مجله مدل‌سازی پیشرفته ریاضی
سال ۱۴۰۳، دوره ۱۴، شماره ۴، ص ۱۰۶-۱۲۷

شاپا چاپی: ۲۲۵۱-۸۰۸۸
DOI: 10.22055/JAMM.2025.47308.2285

آزمون معنی‌داری برآوردگرهای رگرسیون فازی

جلال چاچی[۱] [۱] محمد رضا آخوند[۱] پوران بندانی تَرَشُّکی[۱]

[۱] گروه آمار، دانشکده علوم ریاضی و کامپیوتر، دانشگاه شهید چمران اهواز، اهواز، ایران

دبیر مسئول: حمزه ترابی
تاریخ دریافت: ۱۴۰۳/۴/۱۰ تاریخ پذیرش: ۱۴۰۳/۱۰/۱۸

چکیده: معنی‌داری آماری تعیین کننده این است که آیا رابطه بین دو یا چند متغیر ناشی از عواملی غیر از شانس و تصادف است، به عبارتی آیا نتایج فراهم شده در یک مدلسازی ناشی از داده‌ها است؟ آزمون فرضیه‌های آماری روشی است که توسط آن معنی‌داری آماری تعیین می‌شود. با معرفی رگرسیون فازی، رویکردهای متعددی در برآوردیابی اینگونه مدلها به منظور ارایه برآوردهای دقیق‌تر ارائه شد، اما در مقایسه، توجه بسیار ناچیزی به خواص برآورگردها، فواصل اطمینان، آزمون فرضیه و معنی‌داری مدل شده است. بنابراین هدف اصلی مقاله حاضر این است که در یک چارچوب مدل رگرسیون فازی و با بکار بردن روش m-برآوردگرها، به بررسی معنی‌داری مدل برآورد شده در یک مطالعه کاربردی با داده‌های واقعی فازی-مقدار پرداخته شود. بدین منظور با دسترسی به منبع داده‌های آب و فاضلاب شهر اهواز ابتدا به معرفی متغیرهای مورد نیاز پرداخته شد. از آنجایی که m-برآوردگرها از الگوریتمی مبتنی بر روش وزن-دهی مکرر استفاده می‌کنند، به دنبال تعیین وزن کاربری‌ها از نظر الگوی مصرف هستیم، یعنی وزن کاربری‌های پرمصرف و کاربری‌های کم مصرف در این الگوریتم مشخص می‌شود. اکنون شرکت آب و فاضلاب می‌تواند برمبنای اوزان تعیین شده، ابتدا الگوی مصرف هر کاربری را مشخص و طبقه‌بندی کند و سپس به قیمت‌گذاری پلکانی برای هر کاربری بپردازد. این هدف نیازمند ارایه یک مدل معنی‌دار از رگرسیون فازی است. بنابراین، از آنجاکه m-برآوردگرهای مدل رگرسیون فازی دارای فرم بسته‌ای نیستند، با استفاده از روش بوت استرپ به ارایه شاخص‌های توصیفی برآوردگرها، فواصل اطمینان و آزمون معنی‌داری مدل پرداختیم. در پایان با تحلیل نتایج به‌دست‌آمده و با حفظ پارامترهای معنی‌دار مدل، یک مدل مناسب برای برازش به این داده‌ها معرفی شد.

واژه‌های کلیدی: تحلیل رگرسیون فازی، m-برآوردگرها، داده‌های فازی، معنی‌داری آماری.

رده‌بندی ریاضی: 62A80; 94D05; 03B52

[۱] نویسنده مسئول مقاله

jalal.chachi@scu.ac.ir (J. Chachi)، m.akhond@scu.ac.ir (M.R. Akhond)،
pooran6161@gmail.com (P. Bandani Tarashoki)

۱ مقدمه

تحلیل رگرسیون، استفاده از مجموعه‌ای از روشهای آماری به منظور تعیین رابطه تابعی بین یک متغیر وابسته و تعدادی از متغیرهای مستقل است. برای مدل‌های رگرسیون متداول کلاسیک (غیر فازی)، مقادیر مشاهده شده متغیرهای وابسته و مستقل اعداد دقیق [۲] فرض می‌شوند. با این حال، در کاربردهای دنیای واقعی، به دلیل عدم قطعیت [۳] پارامترها، با وجود خطاهای ابزاری، روش شناختی، محیطی و انسانی، داده‌ها معمولاً اعداد دقیق نیستند بلکه اعداد مبهم [۴] (فازی [۵]) هستند [۴۹]. به عبارتی، زمانی که به عنوان مثال: حجم نمونه خیلی کوچک است، و/یا مفروضات توزیع‌های احتمالاتی برقرار نمی‌باشند، و/یا رابطه بین متغیرهای مستقل و وابسته مبهم/فازی است، و/یا زمانی که رویدادها و پیامدها دقیق نیستند و ابهام دارند، اینگونه روش‌ها دچار اشکال خواهند شد. روش‌های رگرسیونی همچنین بیشترین استفاده را در بین تمام روش‌های آماری دارند و کاربرد وسیعی برای مسائل عملی متعدد پیدا نموده‌اند. علاوه بر این موارد، پیچیدگی مسایل دنیای واقعی به راحتی استفاده مدل‌های اساسی و پایه را جایز نمی‌داند، زیرا اطلاعات و داده‌هایی که اغلب در دنیای واقعی ثبت می‌شوند، از بسیاری جهات نادقیق/مبهم/فازی هستند. بنابراین، مدل‌های رگرسیون کلاسیک در مدلسازی چنین داده‌هایی محدودیت‌هایی را تجربه می‌کنند که برای برطرف کردن آنها اساسا باید روش‌های جدید و نوینی معرفی شوند. برای رفع این انتقادات و برطرف کردن این مسایل، بسیاری از محققان مفاهیم تحلیل رگرسیون آماری را با استفاده از مفاهیم تئوری مجموعه‌های فازی [۱۰، ۱۲، ۴۹] اصلاح، تعمیم و گسترش داده‌اند. بر این اساس پس از معرفی رگرسیون فازی توسط تاناکا و همکاران [۴۸]، رویکردهای متنوعی در مدل‌های رگرسیون فازی، (شامل رویکردهای امکانی، رویکردهای مبتنی بر شیوه کمترین مربعات خطا و دیگر رویکردهای تلفیقی)، به منظور برآورد بهتر و دقیق‌تر پارامترها، ارایه و مورد بررسی و تحلیل قرار گرفته است [۲، ۱۵]. تنوع اینگونه مدل‌ها و رویکردهای برآوردیابی آنها را می‌توان به طور عمده به سه رده اصلی زیر دسته‌بندی نمودند:

۱) تنوع در ساختاربندی و فرمول‌بندی مدل با توجه به فازی بودن (و/یا فازی نبودن) پارامترها و فازی بودن (و/یا فازی نبودن) متغیرهای مدل،

۲) تنوع در نحوه برآوردیابی پارامترهای مدل که در نهایت منجر به تنوع گسترده در معرفی مدل‌های رگرسیون امکانی و مدل‌های رگرسیون کمترین مربعات خطا شده است،

۳) تنوع در رویکردهای ترکیبی [۶] که در آنها چندین روش برآوردیابی متداول در آمار کلاسیک با روش‌های نوین یادگیری ماشین مانند شبکه‌های عصبی، سامانه‌های استنتاج فازی و ... ترکیب می‌شوند.

در حال حاضر، تحقیقات متعدد و فراوانی در زمینه تحلیل رگرسیون فازی، به ویژه در رویکردهای امکانی و حداقل مربعات فازی و رویکردهای تلفیقی (مبتنی بر روش‌های یادگیری ماشین) وجود دارد. علاوه بر این، در روش‌های دیگری از جمله روش‌های احتمالی–امکانی، رگرسیون لجستیک و رگرسیون فازی خوشه‌ای و ... نیز تحقیقات گسترده‌ای صورت گرفته است. همچنین علاوه بر تحقیقات و مقالاتی که عمدتاً به پیشرفت در روش‌شناسی و مسایل نظری اختصاص یافته‌اند، مقالات متعددی نیز وجود دارند که به کاربردهایی از روشهای رگرسیون فازی در زمینه‌های مختلف تحقیقاتی پرداخته‌اند. اما در این تحقیقات به معنی‌داری آماری مدل برآورد شده پرداخته نشده است. به عبارتی تعیین نشده است که آیا رابطه بین دو یا چند متغیر ناشی از عواملی غیر از شانس و تصادف است. بنابراین بررسی خواص برآوردگرها و فرمولبندی آنها به منظور ارایه روابطی که معنی‌داری آماری مدل برآورد شده را نتیجه دهد، نیازمند توجه و بررسی بیشتر در چارچوب مدل‌های رگرسیون فازی است. از جهتی دیگر بسیاری از برآوردگرهای مدل‌های رگرسیون فازی به صورت عددی به دست می‌آیند و دارای فرم بسته‌ای نیستند، لذا بررسی خواص این برآورگرها نیز به صورت فرم‌های بسته امکان‌پذیر نیست و باید از روش‌های عددی استفاده نمود.

در این راستا یکی از روشهای برآوردیابی که اخیرا بسیار مورد توجه قرار گرفته است، روش برآوردیابی وزنی است که یک رویکرد استوار در برآوردیابی مدل رگرسیون فازی است [۱۵]. از آنجا که مدل‌های رگرسیون فازی (وزنی) در کاربرد بسیار مورد توجه هستند، بخصوص در تعیین نقاط پرت و ارایه برآوردگرهای استوار، بنابراین در این مقاله به معرفی m–برآوردگرهای رگرسیون فازی در تحلیل‌داده‌های واقعی آب و فاضلاب شهر اهواز خواهیم پرداخت. بدین منظور با دسترسی به منبع داده‌های آب و فاضلاب شهر اهواز ابتدا به معرفی متغیرهای مورد نیاز می‌پردازیم. اینکار پس از مرتب‌سازی و استخراج شاخص‌های توصیفی داده‌ها صورت می‌گیرد. از آنجا که در محاسبه چنین برآوردگرهایی از الگوریتمی که مبتنی بر روش وزن–دهی تکرار شونده است، استفاده می‌شود، به دنبال این هستیم که در انتها وزن کاربران را از منظر الگوی مصرف مشخص شود. به عبارتی وزن کاربری‌های پرمصرف و وزن کاربری‌های کم مصرف مشخص شود. از جهتی دیگر m–برآوردگرهای مدل‌های رگرسیون فازی به صورت عددی و توسط الگوریتم‌های تکرار شونده حاصل می‌شوند، لذا دارای فرم بسته‌ای نیستند و برای بررسی خواص این برآورگرها از روش بوت‌استرپ استفاده می‌کنیم. روش بوت‌استرپ از جمله روش‌هایی است که ما را قادر می‌سازد تا به بررسی خواص چنین برآوردگرهایی بپردازیم. بنابراین با استفاده از روش بوت استرپ به محاسبه خواص عددی این برآورگرها از جمله میزان اریبی و انحراف معیار پرداخته می‌شود. همچنین با استفاده از توزیع بوت‌استرپی برآوردگرها به انجام آزمون فرضیه و محاسبه فواصل اطمینان برای پارامترها پرداخته می‌شود. در پایان بر اساس تفسیر نتایج به‌دست‌آمده و با انجام آزمون‌های معنی‌داری مدل، به حفظ پارامترهای معنی‌دار در

[۲] Crisp
[۳] Uncertainty
[۴] Imprecise or Vague
[۵] Fuzzy
[۶] Hybrid methods

مدل می‌پردازیم و یک مدل معنی‌دار برای برازش این داده‌ها معرفی می‌شود.

مطالب این مقاله به صورت زیر تدوین شده است. در بخش ۲، برخی از مفاهیم و تعاریف مورد نیاز از مجموعه‌های فازی بیان می‌شوند. در بخش ۳، کلیاتی از مدل‌های رگرسیون فازی بیان می‌شود و در بخش ۴، m-برآورگرها در مدل رگرسیون فازی بیان می‌شود. در بخش ۵، به فواصل اطمینان و آزمون فرضیه بوت استرپی پرداخته می‌شود. در بخش ۶، به تحلیل نتایج مدل‌بندی روی داده‌های آب و فاضلاب شهر اهواز پرداخته می‌شود. در بخش ۷ به آنالیز حساسیت برآورد ضرایب در حضور مشاهدات پرت می‌پردازد. در انتها بحث و نتیجه‌گیری بیان می‌شود.

۲ مفاهیم پایه و نمادها

در این مقاله فرض کنید مجموعه اعداد حقیقی $\mathbb{R}$ مجموعه مرجع باشد. مجموعه فازی $\widetilde{A}$ از $\mathbb{R}$ با تابع عضویت $\widetilde{A}(x):\mathbb{R}\to[0,1]$ مشخص می‌شود [۵۰]. α-برش مجموعه فازی $\widetilde{A}$ برای هر $\alpha\in(0,1]$ به صورت مجموعه معمولی

$$A_\alpha=\{x\in\mathbb{R}:\widetilde{A}(x)\geq\alpha\}$$

تعریف می‌شود و A_0 بستار مجموعه $\{x\in\mathbb{R}:\widetilde{A}(x)>0\}$ است. مجموعه فازی $\widetilde{A}$ از $\mathbb{R}$ را یک عدد فازی گوییم هرگاه برای هر $\alpha\in(0,1]$ مجموعه‌های $A_\alpha=\{x\in\mathbb{R}:\widetilde{A}(x)\geq\alpha\}$ ناتهی، بسته و کراندار باشند. یک رده خاص از اعداد فازی در $\mathbb{R}$ اعداد فازی LR هستند [۵۲]. یک عدد فازی LR به‌صورت $\widetilde{N}=(n,l,r)_{LR}$ نشان داده می‌شود که در آن $n\in\mathbb{R}$ مرکز، $l\in\mathbb{R}^+$ و $r\in\mathbb{R}^+$ به ترتیب پهناهای چپ و راست عدد فازی و $L:\mathbb{R}^+\to[0,1]$ و $R:\mathbb{R}^+\to[0,1]$ به‌ترتیب توابع شکل نزولی چپ و راست، با شرط $L(0)=R(0)=1$ هستند. تابع عضویت و α-برش عدد فازی $\widetilde{N}=(n,l,r)_{LR}$ به صورت زیر تعریف می‌شود

$$\begin{aligned}\widetilde{N}(x)&=\begin{cases}L(\frac{n-x}{l}) & \text{if}\quad x\leq n,\\ R(\frac{x-n}{r}) & \text{if}\quad x\geq n.\end{cases}\\ N_\alpha&=[n-L^{-1}(\alpha)l,n+R^{-1}(\alpha)r],\quad \alpha\in[0,1].\end{aligned}$$

نوع خاصی از اعداد فازی LR، اعداد فازی مثلثی هستند. تابع عضویت و α-برش عدد فازی مثلثی $\widetilde{N}=(n,l,r)_T$ به صورت زیر تعریف می‌شود

$$\begin{aligned}\widetilde{N}(x)&=\begin{cases}\frac{x-(n-l)}{l} & \text{if}\quad x\in[n-l,n],\\ \frac{(n+r)-x}{r} & \text{if}\quad x\in(n,n+r].\end{cases}\\ N_\alpha&=[n-(1-\alpha)l,n+(1-\alpha)r],\qquad \alpha\in[0,1].\end{aligned}$$

