

الحمد لله
الذي هدانا لهذا
الذي كنا لا ندر



دانشکده علوم ریاضی

رشته آمار ، گرایش آمار ریاضی

پایان نامه کارشناسی ارشد

مطالعه‌ی یک روش رگرسیون استوار در محیط فازی

نگارنده: الهم رحیمیان

استادان راهنما

دکتر محمد رضا ربیعی
دکتر داود شاهسونی

شهریور ۱۳۹۶

شماره:

تاریخ:

باسمه تعالی



مدیریت تحصیلات تکمیلی

فرم شماره (۳) صورتجلسه نهایی دفاع از پایان نامه دوره کارشناسی ارشد

با نام و یاد خداوند متعال، ارزیابی جلسه دفاع از پایان نامه کارشناسی ارشد خانم الهام رحیمیان با شماره دانشجویی: ۹۳۰۸۰۹۴ رشته: آمار گرایش: آمار ریاضی تحت عنوان: مطالعه یک روش رگرسیون استوار در محیط فازی که در تاریخ ۹۶/۰۶/۱۳ با حضور هیأت محترم داوران در دانشگاه صنعتی شاهرود برگزار گردید به شرح ذیل اعلام می گردد:

قبول (با امتیاز) درجه ☒	مردود ☐
نوع تحقیق: نظری ☒ عملی ☐	

عضو هیأت داوران	نام و نام خانوادگی	مرتبه علمی	امضاء
۱- استاد راهنمای اول	دکتر محمد رضا ربیعی	استادیار	
۲- استاد راهنمای دوم	دکتر داود شاهسونی	دانشیار	
۳- استاد مشاور			
۴- نماینده تحصیلات تکمیلی	دکتر میثم علیشاهی	استادیار	
۵- استاد ممتحن اول	دکتر محمد آرشی	دانشیار	
۶- استاد ممتحن دوم	دکتر جواد قاسمیان	استادیار	

نام و نام خانوادگی رئیس دانشکده: ابراهیم هاشمی
تاریخ و امضاء و مهر دانشکده: ۹۶/۰۶/۱۳

تبصره: در صورتی که کسی مردود شود حداکثر یکبار دیگر (در مدت مجاز تحصیل) می تواند از پایان نامه خود دفاع نماید (دفاع مجدد نباید زودتر از ۴ ماه برگزار شود).

تقدیم به پدر و مادر عزیزتر از جانم؛

خدای را سپاس که از روی کرم پدر و مادری فداکار نصیصم ساخته تا در سایه‌ی درخت
پر بار وجودشان بیاسیم و از ریشه‌ی آن ها شاخ و برگ گیرم و در سایه‌ی وجودشان
مسیر کسب علم و دانش را بیامیم.

والدینی که نامشان تاج افتخاری است بر سرم.

آموزگارانی که برایم زندگی و انسان بودن را معنا کردند.

حال این مجموعه برگ سبزی است تحفه‌ی درویش تقدیم آنان ...

به پاس محبت‌های بی‌دینشان که هرگز فروکش نمی‌کند.

سپاس

خدایی که آفرید
جهان را،

انسان را،

عقل را،

معرفت را،

عشق را

و کسانی که عشقان را در وجودم دمید.

الهام رحیمیان
شهریور ۱۳۹۶

تعهد نامه

اینجانب **الهام رحیمیان** دانشجوی کارشناسی ارشد رشته آمار دانشکده علوم ریاضی دانشگاه شاهرود، نویسنده پایان نامه با عنوان **مطالعه یک روش رگرسیون استوار در محیط فازی**، تحت راهنمایی **محمد رضا ربیعی و داود شاهسونی** متعهد می شوم:

- تحقیقات در این پایان نامه توسط اینجانب انجام شده است و از صحت و اصالت برخوردار است.
- در استفاده از نتایج پژوهش های دیگر پژوهش گران، به مرجع مورد استفاده استناد شده است.
- مطالب این پایان نامه، تا کنون توسط خود، یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ جا ارایه نشده است.
- حقوق معنوی این اثر، به دانشگاه صنعتی شاهرود تعلق دارد، و مقالات مستخرج با نام “دانشگاه صنعتی شاهرود” یا “Shahrood University of Technology” به چاپ خواهد رسید.
- حقوق معنوی تمام افرادی که در به دست آوردن نتایج اصلی پایان نامه تاثیرگذار بوده اند، در مقالات مستخرج از پایان نامه رعایت می گردد.
- در تمام مراحل انجام این پایان نامه، در مواردی که از موجود زنده (یا بافت های آنها) استفاده شده است، ضوابط و اصول اخلاقی رعایت شده است.
- در تمام مراحل انجام این پایان نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته (یا استفاده شده است)، اصل رازداری و اصول اخلاق انسانی رعایت شده است.

الهام رحیمیان

شهریور ۱۳۹۶

مالکیت نتایج و حق نشر

- تمام حقوق معنوی این اثر و محصولات آن (مقالات مستخرج، کتاب، برنامه های رایانه ای، نرم افزارها و تجهیزات ساخته شده) متعلق به دانشگاه صنعتی شاهرود می باشد. این مطلب باید به نحو مقتضی، در تولیدات علمی مربوطه ذکر شود.
- استفاده از اطلاعات و نتایج موجود در این پایان نامه بدون ذکر منبع مجاز نمی باشد.

چکیده

اگر مفروضات روش کمترین توان‌های دوم برای یک مجموعه داده برقرار باشد، برآوردهای این روش به‌عنوان بهترین برآوردها شناخته می‌شوند اما اگر مشاهدات دورافتاده در مجموعه داده‌ها وجود داشته باشند روش‌های متعارف پیش‌بینی‌های خوبی را نتیجه نمی‌دهد و باید از روش‌های رگرسیون جایگزین استوار استفاده کرد. از طرفی 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. زمانی که هم مشاهدات، فازی هستند و هم در مجموعه داده‌ها مشاهدات دورافتاده وجود دارند باید از روش‌های جایگزین استوار فازی استفاده کرد. پس از معرفی رگرسیون خطی فازی امکانی تحلیل رگرسیون فازی به‌طور گسترده‌ای مورد مطالعه و بررسی قرار گرفت و در زمینه‌های مختلفی به کار رفت. همچنین، روش کمترین توان‌های دوم در محیط فازی تعمیم داده شد. این روش‌ها به حضور داده‌های دورافتاده حساس هستند.

در این پایان نامه، در حالتی که متغیرهای مستقل، دقیق و متغیر وابسته یک عدد فازی بوده و در مجموعه داده‌ها مشاهدات دورافتاده حضور دارند، یک تحلیل رگرسیون استوار فازی را به دو روش مطرح کرده و نتایج را مقایسه می‌کنیم. روش اول، تحلیل رگرسیون کمترین توان‌های دوم فازی اصلاح شده و روش دوم، تحلیل رگرسیون کمترین قدر مطلق انحرافات اصلاح شده می‌باشند. در هر دو روش، باقیمانده‌ها رتبه‌بندی می‌شوند و ماتریس‌های وزن توسط توابع عضویت تعریف شده برای باقیمانده‌ها تشکیل می‌شوند. سپس برآوردهای کمترین توان‌های دوم موزون فازی و کمترین توان‌های قدر مطلق موزون فازی با استفاده از این ماتریس وزن به‌دست می‌آیند. نتایج حاصل از مثال‌ها، برتری روش دوم به روش اول را بیان می‌کنند.

کلمات کلیدی: رگرسیون استوار، داده‌های دورافتاده، رگرسیون فازی، تابع عضویت، رتبه‌بندی مجموعه‌های فازی.

لیست مقالات مستخرج از پایان نامه

۱. رحیمیان ا.، ربیعی م. و شاهسونی د.، (۱۳۹۵)، تحلیل رگرسیون فازی با داده‌های خروجی و پارامترهای فازی بر پایه‌ی رتبه‌بندی مجموعه‌های فازی، ششمین کارگاه کشوری آمار و احتمال فازی، شاهرود، ایران.
۲. رحیمیان ا.، ربیعی م. و شاهسونی د.، (۱۳۹۶)، تحلیل رگرسیون استوار فازی با داده‌های خروجی و پارامترهای فازی بر پایه‌ی رتبه‌بندی مجموعه‌های فازی، مجله اندیشه آماری ، در دست چاپ.
۳. ربیعی م. ، رحیمیان ا. و شاهسونی د.، (۱۳۹۶)، تحلیل رگرسیون فازی بر اساس کمترین قدر مطلق خطاها و رتبه‌بندی مجموعه‌های فازی ، هفتمین سمینار کشوری آمار و احتمال فازی، بیرجند، ایران.

فهرست مطالب

۳	فهرست تصاویر
---	--------------

۵	فهرست جداول
---	-------------

۱	۱ مفاهیم و تعاریف اولیه
۱	۱.۱ مجموعه‌های فازی
۶	۲.۱ اعداد فازی
۹	۳.۱ حساب اعداد فازی
۱۱	۴.۱ فاصله بین دو عدد فازی
۱۱	۱.۴.۱ فاصله موزون
۱۳	۲.۴.۱ فاصله دونسو
۱۳	۳.۴.۱ فاصله دیاموند
۱۳	۴.۴.۱ فاصله همینگ
۱۴	۵.۴.۱ فاصله حسن‌پور
۱۴	۵.۱ رتبه‌بندی مجموعه‌های فازی
۱۶	۱.۵.۱ تابع عضویت در برابر تکیه‌گاه
۱۷	۲.۵.۱ مفهوم حضور سراسری
۱۹	۳.۵.۱ مجموعه‌های فازی نرمال و غیر نرمال
۲۱	۴.۵.۱ یک روش رتبه‌بندی فازی با مداخله تصمیم‌گیرنده

۲۷	۲ رگرسیون استوار
۲۹	۵.۰.۲ محاسبه‌ی برآورد کمترین توان‌های دوم
۲۹	۱.۲ محدودیت‌های برآوردگر کمترین توان‌های دوم
۳۰	۱.۱.۲ دسته‌بندی داده‌های دورافتاده
۳۲	۲.۱.۲ حذف داده‌های دورافتاده قبل از تحلیل
۳۳	۲.۲ برآوردگرهای رگرسیون استوار

۳۵		روش کمترین توان‌های دوم موزون	۱.۲.۲
۳۶		M - برآوردگرها	۲.۲.۲
۴۱		۳ تحلیل رگرسیون در محیط فازی	
۴۲		۱.۳ انواع رگرسیون فازی	
۴۲		۱.۱.۳ رگرسیون کمترین توان‌های دوم فازی	
۴۷		۲.۱.۳ نیکویی برازش	
۴۹		۴ تحلیل رگرسیون استوار فازی بر پایه‌ی رتبه‌بندی مجموعه‌های فازی	
۴۹		۱.۴ مقدمه	
۵۱		۲.۴ رگرسیون کمترین مجموع فاصله‌های خطای فازی	
		۱.۲.۴ الگوریتم کولا و آپایدین برای به‌دست آوردن ضرایب رگرسیون	
۵۲		استوار فازی	
۵۴		۲.۲.۴ مثال‌های عددی	
		۳.۲.۴ الگوریتم پیشنهادی برای به‌دست آوردن ضرایب رگرسیون استوار	
۵۸		فازی	
۵۹		۴.۲.۴ مثال کاربردی	
۶۳		آ برنامه‌های نوشته شده در متمتیکا	
۷۵		مراجع	
۸۱		واژه‌نامه فارسی به انگلیسی	
۸۳		واژه‌نامه انگلیسی به فارسی	

فهرست تصاویر

۱.۱	پروفسور زاده، دانشمند ایرانی تبار و مبدع نظریه فازی (تولد: ۱۵ بهمن ۱۲۹۹ برابر با ۴ فوریه ۱۹۲۱ – وفات: ۱۶ شهریور ۱۳۹۶ برابر با ۶ سپتامبر ۲۰۱۷)	۲
۲.۱	h – برش مجموعه فازی N	۵
۳.۱	h – برش عدد فازی N	۷
۴.۱	نمایش عدد فازی LR	۸
۵.۱	h – برش‌های دو عدد فازی	۱۲
۶.۱	نمایش فاصله‌های دونسو، دیاموند و حسن‌پور	۱۵
۷.۱	مقایسه‌ی دو مجموعه فازی A و B	۱۶
۸.۱	روش آدامو ($h = ۰/۹$)	۱۸
۹.۱	شاخص F_3 یاجر	۱۹
۱۰.۱	شاخص $d(A)$ کافمن	۲۰
۱۱.۱	الف. اختلاف فازی و ب. ماکزیمم فازی	۲۱
۱۲.۱	نمودار دو عدد فازی تقریباً m با مشخصات $A_1 = (m_{A_1}, l_{A_1}, r_{A_1})_T$ و $A_2 = (m_{A_2}, l_{A_2}, r_{A_2})_T$	۲۲
۱۳.۱	شرحی از وزن‌دهی χ – ثابت: $[\chi_1 x'(h) + \chi_2 x''(h)]$	۲۴
۱۴.۱	نمایش رابطه‌ی بین دو عدد فازی مثلثی A و B	۲۴
۱.۲	خطوط رگرسیون کمترین توان‌های دوم در حضور داده‌های دورافتاده	۳۱
۲.۲	توابع ρ و Ψ هوبر، همپل، اندروز وتوکی	۴۰

فهرست جداول

۴۶	داده‌های مربوط به مثال ۱.۳	۱.۳
	مجموعه داده‌ها، وزن‌ها (W_i) و مقایسه مدل FLS و مدل پیشنهادی با	۱.۴
۵۵	توجه به شاخص MPE برای داده‌های شبیه سازی شده	
۵۶	برآورد مدل رگرسیونی داده‌های شبیه سازی شده	۲.۴
	مجموعه داده‌ها، وزن‌ها (W_i) و مقایسه مدل FLS و مدل کولا و آپایدین با	۳.۴
۵۷	توجه به شاخص برای داده‌های مثال سیمان پورتلند	
۵۷	برآورد مدل رگرسیونی داده‌های مثال سیمان پورتلند	۴.۴
	مجموعه داده‌های سیمان پورتلند و مقایسه چند روش مدل سازی رگرسیونی	۵.۴
۵۹	فازی بر اساس وزن‌ها و شاخص MPE	
۶۰	برآورد مدل رگرسیونی داده‌های مثال سیمان پورتلند به شیوه های مختلف	۶.۴

فصل ۱

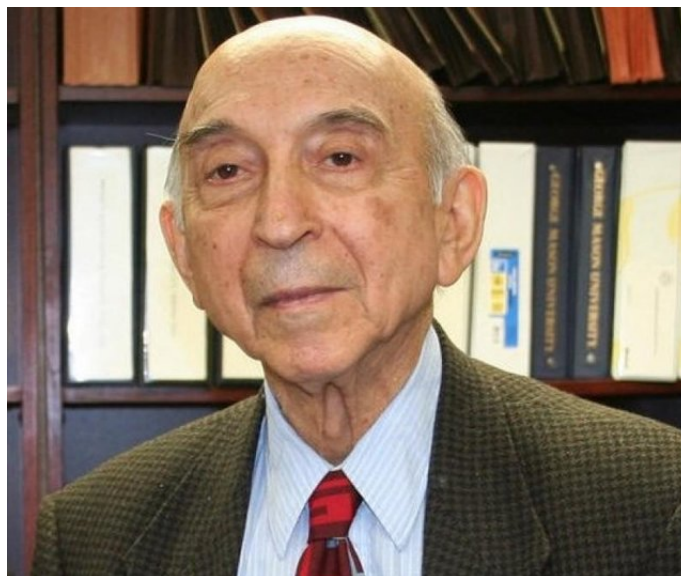
مفاهیم و تعاریف اولیه

مقدمه

در این فصل، به بیان برخی مفاهیم اساسی که در فصل‌های آینده به آن نیاز داریم می‌پردازیم. ابتدا مفاهیمی مربوط به مجموعه‌های فازی، اعداد فازی و فواصل بین اعداد فازی را بیان کرده و سپس به توضیح مطالبی در زمینه‌ی رتبه‌بندی مجموعه‌های فازی خواهیم پرداخت. این موضوع که یکی از روش‌های مورد استفاده در مواجهه با داده‌های دورافتاده در مجموعه داده‌هاست که از آن در رگرسیون استوار استفاده می‌شود.

۱.۱ مجموعه‌های فازی

نظریه مجموعه‌های فازی در سال ۱۹۶۵ توسط پروفسور لطفی عسکر زاده دانشمند ایرانی تبار ارائه شد [۷۲]. این نظریه از زمان ارائه آن تا کنون، گسترش و تعمیق زیادی یافته و کاربردهای گوناگونی در زمینه‌های مختلف پیدا کرده است. در نظریه مجموعه‌ها، یک مجموعه معمولی با یک ویژگی دقیق و معین مشخص می‌شود. بدین صورت که اگر یک عضو مجموعه مرجع، آن ویژگی را دارا باشد عضو مجموعه مورد نظر است و اگر فاقد آن ویژگی باشد، عضو آن مجموعه نیست. برای مثال، ویژگی اول بودن در اعداد طبیعی، یک ویژگی خوش تعریف است. به این



شکل ۱.۱: پروفیسور زاده، دانشمند ایرانی تبار و مبدع نظریه فازی (تولد: ۱۵ بهمن ۱۲۹۹ برابر با ۴ فوریه ۱۹۲۱ – وفات: ۱۶ شهریور ۱۳۹۶ برابر با ۶ سپتامبر ۲۰۱۷)

صورت که برای هر عدد طبیعی می‌توان گفت که یا یک عدد اول است یا یک عدد اول نیست. با فرض این که X یک مجموعه مرجع باشد، برای هر زیرمجموعه معمولی A از مجموعه مرجع که با یک ویژگی خوش تعریف مشخص می‌شود، می‌توان یک تابع نشان‌گر تعریف کرد که برای هر عضو متعلق به A مقدار یک و برای هر عضو غیر متعلق به A ، مقدار صفر بگیرد، یعنی

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

اما نکته مهم این است که اکثر مفاهیمی که انسان با آن‌ها مواجه می‌شود و بر اساس آن‌ها زندگی روزمره خود را پیش می‌برد، مفاهیمی دقیق و خوش تعریف نیستند، مثلاً هوای خیلی سرد، لامپ با طول عمر زیاد، مجموعه اعداد بزرگ، افراد قد بلند و غیره. برای این ویژگی‌ها مرز مشخص و دقیقی را نمی‌توان تعیین کرد. برای مثال فردی ممکن است ۱۸۰ سانتی‌متر را قدی بلند بداند در حالی که دیگری ممکن است آن را نسبتاً بلندفرض کند.

نظریه مجموعه فازی، یک قالب جدید ریاضی به منظور به الگو درآوردن این مفاهیم و ویژگی‌ها و تحلیل آن‌ها ارائه می‌دهد. این نظریه یک تعمیم طبیعی از نظریه مجموعه‌های قطعی است. برای توضیح مفهوم مجموعه فازی، به یک مثال توجه کنید. فرض کنید می‌خواهیم مجموعه اعداد بزرگ را تعریف کنیم. طبیعی است که بزرگ بودن یک ویژگی خوش تعریف نیست. فرض می‌شود که هر x در مجموعه اعداد حقیقی، به اندازه یک مقدار بین صفر و یک، عضو مجموعه اعداد بزرگ باشد، به‌طوری‌که هر چه x بزرگ‌تر باشد، عدد مورد نظر برای

عضویت آن در مجموعه اعداد بزرگ، به یک نزدیک‌تر باشد و بالعکس هرچه آن x کوچک‌تر باشد، عدد مربوط به عضویت آن در مجموعه اعداد بزرگ، به صفر نزدیک‌تر باشد. در این صورت به جای این که گفته شود یک عدد حقیقی بزرگ است یا بزرگ نیست، گفته می‌شود این عدد به اندازه مثلاً 0.9 عضو مجموعه اعداد بزرگ است. بنابراین اساس کار نظریه مجموعه‌های فازی، گسترش مفهوم تابع نشان‌گر یک مجموعه، با برد $[0, 1]$ به جای $\{0, 1\}$ است.

تعریف ۱.۱. یک زیرمجموعه فازی N ^۱ از مجموعه مرجع X توسط یک تابع عضویت $\mu_N(x)$: $X \rightarrow [0, 1]$ مشخص می‌شود که در آن برای هر $x \in X$ ، $\mu_N(x)$ مجموعه فازی N است. نزدیکی این عدد به یک، نشان‌دهنده تعلق بیشتر x به مجموعه فازی N و نزدیکی آن به صفر، نشان‌دهنده تعلق کمتر x به N است.

توجه کنید که اگر x کاملاً در N عضو باشد، $\mu_N(x) = 1$ و اگر اصلاً در N عضو نباشد، $\mu_N(x) = 0$ است؛ لذا مجموعه‌های معمولی و توابع نشان‌گر آن‌ها حالت خاصی از مجموعه‌های فازی و توابع عضویت آن‌ها هستند. برای نشان دادن یک مجموعه فازی روش‌های مختلفی رایج است. یک روش به کار بردن مستقیم تابع عضویت مجموعه فازی است.

مثال ۱.۱. فرض کنید $X = \{1, 2, \dots, 7\}$. می‌خواهیم یک مجموعه فازی تعریف کنیم که اعضای آن ویژگی نادقیق کوچک را داشته باشند. برای مدل‌سازی این مجموعه کافی است تابع عضویت این مجموعه فازی مشخص شود. تعیین این تابع، بستگی به نظر تصمیم‌گیرنده دارد. مثلاً یک تابع عضویت می‌تواند به صورت زیر تعریف شود

$$\mu_N(x) = \begin{cases} 1 & x = 1 \\ 0.9 & x = 2 \\ 0.7 & x = 3 \\ 0.5 & x = 4 \\ 0.3 & x = 5 \\ 0.1 & x = 6 \\ 0 & x = 7 \end{cases}$$

مجموعه فازی N به صورت‌های زیر نمایش داده می‌شود،

$$N = \{(1, 1), (2, 0.9), (3, 0.7), (4, 0.5), (5, 0.3), (6, 0.1), (7, 0)\}$$

$$N = \left\{ \frac{1}{1}, \frac{0.9}{2}, \frac{0.7}{3}, \frac{0.5}{4}, \frac{0.3}{5}, \frac{0.1}{6}, \frac{0}{7} \right\} \quad \text{یا}$$

^۱ ازین پس به اختصار آن را مجموعه فازی می‌نامیم

به عنوان نمونه $\mu_N(3) = 0.7$ بدین معنی است که از نظر تصمیم گیرنده عدد ۳ به اندازه ۰.۷ به مجموعه فازی اعداد کوچک تعلق دارد. به سخن دیگر، «گزاره ۳ عددی کوچک است» از نظر وی به اندازه ۰.۷ درست است.

نکته ۱. تعریف تابع عضویت یک مجموعه فازی بستگی به نظر تصمیم گیرنده دارد. در واقع این تعریف گاهی جنبه ذهنی و شخصی دارد؛ لذا می توان توابع عضویت مختلفی را برای یک مجموعه فازی که بیان گر یک ویژگی فازی باشد، متصور بود.

تعریف ۲.۱. فرض کنیم X یک مجموعه مرجع و N یک زیرمجموعه فازی از آن باشد، تکیه گاه N به صورت زیر تعریف می شود:

$$\text{supp}(N) = \{x \in X | \mu_N(x) > 0\}$$

همچنین $M = \sup_x \mu_N(x) = \text{hgt}(N)$ ارتفاع یک مجموعه فازی نامیده می شود. اگر این عدد برابر یک باشد آن گاه مجموعه فازی N ، نرمال و در غیر این صورت زیر نرمال نامیده می شود.

مثال ۲.۱. در مثال (۱.۱)، $\text{supp}(N) = \{1, 2, \dots, 6\}$ ، همچنین مجموعه فازی N مجموعه فازی نرمال است.

تعریف ۳.۱. زیرمجموعه معمولی از عناصر X که درجه عضویت آن ها در مجموعه فازی N ، حداقل به بزرگی h باشد، h -برش N (مجموعه تراز h ام وابسته به N) نامیده و با N_h نشان داده می شود (شکل ۲.۱)، یعنی

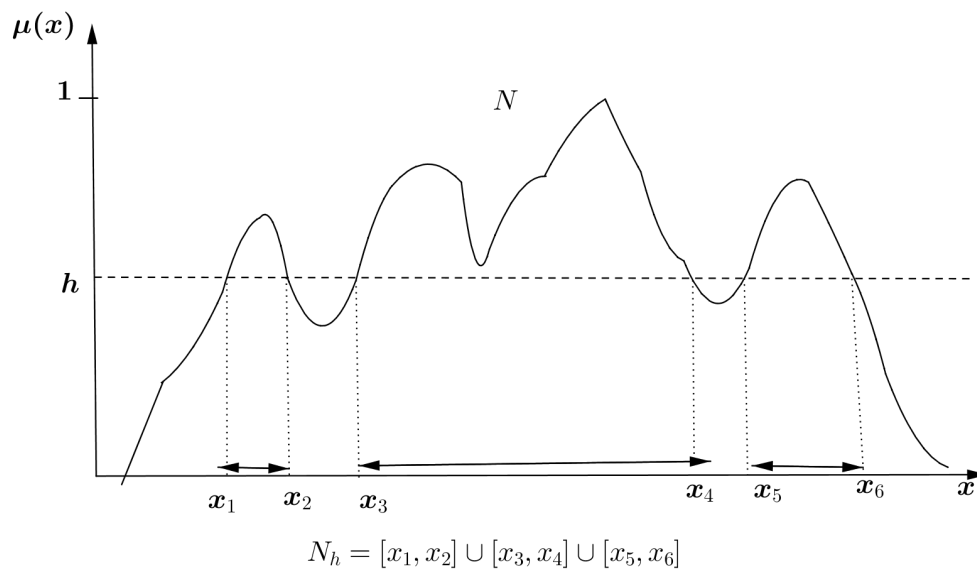
$$N_h = \{x \in X | \mu_N(x) \geq h\} \quad 0 < h \leq 1$$

مثال ۳.۱. فرض کنید $X = \{0, 1, 2, \dots\}$ مجموعه مقادیر یک متغیر تصادفی پواسن مربوط به تعداد تصادفات شبانه روزی در یک ناحیه باشد و فرض کنید مجموعه فازی N بیان گر **تعداد تصادفات کم** با تابع عضویت زیر تعریف شود،

$$\mu_N(x) = \begin{cases} 1 & x = 0 \\ 0.8 & x = 1 \\ 0.6 & x = 2 \\ 0.4 & x = 3 \\ 0.2 & x = 4 \end{cases}$$

در این صورت چند h -برش N عبارت است از:

$$N_{0.2} = \{0, 1, 2, 3, 4\}, \quad N_{0.5} = \{0, 1, 2\}$$



شکل ۲.۱: h -برش مجموعه فازی N

$$N_{\circ/6} = \{\circ, 1, 2\}, \quad N_{\circ/8} = \{\circ\}$$

به‌طور خلاصه می‌توان نوشت:

$$N_h = \begin{cases} \{\circ, 1, 2, 3, 4\} & \circ < h \leq \circ/2 \\ \{\circ, 1, 2, 3\} & \circ/2 < h \leq \circ/4 \\ \{\circ, 1, 2\} & \circ/4 < h \leq \circ/6 \\ \{\circ, 1\} & \circ/6 < h \leq \circ/8 \\ \{\circ\} & \circ/8 < h \leq 1 \end{cases}$$

قضیه ۱.۱. [۷۳] (اتحاد تجزیه^۲) فرض کنید N مجموعه‌ای فازی از $\mathbb{R}$ با تابع عضویت $\mu_N(\cdot)$ باشد. در این صورت

$$\mu_N(x) = \sup_{h \in (\circ, 1]} h I_{N_h}(x)$$

که در آن $I_{N_h}(\cdot)$ تابع نشان‌گر h -برش N است.

اکنون اصل مهمی بیان می‌شود که برای تعیین حاصل عمل یک تابع بر مجموعه‌های فازی به‌کار می‌رود.

^۲Decomposition Identity

تعریف ۴.۱. [۴] (اصل توسیع یا اصل گسترش)^۳ فرض کنید $X_1, \dots, X_n$ ، مجموعه مرجع و

$$X = X_1 \times \dots \times X_n = \{(x_1, \dots, x_n) | x_1 \in X_1, x_2 \in X_2, \dots, x_n \in X_n\}$$

حاصل ضرب دکارتی آنها باشد. همچنین $N_1, \dots, N_n$ ، مجموعه فازی به ترتیب از $X_1, \dots, X_n$ باشند. به علاوه $y = f(x_1, x_2, \dots, x_n)$ یک نگاشت از X به Y باشد. حاصل عمل f بر n مجموعه فازی $N_1, \dots, N_n$ به صورت مجموعه فازی B از Y با تابع عضویت زیر تعریف می شود

$$B(y) = f(N_1, \dots, N_n)(y) = \begin{cases} \sup_{y=f(x_1, \dots, x_n)} \min\{\mu_{N_1}(x_1), \dots, \mu_{N_n}(x_n)\} & f^{-1}(y) \neq \emptyset \\ 0 & f^{-1}(y) = \emptyset \end{cases}$$

که در آن $f^{-1}(y)$ معکوس نگاشت f در y می باشد.

مثال ۴.۱. فرض کنید X_1 و X_2 مجموعه اعداد حسابی، N_1 مجموعه فازی اعداد خیلی کوچک و N_2 مجموعه فازی اعداد تقریباً ۲ با توابع عضویت زیر باشند:

$$N_1 = \left\{ \frac{1}{0}, \frac{0.8}{1}, \frac{0.6}{2}, \frac{0.4}{3}, \frac{0.2}{4} \right\}, \quad N_2 = \left\{ \frac{0.5}{1}, \frac{1}{2}, \frac{0.5}{3} \right\}$$

آن گاه بر اساس اصل توسیع، حاصل عمل $f(x_1, x_2) = x_1 + x_2$ بر N_1 و N_2 به صورت مجموعه فازی

$$\left\{ \frac{0.5}{1}, \frac{1}{2}, \frac{0.8}{3}, \frac{0.6}{4}, \frac{0.5}{5}, \frac{0.4}{6}, \frac{0.2}{7} \right\}$$

به دست می آید.

۲.۱ اعداد فازی

اعداد فازی که زیرمجموعه های فازی خاصی از مجموعه اعداد حقیقی هستند، در بیشتر مسائل کاربردی استفاده می شوند. در ادامه، مفاهیم و نتایج اصلی درباره اعداد فازی بر اساس مراجع [۴]، [۶] و [۲۴] یادآوری می شود.

