مدل بندی رگرسیون لوژستیک برای داده‌های زبانی تحت خطاهای تصادفی فازی

مریم مالکی[۱]، حمیدرضا نیلی ثانی[۲]، محمدقاسم اکبری [۳]

تاریخ دریافت: ۱۴۰۲/۰۹/۱۹
تاریخ پذیرش: ۱۴۰۳/۰۷/۰۸

چکیده:
در این مقاله، مدل‌های رگرسیونی لوژستیک که در آن متغیرهای پاسخ به‌صورت دو (یا چند) ارزشی و متغیرهای توضیحی (پیشگو یا مستقل) متغیرهای معمولی هستند، اما خطاها علاوه بر ماهیتی تصادفی، ماهیتی ابهامی، نیز دارند، موردمطالعه قرار می‌گیرند. بر این اساس مدل پیشنهادی را صورت‌بندی کرده و برآورد ضرایب را برای حالتی با تنها یک متغیر توضیحی و با استفاده از روش کمترین توان‌های دوم تعیین می‌کنیم. در پایان با یک مثال به تشریح نتایج حاصله می‌پردازیم.
واژه‌های کلیدی: رگرسیون لوژستیک، متغیر تصادفی فازی، متغیر تصادفی زبانی فازی.

۱ مقدمه

تحلیل مدل‌های رگرسیونی لوژستیک، زمانی که متغیرهای توضیحی و وابسته به‌صورت فازي گزارش شوند، توسط تاکمورا (۲۰۰۴) موردمطالعه قرارگرفته است. تسینگ و لین (۲۰۰۵) برای پیش‌بینی ورشکستگی در مؤسسات مالی انگلستان از رگرسیون لوژستیک دو ارزشی بازه‌ای درجه دوم با ضرایب فازی و بر اساس روش برنامه‌ریزی درجه دوم استفاده کرده‌اند. ناگار و سریواستاوا (۲۰۰۸) مدلي برای رگرسیون لوژستیک فازی معرفی و مدل را برای پیش‌بینی یک نوع سرطان موردبررسی قرار داده و نتایج را با شبکه‌های عصبی فازی مقایسه نمودند. دام (۲۰۰۹) به بررسی عوامل خطر مرتبط با سرطان دهان از طریق رگرسیون فازی پرداخت. بررسی رگرسیون لوژستیک دو ارزشی بازه‌ای و رگرسیون لوژستیک دو ارزشی امکانی توسط طاهری و میرزایی (۲۰۰۹) و میرزایی و طاهري (۲۰۱۰) انجام شده است. پوراحمد و همکاران (۲۰۱۱(a و (۲۰۱۱(b رگرسیون لوژستیک دو ارزشی را براي حالتی که ورودی‌ها اعدادي دقیق و خروجی‌ها و ضرایب اعداد فازی باشند، ارائه داده‌اند. نامداری و همکاران (۲۰۱۳) مدل‌های رگرسیونی لوژستیک فازي را زمانی که متغیرهای توضیحی به‌صورت معمولی و متغیرهاي پاسخ به‌صورت متغیرهای زبانی بیان شوند، موردبررسی و مطالعه قرار داده‌اند. یو و تسینگ (۲۰۱۴) مدل‌های رگرسیونی لوژستیک فازي را براي پردازش نور در یک کارخانه تلویزیون به کار گرفتند. نامداري و همکاران (۲۰۱۵) مدل‌های رگرسیونی لوژستیک فازي را بر اساس روش کمترین قدر مطلق انحرافات بررسی کرده‌اند. حسامیان و اکبري (۲۰۱۷) به بررسی مدل‌های رگرسیون لوژستیک نیمه پارامتري با متغیر وابسته فازي شهودي و متغیرهاي توضیحی دقیق پرداخته‌اند. سلمانی و همکاران (۲۰۱۷) مدل‌هاي رگرسیون لوژستیک، زمانی که متغیرهاي مستقل و وابسته به‌صورت فازي گزارش شوند، موردمطالعه قرار دادند.

در تحقیقاتی که در زمینه رگرسیون لوژستیک فازي تاکنون انجام شده است (تا آنجایی که مؤلفین اطلاع دارند) خطاهای تصادفی که ناشی از انتخاب تصادفی نمونه‌ها یا اندازه‌گیری‌ها می‌باشند، یا در نظرگرفته نشده‌اند یا خطاهای موجود در مدل با فازی بودن ضرایب یا فازی بودن متغیرها تلفیق‌شده‌اند. درصورتی‌که این خطاها در بسیاری از موارد مشهود و با ماهیت فازی متفاوت است. به‌عنوان‌مثال هنگامی‌که وجود یا عدم وجود درد را در بیماران خاص (مثلاً تحت یک عمل جراحی) موردبررسی قرار می‌دهیم، میزان دردی که توسط فرد بیمار گزارش می‌شود از اهمیت خاصی برخوردار است. میزان درد هرچند که ماهیتی ابهامی دارد، آستانه و میزان درد از فردی به فرد دیگر

[۱] گروه آمار دانشگاه بیرجند
[۲] گروه آمار دانشگاه بیرجند (نویسنده مسئول: nilisani@gmail.com)
[۳] گروه آمار دانشگاه بیرجند

می‌تواند متفاوت باشد، بنابراین ماهیتی تصادفی نیز دارد. در صورت نادیده گرفتن ماهیت تصادفی داده‌ها، همواره این امکان وجود دارد که مدل خوبی به داده‌ها برازش شود (پور احمدی و همکاران (۲۰۱۱(a)) و نیکویی برازش با معیارهای مختلفی مانند قابلیت یا تعمیم مفاهیمی مانند دقت اندازه‌گیری شود. به‌هرحال، در چنین شرایطی تضمینی وجود ندارد که مدل منتخب به‌خوبی به مجموعه دیگری از داده‌ها برازش گردد، زیرا مشخص نیست که ساختار احتمالی دو مجموعه داده‌ها یکسان هستند یا خیر.

در این مقاله تلاش می‌شود که با ورود جمله خطا به مدل که ماهیتی تصادفی و فازی دارد، مدل مناسب به داده‌ها برازش شود. در بخش دوم تعاریف و مفاهیم مقدماتی به همراه برخی گزاره‌های موردنیاز برای ادامه مقاله آورده شده‌اند. بخش سوم را به ارائه نتایج اصلی به همراه یک مثال جهت تشریح نتایج اختصاص داده‌ایم.