عدد فازی $\widetilde{N}=(n,l,r)_{LR}$ با شرایط $L=R$ و $l=r=\lambda$ یک عدد فازی LR متقارن نامیده می‌شود و به‌صورت $\widetilde{N}=(n,\lambda)_L$ نشان داده می‌شود. فرض کنید $\widetilde{M}=(m,l_m,r_m)_{LR}$ و $\widetilde{N}=(n,l_n,r_n)_{LR}$ دو عدد فازی LR باشند و $\lambda\in\mathbb{R}-\{0\}$ یک عدد حقیقی باشد. در اینصورت حساب اعداد فازی به‌صورت زیر تعریف می‌شود

$$\begin{aligned}\lambda\otimes\widetilde{M}&=\begin{cases}(\lambda m,\lambda l_m,\lambda r_m)_{LR} & \text{if}\quad \lambda>0,\\ (\lambda m,|\lambda|r_m,|\lambda|l_m)_{RL} & \text{if}\quad \lambda<0,\end{cases}\\ \widetilde{M}\oplus\widetilde{N}&=(m+n,l_m+l_n,r_m+r_n)_{LR}.\end{aligned}$$

۳ رگرسیون فازی

در حالت کلی، یک مدل رگرسیون فازی به صورت $\widetilde{y}=\widetilde{f_\beta(x)}$ برای مدل‌سازی متغیرهای خروجی-فازی $(\widetilde{y})$ و k متغیر ورودی-دقیق/فازی $(\widetilde{x}/x)$ و پارامترهای دقیق/فازی $(\widetilde{\beta}/\beta)$ تعریف می‌شود. مدل رگرسیون فازی را می‌توان در حالات متنوع زیر (ولی نه فقط محدود به آنها) مورد بررسی قرار داد [۱۵]:

۱) مدل با ورودی‌های–دقیق (x)، خروجی–فازی ($\widetilde{y}$) و پارامترهای فازی $\widetilde{\beta}$.
۲) مدل با ورودی‌های–فازی ($\widetilde{x}$)، خروجی–فازی ($\widetilde{y}$) و پارامترهای دقیق β.
۳) مدل با ورودی‌های–دقیق (x)، خروجی–فازی ($\widetilde{y}$)، و با در نظر گرفتن جمله خطای فازی ($\widetilde{\epsilon}$).
۴) مدل با ورودی‌های–دقیق (x) و خروجی–فازی ($\widetilde{y}$) که در آن فقط پارامتر عرض از مبدا فازی است.
۴) مدل با ورودی‌های–فازی ($\widetilde{x}$) خروجی–فازی ($\widetilde{y}$) و پارامترهای فازی $\widetilde{\beta}$.

ملاحظه ۱.۳. برخی از رویکردهای امکانی در مدل‌بندی رگرسیون فازی معرفی شده‌اند که در آنها خروجی مقادیر دقیق دارند و دیگر کمیت‌ها از قبیل ورودی‌ها و/یا پارامترهای مدل می‌تواند با توجه به ماهیت و ساختار مساله به صورت دقیق و/یا فازی اختیار می‌شود [۴۸].

۱.۳ رویکردهای امکانی

رویکرد امکانی در تحلیل رگرسیون فازی توسط [۴۸] معرفی شد. این روش ابهام کل مدل را با به حداقل رساندن مجموع پهناهای پارامترهای فازی آن به حداقل می‌رساند، مشروط به اینکه دامنه مقادیر برآورد شده فازی در یک سطح خاص α_{fixed}–برش، پوشش کاملی از دامنه مقادیر مشاهده شده فازی در همان سطح خاص α_{fixed}–برش، ارایه نماید. یعنی α_{fixed}–برش خاصی از مشاهدات زیر مجموعه در α_{fixed}–برش مقادیر برآورد شده باشد. در این رویکرد، یک سیستم معادلات خطی–فازی به عنوان یک مدل رگرسیونی فازی پیشنهاد می‌شود و یک مسئله برنامه‌ریزی خطی ریاضی[۷] را با پارامترهای فازی برای تخمین پارامترهای رگرسیون فازی فرمولبندی می‌کنند. از آنجایی که توابع عضویت مجموعه‌های فازی را می‌توان به عنوان توزیع امکان مشاهده کرد، رویکردهای این دسته به عنوان "تحلیل رگرسیون امکانی" نامیده می‌شوند. به دلیل سادگی محاسبات در رویکرد امکانی، این رویکرد توسط محققان متعددی مورد بررسی، بهبود، تعمیم و گسترش قرار گرفت [۱۵]. از جمله انتقاداتی که به رویکرد امکانی می‌شود، این است که ممکن است که خروجی مساله برنامه‌ریزی خطی برای کمینه کردن مجموع ابهام مدل، صفر نتیجه شود. البته این انتقاد یکی از ویژگی‌های ذاتی مساله برنامه‌ریزی خطی است که منجر به برآورد دقیق و بدون ابهام پارامترهایی می‌شود که از ابتدا در مدل رگرسیون فازی ماهیت فازی و همراه با ابهام داشتند. دیگر اینکه به منظور برقراری شروطی که در مساله برنامه‌ریزی خطی است، ممکن است کلیه مشاهدات در برآوردیابی مورد استفاده قرار نگیرند و یا برآوردهایی با ابهام زیاد و غیر قابل قبول حاصل شوند. این رویکرد همچنین در حضور مشاهدات پرت، پیش‌بینی مدل برای مقادیر جدید و وجود همخطی در مشاهدات بسیار ضعیف و ناتوان عمل می‌کند.

۲.۳ رویکردهای حداقل مربعات خطای فازی

تحلیل رگرسیون فازی از دیدگاه روش حداقل مربعات اینگونه است، که در آن فاصله بین مقادیر فازی پیش‌بینی‌شده و مقادیر فازی مشاهده شده با توجه به فاصله بین دو عدد فازی به حداقل می‌رسد. بنابراین رویکرد حداقل مربعات فازی [۹، ۲۰] شامل ترکیب مناسبی از نیکویی برازش و باقیمانده‌ها است که می‌تواند منجر به بررسی دقت مدل شود [۱۰، ۱۲، ۱۳، ۱۵].

۳.۳ رویکردهای مبتنی بر تکنیک‌های یادگیری ماشین

قابلیت تعمیم تحلیل رگرسیون فازی با ترکیب تکنیک‌های یادگیری ماشین [۲۲، ۳۲]، مانند الگوریتم‌های تکاملی، شبکه‌های عصبی یا ماشین‌های بردار پشتیبان، افزایش یافته است [۷، ۱۴، ۲۳، ۳۰، ۳۳].

۴.۳ تاریخچه‌ای از مطالعات رگرسیون فازی و آزمون فرضیه‌های فازی

طی دهه‌های گذشته، مطالعات بسیاری به مدل‌سازی رگرسیون فازی اختصاص یافته است [۱۵]. به عنوان نمونه، چاچی [۱۰] یک مدل رگرسیون فازی کمترین مربعات وزنی را برای داده‌های وردی دقیق و خروجی فازی معرفی کرد (همچنین رجوع شود به [۲۲، ۲۳]). حسامیان و اکبری [۲۶] یک مدل رگرسیون جزئی–نیمه پارامتری با داده‌های ورودی دقیق و ورودی شهودی معرفی نموند. حسامیان و اکبری [۲۷] یک مدل رگرسیون جمعی فازی را برای داده‌های ورودی دقیق و خروجی فازی معرفی کردند. حسامیان و اکبری [۲۸] یک روش استوار با ضرایب متغیر را در مدل‌سازی رگرسیون فازی معرفی کردند. حسامیان و اکبری [۲۹] یک مدل رگرسیون قطعه‌ای فازی را برای داده‌های ورودی دقیق و خروجی فازی معرفی نمودند. خَمّر و همکاران [۳۴–۳۶] مدلهایی از رگرسیون فازی کمترین مربعات استوار را بر پایه استفاده از توابع کرنل معرفی نمودند. عارفی [۵] یک مدل رگرسیون فازی چندکی را برای داده‌های فازی و پارامترهای فازی معرفی نمود. عارفی و خَمّر [۶] پیش‌بینی غیرخطی مدل رگرسیون فازی بر اساس تابع زیان چندکی پرداختند. حسامیان و اکبری [۳۰] یک مدل رگرسیون لجستیک

[۷] Linear Programing

بردار–پشتیبان را برای داده‌های ورودی دقیق و خروجی فازی معرفی کردند. اسداللهی و همکاران [۷] یک مدل رگرسیون بردار–پشتیبان استوار را برای داده‌های ورودی دقیق و خروجی فازی معرفی کردند. باباراسو و مگری [۸] رویکرد جدید مبتنی بر مدل رگرسیون فازی برای یک سیستم فتوولتائیک ارایه نمودند. پاندلارا و همکاران [۴۴] رویکرد علیتی رگرسیون فازی را برای تحلیل رابطه بین مصرف برق و تولید ناخالص داخلی به کار بردند. سلمان و همکاران [۴۷] شواهد جدیدی با استفاده از طراحی ناپیوستگی رگرسیون فازی درتوافقنامه آب و هوای پاریس و کارایی زیست محیطی جهانی ارایه دادند. خاسانزدا و همکاران [۳۷] مدل رگرسیونی برای پیش بینی سرعت جریان باد برای نیازهای انرژی را بر اساس منطق فازی به کار بردند. ژانگ و همکاران [۵۱] رگرسیون فرآیند گاوسی هیبریدی و رویکرد مبتنی بر سیستم استنتاج فازی را برای پایش وضعیت در قسمت روتور یک ژنراتور القایی با تغذیه دوبل به کار بردند.

طی دهه‌های گذشته، مطالعات بسیاری به آزمون فرضیه فازی اختصاص یافته است [۱۷]. چوخروا و جوهانسن آزمون فرضیه فازی برای نسبت جمعیت بر اساس اطلاعات مجموعه‌ای–مقدار [۱۶] و همچنین آزمون فرضیه دو طرفه برای میانه با طبقه‌های فازی [۱۸] را مورد بررسی قرار دادند. آنها همچنین از آزمون فرضیه فازی برای بهبود نمودارهای p–چارتهای شده استفاده کردند [۱۹]. لوبیانو و همکاران [۴۰] به بررسی حساسیت میانگین داده‌های فازی در آزمون فرضیه با توجه به شکل تابع عضویت داده‌ها پرداختند. لوبیانو و همکاران [۳۹] به آزمون فرضیه برای میانگین داده‌های مبتنی بر مقیاس رتبه‌بندی فازی پرداختند و الگوریتم‌ها و کاربردهای مورد نیاز را معرفی کردند. محمدی و همکاران [۴۱] آزمون فرضیه فازی بیزی در شبکه‌های سنسوری بی‌سیم با عدم قطعیت نویزی را مورد مطالعه قرار دادند. اصغری و همکاران [۱] به آزمون آماری براساس فرضیه‌های فازی شهودی پرداختند. پرچمی [۴۵] تصمیم‌گیری فازی در آزمون فرضیه‌ها را مورد مطالعه و بررسی قرار داد. میلوناس و پاپادوپلوس [۴۳] به آزمون فرضیه‌های فازی با داده‌های دقیق بر اساس برآوردگرهای فازی جدید و مقادیر بحرانی فازی پرداختند.

۴ m-برآوردها در رگرسیون فازی

در رگرسیون فازی m–برآوردگرها، هدف این است که با در نظر گرفتن تابعی از خطاها، به‌صورت زیر به برآوردیابی پارامترها، بپردازیم

$$\min_{\beta} \sum_{i=1}^{n} \rho\left(\mathcal{D}(\widetilde{y}_i, \widetilde{f_\beta(x_i)})\right),$$

که در آن

$$\rho(e) = \begin{cases} \frac{e^2}{2}, & \text{اگر } |e| < c, \\ c(|e| - \frac{c}{2}), & \text{اگر } |e| \geq c. \end{cases} \tag{۱.۴}$$

تابع هوبر است [۳۱، ۴۲]، و برای هر $i = 1, \ldots, n$

$$\begin{aligned} e_i &= \mathcal{D}(\widetilde{y}_i, \widetilde{f_\beta(x_i)}) \\ &= \sqrt{\left[y_i - \sum_{j=0}^{k} x_{ij}\,\beta_j\right]^2 + \left[g(l_i) - \sum_{j=0}^{k} x_{ij}\,\gamma_j\right]^2 + \left[g(r_i) - \sum_{j=0}^{k} x_{ij}\,\theta_j\right]^2}. \end{aligned}$$

در این رویکرد به مدل‌سازی مجموعه مشاهدات $(\widetilde{y}_1, \boldsymbol{x}_1), \ldots, (\widetilde{y}_n, \boldsymbol{x}_n)$ پرداخته می‌شود، که مربوط به یک یا چند متغیر مستقل و یک متغیر وابسته می‌باشد. در این مشاهدات، $\boldsymbol{x}_i = (x_{0i}, x_{1i}, \ldots, x_{ki})$، i–امین مشاهده متغیرهای مستقل است، و $\widetilde{y}_i = (y_i, l, r)_{LR}$، $(i = 1, \ldots, n, k < n, x_{0i} = 1)$، i–امین مشاهده متغیر وابسته است. اکنون به منظور مدل‌سازی داده‌های بیان شده، ابتدا تغییر متغیر زیر را بر روی مشاهدات متغیر وابسته اعمال می‌کنیم $\vec{Y} = (y, g(l), g(r))$، که در آن $g(0, \infty) \to \mathbb{R}$ تابعی معکوس‌پذیر است. سپس مدل تبدیل یافته زیر در نظر گرفته می‌شود،

$$\begin{aligned} (y_i, g(l_i), g(r_i)) &= \vec{\Xi}_0 \oplus (\vec{\Xi}_1 \otimes x_{1i}) \oplus \ldots \oplus (\vec{\Xi}_k \otimes x_{ki}) \\ &= (\beta_0, \gamma_0, \theta_0) \oplus ((\beta_1, \gamma_1, \theta_1) \otimes x_{1i}) \oplus \ldots \oplus ((\beta_k, \gamma_k, \theta_k) \otimes x_{ki}) \\ &= (\beta_0 + \beta_1 x_{1i} + \ldots + \beta_k x_{ki}, \gamma_0 + \gamma_1 x_{1i} + \ldots + \gamma_k x_{ki}, \\ &\qquad \theta_0 + \theta_1 x_{1i} + \ldots + \theta_k x_{ki}). \end{aligned} \tag{۲.۴}$$

چون تابع $g:(0,\infty)\to\mathbb{R}$ معکوس‌پذیر است، مدل تبدیل یافته بالا به مدل رگرسیون فازی

$$\begin{aligned}(y,l,r) &= (\beta_0+\beta_1x_1+\ldots+\beta_kx_k, g^{-1}(\gamma_0+\gamma_1x_1+\ldots+\gamma_kx_k),\\ &\qquad g^{-1}(\theta_0+\theta_1x_1+\ldots+\theta_kx_k))_{LR}.\end{aligned}$$

تبدیل می‌شود. در ادامه پارامترهای $\vec{\Xi}_{(k+1)\times 1}=[(\beta_0,\gamma_0,\theta_0),(\beta_1,\gamma_1,\theta_1),\ldots,(\beta_k,\gamma_k,\theta_k)]'$ باید برآورد شوند.