تعریف ۵.۱. [۶] مجموعه فازی N از $\mathbb{R}$ یک عدد فازی گفته می شود، اگر

(۱) N نرمال و تک نمایی باشد،

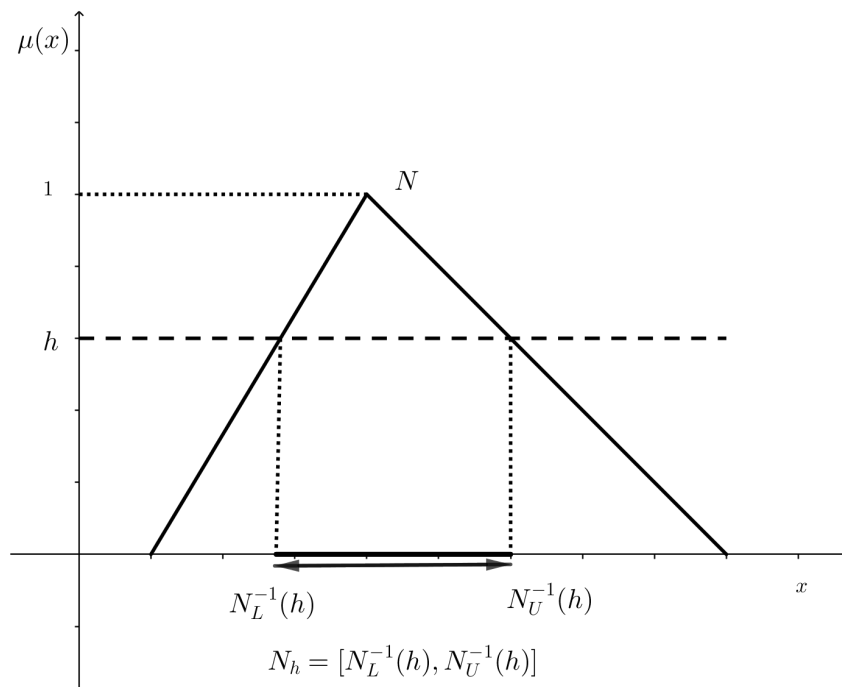
(۲) $-h$ برش های N به ازای هر $h \in (0, 1]$ به صورت بازه های بسته باشند.

منظور از تک نمایی بودن این است که تنها یک $x \in \mathbb{R}$ وجود داشته باشد که برای آن

$$\mu_N(x) = 1$$

^۳ Extension Principle

مجموعه h -برش عدد فازی N که در شکل ۳.۱ نشان داده شده است، به صورت $N_h = [N_L^{-1}(h), N_U^{-1}(h)]$ نمایش داده می‌شود که در آن $N_L^{-1}(h) = \min\{x : N(x) = h\}$ و $N_U^{-1}(h) = \max\{x : N(x) = h\}$ می‌باشد. در حالت خاص N یک عدد دقیق^۴ (غیرفازی) با مقدار m است



شکل ۳.۱: h -برش عدد فازی N

هر گاه تابع عضویت آن به صورت

$$I_m(x) = \begin{cases} 1 & x = m \\ 0 & x \neq m \end{cases}$$

باشد. توجه کنید که برای هر h ، $N_h^L = N_h^U = m$. مجموعه همه اعداد فازی از $\mathbb{R}$ با $\mathcal{F}(\mathbb{R})$ نشان داده می‌شود.

قرارداد ۱.۱. در ادامه برای سادگی تابع عضویت مجموعه فازی N را به جای $\mu_N(x)$ به صورت $N(x)$ نشان می‌دهیم.

تعریف ۶.۱. [۴] عدد فازی N مثبت (منفی) گفته می‌شود اگر برای هر $x \leq 0$ ($x \geq 0$)، $\mu_N(x) = 0$. عدد فازی N نامنفی (نامثبت) گفته می‌شود اگر برای هر $x < 0$ ($x > 0$)، $\mu_N(x) = 0$.

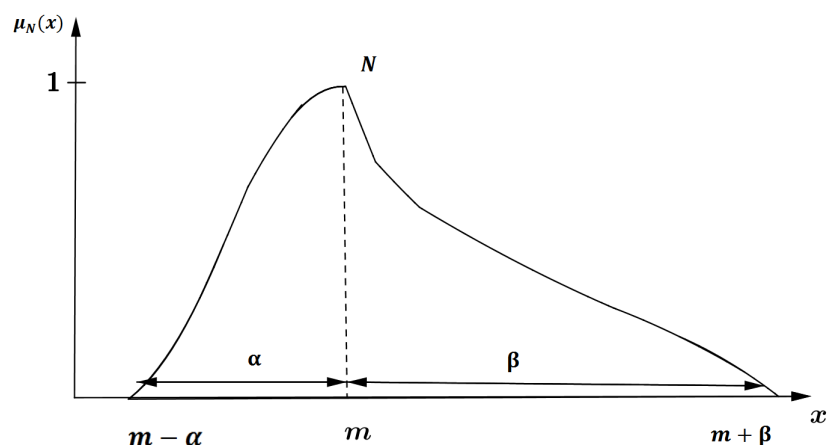
^۴crisp

نوع خاصی از اعداد فازی، اعداد فازی LR هستند که علاوه بر این که ساختار ویژه‌ای دارند، برخی اعمال حسابی بر آن‌ها از قواعد خاصی پیروی می‌کند. این ویژگی‌ها باعث شده است که در کاربردها، عمدتاً از این نوع اعداد فازی استفاده شود.

تعریف ۷.۱. [۲۴] (عدد فازی LR) N را یک عدد فازی LR نامیده و با نماد $N = (m, \alpha, \beta)_{LR}$ نشان می‌دهیم (شکل ۴.۱) اگر ساختار تابع عضویت عدد فازی N به صورت

$$N(x) = \begin{cases} L(\frac{m-x}{\alpha}) & x \leq m \\ R(\frac{x-m}{\beta}) & x > m \end{cases}$$

باشد که در آن توابع شکل L و R ، توابعی غیر صعودی از $\mathbb{R}^+$ به $[0, 1]$ هستند و $L(0) = R(0) = 1$. عدد حقیقی m مقدار نما (یا مرکز و یا میانه) و اعداد مثبت α و β به ترتیب پهنای چپ و پهنای راست N نامیده می‌شوند.



شکل ۴.۱: نمایش عدد فازی LR

در صورتی که $L = R$ و $\alpha = \beta$ ، آن‌گاه عدد فازی را متقارن نامیده و با $N = (m, \alpha)_{LL}$ نشان می‌دهیم.

برای $\alpha = 0$ قرار دهید $L(\frac{m-x}{\alpha}) = 0$ و برای $\beta = 0$ قرار دهید $R(\frac{x-m}{\beta}) = 0$ پس منظور از $N = (m, 0, 0)_{LR}$ یک عدد دقیق با مقدار m است. اعداد فازی LR در تعریف زیر معرفی شده‌اند.

تعریف ۸.۱. فرض کنید $N = (m, \alpha, \beta)_{LR}$ و $L = R$

الف) اگر $L(x) = \max\{0, 1 - |x|\}$ ، آن‌گاه N را یک عدد فازی مثلثی^۵ نامیده و با $N = (m, \alpha, \beta)_T$

^۵Triangular

نشان می‌دهیم.

ب) اگر $L(x) = e^{-x^2}$ ، آن‌گاه N را یک عدد فازی نرمال^۶ نامیده و با $N = (m, \alpha, \beta)_N$ نشان می‌دهیم.

ج) اگر $L(x) = \max\{0, 1 - x^2\}$ ، آن‌گاه N را یک عدد فازی سهموی^۷ نامیده و با $N = (m, \alpha, \beta)_P$ نشان می‌دهیم.

تعریف ۹.۱. (بازه فازی LR) اگر در تعریف عدد فازی شرط تک نمایی بودن برداشته شود، آن‌گاه یک بازه فازی با نماد $N = (m_1, m_2, \alpha, \beta)_{LR}$ و با تابع عضویت زیر به وجود می‌آید

$$N(x) = \begin{cases} L(\frac{m_1 - x}{\alpha}) & x \leq m_1 \\ 1 & m_1 < x \leq m_2 \\ R(\frac{x - m_2}{\beta}) & x > m_2 \end{cases}$$

که در آن اعداد حقیقی m_1 و m_2 به ترتیب نمای اول و نمای دوم و اعداد مثبت α و β به ترتیب پهنای چپ و پهنای راست N می‌باشند.

در حالتی که $L(x) = R(x) = \max\{0, 1 - |x|\}$ ، بازه فازی در برخی متون، عدد فازی دوزنقه‌ای^۸ نیز نامیده می‌شود و با $N = (m_1, m_2, \alpha, \beta)_{Tr}$ نشان داده می‌شود. بازه فازی می‌تواند به عنوان تعمیمی از عدد فازی تلقی شود. در واقع زمانی که $m_1 = m_2 = m$ ، آن‌گاه بازه فازی به عدد فازی تبدیل می‌شود. اکنون چند تعریف و قضیه در مورد حساب اعداد فازی بیان می‌شود.

۳.۱ حساب اعداد فازی

تعریف ۱۰.۱. فرض کنید A و B دو عدد فازی و $X : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ یک عملگر دوتایی بر اعداد حقیقی باشد. اگر تعمیم عملگر $*$ را برای اعداد فازی با $\otimes$ نشان دهیم، با استفاده از اصل توسیع، حاصل $A \otimes B$ به صورت یک مجموعه فازی با تابع عضویت زیر تعریف می‌شود

$$(A \otimes B)(z) = \sup_{z=x \otimes y} \inf[A(x), B(y)] \quad (۱.۱)$$

^۶Normal

^۷Elliptic

^۸Trapezoidal

در حالت‌های خاص برای چهار عمل اصلی، تعریف بالا به صورت‌های زیر در می‌آید

$$(A \oplus B)(z) = \sup_{z=x+y} \inf[A(x), B(y)]$$

$$(A \ominus B)(z) = \sup_{z=x-y} \inf[A(x), B(y)]$$

$$(A \otimes B)(z) = \sup_{z=x \times y} \inf[A(x), B(y)]$$

$$(A \oslash B)(z) = \sup_{z=x/y} \inf[A(x), B(y)]$$

$$(k \odot B)(z) = \sup_{z=k.y} \inf[k, B(y)], k \in \mathbb{R}$$

در این پایان نامه برای سادگی به جای علامت‌های $\oplus, \ominus, \otimes, \odot$ و از $+, -, \times, /$ استفاده می‌کنیم.

برخی اعمال جبری برای اعداد فازی LR بسیار ساده و دارای الگوهای مشخصی است و برای برخی دیگر از اعمال جبری نیز الگوهای تقریبی وجود دارد. در این قسمت برخی روابط مهم در حساب اعداد فازی LR بیان می‌شوند.

قضیه ۲.۱. [۲۴] اگر $A = (m_A, l_A, r_A)_{LR}$ و $B = (m_B, l_B, r_B)_{LR}$ و $\lambda \in \mathbb{R}$ آن‌گاه بر اساس اصل گسترش، داریم

$$A + B = (m_A, l_A, r_A)_{LR} + (m_B, l_B, r_B)_{LR} = (m_A + m_B, l_A + l_B, r_A + r_B)_{LR}$$

$$A - B = (m_A, l_A, r_A)_{LR} - (m_B, l_B, r_B)_{LR} = (m_A - m_B, l_A - l_B, r_A - r_B)_{LR}$$

$$\lambda.A = \begin{cases} (\lambda m_A, \lambda l_A, \lambda r_A)_{LR} & \lambda > 0 \\ (\lambda m_A, -\lambda r_A, -\lambda l_A)_{RL} & \lambda < 0 \end{cases}$$

درباره‌ی عمل ضرب و تقسیم، کار به آسانی جمع و تفریق نیست. در حالتی که A و B دو عدد فازی به صورت LR باشند، نتیجه‌ی ضرب آن‌ها یک عدد فازی LR نخواهد بود. برای ضرب و تقسیم روابط تقریبی پیشنهاد شده که در ادامه بیان می‌شود.

قضیه ۳.۱. [۲۴] (روابط تقریبی برای ضرب) اگر $A = (m_A, l_A, r_A)_{LR}$ و $B = (m_B, l_B, r_B)_{LR}$ آن‌گاه

$$(A \times B) \simeq \begin{cases} (m_A m_B, m_A r_B + m_B l_A, m_A r_B + m_B r_A)_{LR} & A > 0, B > 0 \\ (m_A m_B, m_A l_B - m_B r_A, m_A r_B - m_B l_A)_{RL} & A > 0, B < 0 \\ (m_A m_B, -m_B r_A - m_A r_B, -m_B l_A - m_A l_B)_{RL} & A < 0, B < 0 \end{cases}$$

قضیه ۴.۱. [۲۴] (روابط تقریبی برای تقسیم) اگر $A = (m_A, l_A, r_A)_{LR}$ یک عدد فازی مثبت باشد، آن‌گاه

$$A^{-1} \simeq (m_A^{-1}, r_A m_A^{-2}, l_A m_A^{-2})_{RL}$$

اگر $A = (m_A, l_A, r_A)_{LR}$ و $B = (m_B, l_B, r_B)_{LR}$ دو عدد فازی مثبت باشند، آن گاه

$$(A/B) \simeq \left(\frac{m_A}{m_B}, \frac{r_B m_A + l_A m_B}{m_B^2}, \frac{l_B m_A + r_A m_B}{m_B^2} \right)_{LR}$$

هر چه پهنای اعداد فازی A و B نسبت به مقادیر نمای آن ها کوچک تر باشد، تقریب های فوق دقیق تر است و در همسایگی مقادیر نما، این روابط دقت بیشتری دارند.

۴.۱ فاصله بین دو عدد فازی

در بحث رگرسیون فازی علاقه مند به تخمین و برآورد پارامترهای مدل هستیم و لذا لازم است خطا و به عبارت دقیق تر فاصله بین دو مقدار مشاهده شده و برآورد شده فازی محاسبه شود، لذا در این بخش به معرفی برخی از انواع فاصله بین دو عدد فازی می پردازیم.

۱.۴.۱ فاصله موزون

ژو و لی [۶۳] فاصله ی موزون^۹ بین دو عدد فازی را به صورت زیر تعریف می کنند

تعریف ۱۱.۱. فرض کنید A و B دو عدد فازی باشند، فاصله موزون بین این دو عدد فازی به صورت زیر است

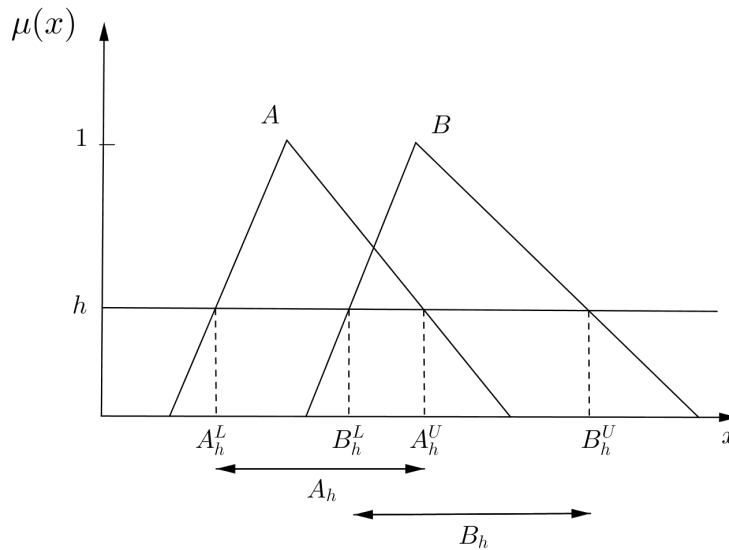
$$d(A, B) = \left[\int_0^1 f(h) d^{\gamma}(A_h, B_h) dh \right]^{\frac{1}{\gamma}} \quad (2.1)$$

که در آن A_h و B_h - برش های دو عدد فازی هستند (شکل ۵.۱) و

$$d^{\gamma}(A_h, B_h) = [A_h^L - B_h^L]^{\gamma} + [A_h^U - B_h^U]^{\gamma} \quad (3.1)$$

و $A_h^L, A_h^U, B_h^L, B_h^U$ کران های بالا و پایین h - برش A و B هستند. همچنین $f(h)$ یک تابع صعودی روی بازه $[0, 1]$ است و $f(0) = 0$ و $\int_0^1 f(h) dh = \frac{1}{\gamma}$ می باشد. تابع $f(h)$ را می توان به عنوان وزن $d^{\gamma}(A_h, B_h)$ تعبیر کرد. ویژگی صعودی بودن $f(h)$ منجر به این می شود که درجات عضویات بالاتر، اهمیت بیشتری در تعیین فاصله ی بین دو عدد فازی A و B داشته باشند. شرط های $f(0) = 0$ و $\int_0^1 f(h) dh = \frac{1}{\gamma}$ نیز این اطمینان را می دهد که فاصله ی فوق، یک تعمیم معمولی از فاصله ی بین دو عدد حقیقی است.

^۹Weighted distance



شکل ۵.۱: h -برش‌های دو عدد فازی

قضیه ۵.۱. [۶۳] تحت مفروضات تعریف ۱۱.۱ فرض کنید A و B دو عدد فازی مثلثی باشند
 آنگاه بر پایه‌ی تابع وزنی $f(h) = h$ داریم

$$\begin{aligned} d^2(A, B) = & (m_A - m_B)^2 \\ & + \frac{1}{3} (m_A - m_B) [(r_A - r_B) - (l_A - l_B)] \\ & + \frac{1}{12} [(l_A - l_B)^2 + (r_A - r_B)^2] \end{aligned} \quad (4.1)$$

نتیجه ۱.۱. اگر $A = (m_A, l_A)_T$ و $B = (m_B, l_B)_T$ دو عدد فازی مثلثی متقارن باشند، آنگاه رابطه (۴.۱) به صورت زیر به دست می‌آید

$$d^2(A, B) = (m_A - m_B)^2 + \frac{1}{6} (l_A - l_B)^2 \quad (5.1)$$

نتیجه ۲.۱. در رابطه (۴.۱) فاصله‌ی هر عدد فازی مثلثی مانند A از مبدأ (یعنی $\circ = (\circ, \circ, \circ)_T$) به صورت

$$\begin{aligned} d^2(A, \circ) &= \|A\|^2 \\ &= m_A^2 + \frac{1}{3} m_A (r_A - l_A) + \frac{1}{12} (l_A^2 + r_A^2) \end{aligned} \quad (6.1)$$

محاسبه می‌شود.

تعریف ۱۲.۱. [۱۱] اگر $\mathbf{W} = (W_1, \dots, W_n)'$ یک بردار فازی باشد به طوری که به ازای $i = 1, \dots, n$ ها W_i عدد فازی باشند آنگاه نرم بردار فازی $\mathbf{W}$ به صورت زیر تعریف می‌شود

$$\|\mathbf{W}\|^2 = \sum_{i=1}^n \|W_i\|^2 \quad (7.1)$$

قضیه ۶.۱. [۱]. تحت مفروضات تعریف ۱۱.۱ فرض کنید $A = (m_A, l_A)_N$ و $B = (m_B, l_B)_N$ دو عدد فازی نرمال متقارن باشند، آنگاه بر پایه‌ی تابع وزنی $f(h) = h$ داریم

$$d^{\gamma}(A, B) = (m_A - m_B)^{\gamma} + \frac{1}{\gamma} (l_A - l_B)^{\gamma} \quad (۸.۱)$$

۲.۴.۱ فاصله دونسو

دونسو و همکاران [۲۲] فاصله‌ی بین دو عدد فازی را به‌صورت زیر تعریف می‌کنند

تعریف ۱۳.۱. فرض کنید $A = (m_A, l_A, r_A)_T$ و $B = (m_B, l_B, r_B)_T$ دو عدد فازی باشند. فاصله بین دو عدد فازی به‌صورت زیر است

$$d^{\gamma}(A, B) = k_1 (m_A - m_B)^{\gamma} + k_2 \left((m_A - l_A - (m_B - l_B))^{\gamma} + (m_A + r_A - (m_B + r_B))^{\gamma} \right) \quad (۹.۱)$$

که در آن k_1 و k_2 اعداد حقیقی بین صفر و یک هستند که توسط محقق تعیین می‌شوند. در صورتی که $k_1 > k_2$ اهمیت مرکز بیشتر از پهنای پاسخ‌ها در نظر گرفته می‌شود و اگر $k_1 < k_2$ پهنای پاسخ‌ها از اهمیت بیشتری برخوردار می‌گردد. همچنین در صورتی که $k_1 = k_2$ ، اهمیت مرکز و پهنای یکسان فرض شده است (شکل ۶.۱).

۳.۴.۱ فاصله دیاموند

تعریف ۱۴.۱. [۲۰] این فاصله حالت خاصی از فاصله دونسو است که در آن $k_1 = k_2 = ۱$. اگر داشته باشیم $A = (m_A, l_A, r_A)$ و $B = (m_B, l_B, r_B)$ که $A, B \in \mathbb{T}(\mathbb{R})$ آنگاه فاصله دیاموند، به صورت زیر تعریف شده و در شکل ۶.۱ نشان داده می‌شود

$$d^{\gamma}(A, B) = [(m_A - l_A) - (m_B - l_B)]^{\gamma} + [(m_A + r_A) - (m_B + r_B)]^{\gamma} + [m_A - m_B]^{\gamma} \quad (۱۰.۱)$$

۴.۴.۱ فاصله همینگ

تعریف ۱۵.۱. [۳۶] اگر $A = (m_A, l_A, r_A)_T$ و $B = (m_B, l_B, r_B)_T$ دو عدد فازی مثلثی باشند، آنگاه فاصله همینگ بین آن‌ها به‌صورت

$$d(A, B) = \int_R |A(x) - B(x)| dx \quad (۱۱.۱)$$

ملاحظه ۱.۱. اگر $A = [A_h^L, A_h^U]$ باشد می‌توان با انتخاب $h = ۱$ مقدار مرکز و با انتخاب $h = ۰$ مقدار پهنای چپ و راست A را مشخص کرد.

ملاحظه ۲.۱. نقاط پایینی و بالایی h -برش A ، یعنی $[A]_h$ ، با $\inf[A]_h$ و $\sup[A]_h$ مشخص شده که برای راحتی به ترتیب با A_h^L و A_h^U نشان داده می‌شوند.

۵.۴.۱ فاصله حسن‌پور

حسن‌پور و همکاران [۳۰] فاصله بین دو عدد فازی را به صورت زیر بیان کردند که در شکل (۶.۱) نشان داده شده است.

تعریف ۱۶.۱. فرض کنید $A = (m_A, l_A, r_A)_T, B = (m_B, l_B, r_B)_T$ دو عدد فازی در $T(\mathbb{R})$ باشند. $d : T(\mathbb{R}) \times T(\mathbb{R}) \rightarrow \mathbb{R}$ را به صورت زیر تعریف می‌کنیم

$$d(A, B) = |m_A - m_B| + |l_A - l_B| + |r_A - r_B| \quad (۱۲.۱)$$

ملاحظه ۳.۱. برای دو عدد دقیق (غیر فازی) فاصله (۱۲.۱) به قدر مطلق تفاوت بین آن‌ها محدود می‌شود.

۵.۱ رتبه‌بندی مجموعه‌های فازی

مقایسه یا رتبه‌بندی اعداد فازی یا به طور عمومی‌تر مجموعه‌های فازی، برای ساختن مدل‌های ریاضی به کار می‌رود. دست کاری‌های متغیرهای فازی یا پارامترهای فازی همواره شامل رتبه‌بندی یا مقایسه مسائل می‌باشد اما متأسفانه اعداد فازی مثل اعداد واقعی، نظم و ترتیب خطی ندارند و مقایسه آن‌ها ساده نیست. غالباً، همپوشانی‌ها و انفصال‌های کوچک در تکیه‌گاه‌های مجموعه‌های فازی، مقایسه را بسیار سخت می‌کند.

چند روش رتبه‌بندی فازی^{۱۰} (FRMs)، که بر پایه رویکردهای متفاوت و نقطه نظرات متفاوتی می‌باشند، در متون معرفی شده و چندین مرور در منابع [۱۳، ۱۷، ۴۳، ۷۵] مشخص شده‌اند.

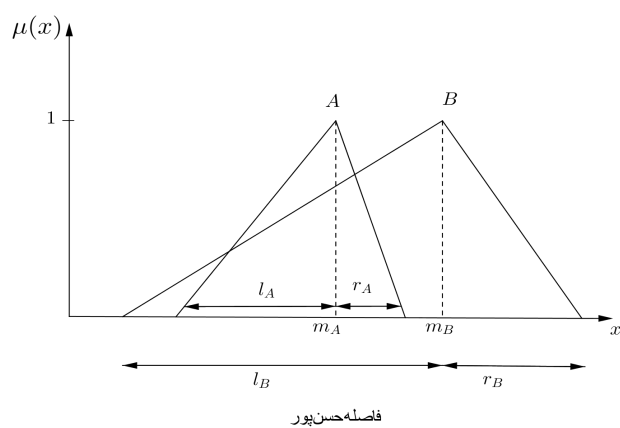
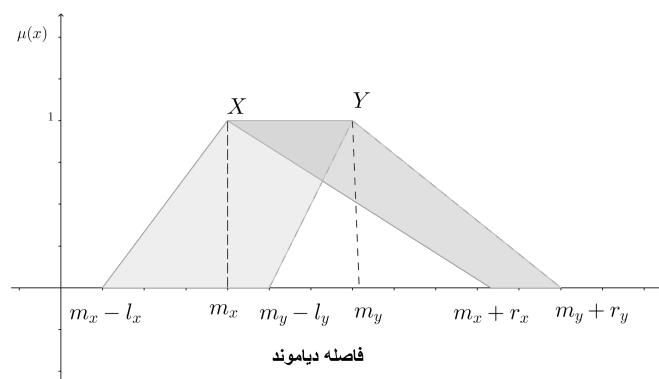
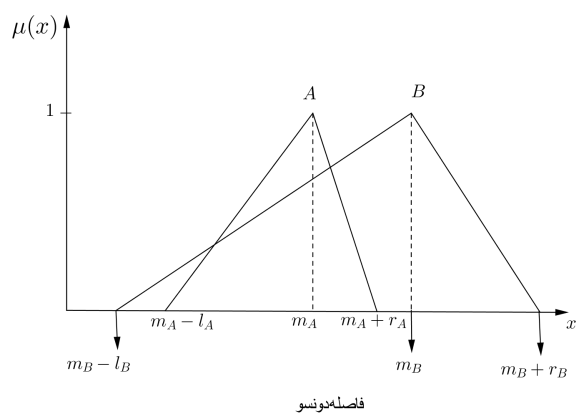
در این پایان نامه، از روش رتبه‌بندی چانگ و لی [۱۷] استفاده می‌کنیم که بر پایه‌ی مفهوم حضور می‌باشد و در ادامه به توضیح آن خواهیم پرداخت.

در روش چانگ و لی (۱۹۹۴)، یک FRM توسعه یافته معرفی شده است و نشان داده شده که چندین روش رتبه‌بندی فازی موجود، از ایده مرکزی این مفهوم استفاده می‌کنند. در گسترش این روش، جنبه‌های زیر مورد تأکید است.

۱. مقایسه مجموعه‌های فازی غیر محدب

۲. مقایسه مجموعه‌های فازی غیر نرمال

^{۱۰}Fuzzy ranking methods



شکل ۶.۱: نمایش فاصله‌های دونسو، دیاموند و حسن پور

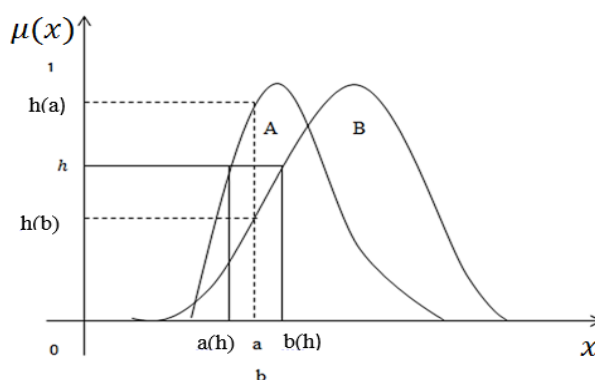
۳. نتیجه گیری‌های زبانی و

۴. فردی بودن نظر تصمیم گیرنده مثلاً خوش بینی یا بد بینی در مقایسه‌ی مجموعه‌های فازی

۱.۵.۱ تابع عضویت در برابر تکیه‌گاه

به منظور رسیدگی به مسائلی که در رتبه بندی اعداد فازی با آن‌ها مواجهیم به مفهوم پایه‌ای زیر مجموعه فازی نگاهی می‌اندازیم. یک زیر مجموعه فازی بر اساس تکیه‌گاه و تابع عضویت تعریف می‌شود. درجه‌ی تعلق یک عنصر معلوم در تکیه‌گاه به مفهومی نادقیق (فازی)، تابع عضویت این عنصر را تشکیل می‌دهد. بنابراین برای مقایسه‌ی چندین مجموعه‌ی فازی، علاقه‌مند به مقایسه‌هایی هستیم که بر پایه‌ی درجه‌ی تعلق (تابع عضویت) عناصر تکیه‌گاه یا مکان این عناصر در تکیه‌گاه باشد. در نظر گرفتن این خصوصیات برای مقایسه‌ی دو مجموعه‌ی فازی، گاهی ویژگی‌های نامطلوبی را برای FRM به همراه دارد.

برای توضیح این مطلب، دو مجموعه‌ی فازی A و B را در نظر بگیرید (شکل ۷.۱). اگر a و b به ترتیب بیان کننده‌ی عناصر موجود در تکیه‌گاه برای مجموعه‌های فازی A و B ، و همچنین h بیان گر درجه‌ی تابع عضویت باشد، آن‌گاه $a(h)$ و $b(h)$ مقادیری از تکیه‌گاه مجموعه‌های فازی A و B را بر پایه‌ی درجه‌ی یکسان h از تابع عضویت، نشان می‌دهند. اگر به ازای $h \in (0, 1]$ ، $a_h < b_h$ بر این اساس نتیجه می‌گیریم که مجموعه‌ی فازی A کوچکتر از مجموعه‌ی فازی B است؛ یعنی $A < B$. این نتیجه‌گیری برای دو عدد کاملاً فازی شاید صحیح باشد اما در این درجه تابع عضویت h ، قطعاً صحیح است.