۲ مفاهیم مقدماتی

تعریف ۱.۲. *تابع $\tilde{u}:\mathbb{R}^n\to[0,1]$ را یک مجموعه فازی از $\mathbb{R}^n$ می‌نامیم. α برش مجموعه فازی $\tilde{u}$ را با $\tilde{u}_\alpha$ نشان داده و به‌صورت زیر تعریف می‌کنیم:*

$$\tilde{u}_\alpha=\begin{cases}\{x\in\mathbb{R}^n:\ \tilde{u}(x)\geq\alpha\}, & 0<\alpha\leq 1\\ supp\ \tilde{u}, & \alpha=0\end{cases}$$

که در آن

$$\tilde{u}_\alpha=supp\ \tilde{u}=cl\{x\in:\tilde{u}(x)>0\}$$

تکیه‌گاه مجموعه فازی $\tilde{u}$ می‌باشد. مجموعه فازی $\tilde{u}$ را نرمال گوییم، اگر $\tilde{u}_1\neq\emptyset$ و محدب نامیم، اگر برای هر $0\leq\alpha\leq 1$، $\tilde{u}_\alpha$ محدب باشد. $\tilde{u}$ را تابع عضویت نیز می‌نامند.

تعریف ۲.۲. مجموعه فازی $\tilde{u}:\mathbb{R}^n\to[0,1]$ *را یک عدد فازی گوییم، اگر $\tilde{u}$ محدب، نرمال و نیم‌پیوسته بالایی و $supp\ \tilde{u}$ فشرده باشد. خانواده همه اعداد فازی را با $F(\mathbb{R})$ نشان می‌دهیم (جو و کیم ۲۰۰۱).*

بنابراین در حالتی که $\tilde{u}$ بر $\mathbb{R}$ تعریف‌شده باشد، $\tilde{u}$ یک عدد فازی است اگر و تنها اگر $\tilde{u}_1\neq\emptyset$ و $\tilde{u}_\alpha$ برای هر $0\leq\alpha\leq 1$، یک فاصله بسته کران‌دار، $\tilde{u}_\alpha=[u_{l\alpha},u_{r\alpha}]$ باشد.

گزارهٔ ۳.۲. *(گوتشل و واکسمن ۱۹۸۶): اگر $\tilde{u}\in F(\mathbb{R})$ و $u_l(\alpha)=u_{l\alpha}$ و $u_r(\alpha)=u_{r\alpha}$ توابعی از $\alpha\in[0,1]$ باشند، آنگاه داریم:*

۱- $u_l(\alpha)$ $(u_r(\alpha))$ *تابعی کران‌دار صعودی (نزولی) بر $[0,1]$ است.*

۲- $u_l(1)\leq u_r(1)$

۳- $u_l(\alpha)$ و $u_r(\alpha)$ *بر بازه $[0,1)$ از چپ پیوسته و در نقطه صفر از راست پیوسته می‌باشند.*

۴. *اگر v_l و v_r در شرایط فوق صدق کنند، آنگاه $\tilde{v}\in F(\mathbb{R})$ یکتایی وجود دارد به‌طوری‌که برای همه $\alpha\in[0,1]$ داشته باشیم $\tilde{v}_\alpha=[v_{l\alpha},v_{r\alpha}]$.*

فرض کنید L و R توابع غیر صعودي از $\mathbb{R}^+$ به $[0,1]$ باشند و $L(0)=R(0)=1$. اگر ساختار تابع عضویت عدد فازی $\tilde{u}$ به‌صورت زیر باشد:

$$\tilde{u}(x)=\begin{cases}L(\frac{m-x}{\alpha}) & x\leq m,\\ R(\frac{x-m}{\beta}) & x\geq m,\end{cases}$$

$\tilde{u}$ را یک عدد فازی LR نامیده و با نماد (m,α,β) نشان می‌دهیم. عدد حقیقی m را مرکز و اعداد مثبت α و β را به ترتیب پهنای چپ و راست می‌نامیم. اگر $L=R$ و $\alpha=\beta$ آنگاه عدد فازی را متقارن و اگر $L=R=\max\{0,1-|x|\}$ آن را عدد فازی مثلثی نامیم و از نماد $(m;\alpha)_T$ برای نمایش آن استفاده می‌کنیم.

تعریف ۴.۲. *(ژو و لی ۲۰۰۱). فرض کنید $\tilde{u}$ و $\tilde{v}$ دو عدد فازی و $\tilde{u}_\alpha=[u_l(\alpha),u_r(\alpha)]$ و $\tilde{v}_\alpha=[v_l(\alpha),v_r(\alpha)]$ به ترتیب α برش‌های $\tilde{u}$ و $\tilde{v}$ باشند. فاصله بین این دو عدد فازی بر پایه تابع وزنی $f(\alpha)$ به‌صورت زیر تعریف می‌شود:*

$$d(\tilde{u},\tilde{v})=\left[\int_0^1 f(\alpha)d^2(\tilde{u}_\alpha,\tilde{v}_\alpha)d\alpha\right]^{\frac{1}{2}},$$

که در آن

$$d^2(\tilde{u}_\alpha,\tilde{v}_\alpha)=[u_l(\alpha)-v_l(\alpha)]^2+[u_r(\alpha)-v_r(\alpha)]^2$$

و $f(\alpha)$ تابعی صعودي روي بازه $[0,1]$ است، به قسمی که $f(0)=0$ و $\int_0^1 f(\alpha)d(\alpha)=1/2$ در این حالت نرم عدد فازی $\tilde{u}$ به‌صورت زیر تعریف می‌گردد:

$$||\tilde{u}||=d(\tilde{u},\tilde{u})=\left[\int_0^1 f(\alpha)\left((u_l(\alpha))^2+(u_r(\alpha))^2\right)d\alpha\right]^{\frac{1}{2}}$$

فرض کنید (Ω,F,P) یک فضای احتمال باشد. پوری و رالسکو در سال ۱۹۸۶ تعریفی از متغیرهای تصادفی فازی را ارائه کردند که مبنای بیشتر مطالعات بعدی است.

تعریف ۵.۲. *تابع $\tilde{X}^*:\Omega\to F(\mathbb{R})$ را متغیر تصادفی فازی نامیم،*

اگر اندازه‌پذیر باشد. متغیر تصادفی فازی $\tilde{X}^*$ را کران‌دار انتگرال‌پذیر گوییم اگر $E||\tilde{X}^*|| < \infty$.

لم زیر ارتباط بین متغیرهای تصادفی فازی و متغیرهای تصادفی معمولی متناظر با آن را نشان می‌دهد.

