سیستم‌های فازی و کاربردها

سال سوم، شماره دوم (پاییز و زمستان ۱۳۹۹) صص. ۱۵۳ تا ۱۶۴

محاسبه تابع عضویت p-مقدار فازی در نرم‌افزار R

محبوبه سادات مدنی، محمدرضا ربیعی، عباس پرچمی

دانشگاه صنعتی شاهرود، دانشکده ریاضی، بخش آمار

دانشگاه شهید باهنر کرمان، دانشکده ریاضی و کامپیوتر، بخش آمار

چکیده

آزمون فرضیه‌های آماری دارای اهمیت زیادی برای تصمیم‌گیری در مسایل علمی و کاربردی است. در روش‌های معمول آزمون فرضیه‌های آماری، داده‌ها، فرضیه‌ها، پارامترها و سایر عناصر مساله دقیق درنظر گرفته می‌شوند. اما در علوم کاربردی مانند اقتصاد، کشاورزی، پزشکی، نجوم و علوم اجتماعی ممکن است با تعاریف مبهم و مفاهیم فازی مانند آستانه‌ی تحمل بیمار، کیفیت محصولات تولیدی کارخانه و درآمد ماهیانه‌ی یک راننده‌ی تاکسی مواجه شویم. در چنین شرایطی روش‌های کلاسیک نیاز به تعمیم برای کارایی در محیط‌های فازی دارند. ورود ابهام در مساله‌ی آزمون فرضیه‌ها می‌تواند از طریق داده‌ها و/یا فرضیه‌ها صورت گیرد. بنابراین سه مساله‌ی عمده‌ی زیر را می‌توان درنظر گرفت: (۱) آزمون فرضیه‌های دقیق براساس داده‌های فازی، (۲) آزمون فرضیه‌های فازی براساس داده‌های دقیق، (۳) آزمون فرضیه‌های فازی براساس داده‌های فازی. در این مقاله به بحث و بررسی رویکرد p-مقدار، در سه مساله بالا، به کمک بسته‌ی نرم‌افزاری Fuzzy.p.value در نرم‌افزار R می‌پردازیم. محاسبه‌ی تابع عضویت p-مقدار فازی، مقایسه‌ی آن با سطح معنی‌داری فازی و تصمیم نهایی فازی در آزمون فرضیه از وظایف اصلی این بسته‌ی نرم‌افزاری است که به همراه چند مثال عددی مورد بررسی قرار گرفته است.

Mathematics Subject Classification (2010): 62A86 , **Email:** rabie1354@yahoo.com.

عبارات و کلمات کلیدی: داده‌های فازی، آزمون فرضیه، p-مقدار فازی، اصل گسترش، فرضیه فازی.

۱ مقدمه

آزمون فرضیه‌های آماری یکی از مهم‌ترین روش‌ها برای تصمیم‌گیری در مسایل علمی و کاربردی است. در علوم کاربردی ممکن است برخلاف روش معمول آزمون فرضیه‌ها که در آن داده‌ها، فرضیه‌ها و پارامترها دقیق هستند، با تعاریف مبهم و مفاهیم فازی مواجه شویم. در چنین شرایطی روش‌های کلاسیک نیاز به تعمیم برای محیط‌های فازی دارند. ورود ابهام در مساله‌ی آزمون فرضیه‌ها می‌تواند از طریق داده‌ها یا فرضیه‌ها صورت گیرد. بنابراین سه مساله‌ی زیر را می‌توان درنظر گرفت:

۱. آزمون فرضیه‌های دقیق براساس داده‌های فازی،

۲. آزمون فرضیه‌های فازی براساس داده‌های دقیق،

۳. آزمون فرضیه‌های فازی براساس داده‌های فازی.

در این مقاله روی سه مساله‌ی بالا بر اساس رویکرد p-مقدار بحث خواهیم کرد. تلاش‌های بسیاری برای تجزیه و تحلیل مساله‌ی آزمون فرضیه‌ها در محیط‌های فازی با استفاده از نظریه‌ی مجموعه فازی انجام شده که در پرچمی و همکاران [۶] آمده است. برای اولین‌بار فیلزموسر و ویرتل [۱] مساله‌ی آزمون فرضیه‌ها و p-مقدار فازی را در آزمون فرضیه‌ها وقتی‌که مشاهدات فازی اما فرضیه‌ها دقیق باشند، معرفی کردند. پرچمی و همکاران [۵، ۶] برای مساله‌ی آزمون فرضیه‌های فازی، p-مقدار فازی را برای داده‌های دقیق و نیز داده‌های فازی بدست آوردند و مفهوم p-مقدار را برای آزمون فرضیه‌های فازی تعمیم دادند.