شکل ۷.۱: مقایسه‌ی دو مجموعه فازی A و B

اما اگر توابع عضویت را بر پایه‌ی عنصر یکسانی از تکیه‌گاه $a(=b)$ مقایسه کنیم، یعنی مقایسه‌ی بین $h(a)$ و $h(b)$ ، نتیجه‌ای که به دست می‌آید کاملاً متفاوت خواهد بود ($B < A$). بنابراین مجموعه‌های فازی باید بر پایه‌ی درجه‌ی یکسانی از عضویت یا سطح حضور مقایسه

شوند نه بر پایه‌ی عنصر یکسان در تکیه‌گاه؛ به عبارتی دیگر نابرابری دو مجموعه‌ی فازی، حداقل در تعدادی عضویت یا سطح حضور، تفاوت عناصر در تکیه‌گاه‌ها را به دنبال خواهد داشت. به‌علاوه، درجات مختلف حضور یا عضویت یک عنصر معلوم در تکیه‌گاه نقش متفاوتی را در مقایسه مجموعه‌های فازی بازی می‌کند.

۲.۵.۱ مفهوم حضور سراسری

برای سطح حضور معلوم h تصویر تابع عضویت بر روی تکیه‌گاه عبارت است از

$$A^{-1}(h) = \{x : A(x) = h\}$$

اگر تابع عضویت مجموعه‌ی فازی A محدب و پیوسته باشد به‌ازای $0 < h < 1$ وقتی $\{A^{-1}(h=1)\}$ مد باشد در نتیجه $\{A^{-1}(h)\}$ همیشه شامل دو گروه از عناصر A می‌باشد. $x_{\min}^h = \min\{x : A(x) = h\}$ و $x_{\max}^h = \max\{x : A(x) = h\}$ که به‌ترتیب به عنوان مراجع راست و چپ تعریف می‌شوند. بر اساس بحث بالا و مفهوم حضور، قضیه زیر را می‌توان بیان کرد.

قضیه ۷.۱. [۶۴] دو مجموعه فازی A و B را با هر نوع تابع عضویتی در نظر بگیرید. اگر در سطح $h \in (0, 1)$ مجموعه A بزرگ‌تر از B باشد آن‌گاه $\{A^{-1}(h)\} > \{B^{-1}(h)\}$.

توجه داشته باشید که معکوس قضیه‌ی بالا همیشه درست نیست. یک مشکل این روش، این است که تنها یک نقطه h را در نظر می‌گیرد بنابراین از اطلاعات کافی استفاده نمی‌کند. روش‌های h -برش‌های آدامو [۹]، باکلی چاناس [۱۴] و ناندا [۵۰] تا حدی شبیه به این روش می‌باشند. برای مثال در روش آدامو، مفهوم حضور در شاخصی مثل

$$F_h(A_i) = \max\{x | A_i(x) \geq h\}$$

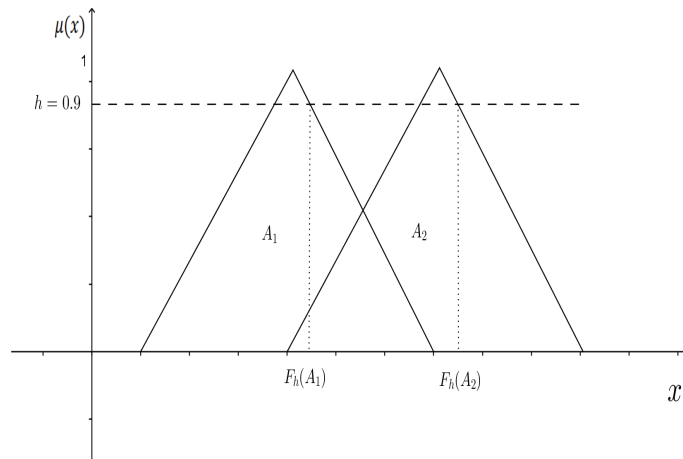
ظاهر می‌شود. شکل ۸.۱ شاخص آدامو را برای $h = 0.9$ نشان می‌دهد.

محدودترین تعریف برای $\{A^{-1}(h)\} > \{B^{-1}(h)\}$ در روش باکلی-چاناس به‌کار رفته است. بر اساس h -برش‌های معلوم A و B که به‌ترتیب با $[a_L^h, a_R^h]$ و $[b_L^h, b_R^h]$ بیان می‌شوند داریم

$$A > B, \quad \text{if} \quad a_L^h > b_R^h$$

در روش ناندا به نظر می‌رسد که مفهوم حضور به شیوه‌ای معقول‌تر ظاهر می‌شود. فرض کنید $X = [x_L, x_R]$ و $Y = [y_L, y_R]$ دو فاصله‌ی محدود بسته روی خط واقعی $\mathbb{R}$ باشند

$$X \leq Y, \quad \text{if} \quad x_L \leq y_L, \quad x_R \leq y_R$$



شکل ۸.۱: روش آدامو ($h = 0.9$)

دو عدد فازی A و B را به ترتیب با h -برش‌های فشرده‌ی A^h و B^h در نظر بگیرید داریم

$$A \leq B \quad \text{if} \quad A_h \leq B_h \quad \text{for any} \quad h \in [0, 1]$$

همچنین تاناکا و همکاران [۶۰] روشی مشابه در مدل کردن حدود «بزرگ‌تر از یا مساوی با» در برنامه نویسی خطی فازی به کار بردند. به منظور استفاده از همه‌ی اطلاعات موجود همه‌ی سطوح حضور h باید در نظر گرفته شوند. این شاخص حضور کلی می‌تواند به صورت

$$I = \int_0^1 g(\{A^{-1}(h)\})dh - \int_0^1 g(\{B^{-1}(h)\})dh \quad (13.1)$$

تعریف شود. A و B به ترتیب توابع عضویت مجموعه‌های فازی A و B هستند و تابع g یک تابع از معکوس این توابع عضویت می‌باشد. برپایه‌ی این شاخص تعریف زیر می‌تواند ارائه شود.

تعریف ۱۷.۱. دو مجموعه‌ی فازی A و B را در نظر بگیرید. گفته می‌شود A بزرگ‌تر از B است هرگاه $I > \epsilon$ باشد که در آن ϵ یک حد آستانه است ($\epsilon \geq 0$).

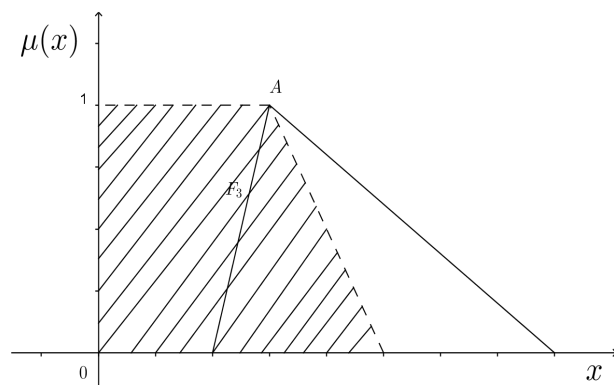
چندین روش FRM موجود، مثل روش‌های F_3 یا جر [۶۶]، کافمن [۳۹] و کمپس و گونزالس [۲۸، ۱۵] تا حدودی شبیه به این مفهوم حضور سراسری هستند. شاخص F_3 یا جر به صورت زیر تعریف می‌شود

$$F_3(A_i) = \int_0^{h_{\max}} M(A_i^h) dh$$

که $h_{\max} = hgt(A_i)$ درجه‌ی عضویت ماکسیمال در A_i است و $M(A_i^h)$ مقدار میانگین عناصر h -برش A_i^h از A_i را بیان می‌کند. برای مجموعه‌های فازی پیوسته و محدب، شاخص F_3 با استفاده از معادله (۱۳.۱) با تعریف تابع g به عنوان یک عمل متوسط گیری روی همه‌ی عناصر $\{A_i^{-1}(h)\} = \{x : A_i(x) = h\}$ به صورت

$$g(A_i^{-1}(h)) = m(\{A_i^{-1}(h)\})$$

محاسبه می‌شود. این شاخص در شکل ۹.۱ توسط ناحیه‌ی هاشور خورده نشان داده شده است. برای هر مقدار درجه عضویت، خط نقطه چین، مقدار متوسط عنصری که حداقل آن درجه‌ی عضویت را دارا می‌باشند نمایش می‌دهد.



شکل ۹.۱: شاخص F_3 یا جگر

شاخص $d(A)$ کافمن به صورت زیر تعریف می‌شود

$$d(A) = \frac{1}{2} [d_L(A) + d_R(A)] = \frac{1}{2} \left[\int_{h=0}^1 |A_h^{(L)}(x)| dx + \int_{h=0}^1 |A_h^{(R)}(x)| dx \right] \quad (14.1)$$

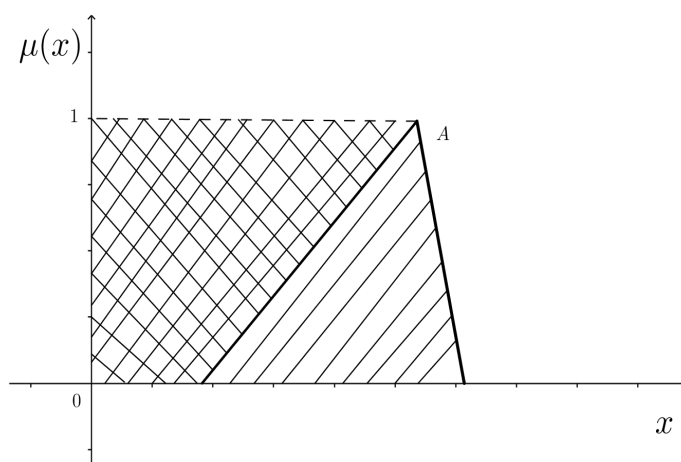
$d_L(A)$ و $d_R(A)$ به ترتیب فاصله‌ی همینگ چپ و راست A را از مبدأ بیان می‌کنند. مشخص است که شاخص کافمن با در نظر گرفتن تابع g به عنوان متوسط $x_{\min}^h$ و $x_{\max}^h$ و با استفاده از معادله‌ی (۱۳.۱) به دست می‌آید. نمایش هندسی $d_L(A)$ و $d_R(A)$ در شکل ۱۰.۱ آمده است. برای مجموعه‌های فازی محدب و پیوسته (مثل اعداد فازی مثلثی) شاخص‌های F_3 یا جگر، $d(A)$ کافمن و معادله‌ی (۱۳.۱) نتایج یکسانی را می‌دهند. همان‌طور که در بالا بیان شد چندین شکل مختلف برای تابع g در معادله‌ی (۱۳.۱) می‌تواند پیشنهاد شود. یک تعریف مفید و ساده‌ی g ، مجموع وزن داده شده بر روی تصویر تابع عضویت به صورت زیر است:

$$I = \int_0^1 \omega(\sum x(h)) dh - \int_0^1 \omega(\sum y(h)) dh \quad (15.1)$$

که در آن ω تابع وزنی از h است، و $x(h)$ نشان‌گر عناصر $\{A^{-1}(h)\}$ ، $y(h)$ نشان‌گر عناصر $\{B^{-1}(h)\}$ و A و B مجموعه‌های فازی پیوسته، محدب و نرمال‌اند. کمپس و گونزالس [۱۵]، مابوچی [۴۵] و یوان [۷۱] شاخص‌های دیگری را مطرح کرده‌اند.

۳.۵.۱ مجموعه‌های فازی نرمال و غیر نرمال

برای مجموعه‌های فازی غیر نرمال، سطح حضور از یک کمتر است و بنابراین مقایسه‌ی مستقیم بین مجموعه‌های فازی نرمال و غیر نرمال نمی‌تواند انجام شود. یک راه برای غلبه



شکل ۱۰.۱: شاخص $d(A)$ کافمن

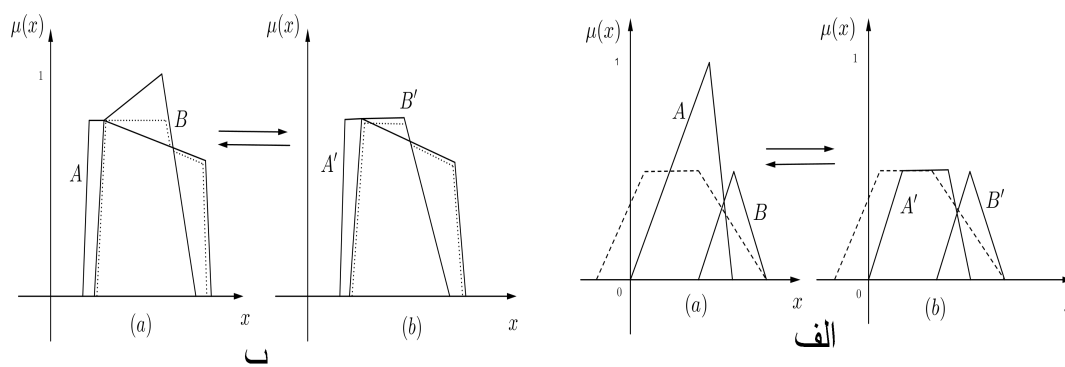
بر این مشکل، نرمال کردن مجموعه‌های فازی غیر نرمال است. ولی این راه اغلب نامطلوب است چون معنی اصلی مسئله از بین می‌رود. به این دلیل که معادله‌ی (۱۳.۱) که تعریفی برای شاخص حضور سراسری است، بر پایه‌ی اختلاف فازی است. لذا روشی منطقی‌تر می‌تواند توسط امتحان کردن تعریف اختلاف فازی بر مبنای اصل توسیع حاصل شود. شکل (۱۱.۱) قسمت الف، شباهت نتایج را هنگامی که اختلاف‌های فازی به‌دست آمده مربوط به دو مجموعه فازی غیر نرمال می‌باشد با وقتی که اختلاف‌های فازی به‌دست آمده مربوط به یک مجموعه فازی نرمال و یک مجموعه فازی غیر نرمال است، نشان می‌دهد؛ که «بالا صاف»^{۱۱} نامیده شده و توسط دوبیوس و پراد [۲۴] مطرح شده است.

برپایه‌ی مفهوم بالا صاف، شاخص حضور سراسری می‌تواند برای مجموعه‌های فازی غیر نرمال به‌صورت زیر اصلاح شود
 فرض کنید $h_{hgt}^* = \min[hgt(A), hgt(B)]$ ، اگر $hgt(A) > hgt(B)$ ، A بالا صاف است که آن را با A^T نمایش می‌دهیم و خواهیم داشت

$$I_{nonnorm} = \int_0^{h_{hgt}^*} g(\{(A^T)^{-1}(h)\})dh - \int_0^{h_{hgt}^*} g(\{(B^T)^{-1}(h)\})dh \quad (۱۶.۱)$$

ماکزیمم یا مینیمم فازی بر پایه‌ی اصل توسیع اثر بالا صاف یکسانی دارد که برای ماکزیمم فازی در شکل (۱۱.۱) قسمت ب، نشان داده شده است. نتایج مشابهی برای مینیمم فازی به‌دست می‌آید. بنابراین ماکزیمم یا مینیمم فازی بر پایه‌ی اصل توسیع نیز تنها کوچکترین درجه‌ی عضویت ماکزیمم مجموعه‌های فازی را مورد بررسی قرار می‌دهد. مینیمم یا ماکزیمم فازی می‌تواند در FRM ها ترکیب شود. برای اطلاعات بیشتر به [۱۳، ۴۳، ۷۵، ۱۷] مراجعه کنید.

^{۱۱}Top flattening



شکل ۱۱.۱: الف. اختلاف فازی و ب. ماکزیمم فازی

۴.۵.۱ یک روش رتبه‌بندی فازی با مداخله تصمیم گیرنده

یک جنبه‌ی مهم مجموعه فازی، ذهنیت^{۱۲} آن است و مخصوصاً زمانی درست است که مجموعه‌های فازی بسیار شبیه، نیازمند رتبه‌بندی هستند. به‌عنوان مثالی برای تشریح این بخش، دو عدد فازی مثلثی A_1 و A_2 را در شکل ۱۲.۱ در نظر بگیرید. هر دو عدد فازی مثلثی $A_1 = (m_{A_1}, l_{A_1}, r_{A_1})_T$ و $A_2 = (m_{A_2}, l_{A_2}, r_{A_2})_T$ ؛ که در آن $m_{A_1} = m_{A_2} = m$ ، 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 مساله ای خاص انتخاب می‌کند. در صورتی که از نگاه ریاضی غیر فازی، به مساله نگاه کنیم و هیچ‌گونه ابهامی را برای این اعداد در نظر نگیریم، هر دو عدد A_1 و A_2 با هم برابرند در صورتی که بنابر نظر تصمیم گیرنده (از نگاه فازی)، این دو عدد با هم متفاوتند. بنابراین نیاز به روشی برای رتبه‌بندی اعداد فازی (یک FRM) است که نظر تصمیم گیرنده را مشارکت دهد.

تعریف ۱۸.۱. فرض کنید که A و B دو مجموعه‌ی فازی با توابع عضویت $A(x)$ و $B(x)$ باشند و فرض کنید که $\{A^{-1}(h)\}$ و $\{B^{-1}(h)\}$ دو زیر مجموعه معمولی باشند که تصویر معکوس توابع عضویت با $h \in (0, 1]$ را مشخص می‌کنند، یعنی

$$\{A^{-1}(h)\} = \{x : A(x) = h\} \text{ and } \{B^{-1}(h)\} = \{x : B(x) = h\}$$

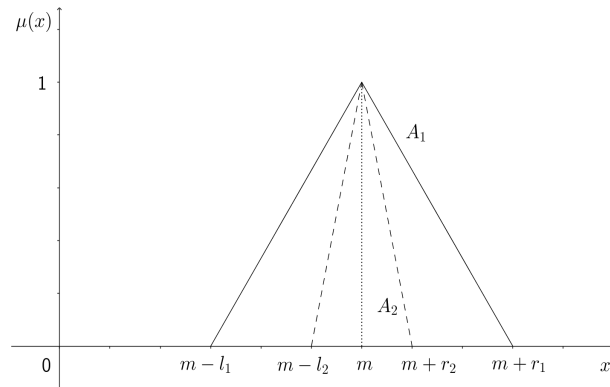
که در آن $x \in \mathbb{R}$. اختلاف بین A و B به‌صورت زیر تعریف می‌شود:

$$d(A, B) = \int_0^{h_{gt}^*} g_A(\{A^{-1}(h)\}) dh - \int_0^{h_{gt}^*} g_B(\{B^{-1}(h)\}) dh \quad (17.1)$$

که در آن $h_{gt}^* = \min[hgt(A), hgt(B)]$ و تابع g عبارت است از

$$g_A(\{A^{-1}(h)\}) = \omega(h)[\chi_1(h)x'_A(h) + \chi_2(h)x''_A(h)]$$

^{۱۲}Subjectivity



شکل ۱۲.۱: نمودار دو عدد فازی تقریباً m با مشخصات $A_1 = (m_{A_1}, l_{A_1}, r_{A_1})_T$ و $A_2 = (m_{A_2}, l_{A_2}, r_{A_2})_T$

$$g_B \left(\{B^{-1}(h)\} \right) = \omega(h) [\chi_1(h)x'_B(h) + \chi_2(h)x''_B(h)]$$

که در آن

$$x''_A(h) = \max\{x_i : A(x_i) = h\}, x'_A = \min\{x_i : A(x_i) = h\}$$

$$x''_B(h) = \max\{x_i : B(x_i) = h\}, x'_B = \min\{x_i : B(x_i) = h\}$$

همچنین $\omega(h)$ ، $\chi_1(h)$ و $\chi_2(h)$ مترهای وزن دهی هستند که به صورت ذهنی توسط تصمیم گیرنده تعیین می شوند.

تعریف ۱۸.۱ می تواند برای n مجموعه ی فازی A_i ، $(i = 1, \dots, n)$ بسط داده شود و اندازه هایی منحصر بفرد برای هر مجموعه ی فازی حاصل می شود. با نماد نویسی ساده تر داریم

$$OM(A_i) = \int_0^{h_{hgt}^*} g_i \left(\{A^{-1}(h)\} \right) dh \quad (18.1)$$

که در آن $h_{hgt}^* = \min_i [hgt(A_i)]$.

بنابراین معادله (۱۷.۱) به صورت زیر در می آید

$$d(A, B) = OM(A) - OM(B) \quad (19.1)$$

تعریف ۱۸.۱ برای همه ی انواع توابع عضویت مثل محدب، غیر محدب، نرمال، غیر نرمال و ... به کار می رود. تعریف بالا برای اعداد فازی نوع LR می تواند ساده شود [۲۴]. فرض کنید که طرف چپ یک عدد فازی LR با A_L و طرف راست آن با A_R نشان داده شود، داریم

$$x'(h) = A_L^{-1}(h), \quad x''(h) = A_R^{-1}(h)$$

$x'(h)$ و $x''(h)$ به‌ترتیب تصویرهای معکوس ضابطه‌ی چپ (L) و ضابطه‌ی راست (R) از تابع عضویت A هستند. حال تعریف ۱۸.۱ به‌صورت زیر در می‌آید

$$OM(A) = \int_0^1 \omega(h) [\chi_1(h) A_L^{-1}(h) + \chi_2(h) A_R^{-1}(h)] dh \quad (20.1)$$

همان‌طور که بیان شد مترهای وزن‌دهی $\omega(h)$ ، $\chi_1(h)$ و $\chi_2(h)$ توسط تصمیم‌گیرنده تعیین می‌شوند. این معیارها برای تأکید بر جنبه‌های مختلف در سطوح حضور مختلف به‌کار می‌روند.

در ابتدا متر وزن‌دهی $\omega(h)$ را بررسی می‌کنیم. این متر می‌تواند به‌عنوان یک تابع وزن‌دهی سطح حضور به‌کار رود و به‌صورت زیر تعریف می‌شود

$$\omega(h) = \frac{h}{\frac{1}{\chi} (h_{hgt}^*)^2} \quad \text{که} \quad \frac{1}{\chi} (h_{hgt}^*)^2 = \int_0^{h_{hgt}^*} h dh \xrightarrow{\text{اگر } h_{hgt}^*=1} \omega(h) = \frac{h}{\int_0^1 h dh} \quad (21.1)$$

وزن‌های مختلف می‌توانند به سطوح حضور مختلف اضافه شوند. برای مثال ممکن است تصمیم‌گیرنده تصمیم بگیرد که عناصر موجود در سطوح حضور میانی باید وزن بیشتری داشته باشند. بنابراین $\omega(h)$ می‌تواند به‌صورت زیر تعریف شود

$$\omega(h) = \frac{h(1-h)}{\Lambda^*}, \quad \text{که} \quad \Lambda^* = \int_0^{h_{hgt}^*} h(1-h) dh \xrightarrow{\text{اگر } h_{hgt}^*=1} \omega(h) = \frac{h(1-h)}{\int_0^1 h(1-h) dh} \quad (22.1)$$

آشکار است که این وزن‌دهی ذهنی $\omega(h)$ با کم وزن‌دهی به سطوح حضور پایین‌تر و بیش وزن‌دهی به سطوح میانی، نظر محافظه‌کارانه‌تر تصمیم‌گیرنده را منعکس می‌کند. وزن‌دهی ذهنی دیگر $\omega(h)$ مثل $\omega(h) = \frac{h(h^r)}{\Lambda^*}$ ، $r \in \mathbb{R}$ می‌باشد.

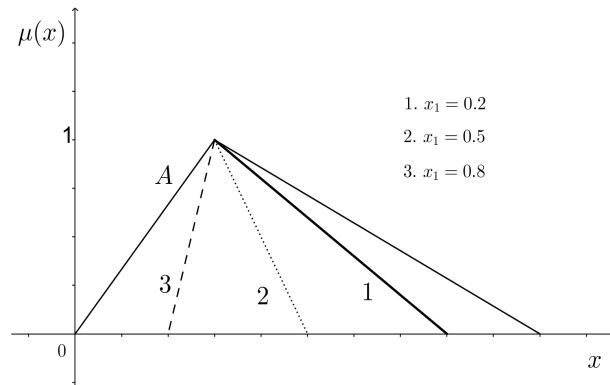
مترهای وزن‌دهی $\chi_1(h)$ و $\chi_2(h)$ ، حتی می‌توانند نقش‌های مهم‌تری را در اولویت تصمیم‌گیرنده ایفا کنند. و توابع مختلفی برای آن‌ها در حالت‌های مختلف می‌توانند امکان‌پذیر باشند. ساده‌ترین توابع عبارتند از

$$\chi_1(h) = \chi_1, \quad \chi_2(h) = \chi_2, \quad \chi_1, \chi_2 \in [0, 1], \quad \chi_1 + \chi_2 = 1 \quad (23.1)$$

که وزن‌دهی نوع $\chi(h)$ - ثابت نامیده می‌شود. این نوع وزن‌دهی، بسته به این‌که تصمیم‌ماکزیم سازی است یا مینیمم سازی، می‌تواند خوش‌بینانه یا بدبینانه باشد. شکل ۱۳.۱ سه وزن‌دهی نوع $\chi(h)$ - ثابت را برای $\chi_1 = 0.8, 0.5, 0.2$ نشان می‌دهد. اگر $\chi_1 = \chi_2 = 0.5$ ، وزن‌دهی نوع $\chi(h)$ - ثابت یکسان نامیده می‌شود. از طرف دیگر اگر $\chi_1 \neq 0.5$ ، وزن‌دهی نوع $\chi(h)$ - ثابت غیر یکسان نامیده می‌شود.

قضیه ۸.۱. اگر $A = (m_A, l_A, r_A)$ و $B = (m_B, l_B, r_B)$ اعداد فازی مثلثی باشند و χ_1 و χ_2 در شرایط (۲۳.۱) صدق کنند، باقی‌مانده مدل، یعنی E_i ، به‌صورت زیر به‌دست می‌آید:

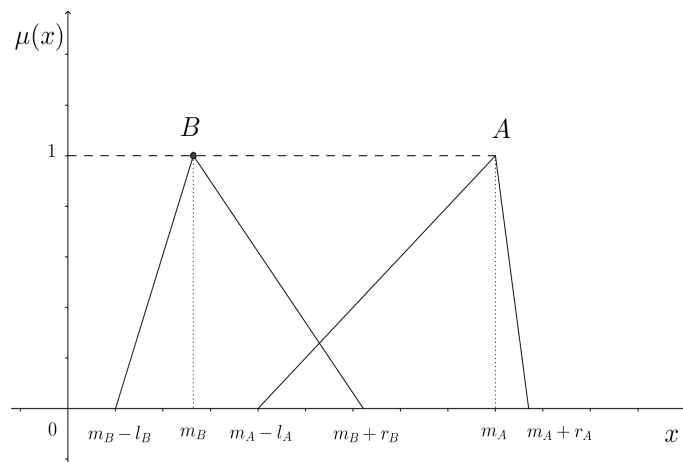
$$E_i = d(A, B)$$



شکل ۱۳.۱: شرحی از وزن دهی χ - ثابت: $[\chi_1 x'(h) + \chi_2 x''(h)]$

$$= \chi_1 \left((m_A - m_B) - \frac{1}{3}(l_A - l_B) \right) + \chi_2 \left((m_A - m_B) + \frac{1}{3}(r_A - r_B) \right) \quad (24.1)$$

برهان. وقتی A و B اعداد فازی مثلثی باشند فاصله‌ی بین A و B با استفاده از شرایط عنوان شده در (۲۳.۱) به صورت زیر محاسبه می‌شود: (شکل ۱۴.۱ را ملاحظه کنید)



شکل ۱۴.۱: نمایش رابطه‌ی بین دو عدد فازی مثلثی A و B

تابع عضویت A عبارت است از:

$$A(x) = \begin{cases} \frac{x - (m_A - l_A)}{l_A} & m_A - l_A \leq x < m_A \\ \frac{(m_A + r_A) - x}{r_A} & m_A \leq x \leq m_A + r_A \\ 0 & \text{otherwise} \end{cases} \quad (25.1)$$

و تصویر معکوس تابع عضویت به شکل زیر می‌باشد:

$$A^{-1}(x) = \begin{cases} hl_A + (m_A - l_A) & m_A - l_A \leq x < m_A \\ (m_A + r_A) - hr_A & m_A \leq x \leq m_A + r_A \\ \circ & \text{ow} \end{cases} \quad (26.1)$$

همچنین تابع عضویت B و تصویر معکوس تابع عضویت نیز به صورت زیر می‌باشد:

$$B(x) = \begin{cases} \frac{x - (m_B - l_B)}{l_B} & m_B - l_B \leq x < m_B \\ \frac{(m_B + r_B) - x}{r_B} & m_B \leq x \leq m_B + r_B \\ \circ & \text{ow} \end{cases} \quad (27.1)$$

$$B^{-1}(x) = \begin{cases} hl_B + (m_B - l_B) & m_B - l_B \leq x < m_B \\ (m_B + r_B) - hr_B & m_B \leq x \leq m_B + r_B \\ \circ & \text{ow} \end{cases} \quad (28.1)$$

شاخص OM برای A و B بر اساس معادلات (۱۸.۱)، (۱۹.۱)، (۲۰.۱)، (۲۱.۱) و شرایط (۲۳.۱) و روابط بالا یعنی (۲۵.۱)، (۲۶.۱)، (۲۷.۱) و (۲۸.۱) به شکل زیر محاسبه می‌شود:

$$\begin{aligned} OM(A) &= \frac{\int_0^1 h (\chi_1(m_A - (1-h)l_A) + \chi_2(m_A + (1-h)r_A)) dh}{\int_0^1 h dh} \\ &= \chi_1 \left(\frac{l_A}{3} + \frac{m_A - l_A}{2} \right) + \chi_2 \left(\frac{m_A + r_A}{2} - \frac{r_A}{3} \right) \\ OM(B) &= \frac{\int_0^1 h (\chi_1(m_B - (1-h)l_B) + \chi_2(m_B + (1-h)r_B)) dh}{\int_0^1 h dh} \\ &= \chi_1 \left(\frac{l_B}{3} + \frac{m_B - l_B}{2} \right) + \chi_2 \left(\frac{m_B + r_B}{2} - \frac{r_B}{3} \right) \end{aligned}$$

پس فاصله‌ی بین A و B ، یعنی $d(A, B)$ ، بر حسب شاخص OM به صورت زیر حاصل می‌شود:

$$\begin{aligned} E_i = d(A, B) &= \int_0^1 g_A(\{A^{-1}_h(x)\}) dh - \int_0^1 g_B(\{B^{-1}_h(x)\}) dh \\ &= OM(A) - OM(B) \\ &= \chi_1 \left((m_A - m_B) - \frac{1}{3}(l_A - l_B) \right) \\ &\quad + \chi_2 \left((m_A - m_B) + \frac{1}{3}(r_A - r_B) \right) \end{aligned} \quad (29.1)$$

□

نتیجه ۳.۱. ، اگر در قضیه ۸.۱، A و B اعداد فازی مثلثی متقارن به شکل $A = (m_A, s_A)_T$ و $B = (m_B, s_B)_T$ آن گاه، E_i به شکل زیر در خواهد آمد:

$$E_i = d(A, B) = (\chi_1 + \chi_2)(m_A - m_B) + \frac{1}{3}(\chi_2 - \chi_1)(s_A - s_B) \quad (30.1)$$

نتیجه ۴.۱. [۴۱] اگر در قضیه ۸.۱، A و B اعداد فازی مثلثی متقارن و $\chi_1 = \chi_2 = \chi = \frac{1}{3}$ باشند، داریم:

$$E_i = d(A, B) = m_A - m_B \quad (31.1)$$

فصل ۲

رگرسیون استوار

در این فصل به تحلیل رگرسیون کلاسیک پرداخته و نحوه‌ی محاسبه‌ی برآورد کمترین توان‌های دوم را بیان می‌کنیم. سپس در مورد رگرسیون جایگزین استوار و دلیل استفاده از آن به جای رگرسیون کلاسیک صحبت می‌کنیم و در نهایت به بررسی اجمالی چندین روش رگرسیون استوار پرداخته و از بین آن‌ها روش کمترین توان‌های دوم موزون و رایج‌ترین روش، که M – برآورد می‌باشد و در این پایان نامه به نوعی از آن‌ها استفاده شده است، را به بحث خواهیم نشست.