لم ۶.۲. *(جو و کیم ۲۰۰۱). فرض کنید $\tilde{X}^* : \Omega \to F((\mathbb{R}))$ و $\tilde{X}^*(w) = \{(X_{l\alpha}(w), X_{r\alpha}(w)) | 0 \le \alpha \le 1\}$ تحققی از $\tilde{X}^*$ باشد، $\tilde{X}^*$ یک متغیر تصادفی فازی است اگر و فقط اگر برای هر $0 \le \alpha \le 1$، $X_{l\alpha}$ (کران پایین آلفا برش $\tilde{X}^*$) و $X_{r\alpha}$ (کران بالای آلفا برش $\tilde{X}^*$) متغیرهای تصادفی معمولی باشند.*

متغیر تصادفی فازی مثلثی را می‌توان با سه متغیر $\tilde{X}^* = (X^L, X, X^U)$ توصیف کرد (حسامیان و اکبری ۲۰۲۲).

تعریف ۷.۲. *امید ریاضی متغیر تصادفی فازی کران‌دار انتگرال‌پذیر $\tilde{X}^*$ یک عدد فازی می‌باشد که به‌صورت $E(\tilde{X}^*) = \{(\int X_{l\alpha}dp, \int X_{r\alpha}dp) | 0 \le \alpha \le 1\}$ تعریف می‌شود. در این صورت*

$$E\left(\tilde{X}^*\right)(x) = sup\left\{\alpha \in [0,1] : x \in E\left(\tilde{X}^*_\alpha\right)\right\}.$$

فرض کنید X متغیر پیشگو و Y متغیر پاسخ نامنفی دو (یا چند) حالتی باشد. به‌عنوان‌مثال، $Y = 1$ نشان دهد فرد به بیماری (یا یکی از انواع بیماری‌ها) مبتلا است و $Y = 0$ نشان دهد فرد به بیماری مبتلا نیست. در بسیاری موارد مایلیم ارتباط بین متغیر پاسخ با متغیر (یا متغیرهای) پیشگو، X ، را مدل‌بندی کنیم؛ اما موارد زیادی وجود دارد که در مدل‌بندی بین متغیرهای پیشگو و پاسخ خطاهایی وجود دارند. این خطاها هم می‌توانند ماهیتی فازی داشته باشند (مانند ابهامات موجود در اندازه‌گیری در آزمایش‌ها) و هم ماهیتی تصادفی داشته باشند، که از برازش مدل به افراد مختلف که به‌تصادف انتخاب می‌شوند، ناشی شده باشند. برای اینکه

$$Y = \beta_0 + \beta_1 X \oplus \tilde{\varepsilon}^* \qquad (1)$$

که در آن β_0 و β_1 پارامترهای مجهول و $\tilde{\varepsilon}^*$ متغیر تصادفی فازی (مثلثی با مرکز 0) است، یک الگوی مناسب برای مدل‌بندی ارتباط بین متغیرهای پاسخ و پیشگو باشد، لازم است پارامترهای مدل برآورد گردند.

لم ۸.۲. *تحت شرایط (۱)، اگر $\tilde{\varepsilon}^*$ متغیر تصادفی فازی مثلثی با مرکز 0 باشد، Y به‌شرط X نیز متغیر تصادفی فازی مثلثی است.*

اثبات. به ازای هر $w \in \Omega$ و بنا به اصل گسترش $Y|X(w)$ یک عدد فازی مثلثی است. با توجه به خاصیت خطی بودن توابع اندازه‌پذیر، $Y|X$ اندازه‌پذیر و لذا یک متغیر تصادفی فازی مثلثی است. □

در مدل رگرسیون لوژستیک مقادیر متغیر پیشگو می‌توانند هر مقدار حقیقی باشند اما مقادیر متغیر پاسخ برچسب‌های رده‌بندی نامنفی (متغیرهای تصادفی فازی) هستند، لذا به‌جای پیشگویی متغیر پاسخ (که ماهیتی ابهامی نیز دارد) امید ریاضی آن، یعنی

$$\tilde{\mu} = E(Y|X) = \beta_0 + \beta_1 X \oplus \tilde{\varepsilon} \qquad (2)$$

را مدل‌بندی می‌کنیم، که در آن $\tilde{\varepsilon}$ یک عدد فازی است. در این صورت $\tilde{\mu}$ نیز یک عدد فازی است، در حالتی که Y متغیر تصادفی فازی باشد که دو عدد فازی را با احتمالاتی انتخاب کند، $\tilde{\mu}$ یک عدد فازی است که نشان‌دهنده احتمال فازی خواهد بود.
به‌عنوان یک نمونه عینی، در مطالعات بالینی و آزمایش‌های پزشکی عموماً هدف بررسی رابطه بین متغیر (متغیرهاي) پیشگو (مستقل) با یک متغیر وابسته (پاسخ) دو (یا چند) حالتی می‌باشد که با استفاده از روش‌های رگرسیون معمولی قابل‌بررسی نمی‌باشند، در این حالت‌ها از رگرسیون لوژستیک استفاده می‌شود؛ اما اغلب آزمایش‌ها یا تشخیص‌ها با خطا همراه می‌باشد. این خطا از فردی به فردی دیگر تغییر می‌کند، لذا ماهیتی تصادفی دارد و در مورد یک فرد هم به دلایل متعدد مانند عدم دقت وسایل اندازه‌گیری یا تشخیصی ماهیتی ابهامی دارد. در چنین حالتی با مواردي برخورد می‌کنیم که متغیر پاسخ نیز ماهیتی تصادفی فازی دارد در این صورت مدل‌های رگرسیون لوژستیک معمولی، قادر به تحلیل این نوع داده‌ها نیستند. استفاده از مدل رگرسیون لوژستیک فازی (عموماً بدون در نظر گرفتن ماهیت تصادفی بودن متغیر پاسخ) به‌خصوص در مطالعات بالینی موردتوجه و استفاده برخی از محققان قرارگرفته است، که در مقدمه به آن اشاره گردید. در این مقاله حالتی را بررسی می‌کنیم که متغیر پاسخ متغیری تصادفی فازی باشد.

۳ نتایج اصلی

تابع خطی سمت راست (۲) یک تابع بیکران است لذا برای به الگو درآوردن مقدار متناهی (سمت چپ) مناسب نخواهد بود. در ادامه از تعمیم نظریه رگرسیون لوژستیک برای مدل‌بندی ارتباط بین $\tilde{\mu}$ و

X استفاده خواهیم کرد. فرض می‌کنیم که $\tilde{\varepsilon}$ یک عدد فازی LR ، $\tilde{\varepsilon}=(m^*;\alpha,\beta)_{LR}$ ، $\alpha,\beta>0$ باشد.

