۰	۰.۳۳	۰
۰.۶	۱۱۸۳	۶۱۲	۲۵۴	۳۱۵	۰.۵۴	۲.۵	۰.۳۳	۰
۰.۸	۱۱۸۳	۱۲۲۴	۲۵۴	۶۳۰	۰.۵۴	۵	۰.۳۳	۰

پس از پیدا کردن مقادیر a_i و s_iها، بایستی مقادیر مرکزی(y^a) و پراکندگی داده‌ها ($y_{s0.8}$ ، $y_{s0.5}$ ، $y_{s0.4}$ ، $y_{s0.2}$ ، $y_{s0.6}$) را مانند مثال بخش قبلی به‌دست بیاوریم که معادلات زیر خواهند بود:

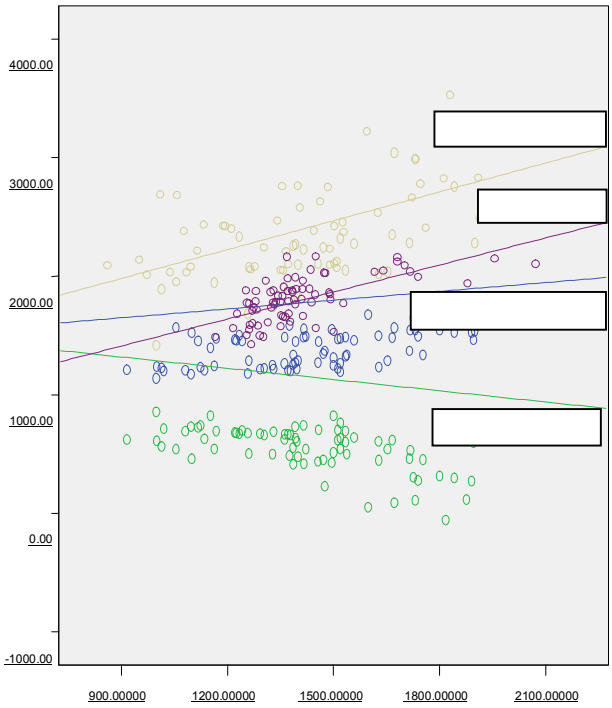
$$y^a = 1183.517 + 254.1639 * jnc + 0.5359716 * ng5 + 0.334554 * ng6.$$
$$y_{s0.2} = 306.2444 + 157.5180 * jnc + 1.269992 * ng5 + 0.000000 * ng6.$$
$$y_{s0.4} = 408.3258 + 210.0240 * jnc + 1.693323 * ng5 + 0.000000 * ng6.$$
$$y_{s0.5} = 489.9910 + 252.0288 * jnc + 2.031987 * ng5 + 0.000000 * ng6.$$
$$y_{s0.6} = 612.4887 + 315.0360 * jnc + 2.539984 * ng5 + 0.000000 * ng6.$$
$$y_{s0.8} = 1224.977 + 630.0720 * jnc + 5.079968 * ng5 + 0.000000 * ng6.$$

پس از پیداکردن کران‌های بالا و پایین برای hهای متفاوت، شاخص اطمینان IC با استفاده از فرمول ۱۵ به شکل جدول شماره ۲ بدست می‌آید:

جدول ۲. مقادیر شاخص اطمینان و ابهام کل برای hهای متفاوت

h	$\frac{SSE}{SST}$	IC	ابهام کل o
۰.۲	۰.۱۳	۰.۸۷	۶۱۲۷۶
۰.۴	۰.۰۷	۰.۹۳	۸۱۷۰۱
۰.۵	۰.۰۵	۰.۹۵	۹۸۰۴۲
۰.۶	۰.۰۴	۰.۹۶	۱۲۲۵۵۲
۰.۸	۰.۰۱	۰.۹۹	۲۴۵۱۰۵

نمودار ۳. نمودار رگرسیون فازی پیش‌بینی‌کننده (شامل تابع کران بالا و پایین و مرکزی) برای h=0.5



به منظور ارزیابی مدل‌های حاصل از رگرسیون فازی، شاخص اطمینان و مقدار ابهام کل مدل برای hهای متفاوت محاسبه و در جدول شماره ۲ آمده است. چنان‌که ملاحظه می‌شود، با بزرگ شدن مقدار h سطح اعتبار مدل بالا می‌رود و در عین حال، این امر باعث افزایش ابهام کل مدل نیز می‌شود. لذا برای انتخاب h صحیح به میزان ابهام کل نیز باید توجه داشت. البته می‌توان سطح اعتبار h=0.5 را به منزلهٔ سطح اعتباری معقول و متداول در نظر گرفت. در داده‌های مورد بررسی سطح اعتبار ۰.۵ به سبب IC بالا و افزایش اندک ابهام مدل معقول به نظر می‌رسد. ضریب همبستگی بین پیش‌بینی‌کننده‌های رگرسیون کلاسیک و فازی مرکزی ۰.۹ است که نشان‌دهندهٔ نزدیک‌بودن نتایج دو روش است. اما رگرسیون فازی این امتیاز را دارد که فاقد فرض‌های محدودکنندهٔ رگرسیون کلاسیک است، مخصوصا در این نمونه که توزیع باقیمانده‌ها نیز نرمال نیست. علاوه بر این، چنان‌که در نمودار ۳ ملاحظه می‌شود، رگرسیون کلاسیک برازش‌شده در داخل فاصلهٔ اطمینان رگرسیون فازی برازش‌شده با h=0.5 قرار می‌گیرد. بنابراین، به نظر می‌رسد در نمونه‌های کوچک، به علت عدم پایایی فرض‌های رگرسیون کلاسیک، به‌کار بردن رگرسیون خطی فازی معقول‌تر است.