مقدمه

روش‌های آماری معمولاً بر چندین فرض متکی هستند که در بسیاری از حالات واقعی، داده‌ها از این مفروضات نظری تبعیت نمی‌کنند. یک حالت رایج، زمانی است که در مجموعه داده‌ها مشاهداتی وجود دارند که از بخش اصلی مجموعه داده‌ها دورند. در چنین شرایطی عملکرد روش‌های استاندارد بسیار ضعیف است. در کل مطلوب است که روش‌های آماری از حضور داده‌های دور افتاده رنج نبرند. در این رابطه ایده‌ی «آماره‌های استوار» به وجود آمدند [۲۶، ۳۵، ۴۷]. در این پایان نامه توجه خود را به مفهوم استواری در تحلیل رگرسیون محدود می‌کنیم.

تحلیل رگرسیون یکی از گسترده‌ترین تکنیک‌های آماری به کار رفته می‌باشد [۳۷]. در کتب

آماري رگرسيون به عنوان روش تعيين و تحليل روابط نادقيق بين متغيرهاي آماري تعريف مي شود. بنابر اين در رگرسيون، تعيين روابط بين متغيرها و تحليل روابط بدست آمده مورد توجه قرار مي گيرد [۴۲]. مدل خطي سنتي، ارتباط بين متغير وابسته يا پاسخ را با متغيرهاي توضيحي $x_{i1}, x_{i2}, \dots, x_{ip}$ براي $i = 1, 2, \dots, n$ به شكل

$$y_i = x_i' \beta + \epsilon_i, i = 1, 2, \dots, n \quad (1.2)$$

بيان مي کند که در آن $x_i' = (x_{i1}, x_{i2}, \dots, x_{ip})$ و $\beta = (\beta_1, \beta_2, \dots, \beta_p)'$ پارامتر و ϵ_i جمله خطا به عنوان یک متغير تصادفي با اميد رياضي صفر مي باشد. اگر مدل شامل عرض از مبدا باشد در اين صورت $x_{i1} = 1$. به منظور برازش چنين مدلي بر روي داده ها بايد از یک برآوردگر رگرسيوني براي برآورد پارامترهاي مجهول β استفاده کنيم که $\hat{\beta} = (\hat{\beta}_1, \dots, \hat{\beta}_p)'$ را توليد کند. مقدار مورد انتظار y_i ، مقدار برازش شده ناميده شده و آن را با $\hat{y}_i = x_i' \hat{\beta}$ نشان مي دهند. باقيمانده i - امين مورد عبارت است از $r_i = y_i - \hat{y}_i$.

اساس آمار کلاسيک، برآوردگر رگرسيوني کمترین توان های دوم (LS) ^۱ می باشد. در ابتدا لازم است که فرض کنيم جملات خطا مستقل هستند و همگي داراي توزيع نرمال به صورت $\epsilon_i \sim N(0, \sigma^2)$ به ازاي هر $i = 1, \dots, n$ می باشند. برآوردگر رگرسيون کمترین توان های دوم، معادل برآوردگر ماکزيمم درستنمايي براي β است که تابع درستنمايي زیر را نسبت به β ماکزيمم می کند

$$L(\beta) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-\epsilon_i^2}{2\sigma^2}\right) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(\hat{y}_i - x_i' \hat{\beta})^2}{2\sigma^2}\right) \quad (2.2)$$

که معادل با ماکزيمم کردن لگاریتم (۲.۲) نسبت به β به صورت

$$\log L(\beta) = \sum_{i=1}^n \left(-\frac{1}{2} \ln(2\pi\sigma^2) - \frac{\epsilon_i^2}{2\sigma^2} \right)$$

است. چون σ ثابت است ماکزيمم کردن عبارت بالا معادل مينيمم کردن $\sum_{i=1}^n \epsilon_i^2$ است؛ بنابر اين برآورد کمترین توان های دوم، مقدار $\hat{\beta}$ است که منجر به کمترین مجموع توان های دوم باقيمانده ها يعني $\sum_{i=1}^n r_i^2$ می شود که در رگرسيون ساده معادل با برازش خط رگرسيون با مينيمم کردن مجموع مربعات فاصله های عمودي نقاط مشاهده شده از خط می باشد [۵۳]. اين مطلب، اثر عميقي روي تأثیر انواع مختلف نقاط، روي برآوردگر رگرسيوني دارد.

^۱Least squares

۵.۰.۲ محاسبه‌ی برآورد کمترین توان‌های دوم

ماتریس طرح X و بردارهای Y و ϵ را به صورت زیر در نظر بگیرید

$$X = \begin{pmatrix} x_{11} & \dots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \dots & x_{np} \end{pmatrix} = \begin{pmatrix} x'_1 \\ \vdots \\ x'_n \end{pmatrix}, \quad Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \quad \text{و} \quad \epsilon = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

مدل خطی کلاسیک، به صورت $Y = X\beta + \epsilon$ در نظر گرفته می‌شود. هدف برآوردگر کمترین توان‌های دوم، مینیمم کردن عبارت زیر است

$$\begin{aligned} \sum_{i=1}^n \epsilon_i^2 &= \epsilon' \epsilon \\ &= (Y - X\beta)'(Y - X\beta) \\ &= Y'Y - Y'X\beta - \beta'X'Y + \beta'X'X\beta \end{aligned}$$

داریم

$$\begin{aligned} \frac{\partial}{\partial \beta} \left(\sum_{i=1}^n \epsilon_i^2 \right) &= \frac{\partial}{\partial \beta} (Y'Y - Y'X\beta - \beta'X'Y + \beta'X'X\beta) \\ &= 0 - X'Y - X'Y + 2(X'X)\beta = 0 \end{aligned}$$

بنابراین برآوردگر کمترین توان‌های دوم $\hat{\beta}$ ، جواب عبارت $X'X\hat{\beta} = X'Y$ می‌باشد که $\epsilon'\epsilon = \sum_{i=1}^n r_i^2$ را مینیمم می‌کند. بنابراین زمانی که $X'X$ غیرمنفرد است، برآوردگر کمترین توان‌های دوم می‌تواند مستقیماً از داده‌ها به دست آید که عبارت است از

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

۱.۲ محدودیت‌های برآوردگر کمترین توان‌های دوم

با وجود این که محاسبه‌ی برآوردگر کمترین توان‌های دوم ساده است ولی شدیداً وابسته به مفروضات مدل است. وقتی رگرسیون کمترین توان‌های دوم را با n مشاهده برای مدل p پارامتری $Y = X\beta + \epsilon$ به کار می‌بریم، فرض‌های بخصوصی را در مورد بردار خطاها داریم. یکی از این فرض‌ها آن است که بردار ϵ دارای توزیع $N(0, \sigma^2 I)$ است. در عمل انحرافات از این فرض‌ها وجود دارد که اگر این انحرافات جدی باشند امیدواریم آن‌ها را در رفتار مانده‌ها تشخیص دهیم. بنابراین این انحرافات منجر می‌شود که تعدیلات مناسبی را در مدل و یا در ماتریس متغیرهای کمکی وارد کنیم.

اگر در تحلیل، متوجه شویم که خطاها توزیع غیر نرمال دارند، مخصوصاً در مواقعی که توزیع خطا‌های پهن‌تری نسبت به توزیع نرمال دارند، یعنی از احتمال‌های بیشتری در دم‌ها نسبت به توزیع نرمال برخوردار باشد، بایستی یک روش رگرسیون استوار را در نظر بگیریم. چنین توزیع‌های دم‌پهن محتمل هستند که خطاهای بزرگتری را نسبت به حالت نرمال تولید کنند. روش کمترین توان‌های دوم در به‌دست آوردن برآورد پارامترها به هر مشاهده وزن یکسان می‌دهد. روش‌های استوار قادرند به مشاهدات وزن‌های نابرابر اختصاص دهند. به‌طور کلی مشاهداتی که مانده‌های بزرگی را تولید می‌کنند به‌وسیله‌ی برآورد استوار، کم وزن‌ترند. تعدادی از روش‌های استوار در بخش ۲.۲ معرفی شده‌اند. به‌طور کلی در روش‌های رگرسیون استوار، نسبت به روش کمترین توان‌های دوم محاسبات بیشتری مورد نیاز است.

مشاهداتی که کاملاً از اکثریت داده‌ها دور هستند می‌توانند اثر قابل توجهی روی برآورد رگرسیون کمترین توان‌های دوم بگذارند. در متون رگرسیونی به این مشاهدات، «داده‌های دورافتاده»^۲ گویند که در بخش ۱.۱.۲ به توضیح آن خواهیم پرداخت. طبقه‌بندی نوع داده‌های دورافتاده مهم است؛ چراکه هر نوع، اثر متفاوتی روی برآوردگر کمترین توان‌های دوم دارد. شکل ۱.۲ این مطلب را روشن‌تر می‌کند.

۱.۱.۲ دسته‌بندی داده‌های دورافتاده

داده‌ای که برای یک متغیر خاص از توزیع دور باشد؛ چه متغیر توضیحی باشد و چه متغیر پاسخ، داده دورافتاده «یک متغیره» می‌باشد [۱۰]. با این وجود این نقطه لزوماً یک دورافتاده در رگرسیون نیست. یک دورافتاده رگرسیونی^۳، نقطه‌ای است که از خط ارتباطی که قسمت عمده‌ی داده‌ها دنبال می‌کنند انحراف دارد؛ طوری که مقدار داده y ، به شرط مقدار x آن، غیر معمول است [۴۲]. نوع دیگر داده‌ی دورافتاده، نقطه‌ی نافذ^۴ نامیده می‌شود و نقطه‌ای است که برای آن متغیرهای توضیحی $x_{k1}, \dots, x_{kp}$ در مقایسه با بقیه داده‌ها $x_{i1}, \dots, x_{ip}$ ، $i = 1, \dots, n$ دورافتاده هستند. گفته می‌شود که چنین داده‌ای روی شیب رگرسیونی نفوذ دارد و اگر این نقطه دور از میانگین‌های متغیرهای توضیحی قرار گیرد، نفوذ بالایی دارد [۱۰].

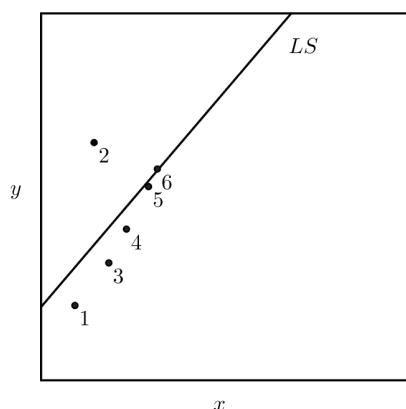
یک نقطه‌ی نافذ لزوماً یک دورافتاده‌ی رگرسیونی نیست چرا که ممکن است باز هم الگوی اصلی در داده‌ها را دنبال کند که در این مورد به این نقطه، نقطه‌ی نافذ خوب^۵ گفته می‌شود؛ درحالی که اگر یک دورافتاده‌ی رگرسیونی باشد نقطه‌ی نافذ بد نامیده می‌شود. یک داده‌ی با نفوذ، داده‌ای است که اگر آن را حذف کنیم، برآورد رگرسیونی به‌طور قابل ملاحظه‌ای تغییر

^۲Outliers

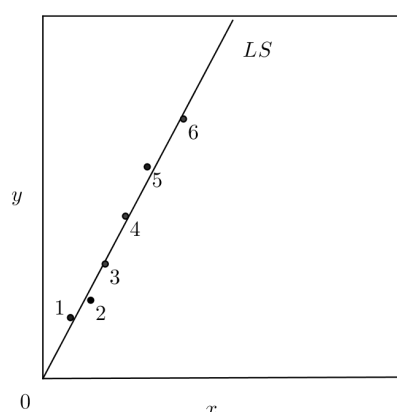
^۳Regression outlier

^۴Leverage point

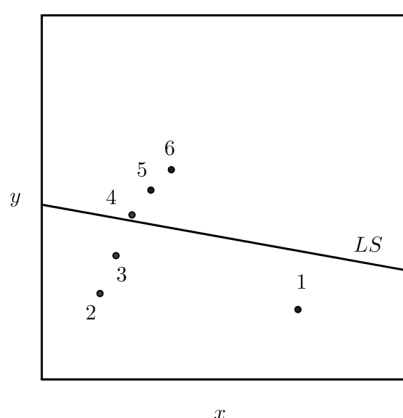
^۵Good leverage point



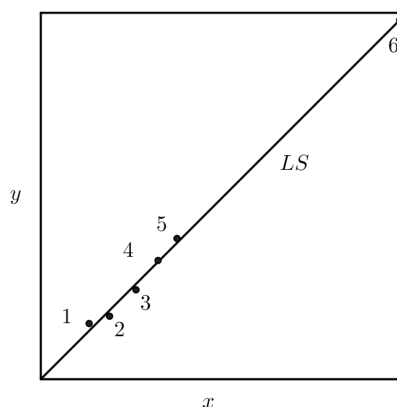
(ب) یک داده با یک دورافتاده در محور y جایگزین می‌شود.



(آ) ۶ داده با رابطه خطی قوی



(ج) یک داده با یک دورافتاده در محور x (د) یک داده با دورافتاده‌ای از نوع متفاوت در محور x جایگزین می‌شود. (یک نقطه‌ی نافذ)



شکل ۱.۲: خطوط رگرسیون کمترین توان‌های دوم در حضور داده‌های دورافتاده

می‌کند. یک نقطه‌ی نافذ بد می‌تواند خط رگرسیون کمترین توان‌های دوم (LS) را به مقدار زیادی تحت تأثیر قرار دهد. بنابراین نقاط نافذ بد اثر و نفوذ زیادی روی برآورد دارند و تهدید جدی‌ای را مطرح می‌کنند.

شکل ۱.۲ نمونه‌های مختلفی از داده‌های دور افتاده را نشان می‌دهد. در شکل ۱.۲ (ب)، تنها نقطه‌ی ۲ در متغیر پاسخ بر حسب متغیر توضیحی‌اش دورافتاده می‌باشد و بنابراین یک دورافتاده‌ی رگرسیونی است. در شکل ۱.۲ (ج) نقطه‌ی ۶ مثالی از یک نقطه‌ی نافذ خوب

است؛ هر چند که در مقایسه با مقادیر دیگر x ، دورافتاده می‌باشد ولی این نقطه با الگوی قسمت عمده‌ی داده‌ها مناسبت دارد. در شکل ۱.۲(د) نقطه‌ی ۱ تنها در محور x دورافتاده است و یک دورافتاده‌ی تک متغیره بر حسب مختص x خود بوده و بنابراین یک نقطه نافذ است. این نقطه، برآورد رگرسیون کمترین توان‌های دوم را تحت تأثیر قرار داده است طوری که در کل، دیگر بیان کننده‌ی روند کلی قسمت عمده‌ی داده‌ها نیست؛ این نقطه، یک نقطه‌ی نافذ بد است.

با توجه به شکل ۱.۲(آ)، مشخص است که هیچ باقیمانده‌ی بزرگی وجود ندارد ولی در شکل ۱.۲(د)، نقطه‌ی ۱ به‌طور افقی به جایی دورتر از خط رگرسیون منتقل شده و در مقایسه با این خط، باقیمانده‌ی منفی بسیار بزرگی دارد که مقدار بزرگی برای عبارت $\sum_{i=1}^n r_i^2$ می‌سازد. به‌منظور به‌دست آوردن برآورد کمترین توان‌های دوم، مجموع توان‌های دوم باقیمانده‌ها باید دوباره مینیمم شود که برای تحقق این امر، خط رگرسیون باید بیشتر به سمت مقدار دورافتاده کج شود. برغم این واقعیت که این کار، باقی‌مانده‌های نقاط دیگر را افزایش می‌دهد ولی در کل موجب کاهش مقدار $\sum_{i=1}^n r_i^2$ می‌شود.

۲.۱.۲ حذف داده‌های دورافتاده قبل از تحلیل

هرچند داده‌های دورافتاده‌ی بسیار پرت در مجموعه‌ی داده‌ها در اصل نادرست در نظر گرفته می‌شوند ولی می‌توانند بسیار جالب باشند. داده‌های دورافتاده می‌توانند ناشی از رخدادهای استثنایی باشند یا گروهی از داده‌های دورافتاده می‌تواند نتایجی از عاملی باشد که هنوز در مطالعه مورد بررسی قرار نگرفته است. حتی می‌تواند به‌طور سیستماتیک رخ دهد که مدل به‌سادگی با آن تطبیق نمی‌یابد. شرایطی وجود دارد که داده‌ها را می‌توان به‌صورت موجه حذف کرد. اما به‌طور کلی، از آن‌جا که مشاهدات غیر معمول، لزوماً مشاهدات بد نیستند به‌طور منطقی نتیجه‌گیری می‌شود که نباید آن‌ها را کم‌اهمیت در نظر گرفته و نمی‌توان آن‌ها را حذف کرد.

تشخیص داده‌های دورافتاده بدون بررسی دقیق می‌تواند بسیار مشکل باشد. نمی‌توان به‌سادگی رگرسیون کمترین توان‌های دوم را به‌کار برد و از نمودارهای باقیمانده برای تشخیص داده‌های دورافتاده استفاده کرد؛ چون این روش نسبت به داده‌های دورافتاده بسیار حساس است. اگر خط رگرسیونی به طرف این داده‌ها کشیده شود لزوماً باقیمانده‌های بزرگی نخواهند داشت. تشخیص داده‌های دورافتاده به‌خصوص در مجموعه داده‌های چندمتغیره مشکل است چراکه نمی‌توانند به‌سادگی در یک نمودار پراکنش شناسایی شوند [۳۱]. در چنین شرایطی که نمی‌توان داده‌های دورافتاده‌ای را که برآورد کمترین مربعات را دچار مشکل می‌کنند با توجه حذف کرد و چنین داده‌های دورافتاده‌ای لزوماً حتی نمی‌توانند تشخیص داده شوند، لازم است

که روش‌های رگرسیون استوار به کار برده شوند.

۲.۲ برآوردگرهای رگرسیون استوار

در آمار، تحلیل رگرسیون خطی ابزاری قدرتمند برای مطالعه‌ی رابطه‌ی خطی بین یک متغیر پاسخ Y (متغیر وابسته یا متغیر خروجی) و مجموعه‌ای از متغیرهای توضیحی $X_1, \dots, X_p$ (متغیرهای مستقل یا متغیرهای ورودی) می‌باشند. یکی از معروف‌ترین روش‌های به کار رفته برای برآورد ضرایب رگرسیونی روش کمترین توان‌های دوم (LS) می‌باشد که پیش‌تر معرفی شد. این روش علی‌رغم داشتن مزیت‌های نظری و عملی دارای یک مشکل اساسی است؛ این که به شدت به حضور داده‌های دورافتاده حساس است. طوری که حتی درصد کوچکی از مشاهداتی که به‌طور قابل ملاحظه‌ای از مجموعه‌ی داده‌ها انحراف دارند، موجب انحراف در برآورد پارامترها می‌شوند.

در تحلیل رگرسیون، مشکل داده‌های دورافتاده به دو صورت زیر مورد بررسی قرار می‌گیرد:

الف. یافتن داده‌های دورافتاده: تعریف معیاری برای تشخیص، حذف یا احتمالاً جایگزین کردن داده‌های غیرعادی.

ب. برآورد استوار: تعریف روش‌های برآورد مناسب برای خنثی کردن اثر حضور داده‌های دورافتاده در روند برآورد.

به دلیل اهمیت تشخیص داده‌های دورافتاده، راه‌حل‌های مختلفی در متون رگرسیون کلاسیک پیشنهاد شده است (برای اطلاعات بیشتر به [۴۲] مراجعه کنید). در مجموع هدف برآوردگرهای رگرسیون استوار، برازش مدلی است که اکثریت یک نمونه را توصیف می‌کند. استواری آن‌ها با دادن وزن‌های مختلف به داده‌ها در حین محاسبه‌ی برآورد به دست می‌آیند. بنابراین داده‌های دور به نسبت اثر کوچکتری روی برآوردگر رگرسیونی دارند [۲۳]، در صورتی که در رگرسیون کمترین توان‌های دوم، همه‌ی داده‌ها وزن یکسان دارند.

در نتیجه برآوردگرهای رگرسیون استوار می‌توانند ابزاری قدرتمند برای تشخیص داده‌های دورافتاده در مجموعه داده‌های پیچیده باشند. روسو و لروی، این برآوردگرها را اعلان کننده به آماردانان به عنوان 'دورافتاده‌ها روی یک «صفحه‌ی نقره‌ای» توصیف کرده‌اند که با کم کردن اثر داده‌های دورافتاده، همچنان قادرند این داده‌ها را آشکار کنند [۴۲]. با این حال، اندرسون، به کارگیری یک رویکرد محتاطانه برای برآوردگرهای رگرسیون استوار را توصیه می‌کند [۱۰]؛ چون اگر داده‌ها رفتار بسیار خوبی داشته باشند، برآوردگرهای استوار مختلف می‌توانند برآوردهای بسیار متفاوتی را ارائه دهند. روش‌های رگرسیون استوار، خودشان یک مدل نهایی را ارائه نمی‌دهند. اندرسن پیشنهاد می‌کند که به عنوان یک مرحله اولیه در تجزیه

و تحلیل رگرسیون، به کارگیری هردو روش استوار چندگانه و کمترین مربعات، برای مقایسه نتایج منطقی است. اگر رسم نمودار باقیمانده‌های روش کمترین مربعات در برابر روش‌های استوار، به‌طور بالقوه داده‌های دورافتاده را نشان دهد لازم است که در صورت امکان قبل از برازش مدل نهایی، این داده‌ها را توضیح دهیم.

با توجه به برآورد استوار که توسط هوپر [۳۳] پیشنهاد شده است یک مدل استوار باید ویژگی‌های زیر را داشته باشد:

- (۱) علی‌رغم حضور داده‌های دورافتاده، مدل باید برازش خوبی را روی داده‌ها نشان دهد.
 - (۲) انحرافات کوچک از مفروضات مدل نباید عملکرد روند برآورد را مختل کند یا اگر می‌کند مقدار آن کم باشد.
 - (۳) انحرافات بزرگ‌تر از مفروضات مدل نباید روند برآورد را خیلی تحت تأثیر قرار دهد.
- در متون مربوط به رگرسیون استوار کلاسیک چندین روش برآورد پیشنهاد شده است. بدون ذکر جزئیات به روش‌های زیر اشاره می‌کنیم:

الف. کمترین توان‌های دوم موزون (WLS): روش WLS ، زمانی مورد استفاده قرار می‌گیرد که فرض ثابت بودن واریانس خطا برقرار نباشد. این روش، به هر مشاهده وزنی را اختصاص می‌دهد؛ وزن خاص هر مشاهده نشان‌دهنده این است که هر مشاهده چقدر در برآوردهای نهایی تأثیر دارد. پس برآورد کمترین توان‌های دوم موزون عبارت است از $\hat{\beta}_{WLS} = \operatorname{argmin}_{\beta} \sum_{i=1}^n w_i \epsilon_i^2$ که در ادامه بیشتر به معرفی این روش خواهیم پرداخت.

ب. کمترین قدر مطلق انحرافات (LAD): [۲۵]^۷: بردار برآورد ضرایب رگرسیونی، β ، به‌منظور مینیمم کردن مجموع قدر مطلق مقادیر باقیمانده‌ها، r_i ، به‌جای مجموع توان‌های دوم باقیمانده‌ها انتخاب می‌شود.

پ. M - برآوردگرها [۳۲]^۸: تابع هدف $\sum_{i=1}^n \rho(r_i)$ مینیمم می‌شود. تابع ρ برای فراهم کردن یک رفتار رضایت بخش انتخاب می‌شود حتی وقتی که داده‌ها از مدلی تولید می‌شوند که از توزیع مفروض انحراف دارد (در ادامه بیشتر به معرفی این روش خواهیم پرداخت).

ت. M - برآوردهای تعمیم یافته [۴۸]^۹: M - برآوردگرها تعمیم می‌یابند تا تابع وزن مناسبی را برای مسئله‌ی مینیمم سازی معرفی کنند.

^۶ Weighted least squares

^۷ Least absolute deviations

^۸ M-estimators

^۹ Generalized M-estimators

ج. کمترین میانه‌ی توان‌های دوم^{۱۰} [۵۱، ۵۴]: این برآورد، شامل تعیین برآوردهای بردار β است که به‌جای مینیمم کردن مجموع، میانه‌ی توان‌های دوم باقیمانده را مینیمم می‌کند، $\min_{\beta} \text{med}_i(r_i^2)$.

چ. کمترین توان‌های دوم پیراسته (LTS) [۵۴]^{۱۱}: این روش، مجموع اولین $h < n$ توان‌های دوم باقیمانده‌های پیراسته را مینیمم می‌کند. $\min_{\beta} \sum_{i=1}^h (r_i^2)_{i:n}$ که $(r^2)_{l:n} \leq \dots \leq (r^2)_{h:n}$ ها اولین باقیمانده‌های h هستند.

روش‌های برآورد قابل توجه دیگر در تحلیل رگرسیون عبارتند از: R - برآوردگرها [۳۳]، S - برآوردگرها [۵۵]، M - برآوردگرهای یک گامی [۲۹] و MM - برآوردگرها [۷۰].

۱.۲.۲ روش کمترین توان‌های دوم موزون

روش LS معمول، فرض می‌کند که واریانس خطاها ثابت هستند. روش WLS، زمانی مورد استفاده قرار می‌گیرد که فرض ثابت بودن واریانس خطا برقرار نباشد. مدل مورد بررسی (۱.۲) را با ماتریس واریانس - کوواریانس غیر ثابت زیر در نظر بگیرید

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

فرض کنید $W_i = \frac{1}{\sigma_i^2}$ و ماتریس قطری W که مبین وزن‌هاست را به‌صورت زیر در نظر بگیرید

$$W = \begin{pmatrix} W_1 & 0 & \dots & 0 \\ 0 & W_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & W_n \end{pmatrix}$$

در این‌صورت برآورد کمترین توان‌های دوم موزون عبارت است از

$$\hat{\beta}_{WLS} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^n W_i e_i^2 = (X'WX)^{-1}X'WY \quad (3.2)$$

پس به‌طور کلی، در برآورد پارامتر به شیوه‌ی کمترین توان‌های دوم موزون، مانند کمترین توان‌های دوم معمول، مقادیر مجهول پارامترها یعنی $\beta_0, \beta_1, \dots$ ، با پیدا کردن مقادیر عددی برای برآورد پارامترها، در تابع رگرسیونی برآورد می‌شوند که این مقادیر مجموع توان‌های دوم

^{۱۰}Least median squares

^{۱۱}Least trimmed squares

انحرافات بین مقادیر مشاهده شده و مقادیر برآورد شده را مینیمم می‌کند. ولی برخلاف روش کمترین توان‌های دوم معمول، در روش کمترین توان‌های دوم موزون هر جمله وزن بخصوص $w_i, i = 1, \dots, n$ را داراست که تعیین می‌کند هر مشاهده در مجموعه‌ی داده‌ها تا چه اندازه بر برآوردهای نهایی پارامتر تأثیر دارند. در اصل روش‌های استوار بر پایه‌ی M – برآوردگرها^{۱۲} شکل می‌گیرند که در ادامه به شرح آن‌ها خواهیم پرداخت (برای مطالعه بیشتر به [۲۳] مراجعه کنید).