گزارهٔ ۱.۳. *اگر در مدل (۲)، $\tilde{\varepsilon}$ ، $\alpha,\beta>0$ یک عدد فازی LR ، $\tilde{\varepsilon}=(m^*;\alpha,\beta)_{LR}$ ، باشد، آنگاه $\tilde{\mu}$ (تحت شرط $X=x$) یک عدد فازی LR ، $\tilde{\mu}=(m;\alpha,\beta)_{LR}$ ، با $m=m^*+\beta_0+\beta_1X$ ، $\alpha,\beta>0$ است.*

اثبات. فرض کنید $\tilde{\varepsilon}_\alpha$ و $\tilde{\mu}_\alpha$ به ترتیب α برش $\tilde{\varepsilon}$ و $\tilde{\mu}$ باشد. بنا به اصل گسترش (طاهری و ماشین چی ۱۳۸۶) $\tilde{\mu}$ یک مجموعه فازی است. چون $\tilde{\mu}$ تابع یک‌به‌یک است، $\tilde{\mu}(y)=\sup_{\varepsilon,y=\beta_0+\beta_1x+\varepsilon}\tilde{\varepsilon}(y-\beta_0+\beta_1x)$

$$=\tilde{\varepsilon}(\varepsilon)$$

و لذا $\tilde{\mu}$ یک عدد فازی LR ، $\tilde{\mu}=(m;\alpha,\beta)_{LR}$، با $m=m^*+\beta_0+\beta_1X$ ، $\alpha,\beta>0$ است. □

بدیهی است که در معادله (۲) تکیه‌گاه طرف چپ بازه صفر و یک (یا در حالت کلی یک بازه مثبت و متناهی) و طرف راست $\mathbb{R}$ است. لذا مدل‌بندی را برای تبدیل زیر بر نمونه تصادفی به حجم n اعمال می‌کنیم:

$$\tilde{w}_i=\ln\left(\frac{\tilde{\mu}_i}{1-\tilde{\mu}_i}\right)=\beta_0+\beta_1X_i\oplus\tilde{\varepsilon}_i, \qquad i=1,2,...,n. \tag{3}$$

در ادامه نخست فرض می‌شود که $\tilde{\mu}$ ، عددی فازی LR و سپس حالت خاص آن عدد فازی مثلثی است.

لم ۲.۳. *تحت شرایط مدل (۳) اگر $\tilde{\varepsilon}$ یک عدد فازی LR ، $\tilde{\varepsilon}=(m^*;\alpha,\beta)_{LR}$ ، $\alpha,\beta>0$ باشد، آنگاه*

$$\tilde{w}=\ln\left(\frac{\tilde{\mu}}{1-\tilde{\mu}}\right)=\beta_0+\beta_1X\oplus\tilde{\varepsilon}$$

یک عدد فازی با تابع عضویت

$$\tilde{w}(y)=\begin{cases} L\left(\frac{m-\frac{1}{1+e^{-y}}}{\alpha}\right) & \frac{1}{1+e^{-y}}\le m\\ R\left(\frac{\frac{1}{1+e^{-y}}-m}{\beta}\right) & \frac{1}{1+e^{-y}}>m\end{cases}$$

خواهد بود.

لازم به ذکر است که $\tilde{\mu}$ ، عددی فازی با تکیه‌گاه مقادیر نامنفی و لذا $m=m^*+\beta_0+\beta_1X$ یک عدد مثبت است.

اثبات. بنا به اصل گسترش

$$\begin{aligned}\tilde{w}(y)&=\sup_{x\in[0,1],y=\ln\frac{x}{1-x}}\tilde{\mu}(x)\\ &=\tilde{\mu}\left(\tfrac{1}{1+e^{-y}}\right).\end{aligned}$$

$\tilde{\mu}$ یک عدد فازی LR ، به مرکز $m=m^*+\beta_0+\beta_1X$ است. چون بر روی تکیه‌گاه، $\ln\frac{x}{1-x}$ تابعی یکنوا می‌باشد و تنها یک x است که در رابطه فوق صدق می‌کند، لذا

$$\tilde{w}(y)=\begin{cases} L\left(\frac{m-\frac{1}{1+e^{-y}}}{\alpha}\right) & \frac{1}{1+e^{-y}}\le m\\ R\left(\frac{\frac{1}{1+e^{-y}}-m}{\beta}\right) & \frac{1}{1+e^{-y}}>m\end{cases}$$

که نتیجه حاصل است. □

نتیجه ۳.۳. *اگر $\tilde{\varepsilon}=E(\tilde{\varepsilon}^*)=(0\ ;\frac{1}{2})_T$ آنگاه $\tilde{\mu}=(\beta_0+\beta_1X;\frac{1}{2})_T$*

$$\tilde{\mu}(x)=\begin{cases}1-2|m-x| & x\le m\\ 1-2|x-m| & x>m\end{cases}$$

یا

$$\tilde{\mu}(x)=\begin{cases}1-2|m-x| & m-2\le x\le m\\ 1-2|x-m| & m<x<m+2\end{cases}$$

و

$$\tilde{\mu}_\alpha=[m+0.5(\alpha-1),m+0.5(1-\alpha)]$$

نتیجه ۴.۳. *اگر $\tilde{\varepsilon}=E(\tilde{\varepsilon}^*)=(0;\frac{1}{2})_T$ آنگاه برای آلفا برش‌های $\tilde{w}$ داریم*

$$\begin{aligned}\tilde{w}_\alpha&=\{y:\tilde{w}(y)\ge\alpha\}\\ &=\{y:\tilde{\mu}(\tfrac{1}{1+e^{-y}})\ge\alpha\}.\end{aligned}$$

اگر $m+0.5(\alpha-1)\ge 0$ آنگاه

$$\begin{aligned}\tilde{w}_\alpha&=\{y:m+0.5(\alpha-1)\le\tfrac{1}{1+e^{-y}}\\ &\qquad\le m+0.5(1-\alpha)\}\\ &=\{y:(m+0.5(1-\alpha))^{-1}-1\le e^{-y}\\ &\qquad\le(m+0.5(\alpha-1))^{-1}-1\}\\ &=\{y:\ln(\tfrac{1-m-0.5(1-\alpha)}{m+0.5(1-\alpha)}\le -y\\ &\qquad\le\ln(\tfrac{1-m-0.5(\alpha-1)}{m+0.5(\alpha-1)}\}\\ &=\{y:\ln(\tfrac{m+0.5(\alpha-1)}{1-m-0.5(\alpha-1)})\le y\\ &\qquad\le\ln(\tfrac{m+0.5(1-\alpha)}{1-m-0.5(1-\alpha)}\}\end{aligned}$$

یعنی

$$\begin{aligned}&\tilde{w}_\alpha=\\ &\left[\ln\tfrac{(\alpha-1)0.5+m}{1-((\alpha-1)0.5+m)},\ln\tfrac{(1-\alpha)0.5+m}{1-((1-\alpha)0.5+m)}\right]\\ &=[A_l,A_r].\end{aligned}$$