تعریف ۱.۴ ([۱۳]). m–برآوردگرها در رگرسیون با کمینه‌سازی تابع هدف $\sum_{i=1}^{n}\rho\left(\frac{e_i}{S}\right)$ به دست می‌آیند که در آن

$$\widehat{S}=\text{median}_{i=1,\ldots,n}\frac{|e_i-\text{median}_{i=1,\ldots,n}\{e_i\}|}{0/6745}.$$

ملاحظه ۲.۴. مقدار ثابت c در تابع هوبر (۱.۴) برای عبارات خطای بی–مقیاس شده $\frac{e_i}{S}$ برای هر $i=1,\ldots,n$ برابر ۱/۳۴۵ در نظر گرفته می‌شود [۳۱].

m–برآوردگرهای رگرسیون فازی از کمینه کردن تابع هدف $\min_{\vec{\Xi}}\sum_{i=1}^{n}\rho\left(\frac{e_i}{S}\right)$ حاصل می‌شوند. حداقل مقدار $M(\vec{\Xi})$ را می‌توان با مشتق گرفتن از تابع هدف نسبت به پارامترهای β_j، γ_j و θ_j برای هر $j=0,1,\ldots,k$، و مساوی صفر قرار دادن آنها، سیستمی از $3(k+1)$ معادله برای هر $j=0,1,\ldots,k$ به صورت زیر حاصل می‌شود [۱۳]

$$\begin{aligned}\sum_{i=1}^{n}x_{ij}\left(y_i-\sum_{j=0}^{k}x_{ij}\,\beta_j\right)\omega_i&=0,\\ \sum_{i=1}^{n}x_{ij}\left(g(l_i)-\sum_{j=0}^{k}x_{ij}\,\gamma_j\right)\omega_i&=0,\\ \sum_{i=1}^{n}x_{ij}\left(g(r_i)-\sum_{j=0}^{k}x_{ij}\,\theta_j\right)\omega_i&=0,\end{aligned}\tag{۳.۴}$$

که در آن

$$\omega_i=\frac{\psi(u_i)}{u_i}=\begin{cases}1 & \text{اگر } |u_i|\le 1/345,\\ \frac{1/345}{|u_i|} & \text{اگر } |u_i|>1/345.\end{cases},\qquad i=1,\ldots,n,$$

همچنین $\psi=\rho'$ مشتق تابع ρ است و $u_i=\frac{e_i}{S}$، برای هر $i=1,\ldots,n$ است .

ملاحظه ۳.۴. در سیستم معادلاتی (۳.۴) تابع ψ یک تابع غیر خطی است و وزن‌های ω_i، برای هر $i=1,\ldots,n$، به باقیمانده‌های u_i بستگی دارد. از طرفی، این باقیمانده‌ها به ضرایب تخمین زده شده β_j، γ_j، و θ_j برای هر $j=0,1,\ldots,k$، بستگی دارند. همچنین ضرایب تخمینی خود به وزن‌ها بستگی دارند. بنابراین، سیستم معادلات (۳.۴) باید با روش‌های تکراری حل شوند. در حالی که چندین روش بهینه‌سازی غیرخطی را می‌توان به کار گرفت [۴۶]، الگوریتم وزن‌دهی مجدد بیشتر مورد استفاده قرار می‌گیرد [۱۳، ۴۲].

ملاحظه ۴.۴. رویکردهای مختلفی از رگرسیون فازی m–برآوردگرها را می‌توان بر حسب انتخابهای متنوع تابع $\rho(\cdot)$، نوع مدل $\widetilde{f_\beta(x)}$ و تعریف فاصله $\mathcal{D}(\cdot,\cdot)$ بین اعداد فازی معرفی و بررسی نمود.

۵ آزمون معنی‌داری مدل رگرسیون فازی

آزمون فرضیه و فاصله اطمینان دو موضوع کلیدی بسیار مهم و مورد توجه در استنباط آماری هستند. اما استنباط آماری در مدلهای رگرسیون فازی کمتر مورد توجه قرار گرفته است. یکی از دلایل عدم توجه کافی به استنباط آماری در مدلهای رگرسیون فازی را می‌توان عدم وجود پیش‌فرض‌هایی برای اینگونه مدلها دانست. به عبارتی در اینگونه مدل‌ها فرضیه‌های زیربنایی مشابه آنچه که در مدل‌های رگرسیون کلاسیک وجود دارد، در نظر گرفته نمی‌شود. البته باید توجه نمود که یکی از شرایط و دلایل استفاده از مدلهای رگرسیون فازی این است که در اینجا فرضیه‌های زیربنایی مدل‌های رگرسیون کلاسیک برقرار نیست و/یا داده‌ها به صورت فازی ثبت شده‌اند. در این بخش، نتایج آزمون فرضیه و فاصله اطمینان بوت استرپی برای ضرایب رگرسیون m–برآوردگرها فازی، با برنامه‌نویسی در نرم افزار R بررسی و ارایه شده‌اند [۲۵].

۱.۵ روش بوت‌استرپ

روش بوت استرپ موردها اغلب در مسائل با داده‌های چندمتغیره به‌صورت زیر اجرا می‌شود

$$\boldsymbol{\Delta}_{n\times(k+2)} = \begin{bmatrix} \mathbf{z}_1 \\ \vdots \\ \mathbf{z}_i \\ \vdots \\ \mathbf{z}_n \end{bmatrix}_{n\times(k+2)} = \begin{bmatrix} y_1 & 1 & x_{11} & \dots & x_{1k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ y_i & 1 & x_{i1} & \dots & x_{ik} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ y_n & 1 & x_{n1} & \dots & x_{nk} \end{bmatrix}_{n\times(k+2)}$$

در این روش ابتدا به برآورد مدل رگرسیون فازی پرداخته می‌شود، یعنی سیستم معادلات (۳.۴) برای داده‌های اصلی $\boldsymbol{\Delta}_{n\times(k+2)}$ حل می‌شود و برآورد $\widehat{\vec{\Xi}}^{(0)}$ ثبت می‌شود. اکنون در روش بوت استرپ موردها، گام‌های زیر به تعداد $b = 1, 2, \dots, B$ اجرا می‌شود:

۱. نمونه بوت استرپ $\boldsymbol{\Delta}^*_{n\times(k+2)} = [\mathbf{z}^*_1, \mathbf{z}^*_2, \dots, \mathbf{z}^*_n]^T_{n\times(k+2)}$ را باجایگذاری از نمونه اصلی $[\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n]^T_{n\times(k+2)}$ انتخاب کنید.

۲. براساس نمونه بوت استرپ $\boldsymbol{\Delta}^*_{n\times(k+2)}$ سیستم معادلات (۳.۴) حل شود و دنباله برآوردهای $\left\{\widehat{\vec{\Xi}}^{(b)}\right\}_{b=1}^{B}$ را ثبت کنید.

۲.۵ فاصله اطمینان پارامترهای مدل رگرسیون فازی

در این بخش هدف این است که برآوردهای فاصله اطمینان $100(1-\alpha)\%$ را برای پارامترها $\vec{\Xi}$ بر اساس دنباله برآوردهای بوت استرپی $\left\{\widehat{\vec{\Xi}}^{(b)}\right\}_{b=1}^{B}$ تعیین کنیم. در این ارتباط، ابتدا بر اساس دنباله‌های برآوردهای بوت استرپی

$$\left\{\widehat{\beta}_j^{(b)}\right\}_{b=1}^{B}, \quad \left\{\widehat{\gamma}_j^{(b)}\right\}_{b=1}^{B}, \quad \left\{\widehat{\theta}_j^{(b)}\right\}_{b=1}^{B}, \qquad j = 0, 1, \dots, k.$$

به ترتیب به تعیین توزیع تجربی برآوردگرهای $\widehat{\beta}_j, \widehat{\gamma}_j, \widehat{\theta}_j$، پرداخته می‌شود. اکنون با تعیین چندک مرتبه $[B \times \frac{\alpha}{2}]$-ام و چندک مرتبه $[B \times (1-\frac{\alpha}{2})]$-ام این توزیع‌ها می‌توان به محاسبه فواصل اطمینان $100(1-\alpha)\%$ برای پارامترهای مورد نظر پرداخت.

ملاحظه ۱.۵. روش‌های متنوعی در تعیین فواصل اطمینان بوت استرپی وجود دارد، که می‌توان از آنها استفاده کرد. در این مقاله از روش پایه‌ای فواصل اطمینان بوت استرپی[8] استفاده می‌شود، که در آن فاصله‌ای که با چندک‌های مرتبه $[B \times \frac{\alpha}{2}]$-ام و $[B \times (1-\frac{\alpha}{2})]$-ام توزیع تجربی، ساخته می‌شود را به عنوان فاصله اطمینان بوت استرپی پارامتر مورد توجه در نظر می‌گیرد [۲۴]. توجه کنید که از سایر فواصل اطمینان بوت استرپی مانند تسریع شده، اریب–تصحیح شده و استیودنت شده نیز می‌توان به‌طور مشابه استفاده نمود و نتایج را تفسیر کرد.

۳.۵ آزمون فرضیه پارمترهای رگرسیون فازی

برای آزمون فرضیه‌های زیر

$$\begin{cases} H_0: & \vec{\Xi}_j = \vec{\Xi}_{j0} \\ H_1: & \vec{\Xi}_j \neq \vec{\Xi}_{j0} \end{cases}, \qquad j = 0, 1, \dots, k.$$

که معادل فرضیه‌های زیر هستند

$$\begin{cases} H_0: & (\beta_j, \gamma_j, \theta_j) = (\beta_{j0}, \gamma_{j0}, \theta_{j0}) \\ H_1: & (\beta_j, \gamma_j, \theta_j) \neq (\beta_{j0}, \gamma_{j0}, \theta_{j0}) \end{cases}, \qquad j = 0, 1, \dots, k.$$

---

[8] Basic Bootstrap Confidence Interval

از دنباله برآوردگرهای بوت‌استرپی به‌دست آمده استفاده می‌کنیم. بدین منظور ابتدا فرضیه‌ها را به‌صورت زیر تقسیم‌بندی می‌کنیم

$$\begin{cases} H'_\circ: & \beta_j = \beta_{j\circ} \\ H'_\backslash: & \beta_j \neq \beta_{j\circ} \end{cases}, \qquad j = \circ, \backslash, \dots, k.$$

$$\begin{cases} H''_\circ: & \gamma_j = \gamma_{j\circ} \\ H''_\backslash: & \gamma_j \neq \gamma_{j\circ} \end{cases}, \qquad j = \circ, \backslash, \dots, k.$$

$$\begin{cases} H'''_\circ: & \theta_j = \theta_{j\circ} \\ H'''_\backslash: & \theta_j \neq \theta_{j\circ} \end{cases}, \qquad j = \circ, \backslash, \dots, k.$$

برای آزمون فرضیه‌های بالا از پی–مقدار به‌صورت زیر استفاده می‌شود. برای برآورد پی–مقدار، نسبت آماره‌هایی که از آماره آزمون فاصله دارد به‌صورت زیر تشکیل می‌شود

$$p - value = ۲ \min \left\{ \frac{\#(t_b \leq t_\circ)}{B + \backslash}, \frac{\#(t_b \geq t_\circ)}{B + \backslash} \right\}, \qquad j = \circ, \backslash, \dots, k.$$

اگر پی–مقدار از سطح آزمون مورد نظر α بزرگتر یا مساوی باشد، فرضیه صفر پذیرفته می‌شود، و در غیر این صورت رد می‌شود.

ملاحظه ۲.۵. برای آزمون فرضیه‌های بالا همچنین می‌توان از فواصل اطمینان بوت‌استرپی ساخته شده استفاده نمود و به آزمون فرضیه‌های بالا پرداخت [۲۴]. در اینجا اگر فاصله اطمینان $\%\backslash\circ\circ(\backslash - \alpha)$ ساخته شده، مقدار مفروض در فرضیه $H_\circ$ را در بر داشت، این فرضیه در اندازه خطای α پذیرفته می‌شود، در غیر اینصورت رد می‌شود.

۶ تحلیل داده‌های آب و فاضلاب شهر اهواز

در ادامه کاربرد و تحلیل روش m–برآوردگرها در برآوردیابی پارامترهای مدل رگرسیون فازی با داده‌های ورودی حقیقی مقدار و داده‌های خروجی فازی مقدار، مورد بررسی قرار خواهد گرفت. داده‌ها در این مدل رگرسیون فازی، داده‌های واقعی هستند که از منابع شرکت آب و فاضلاب شهر اهواز به‌دست آمده‌اند.

۱.۶ معرفی داده‌های آب و فاضلاب شهر اهواز

قیمت مصرف آب در در دهه‌های گذشته و تاکنون بسیار مورد توجه در مجموعه درآمدهای ملی بوده است. همچنین افزایش بی‌رویه و غیر منطقی مصرف آب نه تنها اقتصاد کشور را به خطر می‌اندازد، بلکه ارتباطات اجتماعی و منابع طبیعی کشور را به خطر می‌اندازد. در این ارتباط، طی یکسال اخیر اطلاعات قبوض آب شهر اهواز برای ۳۰۳ مشترک در ۲۵ کاربری مختلف ثبت شده‌اند. در این مطالعه سطح مصرف آب کاربری (بر حسب متر مکعب) یک کمیت فازی و غیر دقیق بود، که به‌صورت اعداد فازی مثلثی ثبت شده است (جدول ۱). به عنوان نمونه در کاربری خشک‌شویی، تعداد قبوض ۷ مشترک با کاربری مصرف خشک‌شویی در طی یکسال در اختیار بوده است (در مجموع حدود ۲۷ قبض برای کاربری خشک‌شویی در اختیار بود است). با میانگین‌گیری روی قبوض هریک از مشترکان خشک‌شویی، متوسط قبض هر یک از مشترکان حاصل شده است. به عنوان نمونه، اولین مشترک از ۷ مشترک خشک‌شویی در طی یک سال مصرف، تعداد ۴ قبض ثبتی طی دوره داشته است. هر قبض شامل تعداد روزهای مصرف به همراه میزان مصرف آب بر حسب مترمکعب است، که به راحتی میانگین مصرف روزانه مشترک (یا میانگین مصرف هر دوره زمانی دلخواه دیگری) را می‌توان محاسبه نمود. این همسان‌سازی برای هریک از ۷ مشترک دیگر انجام می‌شود. در انتها مقدار متوسط روز مصرفی، که میانگین روزهای مصرفی ۷ مشترک با کاربری خشک‌شویی است، به عنوان متغیر مستقل دقیق–مقدار (غیر فازی) ثبت می‌شوند (یعنی مقدار $X = ۲۰۷/۸$ در ستون سوم از جدول ۱). همچنین متغیر وابسته فازی–مقدار، میزان سطح مصرف آب کاربری خشک‌شویی است که بر اساس اطلاعات مصارف همسان‌سازی شده ۷ مشترک خشک‌شویی باید محاسبه و ثبت شود. اکنون ۷ مقدار همسان‌سازی شده مطابق با روش معرفی شده توسط [۳۸] به صورت مجموعه فازی $\widetilde{y} = (y, l, r)_T = (۱۱۷/۰,\ ۷۹/۰,\ ۵۹/۲)_T$ ثبت می‌شود. بر این اساس مطابق الگوی بیان شده، اطلاعات ۳۰۳ مشترک در ۲۵ کاربری مختلف در قالب جدول ۱ ثبت شده است.

ملاحظه ۱.۶. روش بوت استرپ که در ادامه بر روی این داده‌ها اعمال می‌شود، به گونه‌ای است که ابتدا از ۳۰۳ مشترک اصلی، با نسبت‌های "تعداد مشترک مورد بررسی در هر کاربری" در ستون سوم جدول ۱، نمونه باز تولید می‌شود. سپس این نمونه مطابق با فرآیند بالا فازی‌سازی می‌شود و نمونه جدید بوت‌استرپی برای ۲۵ کاربری ثبت می‌شود. نتایجی که در این راستا ارایه خواهد شد، خروجی برنامه‌های اجرا شده در نرم افزار R هستند [۲۵].