۲.۲.۲ M – برآوردگرها

M – برآوردگرها از نوع برآوردگرهای ماکزیمم درست‌نمایی هستند. مدل رگرسیونی (۱.۲) را در نظر بگیرید. فرض کنید خطاها (ϵ_i) مستقل و هم‌توزیع با تابع چگالی احتمال $f(\epsilon)$ باشند. برآوردگر ماکزیمم درست‌نمایی β مقداری است که کمیت $\prod_{i=1}^n f(y_i - x'_i \beta)$ یا به طور معادل $\sum_{i=1}^n \ln f(y_i - x'_i \beta)$ را ماکزیمم می‌کند؛ به‌طوریکه $x'_i, -i$ امین سطر X در مدل (۱.۲) است. در روش کمترین توان‌های دوم که در آن $\epsilon_i \sim N(0, \sigma^2)$ ، تابع درست‌نمایی 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 ماکزیمم کردن تابع درست‌نمایی، منجر به مینیمم کردن مجموع توان‌های دوم خطاها یعنی

$$\sum_{i=1}^n \epsilon_i^2 = \sum_{i=1}^n (y_i - x'_i \beta)^2 \quad (4.2)$$

می‌شود. اگر در رابطه (۴.۲) تعریف کنیم $\rho(\epsilon_i) = (y_i - x'_i \beta)^2$ در این صورت این رابطه به‌صورت زیر به‌دست می‌آید

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

حال اگر خطاها را استاندارد کنیم برآورد ضرایب رگرسیونی را می‌توان از مینیمم کردن $\sum_{i=1}^n \rho(\frac{\epsilon_i}{\sigma})$ به‌دست آورد. از آنجایی که σ معمولاً مجهول است می‌توان از برآورد آن، $\hat{\sigma}$ ، استفاده کرد. شیوه M – برآوردگرها، مینیمم کردن توابع مختلف $\rho(z_i)$ است که در آن $z_i = \frac{\epsilon_i}{\sigma}$ و $e_i = y_i - x'_i \beta$ باقیمانده‌ی i – امین مشاهده است. در حالتی که خطاها به‌طور نرمال توزیع شده باشند انحرافات معیار نمونه، $\hat{\sigma}$ ، اندازه‌های مناسبی برای تغییرپذیری می‌باشند ولی زمانی که در

^{۱۲}”M” for ”Maximum likelihood-type”

مجموعه داده‌ها مشاهدات دورافتاده داریم به اندازه‌های استوار دیگری از تغییرپذیری، نیاز است. یکی از اندازه‌های استوار تغییرپذیری، به صورت زیر تعریف می‌شود

$$\hat{\sigma}_{MAD} = d = \frac{\text{median}|e_i - \text{median}(e_i)|}{0.6745}, \quad i = 1, \dots, n \quad (5.2)$$

بنابراین گوییم یک برآوردگر استوار است اگر عبارت

$$\sum_{i=1}^n \rho\left(\frac{e_i}{d}\right) \quad (6.2)$$

را مینیمم کند. پس، باقیمانده‌های استاندارد شده جدید، به صورت $z_i = \frac{e_i}{d}$ تعریف می‌شوند. ملاحظه می‌شود که اگر $\rho(z) = z^2$ آن گاه با مینیمم کردن رابطه (۶.۲) به همان مدل کمترین توان‌های دوم معمول می‌رسیم.

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 زیر

$$\sum_{i=1}^n x_{ij} \Psi \left[\frac{y_i - \sum_{j=1}^p x_{ij} \beta_j}{d} \right] = 0, \quad j = 1, 2, \dots, n \quad (7.2)$$

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

توابع زیادی از Ψ ، توسط افراد مختلفی تعریف شده است که از آن جمله می‌توان به توابع Ψ هوبر، همپل، اندروز و توکی که بیشتر مورد استفاده قرار می‌گیرند اشاره کرد که در زیر آمده است و همان پیشنهاداتی هستند که پیش‌تر در مورد آن‌ها بحث شد

– تابع Ψ هوبر^{۱۳}

$$\Psi(z) = \begin{cases} -k & z < -k \\ z & |z| \leq k \\ k & z > k \end{cases} \quad (8.2)$$

با $k = 1/5$.

– تابع Ψ همپل^{۱۴}

^{۱۳} Huber

^{۱۴} Hampel

$$\Psi(z) = \begin{cases} |z| & 0 \leq |z| < a \\ a & a \leq |z| \leq b \\ a \left(\frac{c-|z|}{c-b} \right) & b \leq |z| < c \\ 0 & c \leq |z| \end{cases} \quad (9.2)$$

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

– تابع Ψ اندورز^{۱۵}

$$\Psi(z) = \begin{cases} \sin\left(\frac{z}{k}\right) & |z| \leq k\pi \\ 0 & |z| > k\pi \end{cases} \quad (10.2)$$

با $k = 1/5$ یا $k = 2/1$.

– تابع Ψ توکی^{۱۶}

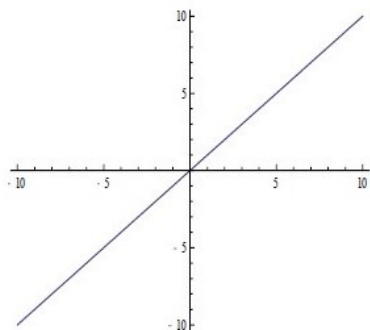
$$\Psi(z) = \begin{cases} z \left(1 - \left(\frac{z}{k} \right)^2 \right)^2 & |z| \leq k \\ 0 & |z| > k \end{cases} \quad (11.2)$$

با $k = 5$ یا $k = 6$.

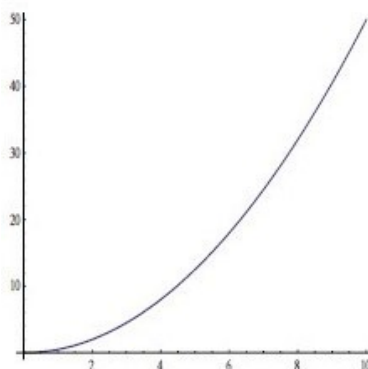
نمودارهای توابع ρ و توابع Ψ مدل LS و موارد فوق، در ۲.۲ نشان داده شده است

^{۱۵} Andrews

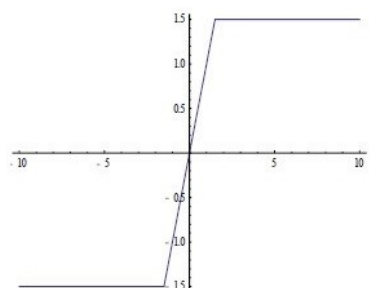
^{۱۶} Tukey



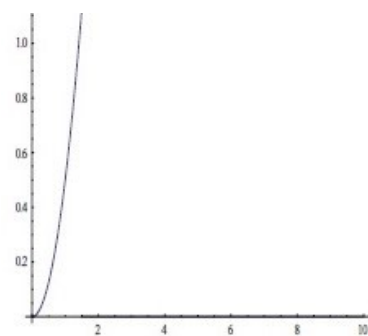
(ب) تابع Ψ برای LS



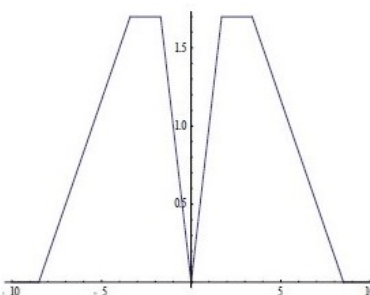
(آ) تابع ρ برای LS



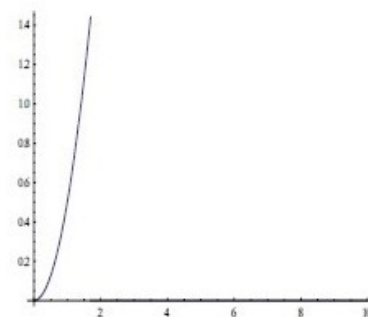
(د) تابع Ψ برای هوبر



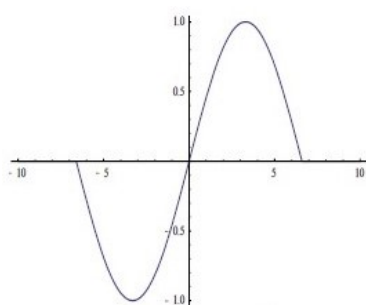
(ج) تابع ρ برای هوبر



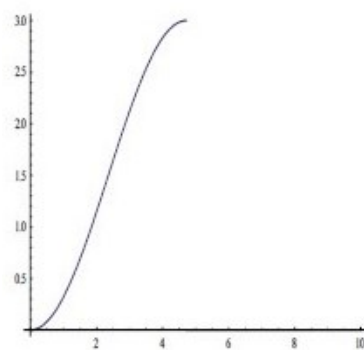
(و) تابع Ψ برای همپل



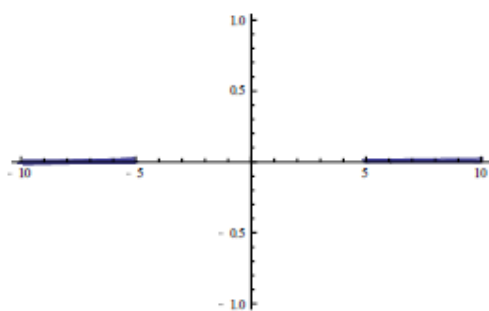
(ه) تابع ρ برای همپل



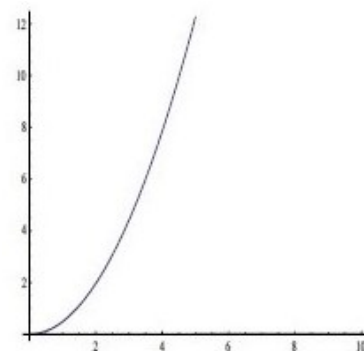
(ح) تابع Ψ برای اندروز



(ز) تابع ρ برای اندروز



(ی) تابع Ψ برای توکی



(ط) تابع ρ توکی برای

شکل ۲.۲: توابع ρ و Ψ هوپر، همپل، اندروز و توکی

فصل ۳

تحلیل رگرسیون در محیط فازی

در این فصل، رگرسیون در محیط فازی را مورد بحث قرار داده و رگرسیون کمترین توان‌های دوم فازی (FLS)^۱ را با استفاده از فاصله‌های موزون و دیاموند بیان می‌کنیم؛ سپس برای ارزیابی مدل‌های رگرسیون فازی، معیار نیکویی برازش MPE را معرفی می‌کنیم. مطالب این فصل عمدتاً برگرفته از طاهری [۶] و اسداللهی [۱] است.

مقدمه

در رگرسیون کلاسیک فرض می‌شود که متغیرها و مشاهدات مربوط به آنها دقیق هستند. همچنین خطای مدل یعنی اختلاف بین مقدار مشاهده شده برای متغیر وابسته و مقداری که از طریق مدل حاصل می‌شود، به خطای تصادفی مربوط به مشاهدات و اندازه‌گیری‌ها، عدم حضور برخی از متغیرها و غیره نسبت داده می‌شود. دربارهی خطای تصادفی و توزیع احتمالی آن، مفروضاتی مانند نرمال بودن، ناهمبسته بودن، ثبات واریانس و ... در نظر گرفته می‌شود؛ به‌گونه‌ای که بتوان بر پایه این مفروضات، تجزیه و تحلیل‌های آماری مانند برآورد پارامترها، پیش‌بینی مقدار متغیر وابسته و آزمون‌های فرض مرتبط با مدل را انجام داد. اما در بسیاری از اوقات ممکن است یک یا چند فرض از فرض‌های بالا برقرار نباشد، یا اینکه مثلاً به دلیل حجم نمونه اندک نتوان از درستی برخی فرض‌ها اطمینان حاصل کرد. برای مثال ممکن است در

^۱Fuzzy least squares

یک بررسی، مشاهدات مربوط به متغیرها نادقیق باشند یا نادقیق گزارش شده باشند؛ یا ممکن است متغیرهای تحت مطالعه دارای ارتباطی نادقیق باشند. همچنین ممکن است مفروضاتی مانند نرمال بودن و ناهمبسته بودن جملات خطای تصادفی برقرار نباشد. در چنین شرایطی ابزارهای کلاسیک نمی‌توانند معیارهای مناسبی را برای مدل‌سازی داده‌ها ارائه دهند. یکی از راه‌های ممکن، استفاده از مفهوم مجموعه‌های فازی، برای مدل‌سازی داده‌ها در چنین شرایطی است.

۱.۳ انواع رگرسیون فازی

در یک تقسیم‌بندی کلی، انواع رگرسیون فازی را می‌توان به سه حالت زیر تقسیم کرد:

- (۱) رگرسیون فازی در حالتی که ارتباط بین متغیرها فازی فرض می‌شود. به عبارت دیگر ضرایب معادله رگرسیونی، فازی در نظر گرفته می‌شوند.
- (۲) رگرسیون فازی در حالتی که متغیرها (اعم از پیش‌بین یا پاسخ) نادقیق و فازی هستند.
- (۳) رگرسیون فازی در حالتی که هم متغیرها و هم ضرایب مدل، فازی در نظر گرفته می‌شوند.

لازم به ذکر است که تنوع در رگرسیون فازی، محدود به حالت‌های فوق نیست بلکه شیوه‌هایی که برای هر حالت پیشنهاد شده است، تنوع بسیاری را در روش‌های رگرسیون فازی ایجاد کرده است. مانند رگرسیون کلاسیک که برحسب اصل یا اصولی که براساس آن‌ها برآورد پارامترهای مدل انجام می‌پذیرد، رگرسیون فازی را هم می‌توان تقسیم بندی کرد که می‌توان به روش‌های رگرسیون امکانی، رگرسیون کمترین توان‌های دوم خطای فازی و رگرسیون کمترین قدرمطلق خطای فازی اشاره کرد. در ادامه علاوه بر معرفی این روش‌ها، به شیوه‌های دیگری از محققینی که بر پایه‌ی یکی از روش‌های فوق، مساله‌ی رگرسیون را مورد بررسی قرار داده‌اند می‌پردازیم.

۱.۱.۳ رگرسیون کمترین توان‌های دوم فازی

اولین پژوهش در مورد تحلیل رگرسیون خطی فازی توسط تاناکا و همکاران [۵۸] انجام شد. تعمیمی از این روش، برای هنگامی که اعداد فازی از نوع LR^2 هستند، توسط باردوسی [۱۲] توسعه داده شد. دیاموند [۲۰]، روش کمترین توان‌های دوم فازی (FLS)^۳ را پیشنهاد کرد. میرزایی و ارقامی [۸] مروری بر چند رویکرد در رگرسیون فازی داشته‌اند. کلکین نما و طاهری [۷] رگرسیون فازی براساس کمترین قدر مطلق انحرافات را زمانی که داده دور افتاده وجود داشته باشد پیشنهاد کردند. چانگ و لی [۱۸] برای مواردی که در آن‌ها داده‌های دورافتاده

^۲ LR type fuzzy numbers

^۳ Fuzzy least squares

حضور دارند، تعمیمی از روش کمترین توان‌های دوم موزون فازی را پیشنهاد کردند که به هر درجه عضویت وزنی را اختصاص می‌دهد و بر پایه‌ی یک اثر متقابل با نظر تصمیم گیرنده است. برای رگرسیون فازی خطی ساده، یانگ و کو [۶۷] یک الگوریتم تکراری را برای تحلیل کمترین توان‌های دوم موزون فازی (WFLS) ^۴ 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.

روش کمترین توان‌های دوم فازی، نخستین بار توسط دیاموند [۲۰] و سلمینس [۱۶] ارائه شد. این روش تعمیمی از رگرسیون کمترین توان‌های دوم در حالت کلاسیک است که بر پایه‌ی تعریف‌هایی برای فاصله‌ی بین اعداد فازی می‌باشد. رگرسیون کمترین توان‌های دوم، مبتنی بر فاصله بین متغیر پاسخ و متغیر پیش‌بین است. دیاموند [۲۰] با تعریف فاصله‌ای بر اعداد فازی، مدل‌هایی را برای برازش کمترین توان‌های دوم برای مدل با متغیر مستقل دقیق و متغیر وابسته فازی و همچنین برای مدل با متغیر مستقل و وابسته فازی، پیشنهاد کرد. دیاموند و کورنر [۱۹] رگرسیون فازی را بر اساس متغیرهای تصادفی فازی ارائه کردند. همچنین، ژو و لی [۶۵] روش کمترین توان‌های دوم را برای برآورد ضرایب فازی مدل با استفاده از متری بر اساس سطوح تراز ارائه کردند و با یک معیار نیکویی برازش، مناسبیت مدل رگرسیونی را اندازه گرفتند. در روش کمترین توان‌های دوم فازی، پارامترها به‌گونه‌ای برآورد می‌شوند که به ازای هر مقدار، فاصله‌ی بین متغیر وابسته‌ی مشاهده شده و متغیر وابسته برآورد شده بر اساس فاصله d مینیمم گردد. به‌عبارتی مانند روش کلاسیک به دنبال مینیمم کردن مجموع توان‌های دوم خطا هستیم. برای مثال، اگر مشاهدات متغیر وابسته، فازی باشند، می‌خواهیم مجموع توان‌های دوم $\sum_{i=1}^n d^2(Y_i, \hat{Y}_i)$ را مینیمم کنیم که $d(.,.)$ می‌تواند فاصله بین اعداد فازی باشد.

قرارداد ۱.۳. برای سادگی در نوشتار مباحث مربوط به رگرسیون فازی در این پایان نامه، حروف بزرگ انگلیسی را به‌عنوان اعداد فازی و حروف کوچک را به‌عنوان اعداد غیر فازی نمایش می‌دهیم. همچنین حروف پررنگ به معنای ماتریس یا بردار و در غیر اینصورت اسکالر هستند. ماتریسی که ستون‌های آن شامل مقادیر متغیرهای پیش‌بین است را ماتریس طرح نامیده و با X نشان می‌دهند.

در اینجا می‌خواهیم یک مدل خطی با ضرایب فازی $A_j, j = 0, \dots, p$ به‌صورت

$$\hat{Y}_i = A_0 + A_1 x_{i1} + \dots + A_p x_{ip}, \quad i = 1, \dots, n \quad (1.3)$$

را بر پایه‌ی مشاهداتی به‌صورت

$$(Y_i, x_{i1}, \dots, x_{ip}), \quad i = 1, \dots, n$$

^۴ Weighted fuzzy least square

که در آن x_{ij} ها اعداد حقیقی و Y_i ها اعداد فازی هستند، برازش دهیم. بدون کاستن از کلیت مسأله فرض می‌کنیم $x_{ij} > 0$ (در صورتی که x_{ij} ها اعدادی منفی باشند می‌توان با یک انتقال آن‌ها را به اعدادی مثبت تبدیل کرد). با استفاده از روش کمترین توان‌های دوم خطا و با توجه به فاصله (۵.۱) مدل (۱.۳) را به داده‌ها برازش می‌دهیم. یعنی ضرایب فازی را به گونه‌ای می‌یابیم که مجموع توان‌های دوم خطا (در زیر) مینیمم شود

$$SSE(A_0, A_1, \dots, A_p) = \sum_{i=1}^n d^2(Y_i, A_0 + A_1 x_{i1} + \dots + A_p x_{ip}) \quad (2.3)$$

بحث را به حالتی که $Y_i = (m_{Y_i}, l_{Y_i})_T$ و $A_j = (m_{A_j}, l_{A_j})_T$ اعداد فازی مثلثی متقارن باشند، محدود می‌کنیم. بر اساس عملگرهای بین اعداد فازی داریم

$$A_0 + A_1 x_{i1} + \dots + A_p x_{ip} = (m_{A_0} + m_{A_1} x_{i1} + \dots + m_{A_p} x_{ip}, l_{A_0} + l_{A_1} x_{i1} + \dots + l_{A_p} x_{ip})_T$$

با توجه به آنچه گفته شد مجموع توان‌های دوم خطا با استفاده از فاصله‌ی موزون که در بخش ۱.۴.۱ و نتیجه‌ی (۵.۱) مطرح شد، به صورت زیر حاصل می‌شود

$$SSE(A_0, A_1, \dots, A_p) = \sum_{i=1}^n (m_{A_0} + m_{A_1} x_{i1} + \dots + m_{A_p} x_{ip} - m_{Y_i})^2 + \frac{1}{\alpha} \sum_{i=1}^n (l_{A_0} + l_{A_1} x_{i1} + \dots + l_{A_p} x_{ip} - l_{Y_i})^2$$

برای مینیمم سازی تابع $SSE(A_0, A_1, \dots, A_p)$ نسبت به m_{A_j} و l_{A_j} مشتق گرفته و برابر صفر قرار می‌دهیم، آنگاه معادلات نرمال زیر حاصل می‌شوند

$$\sum_{i=1}^n (m_{A_0} + m_{A_1} x_{i1} + \dots + m_{A_p} x_{ip} - m_{Y_i}) x_{ij} = 0$$

$$\sum_{i=1}^n (l_{A_0} + l_{A_1} x_{i1} + \dots + l_{A_p} x_{ip} - l_{Y_i}) x_{ij} = 0$$

که در آن $j = 1, \dots, p$ و به ازای $i = 1, \dots, n$ ، $x_{i0} = 1$ منظور شده است. حال از معادلات بالا به ازای $j = 0, 1, \dots, p$ روابط زیر بدست می‌آیند

$$m_{A_0} \sum_{i=1}^n x_{i0} x_{ij} + m_{A_1} \sum_{i=1}^n x_{i1} x_{ij} + \dots + m_{A_p} \sum_{i=1}^n x_{ip} x_{ij} = \sum_{i=1}^n m_{Y_i} x_{ij} \quad (3.3)$$

$$l_{A_0} \sum_{i=1}^n x_{i0} x_{ij} + l_{A_1} \sum_{i=1}^n x_{i1} x_{ij} + \dots + l_{A_p} \sum_{i=1}^n x_{ip} x_{ij} = \sum_{i=1}^n l_{Y_i} x_{ij} \quad (4.3)$$

فرض کنید

$$\mathbf{X} = \begin{pmatrix} x_{10} & x_{11} & \dots & x_{1p} \\ x_{20} & x_{21} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n0} & x_{n1} & \dots & x_{np} \end{pmatrix}$$

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} \sum_{i=1}^n x_{i0}x_{i0} & \sum_{i=1}^n x_{i1}x_{i0} & \cdots & \sum_{i=1}^n x_{ip}x_{i0} \\ \sum_{i=1}^n x_{i0}x_{i1} & \sum_{i=1}^n x_{i1}x_{i1} & \cdots & \sum_{i=1}^n x_{ip}x_{i1} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n x_{i0}x_{ip} & \sum_{i=1}^n x_{i1}x_{ip} & \cdots & \sum_{i=1}^n x_{ip}x_{ip} \end{pmatrix}$$

$$\mathbf{m}_A = (m_{A_0}, \dots, m_{A_p})' \quad , \quad \mathbf{m}_Y = \left(\sum_{i=1}^n m_{Y_i} x_{i0}, \dots, \sum_{i=1}^n m_{Y_i} x_{ip} \right)'$$

$$\mathbf{l}_A = (l_{A_0}, \dots, l_{A_p})' \quad , \quad \mathbf{l}_Y = \left(\sum_{i=1}^n l_{Y_i} x_{i0}, \dots, \sum_{i=1}^n l_{Y_i} x_{ip} \right)'$$

در این صورت معادلات (۳.۳) و (۴.۳) را می‌توان با استفاده از نماد ماتریسی به صورت زیر نمایش داد

$$(\mathbf{X}'\mathbf{X})\mathbf{m}_A = \mathbf{m}_Y \quad , \quad (\mathbf{X}'\mathbf{X})\mathbf{l}_A = \mathbf{l}_Y \quad (۵.۳)$$

اکنون نتایجی را درباره‌ی وجود و یکتایی جواب دستگاه (۵.۳) بیان می‌کنیم.

قضیه ۱.۳. [۶۵] اگر $Rank(\mathbf{X}) = p + 1$ ، آنگاه $\mathbf{X}'\mathbf{X}$ یک ماتریس معین مثبت است.

بر پایه‌ی قضیه ۱.۳، و با توجه به اینکه اگر $\mathbf{X}'\mathbf{X}$ همیشه مثبت باشد دارای معکوس $(\mathbf{X}'\mathbf{X})^{-1}$ است، نتیجه زیر بدست می‌آید.

نتیجه ۱.۳. اگر $Rank(\mathbf{X}) = p + 1$ ، آنگاه دستگاه (۵.۳) دارای جواب یکتا به صورت زیر است

$$\hat{\mathbf{m}}_A = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{m}_Y \quad , \quad \hat{\mathbf{l}}_A = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{l}_Y \quad (۶.۳)$$

بر اساس نتیجه‌ی ۱.۳، قضیه زیر را در مورد مینیمم‌سازی عبارت (۲.۳) داریم.

قضیه ۲.۳. [۶۵] اگر $Rank(\mathbf{X}) = p + 1$ و $(\mathbf{X}'\mathbf{X})^{-1} \mathbf{l}_Y \geq 0$ ، آنگاه مسئله مینیمم‌سازی عبارت (۲.۳) دارای جواب یکتاست که از رابطه (۶.۳) به دست می‌آید.

ملاحظه ۱.۳. شرط $(\mathbf{X}'\mathbf{X})^{-1} \mathbf{l}_Y \geq 0$ در قضیه ۲.۳ تضمین می‌کند که مقادیر l_{A_j} ها، یعنی پهنای ضرایب فازی مدل، نامنفی هستند.

مثال ۱.۳. داده‌های جدول ۱.۳ مربوط به یک بررسی خاک شناسی هستند که درباره‌ی تأثیر x : (SAR - نسبت جذب سدیم) بر y : (ESP - درصد تبادل سدیم) در ناحیه‌ی سیلاخور لرستان در سال ۱۳۸۰ انجام شده است [۴۹]. به دلیل عدم نتیجه‌گیری با دقت کافی در اندازه‌گیری‌های ESP ، مشاهدات مربوط به متغیر y مبهم هستند. در واقع، همانطور که در جدول ۱.۳ درج شده است، این مشاهدات اعداد فازی مثلثی متقارن هستند که ابهام هر عدد نیز برابر 0.1 مقدار نمای آن است که مقدار k عددی تصادفی از توزیع $U(0, 1)$ است. می‌خواهیم یک مدل بهینه برای رابطه ESP بر حسب SAR بیابیم.

جدول ۱.۳: داده‌های مربوط به مثال ۱.۳

ردیف	x_i	$y_i = (m_{y_i}, l_{y_i})_T$	ردیف	x_i	$y_i = (m_{y_i}, l_{y_i})_T$
۱	۰/۷۸	(۳/۰۸, ۰/۳۱)	۱۳	۰/۷۱	(۵/۲۳, ۰/۵۲)
۲	۰/۶۴	(۲/۸۶, ۰/۲۹)	۱۴	۰/۵	(۵/۱۶, ۰/۵۲)
۳	۰/۶۲	(۶/۲۵, ۰/۶۳)	۱۵	۰/۷۷	(۱۱/۱۰, ۱/۱۱)
۴	۰/۴۹	(۴/۱۱, ۰/۴۱)	۱۶	۰/۹۹	(۴/۴۷, ۰/۴۵)
۵	۱/۱۰	(۱/۰۴, ۰/۱۰)	۱۷	۳/۵۶	(۲۸/۸۴, ۲/۸۸)
۶	۰/۶۱	(۲/۷۱, ۰/۲۷)	۱۸	۰/۸۶	(۹/۴۳, ۰/۹۴)
۷	۰/۷۴	(۴/۴۵, ۰/۴۵)	۱۹	۰/۶۱	(۴/۵, ۰/۴۵)
۸	۱/۱۵	(۶/۹۲, ۰/۶۹)	۲۰	۰/۶۴	(۹/۳۰, ۰/۹۴)
۹	۱/۰۸	(۷/۴۱, ۰/۷۴)	۲۱	۰/۷۱	(۹/۴۸, ۰/۹۵)
۱۰	۰/۳۸	(۹/۰۸, ۰/۹۱)	۲۲	۰/۶۱	(۳/۶۵, ۰/۳۷)
۱۱	۰/۶۱	(۶/۵۶, ۰/۶۶)	۲۳	۰/۶۳	(۱۰/۱۴, ۱/۰۱)
۱۲	۰/۹۸	(۵/۰۵, ۰/۵۱)	۲۴	۱/۱۳	(۳, ۰/۳)

برای یافتن مدل بهینه لازم است ماتریس‌های X و $X'X$ و بردارهای m_Y و l_Y را به دست آوریم. با انجام یک سری عملیات جبری ماتریسی داریم:

$$X = \begin{pmatrix} 1 & x_{11} \\ 1 & x_{21} \\ \vdots & \vdots \\ 1 & x_{n1} \end{pmatrix} = \begin{pmatrix} 1 & 0/78 \\ 1 & 0/64 \\ \vdots & \vdots \\ 1 & 1/13 \end{pmatrix}_{24 \times 2}$$

$$X'X = \begin{pmatrix} 24 & 20/9 \\ 20/9 & 26/81 \end{pmatrix}$$

$$m_Y = \left(\sum_{i=1}^{24} m_{y_i} x_{i0}, \sum_{i=1}^{24} m_{y_i} x_{i1} \right)^T = (164/09, 202/32)^T$$

$$l_Y = \left(\sum_{i=1}^{24} l_{y_i} x_{i0}, \sum_{i=1}^{24} l_{y_i} x_{i1} \right)^T = (16/41, 20/23)'$$

چون $|X'X| \neq 0$ ، ماتریس $X'X$ معکوس‌پذیر است و لذا دستگاه (۵.۳) یعنی

$$(X'X)m_A = m_Y, \quad (X'X)l_A = l_Y$$

دارای جواب یکتاست. طبق محاسبات بالا این جواب‌ها عبارتند از

$$\hat{\mathbf{m}}_A = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{m}_Y = (0/8265, 6/5021)'$$

$$\hat{\mathbf{l}}_A = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{l}_Y = (0/826, 0/6902)'$$

بنابراین مدل بهینه برای داده‌های جدول ۱.۳ عبارت است از

$$\begin{aligned} Y &= A_0 + A_1x \\ &= (0/8265, 0/826)_T + (6/5021, 0/6902)_Tx \end{aligned}$$

مدل بالا علاوه بر اینکه یک الگو برای رابطه‌ی بین دو متغیر x و y را ارائه می‌دهد، برای پیش‌گویی مقدار y بر پایه‌ی x نیز قابل استفاده است. برای مثال فرض کنید مقدار x برابر ۰/۷۵ مشاهده شود، آنگاه پیش‌گویی می‌کنیم که مقدار y یعنی ESP برابر با

$$\begin{aligned} \hat{Y} &= (0/8265, 0/826)_T + (6/5021, 0/6902)_T \times 0/75 \\ &= (5/70, 0/60)_T \end{aligned}$$

باشد. به سخن دیگر مقدار y را حدوداً ۵/۷۰ (با ابهام ۰/۶) پیش‌گویی می‌کنیم.

اگر با توجه به فاصله‌ی دیاموند که در بخش ۳.۴.۱ مطرح شد مدل (۱.۳) را به داده‌ها برازش دهیم FLS ، برحسب برآورد پارامترها به شکل زیر به دست می‌آید:

$$\begin{aligned} &d(A_0 + A_1x_{i1} + \dots + A_px_{ip}, Y_i)^2 \\ &= (m_{A_0} + m_{A_1}x_{i1} + \dots + m_{A_p}x_{ip} - m_{Y_i})^2 \\ &+ (m_{A_0} + m_{A_1}x_{i1} + \dots + m_{A_p}x_{ip} - l_{A_0} - l_{A_1}x_{i1} - \dots - l_{A_p}x_{ip} - m_{Y_i} + l_{Y_i})^2 \\ &+ (m_{A_0} + m_{A_1}x_{i1} + \dots + m_{A_p}x_{ip} + r_{A_0} + r_{A_1}x_{i1} + \dots + r_{A_p}x_{ip} - m_{Y_i} - r_{Y_i})^2 \quad (7.3) \end{aligned}$$

می‌توان پارامترها را توسط مشتق جزئی نسبت به m_{A_j} ، l_{A_j} و r_{A_j} ها ($j = 0, 1, \dots, p$) و برابر صفر قراردادن آن‌ها یا به عبارت دیگر به وسیله مینیمم سازی رابطه‌ی (۲.۳)، برآورد کرد. در نتیجه FLS ، بر حسب برآورد پارامترها به دست می‌آید.