نتیجه ۵.۳. *فرض کنید $\tilde{\varepsilon}^*$ متغیر تصادفی فازی مثلثی با مرکز ۰ باشد. تحت شرایط مدل (۲) امید ریاضی مقدار مشاهده‌شده Y به‌شرط $X=x$ مقدار مشاهده‌شده از یک متغیر تصادفی فازی مثلثی به مرکز*

$\beta_0+\beta_1 x$ است. تحت شرایط مدل (۳)، مقدار مشاهده‌شده، مقداری از یک متغیر تصادفی فازی به مرکز $m=\beta_0+\beta_1 x$ و پهنای متناظر با مقدار مشاهده‌شده از متغیرهای تصادفی X^L و X^U مثلاً h_l و h_u است؛ یعنی در این حالت

$$\tilde{w}_\alpha=\left[\ln\frac{(\alpha-1)h_l+m}{1-((\alpha-1)h_l+m)},\ln\frac{(1-\alpha)h_u+m}{1-((1-\alpha)h_u+m)}\right]=[A_l,A_r].$$

اگر فرض شود که متغیر فازی مثلثی متقارن نیز هست در این صورت

$$\tilde{w}_\alpha=\left[\ln\frac{(\alpha-1)h+m}{1-((\alpha-1)h+m)},\ln\frac{(1-\alpha)h+m}{1-((1-\alpha)h+m)}\right]=[A_l,A_r].$$

قضیهٔ ۶.۳. فرض کنید (x_1,y_1) ، ... ، (x_n,y_n) ، n مشاهده مستقل از مدل (۱) و $\tilde{\varepsilon}_i^*$ ها متغیرهای تصادفی فازی مثلثی متقارن، باشند که در آن $\tilde{\varepsilon}_i^*=(e_i,h_i)_T$ مرکز خطا (و برابر صفر) و کران h_i متغیری تصادفی دارای توزیع یکنواخت در بازه (۰،۱) می‌باشد. برآوردهای کمترین توان‌های دوم پارامترها β_0 و β_1 مدل (۳) و نسبت به متر (۱) به ازای $f(\alpha)=\alpha$ عبارت‌اند از

$$\hat{\beta}_0=\bar{z}_i-\hat{\beta}_1\bar{x}_i$$

$$\hat{\beta}_1=\frac{\sum_{i=1}^{n}(z_ix_i-n\bar{x}\bar{z})}{\sum_{i=1}^{n}(x_i^2-n\bar{x}^2)}$$

که در آن m_i مرکز μ_i است و $z_i=z_{i1}+z_{i2}$.

اثبات. تحت شرایط قضیه $X_\alpha\sim U(0,1-\alpha)$ ،$\tilde{\varepsilon}^*_\alpha=(-X_\alpha,0,X_\alpha)$ است و متغیر تصادفی فازی مثلثی را با سه متغیر $\tilde{\varepsilon}^*_\alpha=(-X,0,X)$ ، $X\sim U(0,1)$ می‌توان توصیف کرد؛ بنابراین عدد فازی $E(\tilde{\varepsilon}^*_\alpha)$ به‌صورت $E(\tilde{\varepsilon}^*_\alpha)=(0,\frac{1}{2})_T$ با آلفا برش‌های زیر خواهد بود:

$$(E(\tilde{\varepsilon}^*))_\alpha\sim(\frac{\alpha-1}{2},\frac{1-\alpha}{2}).$$

کران‌های آلفا برش عدد فازی مشاهده‌شده از متغیر تصادفی فازی مثلثی متقارن با توزیع یکنواخت، $\tilde{\varepsilon}^*=(e_i,U(0,1))$ ، به‌صورت زیر می‌باشند:

$$\tilde{\varepsilon}_{l\alpha}=(\alpha-1)h_i+e_i,$$

$$\tilde{\varepsilon}_{r\alpha}=(1-\alpha)(1-h_i)+e_i$$

که در آن h_i مقدار مشاهده از توزیع یکنواخت $U(0,1)$ است. در این صورت

$$\tilde{\mu}_i=E(Y|X)=\beta_0+\beta_1X_i\oplus\tilde{\varepsilon}_i$$

عددی فازی مثلثی متقارن است. مراکز $\tilde{\mu}_i$ را به m_i نمایش می‌دهیم . تحت شرایط مدل (۳)،

$$\tilde{w}_i=\ln(\frac{\tilde{\mu}_i}{1-\tilde{\mu}_i})=\beta_0+\beta_1X_i\oplus\tilde{\varepsilon}_i$$

یک عدد فازی است که آلفا برش آن در نتیجه ۳ داده شده است. اگر $X_i=x_i$ آنگاه

$$\begin{aligned}\hat{w}_i&=\beta_0+\beta_1X_i+E(\tilde{\varepsilon}_i^*)\\ \Rightarrow\hat{w}_{i_\alpha}&=[\{\beta_0+\beta_1x_i+\tfrac{(\alpha-1)}{2},\\ &\beta_0+\beta_1x_i+\tfrac{(1-\alpha)}{2}\}]\\ ,\quad&=[B_{il\alpha},B_{ir\alpha}].\end{aligned}$$

می‌خواهیم پارامترهاي مدل را به روش کمترین مجموع توان دوم خطاها، به‌گونه‌ای برآورد نماییم که فاصله بین بازه مشاهده‌شده $\tilde{w}_{i_\alpha}=[A_{il\alpha},A_{ir\alpha}]$ و بازه برآورد شده $\hat{w}_{i_\alpha}=[B_{il\alpha},B_{ir\alpha}]$ ، (که به‌اختصار از ذکر اندیس α خودداری می‌گردد) مینیمم شود. در این مرحله از فاصله (۱) استفاده می‌کنیم. در این صورت داریم