جدول ۱: داده‌های آب و فاضلاب شهر اهواز

شماره	نوع کاربری	تعداد مشترک مورد بررسی در هر کاربری	X	$\tilde{y}=(y,\ l,\ r)_T$
۱	خشک‌شویی	۷	۲۰۷٫۸	(۱۱۷٫۰, ۷۹٫۰, ۵۹٫۲)$_T$
۲	نانوایی	۱۸	۲۴۶٫۸	(۱۴۰٫۰, ۱۱۵٫۹, ۱۱۲٫۷)$_T$
۳	باشگاه ورزشی	۱۰	۲۸۶٫۲	(۱۱۵٫۰, ۷۴٫۳, ۶۱۲٫۲)$_T$
۴	کترینگ	۷	۲۴۰٫۲	(۱۵۷٫۰, ۶۱٫۶, ۵۸٫۸)$_T$
۵	مرغ فروشی	۱۰	۲۱۴٫۱	(۵۸٫۵, ۳۲٫۷, ۵۳٫۱)$_T$
۶	کافی شاپ	۱۶	۳۹۸٫۶	(۶۴٫۵, ۳۰٫۰, ۴۷۶٫۰)$_T$
۷	لبنیات	۱۳	۲۱۶٫۶	(۵۳٫۰, ۴۰٫۰, ۴۶٫۰)$_T$
۸	آرایشگاه/سلمانی	۱۲	۲۲۳٫۸	(۵۵٫۰, ۳۱٫۵, ۱۳۱٫۲)$_T$
۹	اغذیه فروشی	۱۵	۲۹۰٫۹	(۵۸٫۰, ۴۳٫۸, ۱۹۰٫۸)$_T$
۱۰	خانه بازی و سرگرمی	۱۱	۲۵۵٫۹	(۱۰۷٫۰, ۷۷٫۰, ۳۲۳٫۰)$_T$
۱۱	کلینیک/درمانگاه	۷	۱۹۴٫۷	(۵۹٫۰, ۴۷٫۸, ۲۳۵٫۲)$_T$
۱۲	کارواش	۶	۲۱۶٫۸	(۵۵۶٫۵, ۴۰۰٫۵, ۵۳۱٫۰)$_T$
۱۳	قنادی	۱۱	۲۰۸٫۸	(۵۳٫۰, ۴۶٫۰, ۷۲٫۰)$_T$
۱۴	قصابی	۱۷	۲۴۷٫۴	(۵۴٫۰, ۳۸٫۴, ۸۰٫۰)$_T$
۱۵	ماهی فروشی	۴	۲۰۰٫۰	(۴۶٫۰, ۲۱٫۰, ۴۶٫۹)$_T$
۱۶	آب تصفیه	۵	۳۳۷٫۰	(۴۴۲٫۰, ۲۴۴٫۴, ۹۴۵٫۶)$_T$
۱۷	شیرینی‌پزی	۱۳	۲۰۸٫۹	(۶۱٫۰, ۴۶٫۸, ۲۲۷٫۴)$_T$
۱۸	قهوه‌خانه	۱۷	۳۸۸٫۷	(۶۴٫۰, ۳۱٫۲, ۴۷۱٫۶)$_T$
۱۹	شرکت ملی حفاری	۱۸	۲۳۵٫۰	(۱۱۶۲٫۵, ۶۸۱٫۸, ۱۹۲۷٫۷)$_T$
۲۰	آشپزخانه‌ها	۱۷	۲۶۲٫۵	(۱۴۸٫۰, ۱۱۷٫۰, ۵۷۳٫۸)$_T$
۲۱	تالار	۲۰	۲۹۱٫۹	(۲۳۶٫۰, ۲۱۷٫۴, ۷۵۲٫۴)$_T$
۲۲	فست‌فود	۱۷	۲۸۰٫۰	(۵۸٫۰, ۴۷٫۴, ۱۸۷٫۲)$_T$
۲۳	بانک	۱۳	۲۰۸٫۷	(۱۲۱٫۰, ۸۲٫۲, ۸۹٫۴)$_T$
۲۴	مجموعه ورزشی	۱۳	۲۴۵٫۶	(۸۱٫۰, ۵۷٫۴, ۵۶٫۲)$_T$
۲۵	آموزشگاه‌های علمی	۶	۲۶۳٫۰	(۴۸٫۰, ۳۷٫۵, ۱۸۷٫۵)$_T$
	جمع	۳۰۳		

۲.۶ مدل اول به صورت $\vec{y}=(\beta_0,\gamma_0,\theta_0)\oplus((\beta_1,\gamma_1,\theta_1)\otimes x)$

مدل برآورد شده بر مبنای داده‌های اصلی جدول ۱ به‌صورت زیر به‌دست می‌آید

$$
\begin{aligned}
\widehat{\vec{y}} &= \left(\widehat{y},\widehat{\log(l)},\widehat{\log(r)}\right)\\
&= \widehat{\vec{\Xi}}_0 \oplus \left(\widehat{\vec{\Xi}}_1 \otimes x\right) = \left(\widehat{\beta}_0,\widehat{\gamma}_0,\widehat{\theta}_0\right) \oplus \left(\left(\widehat{\beta}_1,\widehat{\gamma}_1,\widehat{\theta}_1\right)\otimes x\right)\\
&= (74.744,\ 4.189,\ 2.640) \oplus ((0.016,\ -0.001,\ 0.001)\otimes x)\\
&= (74.744+0.016x,\ 4.189-0.001x,\ 2.640+0.001x).
\end{aligned}
\tag{1.6}
$$

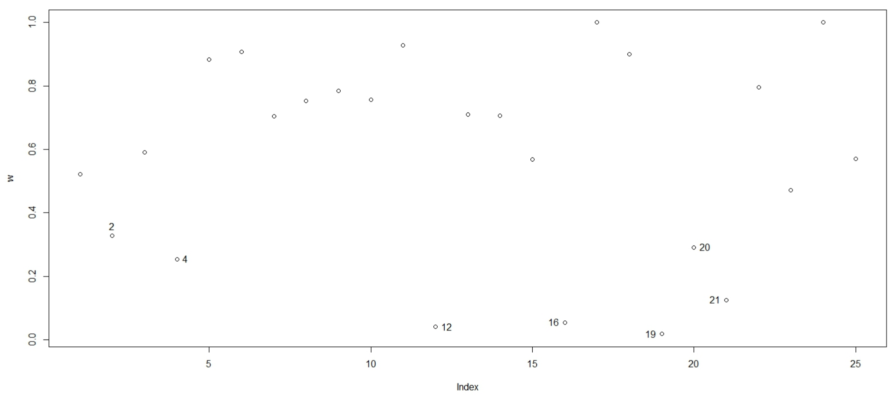
اکنون با توجه به معکوس‌پذیری تابع log، مدل رگرسیون فازی در برازش به مجموعه مشاهدات جدول ۱ به صورت زیر حاصل می‌شود

$$
\widehat{\tilde{y}} = (\widehat{y},\widehat{l},\widehat{r}) = (74.744+0.016x,\ \exp\{4.189-0.001x\},\ \exp\{2.640+0.001x\}),
$$

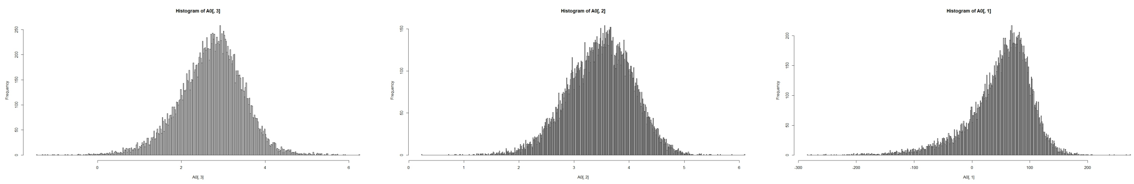
وزنهای حاصل از اجرای الگوریتم در نمودار ۱ نشان داده شده است. در این شکل کاربری‌های با شماره‌های ۱۲، ۱۶، ۱۹، ۲۱ یعنی کارواش، آب تصفیه، شرکت ملی حفاری و تالارها کمترین وزن‌ها را می‌گیرند (بین ۰ تا ۲٫۰). این کاربری‌ها دارای الگوی مصرف بالا دارند. همچنین کاربری‌های با شماره‌های ۲، ۴، ۲۰ یعنی نانوایی‌ها، کترینگ و آشپزخانه‌ها وزن‌هایی (بین ۲٫۰ تا ۴٫۰). می‌توان گفت این کاربری‌ها الگوی مصرف متوسط دارند. دیگر وزن‌های کاربری‌ها، متناسب با نوع الگوی مصرف و تاثیرگذاری آنها در برآوردیابی مدل، در شکل ۱ نشان داده شده است. مقدار وزن نزدیک به ۱ نشان از تاثیر بیشتر و مقدار وزن نزدیک به صفر نشان از تاثیر کمتر آن مشاهده در برآوردیابی دارد.

اکنون با اعمال روش بوت استرپ موردها، به تعداد $B=20000$ بار به باز تولید نمونه‌های بوت‌استرپی از داده‌های جدول ۱ می‌پردازیم (ملاحضه ۱.۶). برای هر نمونه بوت‌استرپی تولید شده، مدل رگرسیون فازی بر مبنای رویکرد m–برآورد به دست می‌آید. توزیع برآوردگرهای $\widehat{\vec{\Xi}}_0$ و $\widehat{\vec{\Xi}}_1$ در شکل‌های ۲ و ۳ رسم شده است. این توزیع‌های تجربی، بر اساس مقادیر حاصل شده $\vec{\Xi}_0$ و $\vec{\Xi}_1$ در نمونه‌های باز–تولید شده، توسط روش بوت استرپ حاصل شده‌اند. بر این اساس در جدول ۲ فواصل بوت استرپ ۹۵٪ پارامترهای $\vec{\Xi}_0$ و $\vec{\Xi}_1$ آمده است. مقادیر پی–مقدار برای آزمون معنی داری مدل؛ یعنی آزمون فرضیه‌های زیر

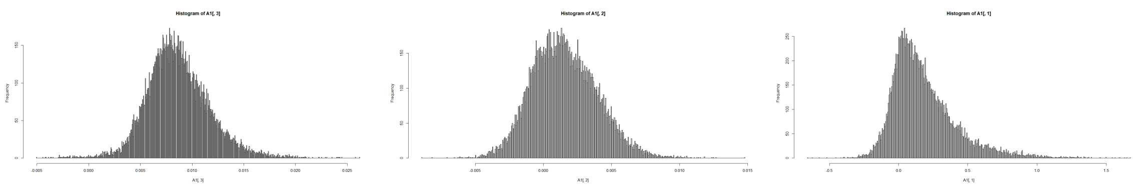
$$
\begin{cases}
H_0: & (\beta_j,\gamma_j,\theta_j)=(0,0,0)\\
H_1: & (\beta_j,\gamma_j,\theta_j)\neq(0,0,0)
\end{cases},\qquad j=0,1.
$$



شکل ۱: وزن‌های مشاهدات، متناسب با تاثیرگذاری آنها در برآوردیابی مدل اول



شکل ۲: توزیع برآوردگرهای $\widehat{\vec{\Xi}}_{\circ} = (\widehat{\beta}_{\circ}, \widehat{\gamma}_{\circ}, \widehat{\theta}_{\circ})$ به ترتیب از چپ به راست برای $\widehat{\beta}_{\circ}$، $\widehat{\gamma}_{\circ}$ و $\widehat{\theta}_{\circ}$.



شکل ۳: توزیع برآوردگرهای $\widehat{\vec{\Xi}}_{1} = (\widehat{\beta}_{1}, \widehat{\gamma}_{1}, \widehat{\theta}_{1})$ به ترتیب از چپ به راست برای $\widehat{\beta}_{1}$، $\widehat{\gamma}_{1}$ و $\widehat{\theta}_{1}$.

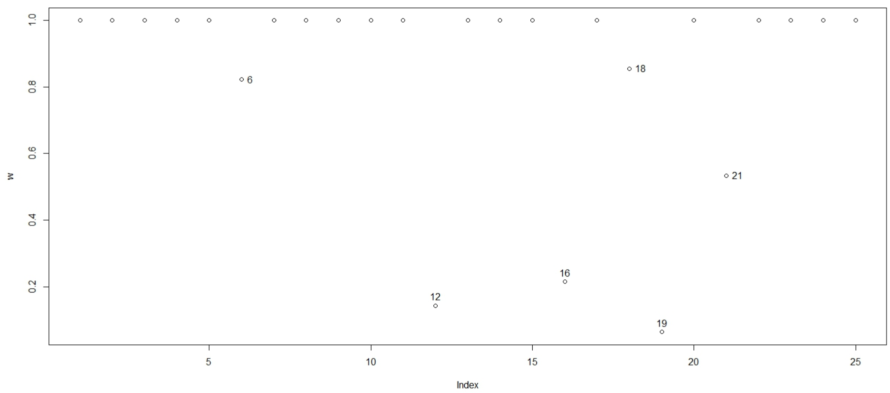
جدول ۲: خصوصیات برآوردگرهای مبتنی بر نمونه اصلی و نمونه های تولید شده توسط بوت‌استرپ در مدل (۱.۶)

پارامترها	برآورد بر مبنای نمونه اصلی	متوسط برآوردهای بوت استرپی	$BIAS$	SSE	MSE
$\beta_{\circ}$	۷۴٫۷۴۴	۴۹٫۷۹۱	−۲۴٫۹۵۲	۲۸۶۴٫۹۱۸	۳۴۸۷٫۵۴۴
$\gamma_{\circ}$	۴٫۱۸۹	۳٫۴۷۵	−۰٫۷۱۳	۰٫۳۴۳	۰٫۸۵۲
$\theta_{\circ}$	۲٫۶۴۰	۲٫۷۱۶	۰٫۰۷۶	۰٫۵۷۳	۰٫۵۷۹
β_{1}	۰٫۰۱۶	۰٫۱۷۹	۰٫۱۶۳	۰٫۰۵۲	۰٫۰۷۹
γ_{1}	−۰٫۰۰۱	۰٫۰۰۱	۰٫۰۰۲	۰٫۰۰۰	۰٫۰۰۰
θ_{1}	۰٫۰۰۹	۰٫۰۰۸	−۰٫۰۰۰	۰٫۰۰۰	۰٫۰۰۰

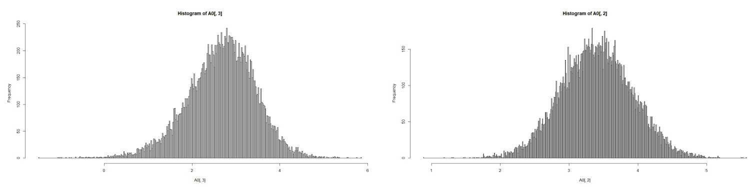
نیز در جدول ۳ آورده شده است. اکنون بر اساس ستون پی–مقدار نتایج بیان شده در جدول ۳ گویای این مطلب است که مدل رگرسیون فازی در تعیین رابطه بین متغیر مستقل "متوسط تعداد روزهای مصرفی آب" با "سطح میزان مصرف کاربری" به طور کامل معنی‌دار نیستند و باید از مدل حذف شوند، زیرا پی–مقدار آزمون‌های مربوط به پارامترهای $\beta_{\circ}$، β_{1} و γ_{1}، بیشتر از $\circ/\circ ۵$ هستند. همچنین فواصل اطمینان ۹۵٪ حاصل شده برای سه پارامتر مذکور شامل عدد صفر هستند، که از این جهت نیز فرضیه $H_{\circ}$ متناظر با این پارامترها پذیرفته می‌شود.