۲.۱.۳ نیکویی برازش

در رگرسیون فازی یک موضوع مهم، توان تشریح مدل می‌باشد. معیار نیکویی برازش برای اندازه‌گیری برازش یک مدل رگرسیونی به کار می‌رود. در این بخش به معرفی معیار نیکویی برازش میانگین توان دوم خطای پیش‌گویی که در این پایان نامه از آن استفاده کرده‌ایم می‌پردازیم.

میانگین توان دوم خطای پیشگویی

میانگین توان دوم خطای پیشگویی^۵ (MPE) به عنوان معیاری برای ارزیابی نیکویی برازش استفاده می شود [۵]. در زیر، معیار MPE را در قالب یک تعریف می آوریم.

تعریف ۱.۳. تحت مفروضات مدل رگرسیونی فازی (۱.۳) اگر $\hat{Y}_i$ مقدار پیشگویی Y_i باشد آنگاه MPE برابر است با

$$MPE = \frac{1}{n} \sum_{i=1}^n d^{\vee}(Y_i, \hat{Y}_i) \quad (۸.۳)$$

که در آن $d(\cdot)$ فاصله بین دو عدد فازی مطرح شده در فصل ۱ است.

^۵Mean squares of prediction error

فصل ۴

تحلیل رگرسیون استوار فازی بر پایه‌ی رتبه‌بندی مجموعه‌های فازی

هنگامی که مشاهدات، فازی بوده و در مجموعه داده‌ها، مشاهدات دور افتاده وجود داشته باشند از روش‌های جایگزین استوار فازی استفاده می‌شود. در این فصل، برای حالتی که متغیرهای وابسته و ضرایب رگرسیونی اعداد فازی بوده و مجموعه داده‌ها حاوی مشاهدات دور افتاده است، یک روش رگرسیون استوار فازی بر پایه‌ی رتبه‌بندی چانگ و لی به دو شیوه‌ی کمترین توان‌های دوم انحرافات فازی اصلاح شده [۳، ۴۱] و کمترین قدر مطلق انحرافات فازی اصلاح شده [۲] مطرح می‌شود و در نهایت با توجه به مثال‌های آمده برتری شیوه‌ی کمترین قدر مطلق انحرافات فازی اصلاح شده نسبت به شیوه‌ی کمترین توان‌های دوم انحرافات فازی اصلاح شده نشان داده می‌شود.

۱.۴ مقدمه

همان‌طور که در فصل ۲ بیان شد، روش کمترین توان‌های دوم (LS)، 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 LS ، پیشگویی‌های

خوبی را نتیجه نمی‌دهد. گاهی می‌توان مشاهدات دور افتاده را از مجموعه داده‌ها حذف کرد و پس از آن روش کلاسیک را به کار برد، ولی در برخی تحقیقات، این مشاهدات حاوی اطلاعات مفیدی هستند؛ لذا حذف آن‌ها امکان پذیر نیست و در این صورت از روش‌های جایگزین رگرسیون استوار استفاده می‌شود. در واقع در این موارد، روش‌های استوار به روش LS ترجیح داده می‌شوند. و همان‌طور که در فصل ۳ بیان شد، 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. بنابراین هنگامی که در مجموعه داده‌ها، مشاهدات دور افتاده وجود داشته باشند و متغیرها یا مشاهدات مربوط به آن‌ها نادقیق باشند از تحلیل رگرسیون استوار فازی استفاده می‌شود.

یانگ و لین [۶۸]، تحلیل رگرسیون FLS را برای ورودی فازی و خروجی فازی پیشنهاد کردند. آن‌ها کاربردی از روش خوشه بندی را برای تعیین ناهمگنی مشاهدات و داده‌های دور افتاده ارائه کردند. یانگ و لیو [۶۹] الگوریتم FLS را برای مدل‌های رگرسیون خطی با اثر متقابل پیشنهاد کردند. این الگوریتم برای برآورد مدل‌های FLR^۱ به‌خصوص برای رگرسیون ساده، استوار است. در این الگوریتم، شرایط تعامد اضافه شده تا مسأله بهینه سازی را محدود کند. هانگ و یانگ [۳۴] به‌منظور بررسی مسأله داده‌های دور افتاده، یک شیوه‌ی حذف برای روش برنامه ریزی خطی تاناکا پیشنهاد کردند. این روش، قدرت امتحان کردن رفتار میزان تغییرات در تابع هدف مدل رگرسیون فازی را زمانی که مشاهدات حذف می‌شوند داشت.

همان‌طور که در فصل در بخش ۵.۱ بیان شد، مقایسه با رتبه‌بندی اعداد فازی یا مجموعه‌های فازی برای اهداف عملی بسیار مهم هستند. زمانی که نظریه‌ی مجموعه فازی برای ساختن مدل‌های ریاضی به کار می‌رود با مسایلی چون مقایسه یا رتبه‌بندی اعداد فازی مواجه می‌شویم. اعداد فازی نظم خطی ندارند و بنابراین 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 h - برش‌ها استفاده می‌کنند بنابراین ممکن است از اطلاعات کافی استفاده نکرده و نتایج ارببی حاصل شوند. به‌منظور استفاده از همه اطلاعات ممکن، باید همه‌ی h - برش‌ها در نظر گرفته شوند. به دلیل پیچیدگی و اشکالات مسأله، یک روش کلی توسط چانگ و لی [۱۷] پیشنهاد شد که در آن با استفاده از مفهوم حضور (وجود)، رتبه‌بندی مجموعه‌های فازی، فرمول‌بندی می‌شود. این روش را برای همه نوع تابع عضویت مثل محدب، غیر محدب، نرمال، غیرنرمال، پیوسته یا پیوسته‌ی تکه‌ای می‌توان بکار برد. در تحلیل رگرسیون فازی بر مبنای کمترین قدر مطلق انحرافات می‌توان به مقالات [۶۲، ۵۷، ۴۰، ۴۴] اشاره کرد. در این فصل، یک تحلیل رگرسیون استوار فازی (FRR)^۲ به دو شیوه کمترین توان‌های دوم انحرافات فازی اصلاح شده [۳] و کمترین قدر مطلق انحرافات فازی اصلاح شده که تعمیمی از شیوه‌ی قبل است [۲] پیشنهاد می‌شود که در آن‌ها از روش رتبه‌بندی فازی

^۱ Fuzzy linear regression^۲ Fuzzy robust regression

چانگ و لی [۱۷] برای رتبه دادن به باقی‌مانده‌ها استفاده می‌شود و با تعریف توابع عضویت برای باقی‌مانده‌ها ماتریس‌های وزن به‌دست می‌آیند. از آن‌جا که هر باقی‌مانده با توجه به درجه عضویت، روی مدل تاثیر خواهد گذاشت مدل‌های پیشنهادی روش FRR خواهند بود.

۲.۴ رگرسیون کمترین مجموع فاصله‌های خطای فازی

هدف آن است که بر پایه‌ی مشاهداتی به‌صورت $(Y_i, x_{i1}, x_{i2}, \dots, x_{ip})$ که $x_{ij} \in \mathbb{R}$ به ازای $i = 1, \dots, n$ و $j = 1, \dots, p$ و Y_i ها اعداد فازی مثلثی به‌شکل $(m_{Y_i}, l_{Y_i}, r_{Y_i})_T$ هستند، یک مدل بهینه با ضرایب فازی به‌صورت

$$Y = A_0 + A_1 x_1 + \dots + A_p x_p \quad (1.4)$$

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 پایه‌ی آن، به‌دست آوریم. با استفاده از قضیه‌ی ۲.۱ که در بخش ۳.۱ بیان شد و با فرض مثبت بودن x_i ها به ازای $i = 1, \dots, n$ مدل (۱.۴) به‌شکل زیر به‌دست می‌آید

$$y = (m_{A_0} + m_{A_1} x_1 + \dots + m_{A_p} x_p, l_{A_0} + l_{A_1} x_1 + \dots + l_{A_p} x_p, r_{A_0} + r_{A_1} x_1 + \dots + r_{A_p} x_p)_T \quad (2.4)$$

روش کمترین مجموع فاصله‌های خطای فازی با استفاده از فاصله بین دوعدد فازی تعریف شده در بخش ۱۲.۱ و مینیمم کردن مجموع فاصله مشاهدات از مقادیر برآورد شده آن‌ها، به‌دست آمده و پارامترهای مدل برآورد می‌شوند. یعنی

$$M(A_0, A_1, \dots, A_p) = \min \sum_{i=1}^n D(A_0 + A_1 x_{i1} + \dots + A_p x_{ip}, Y_i) \quad (3.4)$$

فاصله (D) به‌کار رفته در شیوه‌ی کمترین توان‌های دوم فازی اصلاح شده در این پایان نامه فاصله دیاموند می‌باشد که در بخش ۳.۴.۱ به آن اشاره شد که به‌صورت زیر می‌باشد

$$\begin{aligned} D &= d(A_0 + A_1 x_{i1} + \dots, A_p x_{ip}, Y_i)^2 \\ &= (m_{A_0} + m_{A_1} x_{i1} + \dots + m_{A_p} x_{ip} - m_{Y_i})^2 \\ &\quad + (m_{A_0} + m_{A_1} x_{i1} + \dots + m_{A_p} x_{ip} - l_{A_0} - l_{A_1} x_{i1} - \dots - l_{A_p} x_{ip} - m_{Y_i} + l_{Y_i})^2 \\ &\quad + (m_{A_0} + m_{A_1} x_{i1} + \dots + m_{A_p} x_{ip} + r_{A_0} + r_{A_1} x_{i1} + \dots + r_{A_p} x_{ip} - m_{Y_i} - r_{Y_i})^2 \end{aligned} \quad (4.4)$$

و فاصله (D) به کار رفته در شیوه‌ی کمترین قدر مطلق فازی اصلاح شده فاصله حسن‌پور می‌باشد که در بخش ۵.۴.۱ معرفی شد. بنابراین داریم

$$\begin{aligned} D &= d(A_0 + A_1 x_{i1} + \dots, A_p x_{ip}, Y_i) \\ &= |m_{A_0} + m_{A_1} x_{i1} + \dots + m_{A_p} x_{ip} - m_{Y_i}| + |l_{A_0} + l_{A_1} x_{i1} + \dots + l_{A_p} x_{ip} - l_{Y_i}| \\ &\quad + |r_{A_0} + r_{A_1} x_{i1} + \dots + r_{A_p} x_{ip} - r_{Y_i}| \end{aligned} \quad (5.4)$$

می‌توان پارامترها را به وسیله مینیمم سازی رابطه‌ی (۳.۴)، برآورد کرد. در نتیجه برآورد پارامترها به شیوه‌های کمترین توان‌های دوم انحرافات فازی (FLS) و کمترین قدر مطلق انحرافات فازی ($FLAD$)^۳ به دست می‌آیند.

محاسبه‌ی باقیمانده مدل رگرسیون برای اعداد فازی مثلثی $\hat{Y}_i$ و Y_i

باقیمانده‌ها که فاصله‌ی بین Y_i مشاهده شده و $\hat{Y}_i$ برآورد شده را بیان می‌کنند، نقش مهمی در برآورد مدل رگرسیونی دارند. در این بخش، Y_i مشاهده شده و $\hat{Y}_i$ برآورد شده به عنوان دو عدد فازی مثلثی در نظر گرفته شده و فاصله‌ی بین آن‌ها به عنوان باقیمانده‌ی مدل رگرسیونی به صورت زیر به دست می‌آید.

با استفاده از قضیه‌ی ۸.۱، اگر $Y_i = (m_{Y_i}, l_{Y_i}, r_{Y_i})_T$ و $\hat{Y}_i = (m_{\hat{Y}_i}, l_{\hat{Y}_i}, r_{\hat{Y}_i})_T$ متغیر پاسخ و پیشگوی مدل رگرسیونی باشند، باقی مانده‌ی مدل به صورت

$$\begin{aligned} E_i &= d(\hat{Y}_i, Y_i) \\ &= \chi_1 \left((m_{\hat{Y}_i} - m_{Y_i}) - \frac{1}{3}(l_{\hat{Y}_i} - l_{Y_i}) \right) + \chi_2 \left((m_{\hat{Y}_i} - m_{Y_i}) + \frac{1}{3}(r_{\hat{Y}_i} - r_{Y_i}) \right) \end{aligned} \quad (6.4)$$

به دست می‌آیند. هم‌چنین اگر $Y_i = (m_{Y_i}, s_{Y_i})_T$ و $\hat{Y}_i = (m_{\hat{Y}_i}, s_{\hat{Y}_i})_T$ آن‌گاه

$$E_i = d(\hat{Y}_i, Y_i) = (\chi_1 + \chi_2)(m_{\hat{Y}_i} - m_{Y_i}) + \frac{1}{3}(\chi_2 - \chi_1)(s_{\hat{Y}_i} - s_{Y_i}) \quad (7.4)$$

و در نهایت طبق نتیجه‌ی ۴.۱ باقی مانده به صورت

$$E_i = d(\hat{Y}_i, Y_i) = m_{\hat{Y}_i} - m_{Y_i} \quad (8.4)$$

به ازای $i = 1, \dots, n$ حاصل می‌شود.

۱.۲.۴ الگوریتم کولا و آپایدین برای به دست آوردن ضرایب رگرسیون استوار فازی

کولا و آپایدین [۴۱]، الگوریتمی را برای محاسبه‌ی ضرایب مدل رگرسیون استوار فازی بدین صورت مطرح کردند: ابتدا یک مدل رگرسیون چند گانه برای حالتی که X عددی دقیق و ضرایب

^۳Fuzzy Least Absolute Deviation

رگرسیونی و Y ، اعداد فازی مثلثی متقارن هستند و در مجموعه داده‌ها مشاهده دور افتاده داریم، در نظر گرفته می‌شود. سپس ضرایب مدل رگرسیون از روش کمترین مربعات فازی معمولی محاسبه و مقادیر اعداد فازی $\hat{Y}_i$ را به دست می‌آوریم. برای محاسبه‌ی باقی‌مانده‌ها از فاصله‌ی به دست آمده در روابط (۶.۴) – (۸.۴) استفاده می‌کنیم. سپس با استفاده از تابع عضویت تعریف شده برای باقی‌مانده‌ها، ماتریس وزن به دست آمده و در نهایت، برآوردهای WFLS به وسیله‌ی ماتریس وزن ساخته می‌شوند. برای رگرسیون کلاسیک، برآوردهای WLS در بخش ۱.۲.۲ توضیح داده شد. ما در این پایان نامه نیز برای برآورد پارامترهای فازی، از این روش استفاده می‌کنیم؛ به این صورت که برای مراکز و پهناهای Y_i وزن W را در نظر می‌گیریم؛ که W همان ماتریس وزن است و از طریق محاسبه‌ی تابع عضویت باقی‌مانده‌ها به دست می‌آید. بنابراین مقدار برآورد مراکز و پهناهای پارامترها به ترتیب به صورت بسته می‌آیند که p تعداد پارامترهاست. بنابراین برآورد مدلی به دست می‌آید که تحت تأثیر داده‌های دور افتاده قرار نمی‌گیرد. برای کسب اطلاعات بیشتر در زمینه WFLS به [۱۸] مراجعه کنید. در زیر گام‌های الگوریتم آمده است.

الگوریتم

برای مدل‌های رگرسیون برآورد شده، گام‌های الگوریتم روش FRR به صورت زیر پیشنهاد می‌شود [۴۱، ۳].

گام ۱: وقتی x_i دقیق و $Y_i = (m_{Y_i}, s_{Y_i})_T$ یک عدد فازی مثلثی متقارن باشد، از رابطه‌ی (۲.۳)، برآورد اولیه FLS را می‌یابیم.

گام ۲: مقادیر برآورد مرکز عدد فازی $(m_{\hat{Y}_i})$ و باقیمانده‌ها (E_i) را محاسبه می‌کنیم. باقیمانده‌ها را از رابطه‌ی (۸.۴)، به دست می‌آوریم.

گام ۳: میانه را بر حسب قدر مطلق مقدار باقیمانده‌ها محاسبه می‌کنیم. فاصله‌ها از دستور زیر به دست می‌آیند.

$$D_i = ||abs(E_i) - median(abs(E_i))||, \quad i = 1, 2, \dots, n \quad (9.4)$$

که $||.||$ فاصله اقلیدسی است.

گام ۴: تابع عضویت باقیمانده‌ها را به شکل زیر تعریف می‌کنیم

$$E = \begin{cases} 1 & abs(E) \leq a \\ \frac{b-abs(E)}{b-a} & a < abs(E) < b \\ 0 & OW \end{cases} \quad (10.4)$$

و $a = median(D_i)$ و $b = \max(D_i) + d$ که در آن d ، پارامتر استوار تغییرپذیری معرفی شده در معادله‌ی (۵.۲) 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 به‌دست می‌آوریم.

گام ۶: اگر $|\hat{A}^{k+1} - \hat{A}^k| < \epsilon$ ، الگوریتم را متوقف می‌کنیم. در غیر این‌صورت به گام ۲ می‌رویم. $\epsilon > 0$ ، عددی کوچک است که توسط تصمیم گیرنده تعیین می‌شود.

$\hat{A}$ برآورد پارامترهای رگرسیون، k تعداد تکرارها و $\epsilon > 0$ عدد بسیار کوچکی است که توسط تصمیم گیرنده تعیین می‌شود (در اینجا $\epsilon = 0.05$ ، در نظر گرفته شده است).

شاخص نیکویی برازش

در این فصل از شاخص نیکویی برازش MPE که در بخش ۲.۱.۳ معرفی شد استفاده کرده و d را فاصله حسن پور در نظر می‌گیریم (به بخش ۱۲.۱ رجوع کنید). سپس با تعریف این شاخص و مقایسه‌ی وزن‌های به‌دست آمده، برتری این مدل، نسبت به مدل FLS نشان داده خواهد شد.

۲.۲.۴ مثال‌های عددی

در این قسمت، دو مثال عددی برای نشان دادن کاربرد روش پیشنهادی، به‌کار می‌رود.

داده‌های شبیه‌سازی شده

این مسئله شامل ۳ متغیر مستقل و ۳۰ مشاهده از داده‌های شبیه‌سازی شده با توزیع نرمال است به طوری که $X_3 \sim N(\mu = 32, \sigma = 12)$ و $X_2 \sim N(\mu = 50, \sigma = 12)$ ، $X_1 \sim N(\mu = 20, \sigma = 3)$ (۱۳ می‌باشند). ۱۵-امین مشاهده در متغیرهای وابسته با $y_{15} + 20$ جایگزین شده و به‌عنوان یک داده دور افتاده در نظر گرفته می‌شود. مجموعه داده‌ها در جدول ۱.۴ نشان داده شده است.

برای این مجموعه داده، از طریق مدل رگرسیون فازی FLS با فاصله‌ی دیاموند، 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 پهناها که به‌صورت $s_i = \frac{y_i}{\lambda}$ ($i = 1, 2, \dots, n$) تعریف می‌شوند، محاسبه می‌گردند. باقیمانده‌ها و عناصر روی قطر ماتریس وزن، در جدول ۱.۴ نشان داده شده‌اند. برای مدل رگرسیون فازی FLS و روش پیشنهادی، برنامه‌ای در نرم افزار متمتیکا^۴ نوشته شده و به‌کار می‌رود [۳]. برآوردهای مدل رگرسیونی در جدول ۲.۴ آمده است. در جدول ۱.۴، وزن ۱۵-امین مشاهده که یک داده دور افتاده است، بر اساس

^۴.Mathematica

جدول ۱.۴: مجموعه داده‌ها، وزن‌ها (W_i) و مقایسه مدل FLS و مدل پیشنهادی با توجه به شاخص MPE برای داده‌های شبیه سازی شده

مشاهده	x_1	x_2	x_3	(y_i, s_i)	W_{FLS}	W_{Prop}	$\hat{Y}_{FLS}$	$\hat{Y}_{Prop}$	d_{FLS}	d_{Prop}
۱	۲۴/۰۲	۶۶/۵۷	۳۰/۴۸	(۱۱۹/۱۳, ۱۴/۸۹)	۱	۰/۹۷	(۱۲۳/۱۰, ۱۵/۳۹)	(۱۲۱/۶۷, ۱۵/۲۱)	۴/۹۶	۳/۱۸
۲	۲۱/۱۶	۶۵/۸۴	۳۴/۱۹	(۱۲۰/۱۳, ۱۵/۰۲)	۱	۱	(۱۲۴/۰۴, ۱۵/۵۱)	(۱۲۱/۳۹, ۱۵/۱۷)	۴/۸۹	۱/۵۸
۳	۲۱/۱۸	۳۹/۰۹	۲۵/۴۸	(۹۰/۷۷, ۱۱/۳۵)	۱	۰/۹۶	(۸۸/۲۹, ۱۱/۰۴)	(۸۷/۸۳, ۱۰/۹۸)	۳/۱۰	۳/۶۸
۴	۱۴/۸۸	۲۲/۳۳	۲۲/۸۳	(۵۸/۳۲, ۷/۲۹)	۱	۰/۹۲	(۶۴/۷۷, ۸/۱۰)	(۶۲/۶۴, ۷/۸۳)	۸/۰۷	۵/۴۱
۵	۲۰/۶۸	۷۱/۴۶	۳۸/۶۱	(۱۳۰/۵۲, ۱۶/۳۱)	۱	۱	(۱۳۳/۶۳, ۱۶/۷۰)	(۱۳۰/۳۷, ۱۶/۳۰)	۳/۸۹	۰/۱۸
۶	۲۲/۰۶	۵۴/۶۹	۲۶/۵۳	(۱۰۲/۷۶, ۱۲/۸۴)	۱	۰/۹۹	(۱۰۵/۹۹, ۱۳/۲۵)	(۱۰۴/۵۶, ۱۳/۰۷)	۴/۰۵	۲/۲۶
۷	۱۸/۰۹	۵۰/۲۴	۳۴/۹۸	(۱۰۹/۱۷, ۱۳/۶۵)	۱	۰/۹۰	(۱۰۶/۷۳, ۱۳/۳۴)	(۱۰۴/۰۳, ۱۳/۰۰)	۳/۰۵	۶/۴۳
۸	۱۶/۹۹	۴۵/۱۳	۱۹/۵۳	(۸۲/۸۹, ۱۰/۳۶)	۱	۱	(۸۶/۷۳, ۱۰/۸۴)	(۸۳/۴۴, ۱۰/۴۳)	۴/۸۰	۰/۶۹
۹	۱۹/۴۴	۳۱/۵۸	۳۰/۱۰	(۷۸/۸۸, ۹/۸۶)	۱	۰/۹۲	(۸۳/۶۶, ۱۰/۴۶)	(۸۳/۱۸, ۱۰/۴۰)	۵/۹۷	۵/۳۷
۱۰	۱۶/۸۴	۵۲/۶۶	۴۱/۶۸	(۱۰۸/۸۰, ۱۳/۶۰)	۱	۰/۹۶	(۱۱۴/۶۰, ۱۴/۳۳)	(۱۱۱/۲۷, ۱۳/۹۱)	۷/۲۵	۳/۰۹
۱۱	۱۹/۷۹	۳۳/۵۱	۲۰/۴۲	(۷۹/۹۲, ۹/۹۹)	۱	۰/۹۴	(۷۷/۰۷, ۹/۶۳)	(۷۶/۲۷, ۹/۵۳)	۳/۵۶	۴/۵۷
۱۲	۲۰/۸۴	۳۹/۹۳	۳۳/۸۱	(۱۰۱/۳۹, ۱۲/۶۷)	۱	۰/۹۰	(۹۶/۵۳, ۱۲/۰۷)	(۹۶/۰۹, ۱۲/۰۱)	۶/۰۷	۶/۶۲
۱۳	۲۴/۱۲	۴۷/۵۰	۲۸/۹۳	(۱۰۶/۴۵, ۱۳/۳۱)	۱	۰/۹۳	(۱۰۱/۸۹, ۱۲/۷۴)	(۱۰۲/۲۲, ۱۲/۷۸)	۵/۷۱	۵/۲۹
۱۴	۲۰/۵۴	۵۹/۰۷	۳۱/۰۲	(۱۰۷/۸۶, ۱۳/۴۸)	۱	۰/۹۵	(۱۱۳/۷۵, ۱۴/۲۲)	(۱۱۱/۳۱, ۱۳/۹۱)	۷/۳۶	۴/۳۲
۱۵	۱۸/۳۷	۵۴/۵۱	۲۷/۳۴	(۱۴۰/۵۸, ۱۷/۵۷)	۱	۰	(۱۰۴/۴۰, ۱۳/۰۵)	(۱۰۱/۱۹, ۱۲/۶۵)	۴۵/۲۳	۴۹/۲۴
۱۶	۲۴/۹۰	۳۳/۸۶	۴/۹۹	(۶۴/۱۵, ۸/۰۲)	۱	۰/۹۵	(۶۶/۳۹, ۸/۳۰)	(۶۷/۵۸, ۸/۴۵)	۲/۸۰	۴/۲۹
۱۷	۲۲/۴۸	۶۷/۷۸	۳۰/۱۳	(۱۲۲/۳۱, ۱۵/۲۹)	۱	۱	(۱۲۳/۱۴, ۱۵/۳۹)	(۱۲۰/۸۳, ۱۵/۱۰)	۱/۰۴	۱/۸۵
۱۸	۲۰/۶۹	۵۰/۳۹	۵۰/۱۱	(۱۲۰/۴۸, ۱۵/۰۶)	۱	۱	(۱۲۲/۱۵, ۱۵/۲۷)	(۱۲۱/۲۲, ۱۵/۱۵)	۲/۰۹	۰/۹۲
۱۹	۲۲/۰۱	۷۲/۴۵	۴۰/۴۷	(۱۳۳/۰۰, ۱۶/۶۳)	۱	۱	(۱۳۷/۱۲, ۱۷/۱۴)	(۱۳۴/۵۰, ۱۶/۸۱)	۵/۱۵	۱/۸۷
۲۰	۱۸/۴۸	۳۵/۴۹	۲۷/۱۰	(۸۰/۷۲, ۱۰/۰۹)	۱	۰/۹۸	(۸۴/۴۴, ۱۰/۵۶)	(۸۳/۰۲, ۱۰/۳۸)	۴/۶۴	۲/۸۷
۲۱	۲۲/۵۷	۴۰/۶۱	۲۳/۴۰	(۹۰/۸۳, ۱۱/۳۵)	۱	۰/۹۹	(۸۸/۷۹, ۱۱/۱۰)	(۸۸/۸۱, ۱۱/۱۰)	۲/۵۴	۲/۵۲
۲۲	۲۰/۸۱	۴۰/۷۹	۳۵/۲۴	(۱۰۴/۹۵, ۱۳/۱۲)	۱	۰/۸۶	(۹۸/۷۱, ۱۲/۳۴)	(۹۸/۲۲, ۱۲/۲۸)	۷/۸۱	۸/۴۱
۲۳	۲۱/۸۷	۴۸/۷۱	۲۷/۰۱	(۹۹/۲۹, ۱۲/۴۱)	۱	۱	(۱۰۰/۱۱, ۱۲/۵۱)	(۹۹/۱۵, ۱۲/۳۹)	۱/۰۲	۰/۱۷
۲۴	۱۶/۸۶	۳۸/۲۸	۲۵/۱۳	(۸۰/۴۶, ۱۰/۰۶)	۱	۱	(۸۴/۶۱, ۱۰/۵۸)	(۸۲/۰۷, ۱۰/۲۶)	۵/۱۸	۲/۰۱
۲۵	۲۴/۶۱	۳۸/۴۳	۳۲/۷۲	(۹۸/۸۶, ۱۲/۳۶)	۱	۱	(۹۶/۱۸, ۱۲/۰۲)	(۹۷/۷۱, ۱۲/۲۱)	۳/۳۶	۱/۴۴
۲۶	۲۱/۳۰	۲۱/۴۵	۴۸/۳۰	(۹۰/۵۹, ۱۱/۳۲)	۱	۰/۹۸	(۹۰/۷۳, ۱۱/۳۴)	(۹۲/۷۰, ۱۱/۵۹)	۰/۱۷	۲/۶۳
۲۷	۱۴/۲۵	۳۹/۹۴	-۰/۷۶	(۶۰/۰۲, ۷/۵۰)	۱	۰/۹۴	(۶۱/۳۱, ۷/۶۶)	(۵۶/۴۷, ۷/۰۶)	۱/۶۱	۴/۴۴
۲۸	۲۱/۴۱	۵۳/۰۹	۳۹/۶۰	(۱۱۸/۴۲, ۱۴/۸۰)	۱	۰/۹۴	(۱۱۵/۸۳, ۱۴/۴۸)	(۱۱۴/۶۵, ۱۴/۳۳)	۳/۲۴	۴/۷۱
۲۹	۲۳/۸۲	۴۷/۷۹	۱۸/۹۰	(۸۷/۸۳, ۱۰/۹۸)	۱	۰/۹۱	(۹۲/۹۱, ۱۱/۶۱)	(۹۲/۷۴, ۱۱/۵۹)	۶/۳۵	۶/۱۳
۳۰	۲۱/۹۲	۴۷/۹۹	۴۴/۲۸	(۱۱۳/۱۷, ۱۴/۱۵)	۱	۱	(۱۱۵/۰۶, ۱۴/۳۸)	(۱۱۴/۷۶, ۱۴/۳۴)	۲/۳۶	۱/۹۸
MPE										
۴/۹۳۸۵/۵۷۷										

روش پیشنهادی تقریباً صفر به‌دست آمد. وزن‌های به‌دست آمده از روش پیشنهادی، درجات عضویت هر مشاهده هستند. این عضویت‌ها اثرهای مشاهدات را روی مدل نشان می‌دهند. همان‌طور که در جدول ۱.۴ نشان داده شده، داده‌های دور افتاده با درجات کوچکی از تابع عضویت، روی مدل تأثیر می‌گذارند. درجات عضویت مشاهدات دیگر، یک یا نزدیک به یک بوده و اثرات آن‌ها روی برآورد مدل رگرسیونی، مهم هستند.