خلاصه و نتیجه گیری

در این مقاله، روش‌های رگرسیون کلاسیک و فازی با تاکید بر بعد آموزشی برای علاقه‌مندان رشته‌های علوم اجتماعی به بحث و بررسی گذاشته شد و یک مثال کاربردی در این حوزه به تفضیل تجزیه و تحلیل شد. بایستی توجه داشت که در کاربرد روش‌های فازی در علوم اجتماعی و حتی سایر علوم با توجه به ماهیت علم فازی، مرزبندی‌های تصمیم‌گیری به روشنی روش‌های کلاسیک نیست و در اکثر موارد محقق است که باید تصمیم بگیرد. در این خصوص می‌توان به انتخاب سطح فازی‌بودن یا به‌کاربردن اعداد فازی غیر مثلثی، مانند اعداد فازی نرمال، اشاره کرد.

منابع

آذر، عادل و فرجی، حجت (۱۳۸۶) *علم مدیریت فازی*، تهران: دانشگاه تربیت مدرس.

امینی، امیرمظفر و مهدی خیاطی (۱۳۸۵) "عوامل مؤثر بر عدم موفقیت طرح تشکیل تعاونی‌های آب ایران (استفاده از رگرسیون فازی)" *اقتصاد کشاورزی و توسعه*، شماره ۵۳، ۱۴ (۱): ۶۹-۹۱.

ایمانی، محمدتقی (۱۳۷۶) "ابهام روش‌شناسی و تنگناه‌های پژوهش در ایران"، در مجموعه *مقالات نخستین سمینار آموزش عالی ایران*، تهران: انتشارات دانشگاه علامه طباطبایی.

بارت کاسکو (۱۳۸۰) *تفکر فازی*، ترجمهٔ علی غفاری، عادل مقصودپور، علیرضا پورممتاز و جمشید قسیمی، تهران: دانشگاه خواجه نصیرالدین طوسی.

تیموری، حبیب‌الله (۱۳۸۶) *آمار در علوم اجتماعی*، تهران: کیان مهر.

طاهری، محمود (۱۳۸۷) *مقدمه‌ای بر احتمال و آمار فازی*، کرمان: دانشگاه شهید باهنر کرمان.

لی وانگ (۱۳۷۸) *سیستم‌های فازی و کنترل فازی*، ترجمهٔ محمد تشنه لب، نیما صفارپور، داریوش افیونی، تهران: دانشگاه خواجه نصیرالدین طوسی.

محمدی، جهانگرد؛ طاهری، محمود (۱۳۸۴) "برازش توابع انتقالی خاک با استفاده از رگرسیون فازی"، در *علوم و فنون کشاورزی و منابع طبیعی*، ۹ (۲): ۵۱-۶۱.

Arnold F. Shapiro (2004) *Fuzzy Regression Models*, Pennsylvania: Pen State University.

Chang Ping-Teng Kan (1997) "Fuzzy Regression Analysis" in *Fuzzy Sets and Systems*, volume 90.

E. Pasha, T. Razzaghnia1, T. Allahviranloo, Gh. Yari, H. R. Mostafaei, (2007) "Fuzzy Linear Regression Models with Fuzzy Entropy", *Applied Mathematical Sciences*, 1 (35): 1715 – 1724.

Galton, F. (1885) "Regression toward Mediocrity in Heredity Stature" in *Journal of the Anthropological Institute*, : 15: 246-263.

Modarres. M, E. Nasrabdi and M. M Nasrabadi (2004). "Fuzzy linear Regression Analysis from the Risk Point of view", *International Journal of Uncertainty*, *Fuuiness and Know ledg-Based Systems*, 12 (5): 635-649.

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

Liang.W and Chen. Y (2004) "Estimation of Weibull Parametersun Sing a Fuzzy Least Squares", *International Journal of Uncertainty, Fuzzniss and Know ledge-Based Systems*, 12 (5): 701-711

Zadeh, L. A. (1965) *Fuzzy sets, Inform and control*, 8: 338-353.