جدول ۳: پی-مقدارها و فواصل اطمینان ۹۵٪ بوت استرپی برای پارامترهای $\vec{\Xi}_۰ = (\beta_۰, \gamma_۰, \theta_۰)$ و $\vec{\Xi}_۱ = (\beta_۱, \gamma_۱, \theta_۱)$

پارامترها	فواصل اطمینان ۹۵٪ بوت استرپی	پی-مقدار
$\beta_۰$	(−۸۳٫۸۰۵, ۱۳۱٫۰۲۱)	۰٫۱۶۲
$\gamma_۰$	(۲٫۲۷۷, ۴٫۵۴۱)	۰٫۰۰۰
$\theta_۰$	(۱٫۱۰۷, ۴٫۰۸۵)	۰٫۰۰۰
$\beta_۱$	(−۰٫۱۳۵, ۰٫۷۶۷)	۰٫۹۴۳
$\gamma_۱$	(−۰٫۰۰۲, ۰٫۰۰۶)	۰٫۵۸۳
$\theta_۱$	(۰٫۰۰۳, ۰٫۰۱۵)	۰٫۰۰۱



شکل ۴: وزن‌های مشاهدات، متناسب با تاثیرگذاری آنها در برآوردیابی مدل (۲.۶)



شکل ۵: توزیع برآوردگرهای $\widehat{\vec{\Xi}}_۰ = (۰, \widehat{\gamma}_۰, \widehat{\theta}_۰)$ به ترتیب از چپ به راست برای $\widehat{\gamma}_۰$ و $\widehat{\theta}_۰$.

۳.۶ مدل دوم به صورت $\vec{y} = (۰, \gamma_۰, \theta_۰) \oplus ((\beta_۱, \gamma_۱, \theta_۱) \otimes x)$

مدل برآورد شده جدید که با حذف پارامتر $\beta_۰$ در نظر گرفته شده است، به صورت زیر برآورد می‌شود

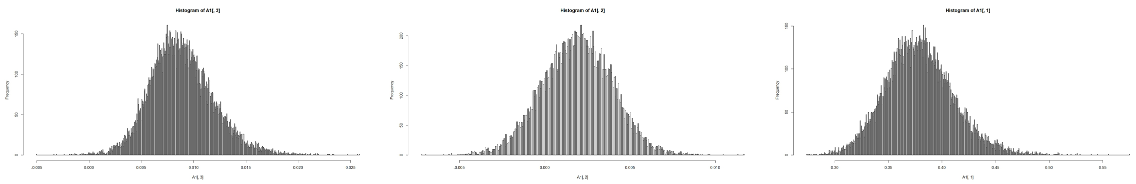
$$\begin{aligned} \vec{y} &= \left(۰, \widehat{\gamma}_۰, \widehat{\theta}_۰\right) \oplus \left(\left(\widehat{\beta}_۱, \widehat{\gamma}_۱, \widehat{\theta}_۱\right) \otimes x\right) \\ &= (۰, ۴٫۰۳۹, ۲٫۲۳۰) \oplus ((۰٫۳۷۰, -۰٫۰۰۰, ۰٫۰۱۱) \otimes x) \\ &= (۰ + ۰٫۳۷۰x, ۴٫۰۳۹ - ۰٫۰۰۰x, ۲٫۲۳۰ + ۰٫۰۱x). \end{aligned} \tag{۲.۶}$$

اکنون با توجه به معکوس‌پذیری تابع $\log$، مدل رگرسیون فازی در برازش به مجموعه مشاهدات جدول ۱ به صورت زیر حاصل می‌شود

$$\widehat{\widetilde{y}} = (۰ + ۰٫۳۷۰x, \exp\{۴٫۰۳۹ - ۰٫۰۰۰x\}, \exp\{۲٫۲۳۰ + ۰٫۰۱۱x\}),$$

وزنهای حاصل از اجرای الگوریتم در نمودار ۴ نشان داده شده است. در این شکل کاربری های کارواش، آب تصفیه و شرکت ملی حفاری کمترین وزن‌ها را می‌گیرند. دیگر وزن‌های کاربری‌ها، متناسب با تاثیرگذاری آنها در برآوردیابی مدل، در شکل ۴ نشان داده شده است.

اکنون با اعمال روش بوت استرپ موردها، به تعداد $B = ۲۰۰۰۰$ بار به باز تولید نمونه‌های بوت‌استرپی از داده‌های جدول ۱ می‌پردازیم. برای هر نمونه بوت‌استرپی باز تولید شده، برآوردهای مدل (۲.۶) بر مبنای رویکرد m-برآورد به دست می‌آید. توزیع برآوردگرهای $\widehat{\vec{\Xi}}_۰$ و $\widehat{\vec{\Xi}}_۱$ در شکل‌های ۵ و ۶ رسم شده است. خصوصیات برآوردگرهای مبتنی بر نمونه اصلی و نمونه های تولید شده توسط بوت‌استرپ در



شکل ۶: توزیع برآوردگرهای $\widehat{\vec{\Xi}}_۱ = (\widehat{\beta}_۱, \widehat{\gamma}_۱, \widehat{\theta}_۱)$ به ترتیب از چپ به راست برای $\widehat{\beta}_۱$، $\widehat{\gamma}_۱$ و $\widehat{\theta}_۱$.

جدول ۴: خصوصیات برآوردگرهای مبتنی بر نمونه اصلی و نمونه های تولید شده توسط بوت‌استرپ در مدل (۲.۶)

پارامترها	برآورد بر مبنای نمونه اصلی	متوسط برآوردهای بوت استرپی	$BIAS$	SSE	MSE
$\beta_۰$	۰/۰۰۰	۰/۰۰۰	۰/۰۰۰	۰/۰۰۰	۰/۰۰۰
$\gamma_۰$	۴/۰۳۸	۳/۴۰۰	−۰/۶۳۸	۰/۲۶۷	۰/۶۷۴
$\theta_۰$	۲/۲۳۰	۲/۶۶۷	۰/۴۳۷	۰/۵۹۰	۰/۷۸۲
$\beta_۱$	۰/۳۶۹	۰/۳۷۹	۰/۰۰۹	۰/۰۰۱	۰/۰۰۱
$\gamma_۱$	−۰/۰۰۰	۰/۰۰۲	۰/۰۰۲	۰/۰۰۰	۰/۰۰۰
$\theta_۱$	۰/۰۱۱	۰/۰۰۸	−۰/۰۰۲	۰/۰۰۰	۰/۰۰۰

جدول ۵: پی-مقدارها و فواصل اطمینان ۹۵٪ برای پارامترهای $\vec{\Xi}_۰ = (۰, \gamma_۰, \theta_۰)$ و $\vec{\Xi}_۱ = (\beta_۱, \gamma_۱, \theta_۱)$ در مدل (۲.۶)

پارامترها	فواصل اطمینان ۹۵٪ بوت استرپی	پی-مقدار
$\beta_۰ = ۰$	(۰/۰۰۰, ۰/۰۰۰)	۰/۰۰۰
$\gamma_۰$	(۲/۴۰۸, ۴/۴۱۰)	۰/۰۰۰
$\theta_۰$	(۱/۰۲۹, ۴/۰۷۴)	۰/۰۰۳
$\beta_۱$	(۰/۳۲۱, ۰/۴۴۸)	۰/۰۰۰
$\gamma_۱$	(−۰/۰۰۲, ۰/۰۰۵)	۰/۹۲۳
$\theta_۱$	(۰/۰۰۳, ۰/۰۱۵)	۰/۰۰۰

مدل (۲.۶) زمانی که $\beta_۰$ در مدل موجود نمی‌باشد، در جدول ۴ آورده شده است. در نهایت، در جدول ۵ فواصل بوت استرپ ۹۵٪ پارامترهای $\vec{\Xi}_۰$ و $\vec{\Xi}_۱$ آمده است. پی-مقدارها برای آزمون معنی‌داری مدل؛ یعنی آزمون فرضیه‌های زیر

$$\begin{cases} H_۰: & (\beta_j, \gamma_j, \theta_j) = (۰, ۰, ۰) \\ H_۱: & (\beta_j, \gamma_j, \theta_j) \neq (۰, ۰, ۰) \end{cases}, \qquad j = ۰, ۱,\ \beta_۰ = ۰.$$

در جدول ۵ آورده شده است. با توجه به اینکه پی-مقدار آزمون مربوط به پارامترهای $\gamma_۱$، بیشتر از ۰/۰۵ است و فواصل اطمینان ۹۵٪ حاصل شده برای این پارامتر شامل عدد صفر است، وجود این پارامتر در مدل معنی‌دار نمی‌باشد و باید از مدل حذف شود.

۴.۶ مدل سوم به صورت $\vec{y} = (۰, \gamma_۰, \theta_۰) \oplus ((\beta_۱, ۰, \theta_۱) \otimes x)$

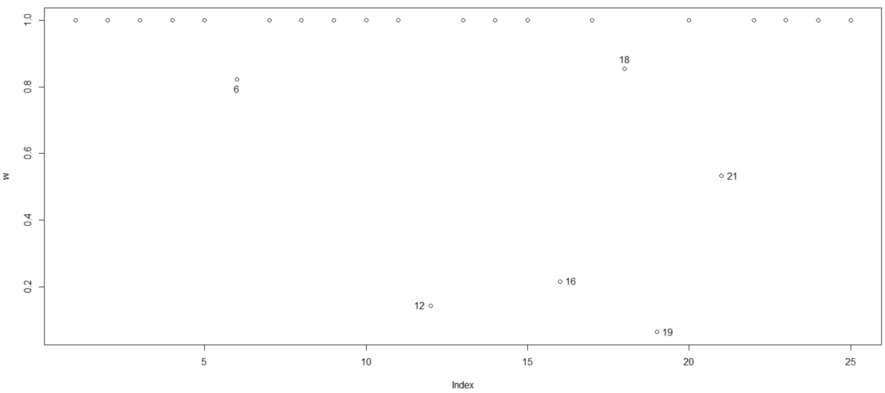
مدل برآورد شده بر مبنای داده‌های اصلی جدول ۱، با حذف پارامترهای $\beta_۰$ و $\gamma_۱$ در مدل (۱.۶)، مدل جدید به‌صورت زیر برآورد می‌شود

$$\begin{aligned} \widehat{\vec{y}} &= \left(۰, \widehat{\gamma}_۰, \widehat{\theta}_۰\right) \oplus \left(\left(\widehat{\beta}_۱, ۰, \widehat{\theta}_۱\right) \otimes x\right) \\ &= (۰, ۴/۲۱۹, ۲/۲۳۰) \oplus ((۰/۳۷۰, ۰, ۰/۰۱۱) \otimes x) \\ &= (۰ + ۰/۳۷۰x, ۴/۲۱۹ + ۰x, ۲/۲۳۰ + ۰/۰۱۱x). \end{aligned} \tag{۳.۶}$$

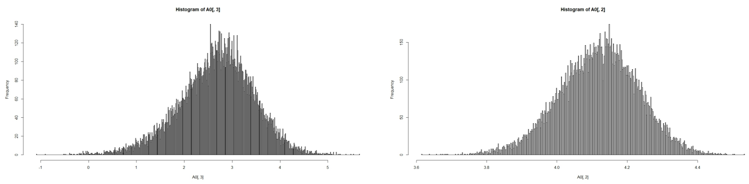
اکنون با توجه به معکوس‌پذیری تابع log، مدل رگرسیون فازی در برازش به مجموعه مشاهدات جدول ۱ به صورت زیر حاصل می‌شود

$$\widehat{\widetilde{y}} = (۰ + ۰/۳۷۰x, \exp\{۴/۲۱۹ + ۰x\}, \exp\{۲/۲۳۰ + ۰/۰۱۱x\}),$$

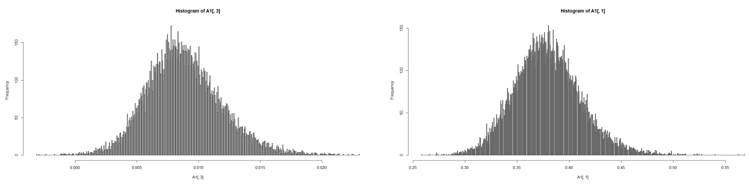
وزنهای حاصل از اجرای الگوریتم در نمودار ۷ نشان داده شده است. در این شکل کاربری‌های کارواش، آب تصفیه و شرکت ملی حفاری کمترین وزن‌ها را می‌گیرند. اکنون با اعمال روش بوت استرپ موردها، به تعداد $B = ۲۰۰۰۰$ بار، توزیع برآوردگرهای $\widehat{\vec{\Xi}}_۰$ و $\widehat{\vec{\Xi}}_۱$ در شکل‌های ۸ و ۹ حاصل می‌شوند. همچنین خصوصیات برآوردگرهای مبتنی بر نمونه اصلی و نمونه های تولید شده توسط بوت‌استرپ در مدل (۳.۶)، زمانی که $\beta_۰$ و $\gamma_۱$ در مدل موجود نمی‌باشد، در جدول ۶ آورده شده است. در نهایت، در جدول ۷ فواصل بوت استرپ ۹۵٪ پارامترهای $\vec{\Xi}_۰$ و $\vec{\Xi}_۱$ آمده است. این نتایج گویای این مطلب است که مدل رگرسیون فازی (۳.۶) در تعیین رابطه بین متغیر مستقل "تعداد روزهای مصرفی آب" با "سطح قیمت مصرف کننده" به طور کامل معنی دار است.