در جدول ۲.۴ 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 ۱.۲.۴ معرفی شد، مقدار MPE برای روش FLS معادل ۵/۵۸ و برای روش پیشنهادی ۴/۹۴ به دست می‌آید. بنابراین با توجه به کمتر شدن مقدار MPE روش پیشنهادی به روش FLS ، برتری روش پیشنهادی مشخص می‌شود (جدول ۱.۴).

جدول ۲.۴: برآورد مدل رگرسیونی داده‌های شبیه سازی شده

روش	برآورد رگرسیونی
FLS	$\hat{y} = (12/1214, 1/5152) + (0/5819, 0/726)x_1 + (1/041, 0/13014)x_2 + (0/9084, 0/1136)x_3$
پیشنهادی	$\hat{y} = (3/9303, 0/4913) + (1/0771, 0/1347)x_1 + (0/94907, 0/1186)x_2 + (0/9425, 0/1177)x_3$

داده‌های سیمان پورتلند

در این مثال اثر ترکیب سیمان پورتلند روی گرمای آزاد شده در حین سفت شدن آن مورد بررسی قرار می‌گیرد [۶۵]. داده‌ها در جدول ۳ نمایش داده شده‌اند. این مسأله با ۱۳ مشاهده شامل ۳ متغیر مستقل می‌باشد (x_1 ، مقدار آلومینات تری کلسیم، x_2 ، مقدار سیلیکات تری کلسیم، x_3 ، مقدار آلومینو فریت تترا کلسیم و (y_i, s_i) مقدار گرمای آزاد شده در هر گرم سیمان است). ۱۱-امین 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 می‌گیریم. مجموعه داده‌ها و وزن‌های روش پیشنهادی در جدول ۳.۴ نشان داده شده و در جدول ۴.۴، برای روش FLS و روش پیشنهادی، برآوردهای مدل رگرسیونی آورده شده‌اند. در جدول ۳.۴ وزن ۱۱-امین مشاهده، در روش پیشنهادی، تقریباً صفر به دست آمد. وزن‌هایی که به عنوان نتیجه روش پیشنهادی پیدا می‌شوند، درجات عضویت هر مشاهده هستند. این عضویت‌ها، اثرات مشاهدات را روی مدل نشان می‌دهند.

جدول ۳.۴: مجموعه داده‌ها، وزن‌ها (W_i) و مقایسه مدل FLS و مدل کولا و آپایدین با توجه به شاخص برای داده‌های مثال سیمان پورتلند

مشاهده	x_1	x_2	x_3	(y_i, s_i)	W_{FLS}	$W_{K.A}$	$\hat{Y}_{FLS}$	$\hat{Y}_{K.A}$	d_{FLS}	$d_{K.A}$
۱	۷	۲۶	۶	(۷۸/۵, ۶/۹)	۱	۱	(۷۹/۸۵, ۶/۸۷)	(۷۸/۷۲, ۶/۸۹)	۱/۴۲	۰/۲۳
۲	۱	۲۹	۱۵	(۷۴/۳, ۶/۴)	۱	۰/۹۳	(۷۰/۷۲, ۶/۲۶)	(۷۲/۲۶, ۶/۲۱)	۳/۸۷	۲/۴۲
۳	۱۱	۵۶	۸	(۱۰۴/۳, ۹/۴)	۱	۱	(۱۰۵/۷۹, ۹/۶۰)	(۱۰۵/۵۳, ۹/۶۰)	۱/۸۹	۱/۶۲
۴	۱۱	۳۱	۸	(۸۷/۶, ۷/۸)	۱	۱	(۸۸/۴۷, ۷/۹۳)	(۸۸/۹۷, ۷/۹۱)	۱/۱۳	۱/۵۸
۵	۷	۵۲	۶	(۹۵/۵, ۸/۶)	۱	۱	(۹۷/۸۷, ۸/۶۰)	(۹۵/۹۴, ۸/۶۵)	۲/۳۸	۰/۵۴
۶	۱۱	۵۵	۹	(۱۰۹/۲, ۹/۹)	۱	۰/۷۳	(۱۰۴/۸۱, ۹/۵۶)	(۱۰۵/۰۳, ۹/۵۴)	۵/۰۷	۴/۸۹
۷	۳	۷۱	۱۷	(۱۰۲/۷, ۹/۳)	۱	۱	(۱۰۲/۱۱, ۹/۴۵)	(۱۰۳/۷۱, ۹/۴۰)	۰/۸۹	۱/۲۱
۸	۱	۳۱	۲۲	(۷۲/۵, ۶/۲)	۱	۱	(۷۰/۰۸, ۶/۵۵)	(۷۴/۷۴, ۶/۴۲)	۳/۱۳	۲/۶۷
۹	۲	۵۴	۱۸	(۹۳/۳, ۸/۳)	۱	۰/۹۰	(۸۸/۶۱, ۸/۱۷)	(۹۰/۹۶, ۸/۱۰)	۴/۹۶	۲/۷۵
۱۰	۲۱	۴۷	۴	(۱۱۵/۹, ۱۰/۶)	۱	۱	(۱۱۵/۰۵, ۱۰/۶۱)	(۱۱۵/۴۵, ۱۰/۵۸)	۰/۸۷	۰/۴۹
۱۱	۱	۴۰	۲۳	(۶۸, ۷/۴)	۱	۰	(۷۶/۰۳, ۷/۱۸)	(۸۰/۸۶, ۷/۰۴)	۸/۴۷	۱۳/۵۹
۱۲	۱۱	۶۶	۹	(۱۱۳/۳, ۱۰/۶)	۱	۱	(۱۱۲/۴۳, ۱۰/۲۹)	(۱۱۲/۳۲, ۱۰/۲۸)	۱/۴۸	۱/۶۲
۱۳	۱۰	۶۸	۸	(۱۰۹/۴, ۹/۹)	۱	۰/۹۰	(۱۱۲/۶۷, ۱۰/۲۳)	(۱۱۱/۸۲, ۱۰/۲۵)	۳/۹۴	۳/۱۱
MPE										
									۳/۰۳	۲/۸۲

همان‌طور که در جدول ۳.۴ قابل مشاهده است داده‌ی دورافتاده با درجه کوچکی از عضویت، روی مدل تأثیر می‌گذارد. 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 می‌باشد (جدول ۳.۴).

جدول ۴.۴: برآورد مدل رگرسیونی داده‌های مثال سیمان پورتلند

روش	برآورد رگرسیونی
FLS	$\hat{y} = (۵۳/۵۳, ۳/۷۹۶) + (۱/۴۳, ۰/۱۷۰)x_1$ $+ (۰/۶۹, ۰/۰۶۶)x_2 + (-۰/۲۸, ۰/۰۲۳)x_3$
کولا و آپایدین	$\hat{y} = (۴۸/۹۳۸۵, ۳/۹۲۴۴) + (۱/۶۵۳۶, ۰/۱۶۳۶)x_1$ $+ (۰/۶۶۲۲, ۰/۰۶۶۷)x_2 + (۰/۱۶۴۲, ۰/۰۱۰۵)x_3$

این الگوریتم تا رسیدن به یک دقت مطلوب برای پارامترها ادامه پیدا می‌کند. در دو مثال مطرح شده در این فصل، وزن‌های به‌دست آمده از روش رگرسیون استوار پیشنهادی کولا و آپایدین و رگرسیون حداقل توان‌های دوم با هم مقایسه شده‌اند (جداول ۱.۴ و ۳.۴). همان‌طور که در این جداول مشاهده می‌شود، داده‌های دور افتاده در روش پیشنهادی کولا و آپایدین بوضوح از وزن‌های کمتری نسبت به روش FLS برخوردارند و همان‌طور که در مثال‌ها بیان شد شاخص نیکویی برازش MPE برتری این روش را نسبت به روش FLS مشخص می‌کند.

۳.۲.۴ الگوریتم پیشنهادی برای به‌دست آوردن ضرایب رگرسیون استوار فازی

این الگوریتم اصلاح شده‌ی الگوریتم کولا و آپایدین است. در این الگوریتم ابتدا بر اساس شیوه‌ی رگرسیون کمترین مجموع قدرمطلق خطای فازی (FLAD) با استفاده از متر حسن‌پور در گام اول الگوریتم کولا و آپایدین، به محاسبه ضرایب اولیه مدل پرداخته و با معرفی تابع عضویت جدید برای باقیمانده‌ها در گام چهارم الگوریتم، یک مدل رگرسیونی به‌دست می‌آوریم. برای نشان دادن عملکرد بهتر این روش پیشنهادی نسبت به روش‌های دیگر مطرح شده در این پایان‌نامه، مثالی مطرح و نتایج و مقایسات حاصل ارائه می‌شود.

الگوریتم

در تحلیل رگرسیون چندگانه، وقتی x_i دقیق و Y_i یک عدد فازی مثلثی متقارن باشد و در مجموعه داده‌ها، مشاهده دور افتاده داشته باشیم، برای مدل‌های رگرسیون برآورد شده، الگوریتم پیشنهادی به‌صورت زیر پیشنهاد می‌شود [۲].

الگوریتم: این الگوریتم دارای گام‌های زیر است:

گام ۱ وقتی x_i دقیق و $Y_i = (m_{Y_i}, l_{Y_i}, r_{Y_i})_T$ یک عدد فازی مثلثی باشد، از رابطه‌ی (۳.۴)، برآورد اولیه FLAD را می‌یابیم.

گام ۲ مقادیر برآورد مرکز عدد فازی ($m_{\hat{Y}_i}$) و باقیمانده‌ها (E_i) محاسبه می‌شوند. باقیمانده‌ها را از رابطه‌ی (۸.۴)، به‌دست می‌آوریم.

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

$$D_i = ||abs(E_i) - median(abs(E_i))||, \quad i = 1, 2, \dots, n \quad (11.4)$$

که $||.||$ فاصله اقلیدسی است.

گام ۴ تابع عضویت جدید را برای باقیمانده‌ها به شکل زیر تعریف می‌کنیم

$$E = \begin{cases} 1 & abs(E) \leq a \\ e^{-\left(\frac{abs(E)-a}{d}\right)^2} & abs(E) > a \end{cases} \quad (12.4)$$

که در آن $a = median(D_i)$.

گام ۵ مقادیر عضویت را با استفاده از گام ۴ محاسبه کرده و ماتریس وزن برای هر تابع عضویت که در گام ۴ معرفی شد را تشکیل می‌دهیم. این ماتریس‌های وزن، ماتریس‌هایی

جدول ۵.۴: مجموعه داده‌های سیمان پورتلند و مقایسه چند روش مدل‌سازی رگرسیونی فازی بر اساس وزن‌ها و شاخص MPE

مشاهده	x_3	x_2	x_1	(y_i, s_i)	W_1	W_2	W_3	d_1	d_2	d_3
۱	۷	۲۶	۶	(۷۸/۵, ۶/۹)	۱	۱	۱	۴۰۷/۱۲۷	۰/۱۸۲	۰/۳۵۰
۲	۱	۲۹	۱۵	(۷۴/۳, ۶/۴)	۱	۰/۸۸۴	۰/۰۰۷	۲۶۰/۱۵۴	۴/۰۵۶	۱۴/۰۸۲
۳	۱۱	۵۶	۸	(۱۰۴/۳, ۹/۴)	۱	۰/۹۶۶	۰/۹۰۱	۰/۳۵۰	۱/۹۹۵	۰/۹۰۹
۴	۱۱	۳۱	۸	(۸۷/۶, ۷/۸)	۱	۰/۹۶۱	۰/۹۸۵	۶۷/۴۵۶	۲/۲۴۶	۰/۳۷۷
۵	۷	۵۲	۶	(۹۵/۵, ۸/۶)	۱	۱	۱	۱۴۱/۲۵۱	۰/۴۹۴	۰/۰۱۶
۶	۱۱	۵۵	۹	(۱۰۹/۲, ۹/۹)	۱	۰/۷۱۸	۰/۰۰۰۳	۲۲/۰۸۱	۱۶/۱۶۱	۲۰/۷۴۸
۷	۳	۷۱	۱۷	(۱۰۲/۷, ۹/۳)	۱	۱	۱	۳۸/۴۵۶	۱/۱۲۰	۰/۰۰۰
۸	۱	۳۱	۲۲	(۷۲/۵, ۶/۲)	۱	۰/۹۲۵	۰/۹۹۳	۰/۰۲۴	۴/۳۱۱	۰/۰۰۱
۹	۲	۵۴	۱۸	(۹۳/۳, ۸/۳)	۱	۰/۸۵۳	۰/۰۰۶	۵/۴۶۳	۵/۵۱۰	۱۴/۳۹۹
۱۰	۲۱	۴۷	۴	(۱۱۵/۹, ۱۰/۶)	۱	۱	۱	۱۳/۲۲۶	۰/۰۷۷	۰/۰۶۱
۱۱	۱	۴۰	۲۳	(۶۸, ۷/۴)	۱	۰/۰۳۱۳	۰	۲۲۵/۷۰۸	۱۶۱/۰۹۸	۱۱۴/۰۱۴
۱۲	۱۱	۶۶	۹	(۱۱۳/۳, ۱۰/۶)	۱	۰/۹۹	۰/۸۱۵۱	۳/۶۳۹	۰/۶۶۹	۱/۳۵۳
۱۳	۱۰	۶۸	۸	(۱۰۹/۴, ۹/۹)	۱	۰/۸۶۱	۰/۱۲	۷/۹۴۴	۶/۹۹۰	۵/۵۱۲
MPE										
								۹۱/۷۶	۱۵/۷۶۲	۱۳/۲۱۷۱

قطری هستند که عناصر روی قطرها، درجه عضویت‌ها می‌باشند. پس از آن، برآوردهای WFLAD را از طریق این ماتریس‌های وزن به دست می‌آوریم.

گام ۶ اگر $|\hat{A}^{k+1} - \hat{A}^k| < \epsilon$ ، الگوریتم را متوقف کنید. در غیر این صورت به گام ۲ بروید.

$\hat{A}$ برآورد پارامترهای رگرسیون، k تعداد تکرارها و $\epsilon > 0$ عدد بسیار کوچکی است که توسط تصمیم گیرنده تعیین می‌شود (در اینجا $\epsilon = 0/2$ در نظر گرفته شد).

معیار نیکویی برازش

در این فصل از شاخص نیکویی برازش MPE با فاصله (d) موزون که در معادله‌ی (۵.۱) بیان شد استفاده می‌کنیم. سپس با تعریف این شاخص و مقایسه‌ی وزن‌های به دست آمده از دو الگوریتم، برتری مدل پیشنهادی، نسبت به مدل FLS و مدل کولا و آپایدین [۳] را نشان خواهیم داد.

۴.۲.۴ مثال کاربردی

در این بخش با بررسی مجدد داده‌های سیمان پورتلند [۶۵] مطرح شده در زیر بخش ۲.۲.۴ و انجام مقایسات بین چند روش، برتری عملکرد مدل ارائه شده در این فصل روشن تر می‌شود. داده‌ها در جدول ۵.۴ نمایش داده شده‌اند. در این جدول (W_1, d_1) ، (W_2, d_2) و (W_3, d_3) ، به

۶۰ تحلیل رگرسیون استوار فازی بر پایه‌ی رتبه‌بندی مجموعه‌های فازی

ترتیب وزن‌ها و فاصله‌های بدست آمده از روش‌های کمترین قدرمطلق خطا، کولا و آپایدین و روش پیشنهادی است.

جدول ۶.۴: برآورد مدل رگرسیونی داده‌های مثال سیمان پورتلند به شیوه‌های مختلف

روش	برآورد رگرسیونی
کمترین قدرمطلق فازی	$\hat{y} = (0/28, 0/15) + (3/15, 0/34)x_1$ $+ (0/96, 0/06)x_2 + (1/77, 0/17)x_3$
رحیمیان و همکاران	$\hat{y} = (49/13, 3/92) + (1/63, 0/15)x_1$ $+ (0/66, 0/07)x_2 + (0/14, 0/01)x_3$
پیشنهادی	$\hat{y} = (47/91, 3/82) + (1/67, 0/16)x_1$ $+ (0/67, 0/07)x_2 + (0/08, 0/001)x_3$

جدول ۶.۴ ضرایب مدل‌های رگرسیونی را با روش‌های مختلف اشاره شده در فوق، نشان می‌دهد. نتایج وزن‌های بدست آمده از شیوه‌های مختلف مدل‌های رگرسیونی، برتری روش پیشنهادی را در کم وزن دهی به داده‌های دور افتاده نشان می‌دهد. همانطور که مشاهده می‌کنید کمترین وزن را در بین ۴ روش به داده یازدهم که از پرت بودن آن مطلع هستیم، روش پیشنهادی می‌دهد. همچنین روش پیشنهادی کمترین مقدار MPE را در میان روش‌های دیگر داراست که تایید دیگری است بر مناسبت این روش استوار، زمانی که در بین داده‌ها، داده دور افتاده وجود دارد.

نتیجه‌گیری و پیشنهادات

نتیجه‌گیری

هر گاه در بین مجموعه داده‌ها مشاهده یا مشاهدات دورافتاده وجود داشته باشد و از طرف دیگر متغیرها یا مشاهدات مربوط به آن‌ها دقیق نباشد، برآوردهای خوبی از روش‌های معمول LS حاصل نخواهد شد و باید از روش‌های جایگزین استوار فازی استفاده کرد. در این پایان نامه برای محاسبه‌ی برآورد مدل رگرسیون خطی چندگانه زمانی که مشاهدات متغیر وابسته، فازی و در مجموعه‌ی داده‌ها دورافتاده وجود داشته باشد از یک تحلیل رگرسیون استوار فازی استفاده شده و یک روش تکراری پیشنهاد می‌شود. در این روش، باقیمانده‌ها با استفاده از شاخص OM محاسبه شده و سپس با استفاده از توابع عضویت استوار تعریف شده برای باقیمانده‌ها، ماتریس وزن W ساخته می‌شود. الگوریتم تا رسیدن به یک دقت مطلوب برای پارامترها ادامه می‌یابد.

در این پایان نامه، از دو فاصله‌ی دیاموند و حسن‌پور به‌ترتیب برای یافتن برآوردها به دو شیوه‌ی FLA و FLS در گام ۱ و از دو نوع تابع عضویت در گام ۴ استفاده شد که به تفصیل در فصل ۴ مورد بررسی قرار گرفت و با بررسی مثال‌های مطرح شده این نتایج حاصل شد که (۱) روش رگرسیون استوار فازی که در قسمت ۱.۲.۴ بیان شد به‌علت وزن کمتری که به باقیمانده‌های مشاهدات دورافتاده می‌دهد و همین‌طور به‌علت کمتر شدن مقدار معیار نیکویی برازش MPE آن نسبت به روش FLS برتری دارد. (۲) روش رگرسیون استوار فازی پیشنهادی که در قسمت ۳.۲.۴ بیان شد نیز به دلیل وزن کمتر باقیمانده‌های مشاهدات دورافتاده و برتری مقدار شاخص MPE ، نسبت به روش FLA برتری دارد. و مهم‌ترین نتیجه‌ای که از مقایسات روش‌ها می‌گیریم این است که (۳) روش رگرسیون استوار پیشنهادی ما در قسمت ۳.۲.۴ به روش رگرسیون استوار بحث شده در قسمت ۱.۲.۴ برتری دارد؛ چون هم باقیمانده‌های داده‌های دورافتاده در این روش وزن کمتری دارند و هم شاخص MPE مقدار کمتری است.

پیشنهادهای

۱. تغییر تابع عضویت باقیمانده‌ها به شکل توابع استوار مطرح شده در بخش ۲.۲ و مقایسه‌ی نتایج.
۲. به کار بردن روش رگرسیون امکانی به جای روش‌های کمترین توان‌های دوم و کمترین قدر مطلق فازی و مقایسه‌ی نتایج.
۳. استفاده این روش رگرسیونی برای حالتی که متغیرهای مستقل، فازی و ضرایب رگرسیونی، دقیق و یا هم متغیرهای مستقل و متغیر پاسخ فازی در نظر گرفته شوند.
۴. به کار بردن فواصل فازی دیگری چون گیلده وگین [۲۷] و محاسبه‌ی برآورد ضرایب مدل.

پیوست آ

برنامه‌های نوشته شده در متمتیکا

برنامه‌ی مربوط به مثال‌های فصل ۴

لازم به ذکر است که مثال‌های ۲.۲.۴ و ۲.۲.۴ برنامه‌های یکسانی با مجموعه داده‌های متفاوت دارند. در این بخش برنامه مربوط به مثال ۲.۲.۴ را می‌آوریم.

```
X = Import["C:\\Users\\TOPCO\\Desktop\\\\
andishe amari\\mathematica and \\
excell\\xu and lee.xlsx", {"Data", 1}];
A = {a0, a1, a2, a3};
L = {10, 11, 12, 13};
R = {r0, r1, r2, r3};
n = 13;
p = 3;

P = NMinimize[
  Sum[(Sum[A[[j]]*X[[i, j]], {j, 1, p + 1}] - X[[i, 5]])^2 +
```

```

        (Sum[A[[j]]*X[[i, j]], {j, 1, p + 1}] -
        Sum[L[[j]]*X[[i, j]], {j, 1, p + 1}] - X[[i, 5]] +
        X[[i, 6]]^2 + (Sum[A[[j]]*X[[i, j]], {j, 1, p + 1}] +
        Sum[R[[j]]*X[[i, j]],
        {j, 1, p + 1}] - X[[i, 5]] - X[[i, 6]]^2, {i, 1,
n}], Join[A, L, R]]

Z = Table[f, {n}];
For[i = 1, i <= n, i++,
  Z[[i]] = (P[[2, 1, 2]]*X[[i, 1]] + P[[2, 2, 2]]*X[[i, 2]] +
    P[[2, 3, 2]]*X[[i, 3]] + P[[2, 4, 2]]*X[[i, 4]]) - X[[i, 5]]]
DD1 = Table[d, {n}];
For[i = 1, i <= n, i++,
  DD1[[i]] = Norm[Abs[Z[[i]]]] - Median[Abs[Z]]]; ]
dd1 = Table[d, {n}];
For[i = 1, i <= n, i++, dd1[[i]] = Abs[Z[[i]]] - Median[Z]];
ddd1 = Median[dd1]/0.6745;
aa1 = Median[DD1];
bb1 = Max[DD1] + ddd1;

W = Table[g, {n}];
For[i = 1, i <= n, i++, W[[i]] = Piecewise[{{1, Abs[Z[[i]]] <= aa1},
  {(bb1 - Abs[Z[[i]]])/(bb1 - aa1),
  aa1 < Abs[Z[[i]]] < bb1}, {0, Abs[Z[[i]]] > bb1}}]]
WW = DiagonalMatrix[W];

XX =
Import[
  "C:\\Users\\TOPCO\\Desktop\\andishe amari\\mathematica and \\
excell\\xu and lee.xlsx",
  {"Data", 2}];

MatrixForm[XX];
TXX = Transpose[XX];

```

```

betahatm1 = Inverse[TXX . WW . XX] . TXX . WW . X[[A11, 5]]
betahats1 = Inverse[TXX . WW . XX] . TXX . WW . X[[A11, 6]]

Z2 = Table[f, {n}];
For[i = 1, i <= n, i++,
  Z2[[i]] = (betahatm1[[1]]*X[[i, 1]] + betahatm1[[2]]*X[[i, 2]] +
    betahatm1[[3]]*X[[i, 3]] + betahatm1[[4]]*X[[i, 4]]) -
    X[[i, 5]]
DD2 = Table[d, {n}];
For[i = 1, i <= n, i++,
  DD2[[i]] = Norm[Abs[Z2[[i]]] - Median[Abs[Z2]]]; ]
dd2 = Table[d, {n}];
For[i = 1, i <= n, i++, dd2[[i]] = Abs[Z2[[i]] - Median[Z2]]];
ddd2 = Median[dd2]/0.6745;
aa2 = Median[DD2];
bb2 = Max[DD2] + ddd2;
W2 = Table[g, {n}];
For[i = 1, i <= n, i++, W2[[i]] = Piecewise[{{1, Abs[Z2[[i]]] <= aa2},
  {(bb2 - Abs[Z2[[i]])/(bb2 - aa2),
    aa2 < Abs[Z2[[i]]] < bb2}, {0, Abs[Z2[[i]]] > bb2}}]]

WW2 = DiagonalMatrix[W2];
betahatm2 = Inverse[TXX . WW2 . XX] . TXX . WW2 . X[[A11, 5]]
betahats2 = Inverse[TXX . WW2 . XX] . TXX . WW2 . X[[A11, 6]]

Norm[betahatm2 - betahatm1]
Norm[betahats2 - betahats1]

```

و این حلقه ادامه می‌یابد تا عبارت بالا کوچکتر از مقدار آستانه‌ی $\epsilon > 0$ در نظر گرفته شده، شود. (گام ۶ الگوریتم ۱.۲.۴).

برنامه‌ی مربوط به محاسبه‌ی شاخص نیکویی برازش MPE با فاصله حسن‌پور به صورت زیر است:

```

yhatfls = Array[sd, {n, 3}];
For[i = 1, i <= n, i++,
  yhatfls[[i, 1]] = P[[2, 1, 2]] + P[[2, 2, 2]]*X[[i, 2]] +
    P[[2, 3, 2]]*X[[i, 3]] + P[[2, 4, 2]]*X[[i, 4]];
  yhatfls[[i, 2]] =
    P[[2, 5, 2]] + P[[2, 6, 2]]*X[[i, 2]] +
    P[[2, 7, 2]]*X[[i, 3]] + P[[2, 8, 2]]*X[[i, 4]];
  yhatfls[[i, 3]] =
    P[[2, 9, 2]] + P[[2, 10, 2]]*X[[i, 2]] + P[[2, 11, 2]]*X[[i, 3]] +
    P[[2, 12, 2]]*X[[i, 4]]; ];

d = Array[k, {n, 1}];
For[i = 1, i <= n, i++,
  d[[i, 1]] =
    Abs[X[[i, 5]] - yhatfls[[i, 1]]] +
    Abs[X[[i, 6]] - yhatfls[[i, 2]]] + Abs[X[[i, 6]] - yhatfls[[i, 2]]]
];

MPEfls = (1/n)*Sum[d[[i, 1]], {i, 1, n}]

yhatprop = Array[f, {n, 2}];
For[i = 1, i <= n, i++,
  yhatprop[[i, 1]] = betahatm3[[1]] + betahatm4[[2]]*X[[i, 2]] +
    betahatm4[[3]]*X[[i, 3]] + betahatm4[[4]]*X[[i, 4]];
  yhatprop[[i, 2]] =
    betahats4[[1]] + betahats4[[2]]*X[[i, 2]] +
    betahats4[[3]]*X[[i, 3]] +
    betahats4[[4]]*X[[i, 4]]; ]

dd = Array[k, {n, 1}];
For[i = 1, i <= n, i++,
  dd[[i, 1]] = Abs[X[[i, 5]] - yhatprop[[i, 1]]] +
    Abs[X[[i, 6]] - yhatprop[[i, 2]]] +

```

```
Abs[X[[i, 6]] - yhatprop[[i, 2]]];
```

```
MPEprop = (1/n)*Sum[dd2[[i, 1]], {i, 1, n}]
```

برنامه‌ی روش بخش ۳.۲.۴

```
X = Import["C:\\Users\\TOPCO\\Desktop\\\\
andishe amari\\mathematica and \\
excell\\xu and lee.xlsx", {"Data", 1}];
A = {a0, a1, a2, a3};
L = {l0, l1, l2, l3};
R = {r0, r1, r2, r3};
n = 13;
p = 3;

P1 = NMinimize[
Sum[Abs[Sum[A[[j]]*X[[i, j]], {j, 1, p + 1}] - X[[i, 5]]] +
Abs[Sum[L[[j]]*X[[i, j]], {j, 1, p + 1}] - X[[i, 6]]] +
Abs[Sum[R[[j]]*X[[i, j]], {j, 1, p + 1}] - X[[i, 6]]], {i,
1, n}], Join[A, L, R]]

Z = Table[f, {n}];
For[i = 1, i <= n, i++,
Z[[i]] = (P[[2, 1, 2]]*X[[i, 1]] + P[[2, 2, 2]]*X[[i, 2]] +
P[[2, 3, 2]]*X[[i, 3]] + P[[2, 4, 2]]*X[[i, 4]]) - X[[i, 5]]]
DD1 = Table[d, {n}];
For[i = 1, i <= n, i++,
DD1[[i]] = Norm[Abs[Z[[i]]] - Median[Abs[Z]]]; ]
dd1 = Table[d, {n}];
For[i = 1, i <= n, i++, dd1[[i]] = Abs[Z[[i]] - Median[Z]]];
ddd1 = Median[dd1]/0.6745;
aa1 = Median[DD1];
bb1 = Max[DD1] + ddd1;
```

```

W = Table[g, {n}];
For[i = 1, i <= n, i++, W[[i]] = Piecewise[{{1, Abs[Z[[i]]] <= aa1},
      {(bb1 - Abs[Z[[i]])/(bb1 - aa1),
      aa1 < Abs[Z[[i]]] < bb1}, {0, Abs[Z[[i]]] > bb1}}]]

WW = DiagonalMatrix[W];

XX =
  Import[
    "C:\\Users\\TOPCO\\Desktop\\andishe amari\\mathematica and \\
    excell\\xu and lee.xlsx",
    {"Data", 2}];

MatrixForm[XX];
TXX = Transpose[XX];
betahatm1 = Inverse[TXX . WW . XX] . TXX . WW . X[[A11, 5]]
betahats1 = Inverse[TXX . WW . XX] . TXX . WW . X[[A11, 6]]

Z2 = Table[f, {n}];
For[i = 1, i <= n, i++,
  Z2[[i]] = (betahatm1[[1]]*X[[i, 1]] + betahatm1[[2]]*X[[i, 2]] +
    betahatm1[[3]]*X[[i, 3]] + betahatm1[[4]]*X[[i, 4]]) -
    X[[i, 5]]]
DD2 = Table[d, {n}];
For[i = 1, i <= n, i++,
  DD2[[i]] = Norm[Abs[Z2[[i]]] - Median[Abs[Z2]]]; ]
dd2 = Table[d, {n}];
For[i = 1, i <= n, i++, dd2[[i]] = Abs[Z2[[i]] - Median[Z2]]];
ddd2 = Median[dd2]/0.6745;
aa2 = Median[DD2];
bb2 = Max[DD2] + ddd2;
W2 = Table[g, {n}];
For[i = 1, i <= n, i++, W2[[i]] = Piecewise[{{1, Abs[Z2[[i]]] <= aa2},
      {(bb2 - Abs[Z2[[i]])/(bb2 - aa2),