$$\begin{aligned}SSE&=\sum_{i=1}^{n}\int_0^1\alpha d^2(\tilde{w}_{i_\alpha},\hat{w}_{i_\alpha})d\alpha\\ &=\sum_{i=1}^{n}\int_0^1\alpha([A_{il}-B_{il}]^2+[A_{ir}-B_{ir}]^2)d\alpha\\ &=\sum_{i=1}^{n}\left[\int_0^1\alpha\left[A_{il}{}^2+A_{ir}{}^2\right]d\alpha+\int_0^1\alpha\left[B_{il}{}^2+B_{ir}{}^2\right]d\alpha\right]\\ &\quad-2\sum_{i=1}^{n}(\int_0^1\alpha\left[A_{il}B_{il}+A_{ir}B_{ir}\right]d\alpha)\\ &=\sum_{i=1}^{n}\left[\int_0^1\alpha\left[A_{il}{}^2+A_{ir}{}^2\right]d\alpha\right]\\ &\quad+\sum_{i=1}^{n}\int_0^1\alpha\left[(\beta_0+\beta_1x_i+\tfrac{(\alpha-1)}{2})^2+(\beta_0+\beta_1x_i+\tfrac{(1-\alpha)}{2})^2\right]d\alpha\\ &\quad-2\int_0^1\alpha\Big[(\ln\tfrac{(\alpha-1)h_i+m_i}{1-((\alpha-1)h_i+m_i)})(\beta_0+\beta_1x_i+\tfrac{(\alpha-1)}{2})\\ &\quad+(\ln\tfrac{(1-\alpha)h_i+m_i}{1-((1-\alpha)h_i+m_i)})(\beta_0+\beta_1x_i+\tfrac{(1-\alpha)}{2})\Big]d\alpha\end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=1}^{n}\Bigg[\int_0^1 \alpha\left[A_{il}{}^2 + A_{ir}{}^2\right]d\alpha + \int_0^1 \alpha\Big(2\beta_0^2 + 2\beta_1^2 x_i^2 + \tfrac{\alpha^2-2\alpha+1}{2} + 4\beta_0\beta_1 x_i\Big)d\alpha \\
&\quad - 2\overbrace{\int_0^1 \alpha\Big(\frac{(\alpha-1)}{2}\Big)A_{il}d\alpha}^{k_{i1}} - 2(\beta_0+\beta_1 x_i)\overbrace{\int_0^1 \alpha A_{il}d\alpha}^{z_{i1}} - 2\overbrace{\int_0^1 \alpha\Big(\frac{(1-\alpha)}{2}\Big)A_{ir}d\alpha}^{k_{i2}} \\
&\quad - 2(\beta_0+\beta_1 x_i)\overbrace{\int_0^1 \alpha A_{ir}d\alpha}^{z_{i2}}
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{i=1}^{n}\Bigg[\int_0^1 \alpha\left[A_{il}{}^2 + A_{ir}{}^2\right]d\alpha + \int_0^1 \alpha\Big(2\beta_0^2 + 2\beta_1^2 x_i^2 + \tfrac{\alpha^2-2\alpha+1}{2} + 4\beta_0\beta_1 x_i\Big)d\alpha \\
&\quad - 2\overbrace{\int_0^1 \alpha\Big(\frac{(\alpha-1)}{2}\Big)A_{il}d\alpha}^{k_{i1}} - 2(\beta_0+\beta_1 x_i)\overbrace{\int_0^1 \alpha A_{il}d\alpha}^{z_{i1}} - 2\overbrace{\int_0^1 \alpha\Big(\frac{(1-\alpha)}{2}\Big)A_{ir}d\alpha}^{k_{i2}} \\
&\quad - 2(\beta_0+\beta_1 x_i)\overbrace{\int_0^1 \alpha A_{ir}d\alpha}^{z_{i2}}
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{i=1}^{n}\Bigg[\int_0^1 \alpha\left[A_{il}{}^2 + A_{ir}{}^2\right]d\alpha + \int_0^1 \alpha\Big(2\beta_0^2 + 2\beta_1^2 x_i^2 + \tfrac{\alpha^2-2\alpha+1}{2} + 4\beta_0\beta_1 x_i^2\Big)d\alpha \\
&\qquad -2[\overbrace{k_{i1}+k_{i2}}^{k_i} + \overbrace{(z_{i1}+z_{i2})}^{z_i}(\beta_0+\beta_1 x_i)\Bigg] \\
&= \sum_{i=1}^{n}\Big[\int_0^1 \alpha\left[A_{il}^2 + A_{ir}^2\right]d\alpha + \big(\beta_0^2 + \beta_1^2 x_i^2 + \tfrac{1}{24} + 2\beta_0\beta_1 x_i\big) \\
&\qquad - 2[k_i + z_i(\beta_0+\beta_1 x_i)]\Big]
\end{aligned}
$$

اکنون با مشتق‌گیری نسبت به β_0 و β_1 داریم:

$$
\begin{aligned}
\frac{\partial SSE}{\partial\beta_0} &= \sum_{i=1}^{n}(2\beta_0 + 2\beta_1 x_i - 2z_i),\\
\frac{\partial SSE}{\partial\beta_1} &= \sum_{i=1}^{n}(2\beta_1 x_i^2 + 2\beta_0 x_i - 2z_i x_i)
\end{aligned}
$$

از روابط

$$\frac{\partial SSE}{\partial\beta_0} = 0$$

و

$$\frac{\partial SSE}{\partial\beta_1} = 0$$

داریم:

$$
\begin{aligned}
&n\beta_0 = \sum_{i=1}^{n}(-\beta_1 x_i + z_i),\\
&\sum_{i=1}^{n}\beta_1 x_i^2 + \frac{1}{n}\sum_{i=1}^{n}(z_i - \beta_1 x_i)\sum_{j=1}^{n}x_j = \sum_{i=1}^{n} z_i x_i
\end{aligned}
$$

ضرایب مدل به‌صورت زیر به دست می‌آیند:

$$\hat\beta_0 = \bar z - \hat\beta_1 \bar x, \qquad \hat\beta_1 = \frac{\sum_{i=1}^{n} z_i x_i - n\bar x\bar z}{\sum_{i=1}^{n} x_i^2 - n\bar x^2}.$$

□

۴ مثال

در مثال زیر متغیر زبانی به‌صورت دوحالتی (خیلی کم و خیلی زیاد) در نظر می‌گیریم که متوسط و بیشتر از آن را خیلی زیاد و کمتر از متوسط را خیلی کم در نظر می‌گیریم.

مثال ۱.۴. *بيماري لوپوس: اين بيماري يک بيماري مزمن است که در آن دستگاه ایمنی بدن، بنا به دلایلی نامعلوم، پادتنی تولید می‌کند که به‌جای مقابله با ویروس‌ها، اندام سالم بدن را مورد هجوم قرار می‌دهد. فایوس و همکاران (۲۰۰۸) تشخیص این بیماري بر اساس علائم آن صورت می‌گیرد. به‌طور متداول مجموعه‌اي از ۱۱ علامت براي بيماري*

لوپوس در نظر گرفته‌شده و فردي با حداقل ۴ علامت از این مجموعه، بیمار در نظر گرفته می‌شود هینلین و همکاران (۲۰۰۷) و نوسینت و همکاران (۲۰۰۵). نظر پزشک به علائم، نتایج آزمایش‌ها و سوابق خانوادگی فرد در بیماري لوپوس وابسته و ماهیتی تصادفی و فازي دارد.