شکل ۷: وزن‌های مشاهدات، متناسب با تاثیرگذاری آنها در برآوردیابی مدل (۳.۶)



شکل ۸: توزیع برآوردگرهای $\widehat{\vec{\Xi}}_0 = (0, \widehat{\gamma}_0, \widehat{\theta}_0)$



شکل ۹: توزیع برآوردگرهای $\widehat{\vec{\Xi}}_1 = (\widehat{\beta}_1, 0, \widehat{\theta}_1)$

جدول ۶: خصوصیات برآوردگرهای مبتنی بر نمونه اصلی و نمونه های تولید شده توسط بوت‌استرپ در مدل (۳.۶)

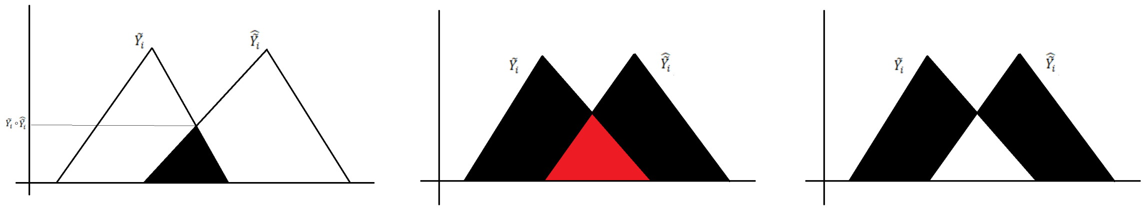
پارامترها	برآورد بر مبنای نمونه اصلی	متوسط برآوردهای بوت استرپی	$BIAS$	SSE	MSE
$\beta_0 = 0$	۰٫۰۰۰	۰٫۰۰۰	۰٫۰۰۰	۰٫۰۰۰	۰٫۰۰۰
γ_0	۴٫۲۱۹	۴٫۱۱۹	−۰٫۱۰۰	۰٫۰۱۲	۰٫۰۲۲
θ_0	۲٫۲۲۹	۲٫۶۷۳	۰٫۴۴۳	۰٫۵۹۴	۰٫۷۹۱
β_1	۰٫۳۶۹	۰٫۳۷۹	۰٫۰۰۹	۰٫۰۰۰	۰٫۰۰۰
$\gamma_1 = 0$	۰٫۰۰۰	۰٫۰۰۰	۰٫۰۰۰	۰٫۰۰۰	۰٫۰۰۰
θ_1	۰٫۰۱۱	۰٫۰۰۸	−۰٫۰۰۲	۰٫۰۰۰	۰٫۰۰۰

جدول ۷: پی‌–مقدارها و فواصل اطمینان ۹۵٪ بوت استرپی برای پارامترهای $\vec{\Xi}_0 = (0, \gamma_0, \theta_0)$ و $\vec{\Xi}_1 = (\beta_1, 0, \theta_1)$

پارامترها	فواصل اطمینان ۹۵٪ بوت استرپی	پی–مقدار
$\beta_0 = 0$	(۰٫۰۰۰, ۰٫۰۰۰)	۰٫۰۰۰
γ_0	(۳٫۸۹۱, ۴٫۳۳۲)	۰٫۰۰۰
θ_0	(۱٫۰۲۱, ۴٫۰۹۲)	۰٫۰۰۳
β_1	(۰٫۳۲۰, ۰٫۴۴۸)	۰٫۰۰۰
$\gamma_1 = 0$	(۰٫۰۰۰, ۰٫۰۰۰)	۰٫۰۰۰
θ_1	(۰٫۰۰۳, ۰٫۰۱۵)	۰٫۰۰۰

جدول ۸: نیکویی برازش مدل‌های (۱.۶)، (۲.۶) و (۳.۶) در برازش به داده‌های جدول ۱

مدل‌ها	IC	S	E_1	E_2	MDM
مدل ۱	۰/۷۴۴	۰/۳۷۸	۲۰۹/۴	۰/۹۹۹	۰/۵۶۸
مدل ۲	۰/۷۴۱	۰/۳۸۵	۲۰۵/۴	۰/۸۹۸	۰/۵۶۹
مدل ۳	۰/۷۴۹	۰/۴۰۴	۲۰۴/۷	۰/۸۸۲	۰/۵۶۹



شکل ۱۰: شکل سمت راست: مساحت قسمت رنگی، مقدار خطای برآورد $E_1(\widetilde{Y}_i, \widehat{\widetilde{Y}}_i)$ را برای مقادیر $\widetilde{Y}_i$ و $\widehat{\widetilde{Y}}_i$ نشان می دهد. شکل وسط: نسبت مساحت قسمت قرمز به جمع دو مساحت مشکی و قرمز به عنوان میزان اندازه تشابه $S(\widetilde{Y}_i, \widehat{\widetilde{Y}}_i)$ دو مقدار $\widetilde{Y}_i$ و $\widehat{\widetilde{Y}}_i$ نشان می‌دهد. شکل سمت چپ: ارتفاع قسمت رنگی مقدار شاخص تطبیق $I(\widetilde{y}_i, \widehat{\widetilde{y}}_i)$ را برای مقادیر $\widetilde{Y}_i$ و $\widehat{\widetilde{Y}}_i$ نشان می دهد.

۵.۶ مقایسه سه مدل برآورد شده بر مبنای مقادیر نیکویی برازش

در این بخش به مقایسه ۳ مدل برآورد شده بر مبنای معیارهای نیکویی برازش می‌پردازیم. در این معیارها فرض کنید $\widetilde{y}_i$ و $\widehat{\widetilde{y}}_i$ به ترتیب مقدار فازی متغیر وابسته‌ی مشاهده شده و مقدار فازی برآورد شده از یک مدل رگرسیونی فازی باشند و $S_{\widetilde{y}_i}$ و $S_{\widehat{\widetilde{y}}_i}$ به ترتیب تکیه‌گاه $\widetilde{y}_i$ و $\widehat{\widetilde{y}}_i$ هستند [۱۱، ۱۲، ۲۱، ۳۳]. تفاضل نسبی بین توابع عضویت بین این دو عدد به‌صورت زیر تعریف می‌شود

$$IC = \frac{1}{n}\sum_{i=1}^{n} \sup\left\{ \min_{x\in S_{\widetilde{y}_i}\cap S_{\widehat{\widetilde{y}}_i}} \left(\widetilde{y}_i(x), \widehat{\widetilde{y}}_i(x)\right)\right\}.$$

$$S = \frac{1}{n}\sum_{i=1}^{n} \frac{\int_{S_{\widetilde{y}_i}\cup S_{\widehat{\widetilde{y}}_i}} \min(\widetilde{y}_i(x), \widehat{\widetilde{y}}_i(x))}{\int_{S_{\widetilde{y}_i}\cup S_{\widehat{\widetilde{y}}_i}} \max(\widetilde{y}_i(x), \widehat{\widetilde{y}}_i(x))} \mathrm{d}x.$$

$$E_1 = \frac{1}{n}\sum_{i=1}^{n} \int_{S_{\widetilde{y}_i}\cup S_{\widehat{\widetilde{y}}_i}} |\widetilde{y}_i(x) - \widehat{\widetilde{y}}_i(x)|\mathrm{d}x.$$

$$E_2 = \frac{1}{n}\sum_{i=1}^{n} \frac{\int_{S_{\widetilde{y}_i}\cup S_{\widehat{\widetilde{y}}_i}} |\widetilde{y}_i(x) - \widehat{\widetilde{y}}_i(x)|\mathrm{d}x}{\int_{S_{\widetilde{y}_i}} \widetilde{y}_i(x)\mathrm{d}x}.$$

$$MDM = \frac{1}{n}\sum_{i=1}^{n} \widetilde{y}_i(\widehat{y}_i), \qquad \widehat{\widetilde{y}}_i = (\widehat{y}_i, \widehat{l}_i, \widehat{r}_i),\ i = 1, \ldots, n.$$

در جدول ۸ نتایج نیکویی برازش سه مدل آمده است. این نتایج گویای این است که مدل سوم، بهترین برازش را به داده‌ها دارد. توجه کنید که این مدل معنی‌دار است و می‌تواند به عنوان یک مدل مناسب گزارش شود.

۶.۶ مدل با دو متغیر x_1 و x_2

در ادامه مدل با دو متغیر x_1 ("متوسط تعداد روزهای مصرفی آب") و x_2 ("تعداد مشترکان") به صورت

$$\vec{y} = \vec{\Xi}_0 \oplus \left(\vec{\Xi}_1 \otimes x_1\right) \oplus \left(\vec{\Xi}_2 \otimes x_2\right) \tag{۴.۶}$$
$$= (\beta_0, \gamma_0, \theta_0) \oplus (\beta_1, \gamma_1, \theta_1) \otimes x_1 \oplus (\beta_2, \gamma_2, \theta_2) \otimes x_2$$

جدول ۹: خصوصیات برآوردگرهای مدل (۴.۶)

پارامترها	برآوردها بر مبنای نمونه اصلی	متوسط برآوردهای بوت استرپی	$BIAS$	SSE	MSE_W	فاصله اطمینان ٪۹۵	پی-مقدار	آزمون معنی‌داری
$\beta_۰$	۷۲٫۲۲۷	۴۷٫۸۱۶	−۲۴٫۴۱۱	۲۶۳۰٫۷۰۷	۳۲۲۶٫۶۳۲	(−۷۴٫۹۵, ۱۲۸٫۹۸)	۰٫۱۵۹۰	رد معنی‌داری
$\gamma_۰$	۴٫۰۴۰	۳٫۲۰۹	−۰٫۸۳۰	۰٫۳۶۵	۱٫۰۵۵	(۲٫۰۰, ۴٫۳۲)	۰٫۰۰۰۰	پذیرش معنی‌داری
$\theta_۰$	۲٫۵۵۳	۲٫۴۹۹	−۰٫۰۵۴	۰٫۵۲۲	۰٫۵۲۵	(۱٫۰۴, ۳٫۸۲)	۰٫۰۰۰۴	پذیرش معنی‌داری
$\beta_۱$	۰٫۰۱۴۳	۰٫۱۸۱۲	۰٫۱۶۶	۰٫۰۰۷۵	۰٫۱۰۳۰	(−۰٫۱۹۱, ۰٫۹۰۷)	۰٫۹۵۸۳	رد معنی‌داری
$\gamma_۱$	−۰٫۰۰۲۳	۰٫۰۰۰۲	۰٫۰۰۲	۰٫۰۰۰۰	۰٫۰۰۰۰	(−۰٫۰۰۴, ۰٫۰۰۶)	۰٫۳۷۴۵	رد معنی‌داری
$\theta_۱$	۰٫۰۱۰۱	۰٫۰۰۷۵	−۰٫۰۰۲	۰٫۰۰۰۰	۰٫۰۰۰۰	(۰٫۰۰۱, ۰٫۰۱۴)	۰٫۰۰۱۸	پذیرش معنی‌داری
$\beta_۲$	۰٫۲۶۷۵	۰٫۱۷۶۹	−۰٫۰۹۰	۵٫۹۵۷۶	۵٫۹۶۵۸	(−۵٫۲۶۷, ۴٫۳۲۹)	۰٫۹۱۲۷	رد معنی‌داری
$\gamma_۲$	۰٫۰۳۳۴	۰٫۰۴۸۴	۰٫۰۱۴	۰٫۰۰۱۰	۰٫۰۰۱۳	(−۰٫۰۱۸, ۰٫۱۱۰)	۰٫۳۰۹۶	رد معنی‌داری
$\theta_۲$	−۰٫۰۱۱۶	۰٫۰۳۹۶	۰٫۰۵۱	۰٫۰۰۱۶	۰٫۰۰۴۲	(−۰٫۰۳۹, ۰٫۱۱۸)	۰٫۷۷۴۹	رد معنی‌داری

جدول ۱۰: خصوصیات برآوردگرهای مدل (۴.۶) با حذف پارامتر $\beta_۰$

پارامترها	برآوردها بر مبنای نمونه اصلی	متوسط برآوردهای بوت استرپی	$BIAS$	SSE	MSE_W	فاصله اطمینان ٪۹۵	پی-مقدار	آزمون معنی‌داری
$\gamma_۰$	۳٫۹۵۰۵	۳٫۱۳۰۲	−۰٫۸۲۰۲	۰٫۲۸۲۶	۰٫۹۵۵۴	(۲٫۱۳, ۴٫۱۹)	۰٫۰۰۰۰	پذیرش معنی‌داری
$\theta_۰$	۲٫۲۱۹۹	۲٫۴۶۱۴	۰٫۲۴۱۵	۰٫۵۷۶۸	۰٫۶۳۵۱	(۰٫۸۸, ۳٫۸۸)	۰٫۰۰۳۴	پذیرش معنی‌داری
$\beta_۱$	۰٫۳۱۵۸	۰٫۳۴۲۷	۰٫۰۲۶۹	۰٫۰۲۴۷	۰٫۰۲۵۴	(۰٫۰۸۵, ۰٫۶۹۲)	۰٫۰۴۴۴	پذیرش معنی‌داری
$\gamma_۱$	−۰٫۰۰۱۹	۰٫۰۰۰۶	۰٫۰۰۲۵	۰٫۰۰۰۰	۰٫۰۰۰۰	(−۰٫۰۰۴, ۰٫۰۰۵)	۰٫۴۳۵۴	رد معنی‌داری
$\theta_۱$	۰٫۰۱۰۶	۰٫۰۰۷۸	−۰٫۰۰۲۸	۰٫۰۰۰۰	۰٫۰۰۰۰	(۰٫۰۰۱, ۰٫۰۱۵)	۰٫۰۰۲۵	پذیرش معنی‌داری
$\beta_۲$	۱٫۰۵۴۲	۰٫۷۹۶۲	−۰٫۲۵۸۰	۸٫۲۰۷۵	۸٫۲۷۴۰	(−۵٫۳۸۰, ۵٫۶۹۸)	۰٫۷۱۲۹	رد معنی‌داری
$\gamma_۲$	۰٫۰۴۲۴	۰٫۰۴۸۶	۰٫۰۰۶۱	۰٫۰۰۱۰	۰٫۰۰۱۰	(−۰٫۰۱۵, ۰٫۱۰۷)	۰٫۱۸۰۸	رد معنی‌داری
$\theta_۲$	۰٫۰۱۱۶	۰٫۰۳۹۷	۰٫۰۲۸۰	۰٫۰۰۱۶	۰٫۰۰۲۳	(−۰٫۰۴۱, ۰٫۱۱۵)	۰٫۷۶۸۵	رد معنی‌داری

به داده‌های اصلی جدول ۱، برازش می‌شود. نتایج این برازش به همراه آزمون معنی‌داری ضرایب مدل در جدول ۹ گزارش شده است. اکنون بر اساس ستون پی-مقدار جدول ۹، نتایج گویای این مطلب است که در مدل رگرسیون فازی (۴.۶) پارامتر $\beta_۰$ معنی‌دار نیست و باید از مدل حذف شود. با حذف این پارامتر از مدل، نتایج جدول ۱۰ حاصل می‌شوند. با توجه به نتایج جدول ۱۰ پارامتر $\vec{\Xi}_۲ = (\beta_۱, \gamma_۱, \theta_۱)$ معنی‌دار نمی‌باشد. لذا این پارامتر را می‌توان از مدل حذف نمود.

۷ تحلیل حساسیت نقاط پرت

در این بخش، آنالیز حساسیت برآورد ضرایب زمانی که داده‌ها دارای نقاط پرت هستند، با استفاده از نتایج شبیه‌سازی بوت استرپ انجام خواهد شد. بر این اساس، هدف این است که کارایی روش پیشنهادی با مدل رگرسیون وزنی فازی که یک روش برآوردیابی استوار است، و مدل رگرسیون کمترین مربعات خطای فازی مقایسه شود [۱۰]. برای این منظور، مدل زیر را با پارامترهای متقارن در نظر بگیرید

$$\begin{aligned}(y, \log(l), \log(r)) &= \vec{\Xi}_۰ \oplus (\vec{\Xi}_۱ \otimes x_۱) \oplus \ldots \oplus (\vec{\Xi}_۵ \otimes x_۵) \\ &= (\beta_۰, \gamma_۰, \gamma_۰) \oplus ((\beta_۱, \gamma_۱, \gamma_۱) \otimes x_۱) \oplus \ldots \oplus ((\beta_۳, \gamma_۳, \gamma_۵) \otimes x_۵),\end{aligned} \quad (۱.۷)$$

که در آن

$$\begin{aligned}x_{i۱} &\sim \text{Uniform}(۱۰, ۲۰), \\ x_{i۲} &\sim |\text{Uniform}(۱۰۰, ۱۰۱) - \text{Uniform}(۱۰۰, ۱۰۱)|, \\ x_{i۳} &\sim \text{Normal}(۳۰, ۱), \\ x_{i۴} &\sim \chi^۲(df = ۲۰), \\ x_{i۵} &\sim \text{Beta}(۳, ۱), \\ y_i &= ۱ + x_{۱i} + x_{۲i} + x_{۳i} + x_{۴i} + x_{۵i} + \varepsilon_i, \\ \log(l_i) &\sim ۰٫۰۱x_{۱i} + ۰٫۰۱x_{۲i} + ۰٫۰۱x_{۳i} + ۰٫۰۱x_{۴i} + x_{۵i} + |N(۰, ۰٫۱)|, \\ \log(r_i) &= \log(l_i),\end{aligned}$$

جدول ۱۱: خصوصیات m-برآوردگرهای مدل رگرسیون فازی مبتنی بر نمونه‌های تولید شده در مدل (۱.۷)