```

```
aa2 < Abs[Z2[[i]]] < bb2}, {0, Abs[Z2[[i]]] > bb2}}}]
```

```
WW2 = DiagonalMatrix[W2];
```

```
betahatm2 = Inverse[TXX . WW2 . XX] . TXX . WW2 . X[[A11, 5]]
```

```
betahats2 = Inverse[TXX . WW2 . XX] . TXX . WW2 . X[[A11, 6]]
```

```
Norm[betahatm2 - betahatm1]
```

```
Norm[betahats2 - betahats1]
```

ادامه‌ی حلقه تا محقق شدن گام ۶ الگوریتم ۳.۲.۴

برنامه‌ی محاسبه‌ی MPE با فاصله‌ی موزون

```
yhatfls = Array[sd, {n, 3}];
```

```
For[i = 1, i <= n, i++,
```

```
  yhatfls[[i, 1]] = P[[2, 1, 2]] + P[[2, 2, 2]]*X[[i, 2]] +  
    P[[2, 3, 2]]*X[[i, 3]] + P[[2, 4, 2]]*X[[i, 4]];
  yhatfls[[i, 2]] =
```

```
    P[[2, 5, 2]] + P[[2, 6, 2]]*X[[i, 2]] +  
    P[[2, 7, 2]]*X[[i, 3]] + P[[2, 8, 2]]*X[[i, 4]];
  yhatfls[[i, 3]] =
```

```
    P[[2, 9, 2]] + P[[2, 10, 2]]*X[[i, 2]] + P[[2, 11, 2]]*X[[i, 3]] +  
    P[[2, 12, 2]]*X[[i, 4]]; ];
```

```
d2 = Array[k, {n, 1}];
```

```
For[i = 1, i <= n, i++, d2[[i, 1]] = (yhatfls[[i, 1]] - X[[i, 5]])^2 +  
  (1/6)*(yhatfls[[i, 2]] - X[[i, 6]])^2]
```

```
MPEfls = (1/n)*Sum[d2[[i, 1]], {i, 1, n}]
```

```
yhatprop = Array[f, {n, 2}];
```

```
For[i = 1, i <= n, i++,
```

```
  yhatprop[[i, 1]] = betahatm3[[1]] + betahatm4[[2]]*X[[i, 2]] +  
    betahatm4[[3]]*X[[i, 3]] + betahatm4[[4]]*X[[i, 4]];
  yhatprop[[i, 2]] =
```

```

betahats4[[1]] + betahats4[[2]]*X[[i, 2]] +
betahats4[[3]]*X[[i, 3]] +
betahats4[[4]]*X[[i, 4]]; ]

dd2 = Array[k, {n, 1}];
For[i = 1, i <= n, i++,
dd2[[i, 1]] = (yhatprop[[i, 1]] - X[[i, 5]])^2 +
(1/6)*(yhatprop[[i, 2]] - X[[i, 6]])^2];

MPEprop = (1/n)*Sum[dd2[[i, 1]], {i, 1, n}]

```

برنامه‌ی روش [۳]

```

X = Import[
"C:\\Users\\TOPCO\\Desktop\\andishe amari\\mathematica and \\
excell\\xu and lee.xlsx",
{"Data", 1}];
A = {a0, a1, a2, a3};
L = {l0, l1, l2, l3};
R = {r0, r1, r2, r3};
n = 13;
p = 3;

P = NMinimize[
Sum[Abs[Sum[A[[j]]*X[[i, j]], {j, 1, p + 1}] - X[[i, 5]]] +
Abs[Sum[L[[j]]*X[[i, j]], {j, 1, p + 1}] - X[[i, 6]]] +
Abs[Sum[R[[j]]*X[[i, j]], {j, 1, p + 1}] - X[[i, 6]]], {i,
1, n}], Join[A, L, R]]

Z = Table[f, {n}];
For[i = 1, i <= n, i++,
Z[[i]] = (P[[2, 1, 2]]*X[[i, 1]] + P[[2, 2, 2]]*X[[i, 2]] +
P[[2, 3, 2]]*X[[i, 3]] + P[[2, 4, 2]]*X[[i, 4]]) - X[[i, 5]]]
DD1 = Table[d, {n}];

```

```

For[i = 1, i <= n, i++,
  DD1[[i]] = Norm[Abs[Z[[i]]] - Median[Abs[Z]]]; ]
dd1 = Table[d, {n}];
For[i = 1, i <= n, i++, dd1[[i]] = Abs[Z[[i]] - Median[Z]]];
ddd1 = Median[dd1]/0.6745;
aa1 = Median[DD1];
bb1 = Max[DD1] + ddd1;

W = Table[g, {n}];
For[i = 1, i <= n, i++, W[[i]] = Piecewise[{{1, Abs[Z[[i]]] <= aa1},
  {Exp[-((aa1 - Abs[Z[[i]])/ddd1)^2], aa1 < Abs[Z[[i]]]}}]]
WW = DiagonalMatrix[W];

XX =
  Import[
    "C:\\Users\\TOPCO\\Desktop\\andishe amari\\mathematica and \\
    excell\\xu and lee.xlsx", {"Data", 2}];

MatrixForm[XX];
TXX = Transpose[XX];

Norm[betahatm2 - betahatm1]
Norm[betahats2 - betahats1]

```

ادامه‌ی حلقه تا محقق شدن گام ۶ الگوریتم ۳.۲.۴

برنامه‌ی محاسبه‌ی MPE با فاصله موزون در این حالت نیز مانند مرحله‌ی قبل می‌باشد.

برنامه‌ی مربوط به رسم شکل‌های ۲.۲

الف. تابع ρ هوپر

```

f2[u_] = Piecewise[{{u^2/2, -1.5 <= u <= 1.5}, {1.5*Abs[u] - 1.5^2/2,
u < -1.5 && u > 1.5}}]

```

ب. تابع ρ همپل

```
e = 1.7; f = 3.4; g = 8.5;
f7[u_] = Piecewise[{u^2/2, -e <= u <= e}, {e*Abs[u] -
e^2/2, -f <= u <= -e &&
Inequality[e, Less, u, LessEqual,
f]}, {e*((e*Abs[u] - u^2/2)/(g - f)) - (7*e^2)/6,
Inequality[-g, LessEqual, u, Less, -f] &&
Inequality[f, Less, u, LessEqual, g]}, {e*(f + g - e),
u < g && u > g}]]
```

ج. تابع ρ اندروز

```
a = 1.5; b = 2.1;
f3[u_] = Piecewise[{a*(1 - Cos[u/a]), (-a)*Pi <= u <= a*Pi}, {2*a,
u < (-a)*Pi && u > a*Pi}]]
f4[u_] = Piecewise[{b*(1 - Cos[u/a]), (-b)*Pi <= u <= b*Pi}, {2*b,
u < (-b)*Pi && u > b*Pi}]]
```

د. تابع ρ توکی

```
c = 5; d = 6;
f5[u_] = Piecewise[{u^2/2 - u^2/(4*c^2), -c <= u <= c}, {c^2/4,
u < -c && u > c}]]
f6[u_] = Piecewise[{u^2/2 - u^2/(4*d^2), -d <= u <= d}, {d^2/4,
u < -d && u > d}]]
```

برنامه‌ی رسم همه‌ی توابع ρ بالا

```
Plot[f1[u], {u, 0, 10}, PlotLegends -> "Expressions"]
Plot[f2[u], {u, 0, 10}, PlotLegends -> "Expressions"]
Plot[f3[u], {u, 0, 10}, PlotLegends -> "Expressions"]
Plot[f5[u], {u, 0, 10}, PlotLegends -> "Expressions"]
Plot[f7[u], {u, 0, 10}, PlotLegends -> "Expressions"]
Plot[{f1[u], f2[u], f4[u], f5[u], f6[u]}, {u, 0, 2.5},
PlotLegends -> "Expressions"]
```

الف. تابع ψ هوبر

```
s2[z_] = Piecewise[{{-1.5, z < -1.5}, {z, Abs[z] <= 1.5}, {1.5,
z > 1.5}}]
```

ب. تابع ψ همپل

```
a = 1.7; b = 3.4; c = 8.5;
s3[z_] = Piecewise[{{Abs[z],
Inequality[0, LessEqual, Abs[z], Less, a]}, {a,
a <= Abs[z] <= b}, {a*((c - Abs[z])/(c - b)),
b <= Abs[z] <= c}, {0, c <= Abs[z]}}]
```

ج. تابع ψ اندروز

```
k1 = 1.5; k2 = 2.1;
s4[z_] = Piecewise[{{Sin[z/k1], Abs[z] <= k1*Pi}, {0, Abs[z] > k1*Pi}}]
s5[z_] = Piecewise[{{Sin[z/k2], Abs[z] <= k2*Pi}, {0,
Abs[z] > k2*Pi}}]
```

د. تابع ψ توکی

```
k3 = 5; k4 = 6;
s6[z_] = Piecewise[{{z*(1 -
TextCell[
RawBoxes[
Cell[BoxData[TraditionalForm` \((z \ / \ k3)\) \^
2)\)\]]])^2, Abs[z] <= k3}, {0, Abs[z] > k3}}]
s7[z_] = Piecewise[{{z*(1 -
TextCell[
RawBoxes[
Cell[BoxData[TraditionalForm` \((z \ / \ k4)\) \^
2)\)\]]])^2, Abs[z] <= k4}, {0, Abs[z] > k4}}]
```

برنامه‌ی رسم همه توابع ψ بالا الف. تابع ψ هوپر

```
Plot[s1[z_], {z, -2, 3}, PlotLegends -> "Expressions"]
Plot[s2[z_], {z, -2, 3}, PlotLegends -> "Expressions"]
Plot[s3[z_], {z, -2, 3}, PlotLegends -> "Expressions"]
```

```
Plot[s5[z_], {z, -2, 3}, PlotLegends -> "Expressions"]  
Plot[s7[z_], {z, -2, 3}, PlotLegends -> "Expressions"]  
Plot[{s1[z_], s2[z_], s3[z_], s5[z_], s7[z_]}, {z, -2.5, 2.5},  
PlotLegends -> "Expressions"]
```

مراجع

- [۱] اسدالهی، م، (۱۳۹۵)، پایان نامه کارشناسی ارشد: ”برآورد ضرایب مدل های رگرسیون خطی بر اساس متغیر وابسته فازی و متغیرهای مستقل فازی و غیر فازی“، دانشگاه بیرجند.
- [۲] ربیعی م.ر.، رحیمیان ا. و شاهسونی د، (۱۳۹۶) ”تحلیل رگرسیون استوار فازی بر اساس کمترین قدر مطلق خطاها و رتبه بندی مجموعه های فازی“، هفتمین سمینار آمار و احتمال فازی کشوری، بیرجند، ایران.
- [۳] رحیمیان ا، ربیعی م و شاهسونی د، (۱۳۹۶)، ”تحلیل رگرسیون استوار فازی با داده های خروجی و پارامترهای فازی“، اندیشه آماری، در دست چاپ.
- [۴] طاهری م، (۱۳۷۵)، ”آشنایی با نظریه مجموعه های فازی“، انتشارات جهاد دانشگاهی مشهد.
- [۵] فرخی، ا، (۱۳۹۶)، پایان نامه کارشناسی ارشد: ”مطالعه چندین روش رگرسیون ریج در محیط فازی“، دانشگاه صنعتی شاهرود.
- [۶] طاهری م و ماشین چی م، (۱۳۹۲)، ”مقدمه ای بر احتمال و آمار فازی“ چاپ دوم، انتشارات دانشگاه شهید باهنر کرمان.
- [۷] کلکین نما، مریم و طاهری، سید محمود، (۱۳۸۶) ”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)، ص ص ۶۷-۵۹.
- [۸] میرزایی، یگانه و همکاران، (۱۳۸۶) ”رگرسیون فازی: مروری بر چند رویکرد“، اندیشه آماری، سال ۱۲ شماره ۱ (شماره پیاپی ۲۳) (شماره ویژه دومین کارگاه آمار و احتمال فازی)، ص ص ۴۷-۳۵.

[9] Adamo J. M. (1980), "Fuzzy decision trees", **Fuzzy Sets and Systems**, 4(3), pp 207-219.

[10] Andersen R. (2008), "**Modern methods for robust regression**", Thousand Oaks: SAGE Publications, 152.

-
- [11] Balasundaram S. Kapil. (2011), "Weighted fuzzy ridge regression analysis with crisp inputs and triangular fuzzy outputs" **International Journal of Advanced Intelligence Paradigms**, 3(1), pp 67-81.
- [12] Bárdossy A. (1990), "Note on fuzzy regression", **Fuzzy Sets and Systems**, 37(1), pp 65-75.
- [13] Bortolan G. Deganin R. (1985), "A review of some methods for ranking fuzzy subsets". **Fuzzy Sets and Systems**, 15(1), pp 1-19.
- [14] Buckley J. Chanas S. (1989), "A fast method of ranking alternatives using fuzzy numbers". **Fuzzy sets and systems**, 30(3), pp 337-338.
- [15] Campos L. M. González A. (1989), "A subjective approach for ranking fuzzy numbers". **Fuzzy Sets and Systems**, 29(2), pp 145-153.
- [16] Celmins A. (1987), "Least squares model fitting to fuzzy vector data", **Fuzzy Sets and Systems**, 22(3), pp 260-269. , 22(3): 260-269.
- [17] Chang P. T. and Lee E. S. (1994) "Ranking of fuzzy sets based on the concept of existence", **Computers and Mathematics with Applications**, 27(9-10), pp 1-21.
- [18] Chang P. T. and Lee E. S. (1996), "A generalized fuzzy weighted least-squares regression", **Fuzzy Sets and Systems**, 82(3), pp 289-298.
- [19] Diamond P. and Korner R. (1997), "Extended fuzzy linear models and least squares estimates", **Computers and Mathematics with Applications**, 33(9), pp 15-32
- [20] Diamond P. (1988), "Fuzzy least squares", **Information Sciences**, 46(3), pp 141-157.
- [21] Donoso S., Marín N and Vila M. A. (2007), "Fuzzy ridge regression with non symmetric membership functions and quadratic models", In International Conference on Intelligent Data Engineering and Automated Learning, pp 135-144.
- [22] Donoso, S., Marín, N. and Vila, M.A. (2007), "Fuzzy ridge regression with non symmetric membership functions and quadratic models", In International Conference on Intelligent Data Engineering and Automated Learning (pp. 135-144). Springer Berlin Heidelberg.
- [23] Draper N. R., Smith H. (2014). "**Applied regression analysis**, John Wiley Sons, New York.
- [24] Dubois D. Prade H. (1980), "**Fuzzy Sets and Systems: theory and applications**", Vol. 144, Academic press, Toulouse, France.

- [25] Edgeworth F. Y. (1888), XXII, "On a new method of reducing observations relating to several quantities", **Philosophical Magazine and Journal of Science**, The London, Edinburgh and Dublin, 25(154), pp 184-191.
- [26] Filzmoser P. (2010) , " Soft Methods in Robust Statistics", **Combining Soft Computing Stats Methods**, AISC, 77, pp 273-280.
- [27] Gildeh, B. S., Gien, D. E. N. I. S. (2002), "A goodness of fit index to reliability analysis in fuzzy model". **In 3rd WSEAS international conference on fuzzy sets and fuzzy systems**. Iterlaken, Switzerland (pp. 11-14).
- [28] González A. (1990), "A study of the ranking function approach through mean values", **Fuzzy Sets and Systems**, 35(1), pp 29-41.
- [29] Hampel F. R. (1978), "Optimally bounding the gross-error-sensitivity and the influence of position in factor space, **Proceedings of the ASA Statistical Computing Section**, ASA, Washington, DC, pp 59–64.
- [30] Hassanpour H. Maleki HR. and Yaghoobi MA. (2010), "Fuzzy linear regression model with crisp coefficients: a goal programming approach", **Iranian Journal of Fuzzy Systems**, 7(2) pp 19-39.
- [31] Hogg, R. V. (1979), "Statistical Robustness: One View of Its Use in Applications Today", **The American Statistician**, Vol. 33, No. 3, pp:108-115.
- [32] Huber P. J. (1973), "Robust regression: asymptotics, conjectures and Monte Carlo", **The Annals of Statistics**, pp 799-821.
- [33] Huber P. J. , Ronchetti E. M. (1981), "Robust Statistics", Wiley, New York.
- [34] Hung W. L., Yang M. S. (2006), "An omission approach for detecting outliers in fuzzy regression models". **Fuzzy sets and systems**, 157(23), pp 3109-3122.
- [35] Huber, P. J., and Ronchetti, E. M. (2009), "**Robust statistics**", Hoboken: John Wiley and Sons, 2.
- [36] Izadikhah, M. (2009), "Using the Hamming distance to extend TOPSIS in a fuzzy environment". **Journal of Computational and Applied Mathematics**, 231(1), pp 200-207.
- [37] Kariya, T., and Kurata, H. (2004), "Generalized least squares", **Theory and Methods**.

-
- [38] Kauffman A. , Gupta, M. M. (1985), "Introduction to Fuzzy Arithmetic", Van Nostrand Reinhold co, New York.
- [39] Kaufmann A. (1986), "Hybrid Data-Variou Associations Between Fuzzy Subsets and Random Variables". **In Fuzzy Sets Theory and Applications**, pp 171-211.
- [40] Kim K. and etal. (2005), "Least absolute deviation estimator in fuzzy regression", **J. Appl. Math. Comput**, 18, pp 649-656.
- [41] Kula K. Ş. , Apaidin A. (2008), "Fuzzy robust regression analysis based on the ranking of fuzzy sets", **International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems**, 16(05), pp 663-681.
- [42] Leroy A. M. , Rousseeuw P. J. (1987), " Robust regression and outlier detection". **Wiley Series in Probability and Mathematical Statistics**, New York: Wiley.
- [43] Li R. J, Lee E. S. (1987), "Ranking fuzzy numbers-A comparison", **In proceedings of North American Fuzzy information Processing Society Workshop**, pp 169-204, West Lafayette, IN.
- [44] Li J. and etal. (2016), "A new fuzzy regression model based on least absolute deviation", **Engineering Applications of Artificial Intelligence**, 52, pp 54-64.
- [45] Mabuchi S. (1988), "An approach to the comparison of fuzzy subsets with an alpha-cut dependent index", **IEEE Transactions on Systems, Man, and Cybernetics**, 18(2), pp 264-272.
- [46] Maeder R. E. (1991), " Programming in mathematica", **Addison-Wesley Longman Publishing Co, Inc**.
- [47] Maronna, R., Martin, D., Yohai, V. (2006), "**Robust Statistics: Theory and Methods**", Wiley, New York.
- [48] Mallows C. L. (1975), "On some topics in robustness unpublished memorandum", Bell Telephone Laboratories, Murray Hill, NJ.
- [49] Mohammadi, J., Taheri, S.M. (2003), "Pedomodels fitting with fuzzy least squares regression", **Iranian J. Fuzzy Systems**, 1 (2): 45-62
- [50] Nanda S. (1989), "On sequences of fuzzy numbers". **Fuzzy Sets and Systems**, 33(1), pp 123-126.

- [51] Nguyen T. D. , Welsch, R. (2010), "Outlier detection and least trimmed squares approximation using semi-definite programming", **Computational Statistics and Data Analysis**, 54(12), pp 3212-3226.
- [52] Pedrych W, Gomide. (2005), " **An introduction to Fuzzy Sets and Systems**", MIT Press.
- [53] Rice, J. A. (1995), " **Mathematical Statistics and Data Analysis**", Duxbury Press, Second edition, Belmont, CA.
- [54] Rousseeuw P. J. (1984), "Least median of squares regression", **Journal of the American Statistical Association**, 79 (388), pp 871–880.
- [55] Rousseeuw P. J, Yohai V. J. (1984), "Robust regression by means of S-estimators", **Robust and Nonlinear Time Series Regression, Springer Lecture Notes in Statistics**, vol. 26, New York, pp 256–272.
- [56] Stigler, S. M., (1986), " The history of statistics: The measurement of uncertainty before 1900", **Harvard University Press**.
- [57] Taheri S. M. and Kelkinnama, M. (2012), "Fuzzy linear regression based on least absolutes deviations", **Iranian Journal of Fuzzy Systems**, 9(1), pp 121-140.
- [58] Tanaka-S, Asai H, Uegima-K.(1982), "Linear regression analysis with fuzzy model", **IEEE Transaction Systems Man and Cybermatics**" 12(6), pp 903-07.
- [59] Tanaka H. , Uegima, S. and Asai, K. (1982), "Linear regression analysis with fuzzy model", **IEEE Transactions on Systems. Man. and Cybematics**, 12, pp 903-907.
- [60] Tanaka H. , etal. (1984), "A formulation of fuzzy linear programming based on comparison of fuzzy numbers", **Control and cybernetics**, 13(3), pp 185-194.
- [61] Tanaka H , Lee. H. (1998), " Interval regression analysis by quadratic programming approach", **IEEE Trans. on Fuzzy Systems**, 6(4), pp 473-481.
- [62] Torabi H. , and Behboodian J. (2007), " Fuzzy least-absolutes estimates in linear models. Communications in Statistics", **Theory and Methods**, 36(10), pp 1935-1944.
- [63] Xu, R. (1991), " A linear regression model in fuzzy environment", **Advances in Modeling , Simulation**, 27, pp 31-40.
- [64] Chang, P-T., and Lee E. S. (1994), "Ranking of fuzzy sets based on the concept of existence", **Computers and Mathematics with Applications** 27, no. 9-10, pp 1-21.

-
- [65] Xu, R. and Li, C. (2001), "Multidimensional least squares fitting with a fuzzy model", **Fuzzy Sets and Systems**, 119, pp 215-223.
- [66] Yager R. (1979), "Ranking fuzzy subsets over the unit interval", In Decision and Control including the 17th Symposium on Adaptive Processes, 1978 IEEE Conference on (pp. 1435-1437).
- [67] Yang M. S. , Ko C. H. (1997), "On cluster-wise fuzzy regression analysis", **IEEE Transactions on Systems, Man, and Cybernetics**, Part B (Cybernetics), 27(1), pp 1-13.
- [68] Yang M. S. , Lin T. S. (2002), "Fuzzy least-squares linear regression analysis for fuzzy input-output data", **Fuzzy sets and systems**, 126(3), pp 389-399.
- [69] Yang M. S. , Liu H. H. (2003), "Fuzzy least-squares algorithms for interactive fuzzy linear regression models", **Fuzzy Sets and Systems**, 135(2), pp 305-316.
- [70] Yohai V. J. (1987), "High breakdown-point and high efficiency robust estimates for regression", **The Annals of Statistics**, pp 642-656.
- [71] Yuan Y. (1991), "Criteria for evaluating fuzzy ranking methods", **Fuzzy Sets and Systems**, 43(2), pp 139-157.
- [72] Zadeh, L. A. (1965), "Fuzzy sets" **Information and control**, 8(3), pp 338-353
- [73] Zadeh, L. A. (1975), "The concept of a linguistic variable and its application to approximate reasoning *I, II and III*" **Information sciences** , 8: pp199-249, 8: 301-357, 9:43-80.
- [74] Zadeh L. etal. (1978), "Fuzzy sets as a basis for a theory of possibility", **Fuzzy Sets and Systems**, 1(3-28), pp 61-72.
- [75] Zhu Q. Lee, E. S. (1992), "Comparison and ranking of fuzzy numbers". **Fuzzy regression analysis**, 1, pp 21-44.

واژه‌نامه فارسی به انگلیسی

Decomposition identity	اتحاد تجزیه
Generalized M-estimators	برآوردگرهای M-تعمیم یافته
Extension principle	اصل توسیع (اصل گسترش)
Additivity principle	اصل جمع پذیری
Fuzzy measure	اندازه فازی
Residual	باقیمانده
Top flattening	بالا صاف
Estimate	برآورد
Estimator	برآوردگر
M-estimates	برآوردگرها-M
h-cut	برش-h
Spread	پهنا
Left spread	پهنای چپ
Right spread	پهنای راست
Likelihood function	تابع درست‌نمایی
Membershib function	تابع عضویت
Objective function	تابع هدف
Least median squares	ترین میانه توان‌های دوم
Decision maker	تصمیم گیرنده
Arithmetic of fuzzy numbers	حساب اعداد فازی
Estimation error	خطای برآورد
Well-defined	خوش تعریف
Estimated data	داده برآورد شده
Outlier	داده دور افتاده (پرت)
Observed data	داده مشاهده شده
Crisp	دقیق

Regression outlier	دورافتاده رگرسیونی
Robust regression	رگرسیون استوار
Linear regression	رگرسیون خطی
Multiple linear regression	رگرسیون خطی چندگانه
Fuzzy regression	رگرسیون فازی
Fuzzy ranking method	روش رتبه‌بندی فازی
Overall existence level	سطح حضور سراسری
Fuzzy number	عدد فازی
Trapezoidal fuzzy number	عدد فازی ذوزنقه‌ای
Elliptic fuzzy number	عدد فازی سهموی
Symmetric fuzzy number	عدد فازی متقارن
Triangular fuzzy number	عدد فازی مثلثی
Normal fuzzy number	عدد فازی نرمال
Fuzzy distance	فاصله فازی
Least squares	کمترین توان‌های دوم
Least trimmed squares	کمترین توان‌های دوم پیراسته
Weighted least squares	کمترین توان‌های دوم موزون
Weighted fuzzy least squares	کمترین توان‌های دوم موزون فازی
Least absolute deviation	کمترین قدرمطلق انحرافات
Fuzzy least absolute deviation	کمترین قدرمطلق انحرافات فازی
Least absolute error	کمترین قدرمطلق خطا
Response variable	متغیر پاسخ
Predictor variable	متغیر پیش‌بینی
Random variable	متغیر تصادفی
Fuzzy random variable	متغیر تصادفی فازی
Output variable	متغیر خروجی
Input value	متغیر ورودی
Fuzzy set	مجموعه فازی
Fuzzy environment	محیط فازی
Fuzzy observation	مشاهده فازی
Mean predicted error	میانگین خطای پیش‌بینی
Leverage point	نقطه نافذ
Goodness of fit	نیکویی برازش

واژه‌نامه انگلیسی به فارسی

Additivity principle	اصل جمع پذیری
Arithmetic of fuzzy numbers	حساب اعداد فازی
Crisp	دقیق
Decision maker	تصمیم گیرنده
Decomposition identity	اتحاد تجزیه
Elliptic fuzzy number	عدد فازی سهموی
Estimate	برآورد
Estimated data	داده برآورد شده
Estimation error	خطای برآورد
Estimator	برآوردگر
Extension principle	اصل توسیع (اصل گسترش)
Fuzzy distance	فاصله فازی
Fuzzy environment	محیط فازی
Fuzzy least absolute deviation	کمترین قدرمطلق انحرافات فازی
Fuzzy measure	اندازه فازی
Fuzzy number	عدد فازی
Fuzzy observation	مشاهده فازی
Fuzzy random variable	متغیر تصادفی فازی
Fuzzy ranking method	روش رتبه‌بندی فازی
Fuzzy regression	رگرسیون فازی
Fuzzy set	مجموعه فازی
Generalized M-estimators	برآوردگرهای M-تعمیم یافته
Goodness of fit	نیکویی برازش
h-cut	برش - h
Input value	متغیر ورودی
Least absolute deviation	کمترین قدرمطلق انحرافات

Least absolute error	کمترین قدرمطلق خطا
Least median squares	ترین میانه توان‌های دوم
Least squares	کمترین توان‌های دوم
Least trimmed squares	کمترین توان‌های دوم پیراسته
Left spread	پهنای چپ
Leverage poin	نقطه نافذ
Likelihood function	تابع درست‌نمایی
Linear regression	رگرسیون خطی
Mean predicted error	میانگین خطای پیش‌بینی
Membershib function	تابع عضویت
M-estimates	برآوردگرها-M
Multiple linear regression	رگرسیون خطی چندگانه
Normal fuzzy number	عدد فازی نرمال
Objective function	تابع هدف
Observed data	داده مشاهده شده
Outlier	داده دور افتاده (پرت)
Output variable	متغیر خروجی
Overall existance level	سطح حضور سراسری
Predictor variable	متغیر پیش‌بینی
Random variable	متغیر تصادفی
Regression outlier	دورافتاده رگرسیونی
Residual	باقیمانده
Respose variable	متغیر پاسخ
Right spread	پهنای راست
Robust regression	رگرسیون استوار
Spread	پهنا
Symetric fuzzy number	عدد فازی متقارن
Top flattening	بالا صاف
Trangular fuzzy number	عدد فازی مثلثی
Trapezodail fuzzy number	عدد فازی ذوزنقه‌ای
Weighted fuzzy least squares	کمترین توان‌های دوم موزون فازی
Weighted least squares	کمترین توان‌های دوم موزون
Well-defined	خوش تعریف

Aabstract

If the regularity assumptions of least squares method are satisfied, the estimates of this method will be known as the best one. If outliers exist in data set, traditional methods won't result in good consequences and we should use the alternative robust regression. On the other hand, when we have fuzzy observations, we can't use the traditional methods to model them. So we should use the alternative fuzzy regression. When we have both fuzzy observations and outliers in data set, we should use the alternative fuzzy robust regression. After introducing the possibilistic fuzzy linear regression, fuzzy regression analysis has been widely studied and applied in various areas. Also the least square methods were generalized in fuzzy environment. These methods are sensitive to the presence of outliers.

In this dissertation, two fuzzy regression analyses are proposed. When independent variables are crisp, the dependent variable is a fuzzy number and outliers exist in data set. First method is a modified fuzzy least squares regression analysis and second method is a modified fuzzy absolute deviations regression analysis. Results demonstrate that the second method is better than the first. In these two methods, the residuals are ranked and then the weight matrices are constructed by defined membership functions. Fuzzy least squares estimations and fuzzy least absolute deviations are obtained by these weight matrices.

key words: Robust regression, outliers, fuzzy regression, Ranking of fuzzy sets.



Shahrood University of Technology

Faculty Of Mathematical Sciences

MSc Thesis in: Mathematical Statistics

**Studying of a robust regression on fuzzy
environment**

By: Elham Rahimian

Supervisors

Dr. Mohammad Reza Rabiei

Dr. Davod Shahsavani

September 2017