نمونه‌اي متشکل از ۱۵ نفر زن مشکوك به بیماري لوپوس در بازه سنی ۱۸-۴۰ سال را که از درمانگاه‌هاي شهر شیراز در مطالعات پیشین، جمع آوري شده بود، موردبررسی قرارگرفته‌اند. هدف از این مطالعه، مدل بندی وضعیت افراد مشکوك به بیماري لوپوس، بر اساس تعدادي عوامل خطر مهم است به‌گونه‌ای که مدل برازش داده شده قادر باشد بخت احتمالی - امکانی ابتلا به بیماري لوپوس را در هر فرد برآورد کند. بر اساس اطلاعات هر یک از ۱۵ نمونه، از پزشک خواسته‌شده است تا امکان بیمار بودن هر فرد را با توجه به علائم و وضعیت عمومی و به‌صورت یکی از واژه‌هاي زبانی «خیلی کم، کم، متوسط، زیاد و خیلی زیاد» بیان کند. می‌پذیریم که نظر پزشک ماهیت تصادفی نیز دارد به‌گونه‌ای که این ماهیت تصادفی در درجه عضویت واژه‌های زبانی تأثیرگذار است. اطلاعات مربوط به ۱۵ فرد نمونه در جدول زیر ارائه‌شده است. در تمام موارد کد یک داشتن مشخصه موردنظر و کد صفر عدم داشتن مشخصه است. (در این قسمت تنها عامل anti DNA را به‌عنوان متغیر مستقل و متغیر پاسخ در نظر می‌گیریم.)

سابقه خانوادگی	قرارگرفتن در معرض خورشید	ANA	$AntiDNA$	ESR	شدت بیماری
۱	۱	۱۱۲	۱۰۵	۱	زیاد
۰	۱	۸۰	۲۳	۰	متوسط
۰	۱	۱۱۵	۱۵	۰	زیاد
۰	۱	۱۰۵	۱۰۷	۱	زیاد
۰	۰	۸۹	۱۵۰	۱	متوسط
۱	۱	۱۶۰	۱۰	۱	خیلی زیاد
۰	۱	۱۰۰	۲۳	۰	متوسط
۰	۰	۱۰۰	۸۵	۱	زیاد
۰	۱	۴۸	۸۳	۰	کم
۱	۰	۱۵	۱۹	۱	خیلی کم
۰	۰	۵۰	۹۱	۰	کم
۰	۱	۵۹	۲۰۰	۱	متوسط
۰	۱	۸۳	۲۰	۱	کم
۰	۰	۱۵	۲۰۰	۰	کم
۱	۰	۸۵	۱۵	۱	متوسط

شکل ۱. داده‌ها

خیلی کم : متغیر تصادفی فازي حدوداً ۰/۲۵ با تابع عضویت:

$$(۰/۲۵; U(۰, ۰/۲۵); U(۰/۲۵, ۰/۵))_T$$

خیلی زیاد : متغیر تصادفی فازي حدوداً ۰/۹۸ با تابع عضویت

$$(۰/۷۵; U(۰/۵, ۰/۷۵); U(۰/۷۵, ۱))_T$$

در این صورت

$$m = \{۰/۷۵, ۰/۷۵, ۰/۷۵, ۰/۷۵, ۰/۷۵, ۰/۷۵, ۰/۷۵, ۰/۷۵, ۰/۲۵, ۰/۲۵, ۰/۲۵, ۰/۷۵, ۰/۲۵, ۰/۲۵, ۰/۷۵\}$$

$$h = \{۰/۶۰, ۰/۶۹, ۰/۵۱, ۰/۵۰, ۰/۶۵, ۰/۷۳, ۰/۵۱, ۰/۵۸, ۰/۱۰, ۰/۱۷, ۰/۱۱, ۰/۵۲, ۰/۲۰, ۰/۰۴, ۰/۷۲\}$$

$$q = \{۰/۸۵, ۰/۸۲, ۰/۹۵, ۰/۷۵, ۰/۹۶, ۰/۸۴, ۰/۸۵, ۰/۹۵, ۰/۳۸, ۰/۲۷, ۰/۳۲, ۰/۸۷, ۰/۲۸, ۰/۴۱, ۰/۹۱\}$$

$$h۱ = m - h = \{۰/۱۵, ۰/۰۶, ۰/۲۴, ۰/۲۵, ۰/۱۰, ۰/۰۲, ۰/۲۴, ۰/۱۷, ۰/۱۵, ۰/۰۸, ۰/۱۴, ۰/۲۳, ۰/۰۵, ۰/۲۱, ۰/۰۳\}$$

$$h۲ = m - q = \{۰/۱۰, ۰/۰۷, ۰/۲۰, ۰/۰۰, ۰/۲۱, ۰/۰۹, ۰/۱۰, ۰/۲۰, ۰/۱۳, ۰/۰۲, ۰/۰۷, ۰/۱۲, ۰/۰۳, ۰/۱۶, ۰/۱۶\}$$

کدهای محاسبات پارامترها به شرح زیرند:

$$z < -function\,(x)$$
$$\{x * (log((x - ۱) * hl[i] +$$
$$m\,[i])\,(۱ - ((x - ۱) * hl\,[i]$$
$$+m\,[i]))) + x * (log\,((x - ۱) * hr\,[i]$$
$$+m\,[i])\,/(۱ - ((x - ۱) * hr\,[i] + m\,[i]))))\}$$
$$integrate(z, ۰, ۱)$$

$$x۱ = c(۱۰۵, ۲۳, ۱۵, ۱۰۷, ۱۵۰, ۱۰, ۲۳, ۸۵,$$
$$۸۳, ۱۹, ۹۱, ۲۰۰, ۲۰, ۲۰۰, ۱۵)$$

$$z۱ = c(-۱٫۱۸۳, -۱٫۱۶۷, -۱٫۲۱, -۱٫۱۸۵,$$
$$-۱٫۱۹۱, -۱٫۶۴, -۱٫۹۶, -۱٫۱۹۹, -۲٫۰۱۲, -۱٫۹,$$
$$-۱٫۹۶۶, -۱٫۱۹۷, -۱٫۸۸۸, -۲٫۰۸۸, -۱٫۱۷۵)$$

$$b۱ = sum((z۱ * x۱) - (۱۵) * (mean(x۱))$$
$$*mean(z۱))/sum((x۱ * x۱) - (۱۵)$$
$$*(mean(x۱)) * mean(x۱))$$

بنابراین برآورد پارامترها عبارت‌اند از

$$b۰ = mean\,(z۱) - b۱ * (mean\,(x۱))$$
$$> b۱$$
$$[۱]\ -\ ۰٫۰۲۱۱۴۷۶$$
$$> b۰$$
$$۰٫۰۸۳۲۳۵۸$$