پارامترها	متوسط برآوردهای بوت استرپی	$BIAS$	SSE	MSE_M	فاصله اطمینان ۹۵٪	آزمون معنی‌داری
β_0	۳٫۶۹۴۳	۲٫۶۹۴۳	۱۱٫۱۰۵۸	۱۸٫۳۶۵۱	(−۴٫۷۰۵۱, ۱۲٫۰۹۳۷)	رد معنی‌داری
γ_0	۰٫۴۲۱۰	۰٫۴۲۱۰	۰٫۲۵۸۴	۰٫۴۳۵۷	(−۰٫۸۷۲۷, ۱٫۷۱۴۷)	رد معنی‌داری
β_1	۱٫۰۰۶۴	۰٫۰۰۶۳	۰٫۰۰۱۸	۰٫۰۰۱۹	(۰٫۹۲۰۹, ۱٫۰۹۱۸)	پذیرش معنی‌داری
γ_1	۰٫۰۱۳۴	۰٫۰۰۳۴	۰٫۰۰۰۱	۰٫۰۰۰۲	(−۰٫۰۱۴۳, ۰٫۰۴۱۱)	رد معنی‌داری
β_2	۰٫۸۳۰۱	−۰٫۱۶۹۸	۰٫۲۲۴۹	۰٫۲۵۳۸	(−۰٫۱۵۷۳, ۱٫۸۱۷۵)	رد معنی‌داری
γ_2	۰٫۱۴۴۱	۰٫۱۳۴۱	۰٫۰۴۱۲	۰٫۰۵۹۲	(−۰٫۳۳۲۷, ۰٫۶۲۰۹)	رد معنی‌داری
β_3	۰٫۹۴۸۱	−۰٫۰۵۱۸	۰٫۰۱۲۲	۰٫۰۱۴۹	(۰٫۷۰۸۸, ۱٫۱۸۷۳)	پذیرش معنی‌داری
γ_3	۰٫۰۱۴۹	۰٫۰۰۴۹	۰٫۰۰۰۲	۰٫۰۰۰۳	(−۰٫۰۱۹۰, ۰٫۰۴۸۸)	رد معنی‌داری
β_4	۰٫۹۹۶۰	−۰٫۰۰۴۰	۰٫۰۰۰۶	۰٫۰۰۰۷	(۰٫۹۴۴۱, ۱٫۰۴۷۸)	پذیرش معنی‌داری
γ_4	۰٫۰۰۹۹	−۰٫۰۰۰۵	۰٫۰۰۰۱	۰٫۰۰۰۱	(−۰٫۰۰۹۷, ۰٫۰۲۹۵)	رد معنی‌داری
β_5	۱٫۰۴۵۲	۰٫۰۴۵۱	۰٫۳۲۵۴	۰٫۳۲۷۵	(−۰٫۰۷۶۴, ۲٫۱۶۶۸)	رد معنی‌داری
γ_5	۱٫۰۱۳۷	۰٫۰۱۳۷	۰٫۱۱۹۹	۰٫۱۲۰۱	(۰٫۳۳۴۴, ۱٫۶۹۲۹)	پذیرش معنی‌داری

جدول ۱۲: خصوصیات برآوردگرهای وزنی مدل رگرسیون فازی [۱۰] مبتنی بر نمونه‌های تولید شده در مدل (۱.۷)

پارامترها	متوسط برآوردهای بوت استرپی	$BIAS$	SSE	MSE_W	فاصله اطمینان ۹۵٪	آزمون معنی‌داری
β_0	۳٫۶۳۰۳	۲٫۶۳۰۲	۱۲٫۳۶۷۳۴	۱۹٫۲۸۵۳	(−۴٫۹۷۷۰, ۱۲٫۲۳۷۶)	رد معنی‌داری
γ_0	۱٫۰۲۲۹	۱٫۰۲۲۹	۴٫۴۰۴۳۷	۵٫۴۵۰۷	(−۳٫۵۵۳۰, ۵٫۵۹۸۸)	رد معنی‌داری
β_1	۱٫۰۱۳۱	۰٫۰۱۳۰	۰٫۰۰۱۳۳	۰٫۰۰۱۵	(۰٫۹۳۷۱, ۱٫۰۸۹۰۸)	پذیرش معنی‌داری
γ_1	۰٫۰۱۰۲	۰٫۰۰۰۲	۰٫۰۰۰۳۹	۰٫۰۰۰۴	(−۰٫۰۲۹۰, ۰٫۰۴۹۴)	رد معنی‌داری
β_2	۰٫۷۴۱۴	−۰٫۲۵۸۵	۰٫۱۶۴۴۷	۰٫۲۳۱۳	(−۰٫۲۰۱۲, ۱٫۶۸۴۰)	رد معنی‌داری
γ_2	−۰٫۰۲۱	−۰٫۰۳۱۹	۰٫۰۶۱۱۸	۰٫۰۶۲۲	(−۰٫۵۰۹۸, ۰٫۴۶۷۸)	رد معنی‌داری
β_3	۰٫۹۴۸۱	−۰٫۰۵۱۸	۰٫۰۱۳۵۱	۰٫۰۱۶۲	(۰٫۶۹۸۶, ۱٫۱۹۷۵)	پذیرش معنی‌داری
γ_3	۰٫۰۰۰۵	−۰٫۰۰۹۴	۰٫۰۰۴۴۱	۰٫۰۰۴۵	(−۰٫۱۳۰۹, ۰٫۱۳۱۹)	رد معنی‌داری
β_4	۰٫۹۹۷۳	−۰٫۰۰۲۶	۰٫۰۰۰۴۹	۰٫۰۰۰۵	(۰٫۹۵۳۴, ۱٫۰۴۱۱)	پذیرش معنی‌داری
γ_4	۰٫۰۰۹۴	−۰٫۰۰۰۵	۰٫۰۰۰۰۹	۰٫۰۰۰۱	(−۰٫۰۱۰۲, ۰٫۰۲۹۰)	رد معنی‌داری
β_5	۱٫۰۴۵۲	۰٫۰۴۵۱	۰٫۲۷۴۱۶	۰٫۲۷۶۲	(۰٫۰۱۵۱, ۲٫۰۷۵۲)	پذیرش معنی‌داری
γ_5	۱٫۰۵۵۲	۰٫۰۵۵۲	۰٫۰۸۹۸۵	۰٫۰۹۲۹	(۰٫۴۵۷۸, ۱٫۶۵۲۵)	پذیرش معنی‌داری

جدول ۱۳: خصوصیات برآوردگرهای کمترین مربعات خطای رگرسیون فازی LS [۱۰] مبتنی بر نمونه‌های تولید شده در مدل (۱.۷)

پارامترها	متوسط برآوردهای بوت استرپی	$BIAS$	SSE	MSE_{LS}	فاصله اطمینان ۹۵٪	آزمون معنی‌داری
β_0	۱۲٫۰۲۳۳	۱۱٫۰۲۳۳	۱۹٫۸۸۵۲	۱۴۱٫۳۹۸۴	(−۱۱٫۲۸۳۲, ۳۵٫۳۲۹۸)	رد معنی‌داری
γ_0	۰٫۹۲۲۷	۰٫۹۲۲۷	۴٫۰۲۲۲	۴٫۸۷۳۶	(−۳٫۴۰۴۲, ۵٫۲۴۹۶)	رد معنی‌داری
β_1	۱٫۰۵۲۶	۰٫۰۵۲۶	۰٫۰۰۱۷	۰٫۰۰۴۵	(۰٫۹۲۱۱, ۱٫۱۸۴۰)	پذیرش معنی‌داری
γ_1	۰٫۰۱۰۲	۰٫۰۰۰۲	۰٫۰۰۰۴	۰٫۰۰۰۴	(−۰٫۰۲۹۰, ۰٫۰۴۹۴)	رد معنی‌داری
β_2	−۰٫۱۵۴۲	−۱٫۱۵۴۱	۰٫۳۱۶۴	۱٫۶۴۸۴	(−۲٫۶۷۰۶, ۲٫۳۶۲۲)	رد معنی‌داری
γ_2	−۰٫۰۱۲۳	−۰٫۰۲۲۳	۰٫۰۵۵۵	۰٫۰۵۶۰	(−۰٫۴۷۶۱, ۰٫۴۵۱۵)	رد معنی‌داری
β_3	۰٫۷۰۱۹	−۰٫۲۹۸۰	۰٫۰۱۹۳	۰٫۱۰۸۱	(۰٫۰۵۷۴, ۱٫۳۴۶۳)	پذیرش معنی‌داری
γ_3	۰٫۰۰۴۳	−۰٫۰۰۵۶	۰٫۰۰۴۰	۰٫۰۰۴۱	(−۰٫۱۲۱۲, ۰٫۱۲۹۸)	رد معنی‌داری
β_4	۰٫۹۸۲۵	−۰٫۰۱۷۵	۰٫۰۰۰۶	۰٫۰۰۰۹	(۰٫۹۲۳۷, ۱٫۰۴۱۳)	پذیرش معنی‌داری
γ_4	۰٫۰۰۹۶	−۰٫۰۰۰۳	۰٫۰۰۰۱	۰٫۰۰۰۱	(−۰٫۰۱۰۰, ۰٫۰۲۹۲)	رد معنی‌داری
β_5	۰٫۸۴۸۳	−۰٫۱۵۱۶	۰٫۳۳۵۰	۰٫۳۵۸۰	(−۰٫۳۲۴۴, ۲٫۰۲۱۰)	رد معنی‌داری
γ_5	۱٫۰۶۶۰	۰٫۰۶۶۰	۰٫۰۹۲۱	۰٫۰۹۶۵	(۰٫۴۵۷۱, ۱٫۶۷۴۸)	پذیرش معنی‌داری

همچنین مقادیر پارامترهای شبیه‌سازی شده، به‌صورت زیر اختیار شده‌اند

$$
\begin{aligned}
\vec{\Xi}_0 &= (1, 0, 0), & \vec{\Xi}_3 &= (1, 0/01, 0/01),\\
\vec{\Xi}_1 &= (1, 0/01, 0/01), & \vec{\Xi}_4 &= (1, 0/01, 0/01),\\
\vec{\Xi}_2 &= (1, 0/01, 0/01), & \vec{\Xi}_5 &= (1, 1, 1).
\end{aligned}
$$

هدف از این مطالعه این است که فقط یک ورودی تغییر کند تا چگونگی تاثیر آن بر مدل سنجیده شود. بنابراین، در هر تولید مجموعه داده‌ها، ضرایب رگرسیون ثابت نگه داشته شد. همچنین، برای بررسی عملکرد مدل‌های رگرسیون فازی مورد مقایسه، مجموعه داده‌های شبیه‌سازی‌شده با افزودن تعداد فزاینده‌ای از نقاط پرت توسط الگوی زیر

$$
\varepsilon_i \sim \begin{cases} N(0, 1) & i = 1, \ldots n_1, \\ t(df = 10, noc = 10) & i = n_1 + 1, \ldots n. \end{cases}
$$

تولید می‌شوند. سپس مدل‌های رگرسیون فازی مورد مقایسه به هر مجموعه داده تولید شده برازش می‌شوند. این برازش‌ها به تعداد ۱۰۰۰۰ بار بر مجموعه داده‌های مختلف شبیه‌سازی شده با حجم‌های نمونه مختلف و تعداد مشاهدات پرت مختلف تکرار می‌شوند. نتایج جداول ۱۱،

جدول ۱۴: مقادیر کارایی m-برآوردگرها (جدول ۱۱) در مقابل برآوردگرهای وزنی (جدول ۱۲) و برآوردگرهای کمترین مربعات خطا (جدول ۱۳)

پارامترها	$\frac{MSE_W}{MSE_M}$	$\frac{MSE_{LS}}{MSE_M}$
β_0	۱٫۰۵۰	۷٫۶۹۹
γ_0	۱۲٫۵۱۰	۱۱٫۱۸۶
β_1	۰٫۷۸۹	۲٫۳۶۸
γ_1	۲٫۰۰۰	۲٫۰۰۰
β_2	۰٫۹۱۱	۶٫۴۹۵
γ_2	۱٫۰۵۱	۰٫۹۴۶
β_3	۱٫۰۸۷	۷٫۲۵۵
γ_3	۱۵٫۰۰۰	۱۳٫۶۶۷
β_4	۰٫۷۱۴	۱٫۲۸۶
γ_4	۱٫۰۰۰	۱٫۰۰۰
β_5	۰٫۸۴۳	۱٫۰۹۳
γ_5	۰٫۷۷۴	۰٫۸۰۳

جدول ۱۵: کارایی مدل‌های مقایسه‌ای در مطالعه شبیه‌سازی با حجم نمونه‌های مختلف و تعداد متفاوت نقاط پرت در ۱۰۰۰۰ چرخه تکرار

حجم نمونه	تعداد نقاط پرت	$\text{eff.}(\hat{\vec{\Xi}}_{LS}, \hat{\vec{\Xi}}_M)$	$\text{eff.}(\hat{\vec{\Xi}}_W, \hat{\vec{\Xi}}_M)$
۱۰۰	۱۰	۷٫۵۸	۱٫۲۹
	۳۰	۸٫۶۶	۱٫۹۱
	۴۰	۱٫۶۳	۱٫۳۴
۳۰۰	۷۰	۷٫۵۶	۱٫۲۴
	۱۰۰	۳۵٫۳۷	۱۳٫۰۴
	۱۲۰	۲٫۹۹	۳٫۱۰
۵۰۰	۵۰	۸٫۸۷	۰٫۹۶
	۱۰۰	۲٫۹۲	۱٫۰۸
	۲۰۰	۳٫۸۱	۴٫۷۹
۷۰۰	۱۰۰	۶٫۱۳	۱٫۳۵
	۲۰۰	۳٫۴۹	۱٫۳۹
	۳۰۰	۱٫۶۰	۲٫۵۴

۱۲ و ۱۳ مربوط به نتایج ۱۰۰۰۰ بار شبیه‌سازی مجموعه داده‌هایی با حجم $n = ۱۰۰$ است که ۱۰ درصد آنها را مشاهدات پرت تشکیل می‌دهد. مقادیر کارایی m-برآوردگرها (جدول ۱۱) در مقابل برآوردگرهای وزنی (جدول ۱۲) و برآوردگرهای کمترین مربعات خطا (جدول ۱۳) در جدول ۱۴ گزارش شده است. این مقادیر نشان می‌دهد که m-برآوردگرها در مقایسه با برآوردگرهای وزنی و برآوردگرهای کمترین مربعات خطا از کارایی بیشتری برخوردار هستند.

اکنون به‌طور مشابه، شبیه‌سازی بر اساس درصدهای فزآینده‌ای از تعداد مشاهدات پرت تکرار شد. مقادیر کارایی m-برآوردگرهای بردار فازی $\hat{\vec{\Xi}}_M$ در مقابل برآوردگرهای وزنی بردار فازی $\hat{\vec{\Xi}}_W$ و برآوردگرهای کمترین مربعات خطا بردار فازی $\hat{\vec{\Xi}}_{LS}$ در جدول ۱۵ نشان داده شده است. برآوردگر $\hat{\vec{\Xi}}_I$ در مقایسه با برآوردگر $\hat{\vec{\Xi}}_{II}$ از کارایی بیشتری برخوردار است، اگر $\text{eff.}(\hat{\vec{\Xi}}_I, \hat{\vec{\Xi}}_{II}) > ۱$، که در آن

$$\text{eff.}(\hat{\vec{\Xi}}_I, \hat{\vec{\Xi}}_{II}) = \frac{||\hat{\vec{\Xi}}_{II} \ominus \vec{\Xi}||_2}{||\hat{\vec{\Xi}}_I \ominus \vec{\Xi}||_2},$$

و $||\cdot||_2$ نرم اقلیدسی است. نتایج جدول ۱۵ نشان از کارایی رویکرد معرفی شده در این مقاله در مقایسه با رویکرد استوار وزنی و رویکرد کمترین توان دوم خطای فازی دارد.