و مدل به‌صورت زیر به دست می‌آید:

$$\beta_۰ = ۰٫۰۸۳۲, \quad \beta_۱ = -0٫0211$$
$$\tilde{w}_i = ۰٫۰۸۳۲ - 0٫0211X \oplus \tilde{\epsilon}$$

۵ نتیجه‌گیری

رگرسیون لوژستیک یک مدل مناسب برای توصیف، مدل بندی یا پیشگویی متغیرهای وابسته کیفی نسبت به متغیرهای مستقل، بر اساس یک نمونه تصادفی از این متغیرها است که توسط محققین مختلفی موردبحث قرارگرفته است. یک دسته از مؤلفین با صرف‌نظر از ماهیت تصادفی متغیرها حالتی را در نظر گرفته‌اند که متغیرها ماهیتی ابهامی داشته و با این فرض متغیرها را مدل بندی کرده‌اند. در این مقاله به مدل بندی مشاهدات تحت شرایط وجود تصادف و ابهام در مشاهدات پرداخته‌شده است. نتایج تنها برای حالتی که یک متغیر مستقل وجود دارد و نسبت به فاصله ژو و لی (۲۰۰۱) استخراج‌شده است و با مثالی نحوه محاسبات تشریح شده‌اند. در ادامه این مدل بندی برای تعداد کلی‌تری از متغیرهای مستقل و تحت شرایط مختلفی در خصوص ابهام در پارامترها یا متغیرها می‌تواند موردمطالعه قرار گیرد.

مراجع

[۱] طاهری، س.م و ماشین چی، م. (۱۸۷۶). مقدمه‌ای بر احتمال و آمار فازی. دانشگاه شهید باهنر کرمان.

[2] Chen, X. W., Anantha, G., and Lin, X. (2008). Improving Bayesian network structure learning with mutual information-based node ordering in the K2 algorithm. *IEEE Transactions on Knowledge and Data Engineering*, **20(5)**, 628-640.

[3] Fauci, A. Braunwald, S.E., Kasper, D.L., Hauser, S. L., Longo, D.L., Jameson, J.L., and Loscalzo, J. (2008). Harrison's principals of internal medicine, John Wiley,New York. 2275- 2279.

[4] Goetschel, R. and Voxman, W. (1986). Elementary fuzzy calculus, *Fuzzy Sets and Systems*, **18**, 31-43.

[5] Heinlen, LD., clain, M.T., and Merrill, J. (2007). Clinical criteria for systemic lupus erythematosus precede diagnosis, and associated autoantibodies are present before clinical symptoms, *Arthritis Rheumatology*, **56**, 2344-2351.

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

[7] Hesamian, G., and Akbari, M.G. (2022). A fuzzy empirical quantile-based regression model based on triangular fuzzy numbers, *Computational and Applied Mathematics*, **41(6)**, 1-26.

[8] Joo, S. Y., and Kim, Y. K. (2001). Kolmogorov's strong law of large numbers for fuzzy random variables, *Fuzzy Sets and Systems*, **120(3)**, 499-503.

[9] Mirzaei Yeganeh, S., and Taheri, S.M (2010). Possibilistic Logistic Regression by Linear Programming Approach, *Proc of 7th probability and random process. Isfahan University of Technology. Isfahan. Iran.* , 139-143.

[10] Mohd Dom, R. (2009). A fuzzy regression model for the prediction of oral cancer susceptibility, *Ph.D. Thesis, Computer Science and Information Technology, University of Malaya, Kualalumpur.*

[11] Nagar, P., and Srivastava, S. (2008).Adaptive fuzzy regression model for the prediction of dichotomous response variables using cancer data: a case study, *Journal of Applied Mathematics Statistics and Informatics.*,**4**, 183-191.

[12] Namdari, M., Taheri, S.M., Abadi, A., Rezaei, M., and Kalantari, N. (2013). Possibilistic logistic regression for fuzzy categorical response data, *International Conference on Fuzzy Systems.*, **8**, 1-6.

[13] Nossent, H.C., and Rekvig, O.P. (2005). Is closer linkage between systemic lupus erythematosus and anti-double-stranded DNA antibodies a desirable and attainable goal, *Fuzzy Sets and Systems*, **2**, 85-87.

[14] Pourahmad, S., Ayatollahi, SMT., Taheri, SM., and Agahi, ZH. (2011a). Fuzzy logistic regression based on the least squares approachs with application in clinical studies, *Computer Mathematics Applications.*, **62**, 3353-3365.

[15] Puri, M. L., and Ralescu, D. A. (1986). Fuzzy random variables, *J. Math. Anal. Appl.* , **114**, 409-422.

[16] Salmani, F., Taheri, S. M., Yoon, J.H., Abadi, A., Alavi Majd, H., and Abbaszadeh, A. (2017). Logistic regression for fuzzy covariates: modeling, inference and Applications, *International Journal of Fuzzy Systems* , **114**, 409-422.

[17] Taheri, S.M. and Mirzaie Yeganeh, S. (2009). Logistic regression with non-precise response, *Proceedings. of the 57th ISI conference, Durban, South Africa.*, 98-101.

[18] Taheri, S.M. and Mirzaie Yeganeh, S. (2009). Logistic regression with non-precise response, *Proceedings. of the 57th ISI conference, Durban, South Africa.*, 98-101.

[19] Takemura, K. (2004). Fuzzy logistic regression analysis for fuzzy input–output data, *International Conference on Soft Computing and Intelligent Systems and the 5th International Symposium on Advanced Intelligent Systems,nJapan.*, 1-6.

[20] Tseng, F. and Lin, L. (2005). A quadratic interval logit model for forecasting bankruptcy, *Omega*, **33**, 85-91.

[21] Yu, J.R., and Tseng, F.M. (2014). Fuzzy piecewise logistic growth model for innovation diffusion: a case study of the TV industry, *International Journal of Fuzzy Systems* , **18**, 1-12.

[22] Xu, R., Li, C. (2001). Multidimensional least-squares fitting with a fuzzy model, *Fuzzy Sets and Systems.*, **119**,**(2)**, 215-223.

Logistic regression modeling for linguistic data under fuzzy random errors

M. Maleki [1], H.R.Nili-Sani [2], M.GH.Akbari[3]

Abstract:

In this article, logistic regression models are studied in which the response variables are two (or multiple) values and the explanatory variables (predictor or independent) are ordinary variables, but the errors have a vagueness nature in addition to being random. Based on this, we formulate the proposed model and determine the estimation of the coefficients for a case with only one explanatory variable using the method of least squares. In the end, we explain the results with an example.

Keywords: Logistic regression, Fuzzy random variables,Fuzzy Linguistic random variables.

[1] **University of Birjand**
[2] **University of Birjand**
[3] **University of Birjand**