بحث و نتیجه‌گیری

هدف ما در این مقاله بر این بود که با اعمال مدل رگرسیون m-برآوردگرهای فازی بر داده‌های واقعی آب و فاضلاب شهر اهواز به استنباط درباره پارامترهای مدل بپردازیم. در این مطالعه، میزان مصرف آب کاربری‌ها به عنوان یک متغیر وابسته فازی در نظر گرفته شد. این انتخاب به دلیل عدم اطمینان و رفتار مصرف کننده است. از این رو، منطقی‌تر است که متغیر وابسته را به عنوان یک عدد فازی، و به طور دقیق‌تر، به عنوان یک عدد فازی مثلثی در نظر گرفته شود. در این راستا با استفاده از مدل رگرسیون m-برآوردگرهای فازی به تحلیل سطوح میزان آب مصرفی از دیدگاه شرکت آب و فاضلاب شهر اهواز پرداخته شد. در انتها، پس از مدل‌سازی داده‌ها به استنباط بوت استرپی درباره پارامترها نیز پرداختیم. در این راستا، نتیجه گرفتیم که می‌توان با برازش مدل رگرسیون m-برآوردگرهای فازی، یک مدل مناسب و معنی‌داری جهت برازش به داده‌ها به‌دست آورد. بر اساس نتایج بیان شده برای هر یک از ضرایب مدل، فاصله اطمینان و آزمون معنی‌داری محاسبه گردید.

این مقادیر به کاربر این امکان را می‌دهد تا پارامترهایی که در مدل معنی‌دار نمی‌باشند را شناسایی کند و با حذف آنها به یک مدل معنی‌دار دست پیدا کند. همچنین، استفاده از استنباط آماری در مطالعات مدلسازی امری ضروری بوده و روش بوت استرپ در نمونه‌های متفاوت نشان می‌دهد که با بکارگیری روش بوت استرپ در مواردی که با حجم نمونه ناکافی مواجه هستیم و/یا شرایط مناسب برای استفاده از روش‌های متداول رگرسیون کلاسیک فراهم نیست، شرایطی را فراهم می‌آورد تا بتوان به استنباط آماری درباره برآوردهای ارایه شده پرداخت. همچنین بر اساس نمونه‌های تولید شده به مقایسه کارایی برآوردگرهای تولید شده توسط رویکرد معرفی شده با دو رویکرد استوار برآوردگرهای وزنی رگرسیون فازی و رویکرد کمترین مربعات خطای فازی پرداخته شد. نتایج بیان شده، حاکی از این است که m-برآوردگرها در مقایسه با دو برآوردگر دیگر، کاراتر هستند. خاطرنشان می‌شود که روش پیشنهادی در این مقاله یک روش کلی در تحلیل رگرسیونی در محیط فازی است و می‌توان آن را در سایر مدل‌های رگرسیونی مانند رگرسیون نیمه پارامتری یا رگرسیون ریج به کاربرد.

فهرست منابع

[۱] اصغری، ز.، زارعی، ح. و اکبری، م. ق. (۱۴۰۱). آزمون آماری براساس فرضیه‌های فازی شهودی. سیستم های فازی و کاربردها، **۵**(۲)، ۲۹۲-۲۷۱.

[۲] چاچی ج. و چاچی ع. (۱۴۰۰)، کاربرد عملگرهای وزنی در مدل رگرسیون قدرمطلق انحرافات مرتب شده، مجله علوم آماری، **۱۵**(۱)، ۶۰-۳۹.

[۳] چاچی، ج.، کاظمی‌فرد، ا. و فهیمی، ح. (۱۴۰۰)، رهیافت تصمیم‌گیری‌های چند معیاره در ارزیابی نیکویی برازش مدل‌های آماری، سیستم‌های فازی و کاربردها، **۴**(۱)، ۲۶۷-۲۴۷.

[۴] چاچی، ج. و چاچی، ع. (۱۳۹۷). رویکردهای وزنی در برازش مدل‌های رگرسیون فازی، سیستم‌های فازی و کاربردها، **۲**(۱)، ۱۰۵-۱۱۷.

[5] Arefi, M. (2020). Quantile fuzzy regression based on fuzzy outputs and fuzzy parameters. *Soft Computing*, **24**(1), 311-320.

[6] Arefi, M., & Khammar, A.H. (2023). Nonlinear prediction of fuzzy regression model based on quantile loss function. *Soft Computing*. https://doi.org/10.1007/s00500-023-09190-w.

[7] Asadolahi, M., Akbari, M. G., Hesamian, G., & Arefi, M. (2021). A Robust Support Vector Regression with Exact Predictors and Fuzzy Responses. *International Journal of Approximate Reasoning*, **132**, 206-225.

[8] Bayarassou, H., & Megri A.F.(2023). New approach based on a fuzzy regression model for a photovoltaic system, *Electric Power Systems Research*, **217**, 109091.

[9] Celmins, A. (1987). Least Squares Model Fitting to Fuzzy Vector Data. *Fuzzy sets and systems*, **22**(3), 245-269.

[10] Chachi, J. (2019). A Weighted Least Squares Fuzzy Regression for Crisp Input-Fuzzy Output Data. *IEEE Transactions on Fuzzy Systems*, **27**(4), 739-748.

[11] Chachi, J., & Kazemifard, A. (2024). A Novel Extended Approach to Evaluate Criteria Weights in MADM Problems in Fuzzy Framework. *Iranian Journal of Fuzzy Systems*, **21**(4), 101-122.

[12] Chachi, J., Kazemifard, A., & Jalalvand, M. (2021). A Multi-Attribute Assessment of Fuzzy Regression Models. *Iranian Journal of Fuzzy Systems*, **18**(4), 131-148.

[13] Chachi, J., Taheri, S. M., & D’Urso, P. (2022). Fuzzy Regression Analysis Based on M-Estimates. *Expert Systems with Applications*, **187**, 115891.

[14] Chakravarty, S., Demirhan, H., & Baser, F. (2020). Fuzzy Regression Functions with a Noise Cluster and the Impact of Outliers on Mainstream Machine Learning Methods in the Regression Setting. *Applied Soft Computing*, **96**, 106535.

[15] Chukhrova, N., & Johannssen, A. (2019). Fuzzy Regression Analysis: Systematic Review and Bibliography. *Applied Soft Computing*, **106**, 107331.

[16] Chukhrova, N., & Johannssen, A. (2020). Fuzzy hypothesis testing for a population proportion based on set-valued information *Fuzzy Sets and Systems*, **387**, 127-157.

[17] Chukhrova, N., & Johannssen, A. (2021). Fuzzy hypothesis testing: Systematic review and bibliography *Applied Soft Computing*, **84**, 105708.

[18] Chukhrova, N., & Johannssen, A. (2022). Two-tailed hypothesis testing for the median with fuzzy categories applied to the detection of health risks. *Expert Systems with Applications*, **192**, 116362.

[19] Chukhrova, N., & Johannssen, A. (2023). Employing fuzzy hypothesis testing to improve modified p charts for monitoring the process fraction nonconforming. *Information Sciences*, **633**, 141-157.

[20] Diamond, P. (1988). Fuzzy Least Squares. *Information Sciences*, **46**(3), 141-157.

[21] D'Urso, P., Chachi, J., Kazemifard, A., & De Giovanni, L. (2024). OWA-Based Multi-Criteria Decision Making Based on Fuzzy Methods. *Annals of Operations Research*, p. 1-35, https://doi.org/10.1007/s10479-024-05926-5.

[22] D'Urso, P., De Giovanni, L., Alaimo, L. S., Mattera, R., & Vitale, V. (2023). Fuzzy Clustering with Entropy Regularization for Interval-Valued Data with an Application to Scientific Journal Citations. *Annals of Operations Research*, p. 1-24, https://doi.org/10.1007/s10479-023-05180-1.

[23] D'Urso, P., & Leski, J. M. (2020). Fuzzy Clustering of Fuzzy Data Based on Robust Loss Functions and Ordered Weighted Averaging. *Fuzzy Sets and Systems*, **389**, 1-28.

[24] Efron, B., & Tibshirani, R. J. (1994). *An Introduction to the Bootstrap*. CRC press.

[25] Fox, J., & Weisberg, S. (2018). *An R companion to applied regression*. Sage publications.

[26] Hesamian, G., & Akbari, M. G. (2017). Semi-parametric partially logistic regression model with exact inputs and intuitionistic fuzzy outputs. *Applied Soft Computing*, 58, 517-526.

[27] Hesamian, G., & Akbari, M. G. (2020). A fuzzy additive regression model with exact predictors and fuzzy responses. *Applied Soft Computing*, **95**, 106507.

[28] Hesamian, G., & Akbari, M. G. (2020). A robust varying coefficient approach to fuzzy multiple regression model. *Journal of Computational and Applied Mathematics*, **371**, 112704.

[29] Hesamian, G., & Akbari, M. G. (2020). Fuzzy spline univariate regression with exact predictors and fuzzy responses. *Journal of Computational and Applied Mathematics*, **375**, 112803.

[30] Hesamian, G., & Akbari, M. G. (2023). Support Vector Logistic Regression Model with Exact Predictors and Fuzzy Responses. *Journal of Ambient Intelligence and Humanized Computing*, **14**, 817-828.

[31] Ronchetti, E. M., & Huber, P. J. (2009). *Robust Statistics*. Hoboken, NJ, USA: John Wiley & Sons.

[32] James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. (2023). *An Introduction to Statistical Learning: with Applications in Python*. Cham: Springer International Publishing.

[33] Kazemifard, A., & Chachi, J. (2022). MADM Approach to Analyse the Performance of Fuzzy Regression Models. *Journal of Ambient Intelligence and Humanized Computing*, **13**(8), 4019-4031.

[34] Khammar, A. H., Arefi, M., & Akbari, M. G. (2020). A Robust Least Squares Fuzzy Regression Model Based on Kernel Function. *Iranian Journal of Fuzzy Systems*, **17**(4), 105-119.

[35] Khammar, A. H., Arefi, M., & Akbari, M. G. (2021a). A general approach to fuzzy regression models based on different loss functions. *Soft Computing*, **25**(2), 835-849.

[36] Khammar, A. H., Arefi, M., & Akbari, M. G. (2021b). Quantile fuzzy varying coefficient regression based on kernel function. *Applied Soft Computing*, 107, 107313.

[37] Khasanzoda, N., Zicmane, I., Beryozkina, S., Safaraliev, M., Sultonov, S., & Kirgizov, A. (2022). Regression model for predicting the speed of wind flows for energy needs based on fuzzy logic. *Renewable Energy*, 191, 723-731.

[38] Lubiano, M. A., de Saa, S. D. L. R., Montenegro, M., Sinova, B., & Gil, M. A. (2016). Descriptive Analysis of Responses to Items in Questionnaires. Why not Using a Fuzzy Rating Scale?. *Information Sciences*, **360**, 131-148.

[39] Lubiano, M. A., Montenegro, M., Sinova, B., de Saa, S. D. L. R., & Gil, M. A. (2016). Hypothesis testing for means in connection with fuzzy rating scale-based data: algorithms and applications. *European Journal of Operational Research*, **251**(3), 918-929.

[40] Lubiano, M. A., Salas, A., & Gil, M. A. (2017). A hypothesis testing-based discussion on the sensitivity of means of fuzzy data with respect to data shape. *Fuzzy Sets and Systems*, **328**, 54-69.

[41] Mohammadi, A., Javadi, S. H., & Ciuonzo, D. (2019). Bayesian fuzzy hypothesis test in wireless sensor networks with noise uncertainty. *Applied Soft Computing*, **77**, 218-224.

[42] Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). *Introduction to Linear Regression Analysis*. John Wiley & Sons.

[43] Mylonas, N., & Papadopoulos, B. (2021). Fuzzy hypotheses tests for crisp data using non-asymptotic fuzzy estimators, fuzzy critical values and a degree of rejection or acceptance. *Evolving Systems*, **12**(3), 723-740.

[44] Pandelara, D., Kristjanpoller, W., Michell, K., & Minutolo, M. C. (2022). A fuzzy regression causality approach to analyze relationship between electrical consumption and GDP. *Energy*, 239, 122459.

[45] Parchami, A. (2020). Fuzzy decision making in testing hypotheses: An introduction to the packages FPV. *Iranian Journal of Fuzzy Systems*, **17**(2), 67-77.

[46] Ruszczynski, A. (2011). *Nonlinear Optimization*. Princeton university press.

[47] Salman, M., Long, X., Wang, G., & Zha, D. (2022). Paris climate agreement and global environmental efficiency: New evidence from fuzzy regression discontinuity design. *Energy Policy*, 168, 113128.

[48] Tanaka, H., Uejima, S., & Asia, K. (1982). Linear Regression Analysis with Fuzzy Model. *IEEE Trans. Systems Man Cybern*, **12**, 903-907.

[49] Viertl, R. (2011). *Statistical Methods for Fuzzy Data*. John Wiley & Sons.

[50] Zadeh, L.A., (1965). Fuzzy Sets. *Information and Control*, **8**, 338-353.

[51] Zhang, S., Robinson, E., & Basu, M. (2022). Hybrid Gaussian process regression and Fuzzy Inference System based approach for condition monitoring at the rotor side of a doubly fed induction generator. *Renewable Energy*, 198, 936-946.

[52] Zimmermann, H. J. (2011). *Fuzzy set theory—and its applications*. Springer Science & Business Media.

Significance test of fuzzy regression model

Jalal Chachi[1, 9] **Mohammad-Reza Akhond**[1], **Pooran Bandani Tarashoki**[1],

[(1)] Department of Statistics, Faculty of Mathematical Sciences and Computer, Shahid Chamran University of Ahvaz, Ahvaz 6135714463, Iran

Communicated by: Hamzeh Torabi
Received:30 June 2024 Accepted: 17 January 2025

Abstract: Statistical significance determines whether the relationship between two or more variables is caused by factors other than chance and randomness. In other words, are the results provided in a modeling caused by data? Statistical hypothesis testing is a method by which statistical significance is determined. By introducing fuzzy regression, several approaches were presented to estimate parameters of such the models more precisely. But until now, most researches in the literature have conducted on the estimation method and in comparison, very little attention has been paid to the properties of estimators, confidence intervals and significance tests of parameters. Therefore, the main goal of this paper is to investigate the significance of estimated parameters in an applied study with fuzzy-valued real data in the framework of fuzzy regression modeling using m-estimators. For this purpose, by accessing the water and sewage data source of Ahvaz city, the required variables were first introduced. Since m-estimators use an algorithm based on reweighted method, we are looking to determine the weight of the users in terms of the consumption pattern, that is, the weight of high consumption users and low consumption users is determined. Now, the company can first identify and classify the consumption patterns of each user based on the determined weights, and then deal with the stepped pricing of each cubic meter of water for each user. This goal requires providing a significant model of fuzzy regression. Therefore, since the m-estimators of fuzzy regression model do not have a closed form, the bootstrap is used to present the numerical indices of the estimators, confidence intervals and the significance test of the model. Finally, by analyzing the results and keeping significant parameters in the model, a suitable model was introduced to fit the dataset.

Keywords: Fuzzy regression analysis, m-estimators, Fuzzy Data, Statistical significance.

[9]Corresponding author.
E-mail addresses: (J. Chachi) jalal.chachi@scu.ac.ir, (M.R. Akhond)m.akhond@scu.ac.ir
(P. Bandani Tarashoki)pooran6161@gmail.com