با توجه به روش‌های گوناگون آزمون فرضیه‌ها، روش p-مقدار به عنوان ساده‌ترین و معمول‌ترین روش آزمون در علوم مختلف شناخته شده است. اما محاسبه‌ی تابع عضویت p-مقدار فازی پیچیده است. لذا تولید نرم‌افزار آماری برای تحقیق در یک محیط فازی می‌تواند تجزیه و تحلیل داده‌های فازی و اطلاعات مبهم را برای کاربران آسان کند. بسته‌ی Fuzzy.p.value یکی از اولین نرم افزارهای آماریست که امکان انجام آزمون فرضیه‌ها را در محیط فازی مبتنی بر رهیافت p-مقدار برای کاربران فراهم می‌سازد.

در بخش ۲ برخی پیش نیازها، مفاهیم و تعاریف اولیه آزمون فرضیه‌ها در محیط فازی بیان می‌شود. سپس نحوه‌ی استفاده از بسته‌ی Fuzzy.p.value در بخش ۳ ذکر شده و در نهایت با چند مثال کاربردی آن را در بخش ۴ توضیح می‌دهیم.

۲ پیش نیازها

۱.۲ تعاریف اولیه

فرض کنید $\mathbb{R}$ مجموعه تمام اعداد حقیقی و

$$\boldsymbol{F}(\mathbb{R}) = \{\tilde{A} | \tilde{A} : \mathbb{R} \to [۰, ۱]\}$$

مجموعه تمام مجموعه‌های فازی روی $\mathbb{R}$ باشد. همچنین فرض کنید $\boldsymbol{F_c}(\mathbb{R})$ مجموعه تمامی مجموعه‌های فازی نرمال، محدب و پیوسته روی $\mathbb{R}$ باشد. تکیه‌گاه $\tilde{A} \in \boldsymbol{F}(\mathbb{R})$ مجموعه غیرفازی $\mathbf{supp}(\tilde{A}) = \{x | \tilde{A}(x) > ۰\}$ و همچنین δ-برش $\tilde{A} \in \boldsymbol{F}(\mathbb{R})$ مجموعه غیرفازی $\tilde{A}_\delta = \{x | \tilde{A}(x) \geq \delta\}$ به‌ازای $\delta \in (۰, ۱]$ است.

تعریف ۱.۲. هر فرضیه به‌صورت « θ تابع H است : $\tilde{H}$» فرضیه فازی نامیده می‌شود به‌طوری که $H : \theta \to (۰, ۱]$، عبارت «θ تابع H است» دلالت بر آن دارد که θ، با درجه $H(\theta)$ متعلق به زیرمجموعه فازی $\tilde{H}$ از فضای پارامتر Θ است. به‌عبارتی هر ادعا در خصوص پارامتر مجهول جامعه که به‌وسیله یک مجموعه فازی بیان گردد، یک فرضیه فازی است.

توجه کنید که فرضیه معمولی « $H_۰ : \theta \in \Theta_۰$» یک فرضیه فازی با تابع عضویت $H_۰ = I_{\Theta_۰}$ است. که در آن $I_{\Theta_۰}$ تابع نشانگر فضای پارامتر می‌باشد.

مسأله اصلی

فرض کنید یک نمونه تصادفی با حجم n از متغیری با تابع چگالی (یا جرم) احتمال $f_\theta(x)$ داشته باشیم که بجای داده‌های دقیق، مشاهدات فازی $\tilde{x}_۱, \ldots, \tilde{x}_n \in \boldsymbol{F_c}(\mathbb{R})$ برای این نمونه

تصادفی گزارش یا ثبت شده باشد. براساس اصل گسترش می‌توان تابع عضویت آماره آزمون $t = h(\tilde{x}_1, \ldots, \tilde{x}_n)$، را بدست آورد که آن‌را با نماد $\mathbf{t}(t)$ نشان می‌دهیم. مساله اصلی مورد بحث در این مقاله آزمون فرضیه‌های فازی:

$$\theta \text{ تابع } H_0 \text{ است } : \tilde{H}_0 \quad \text{درمقابل} \quad \theta \text{ تابع } H_1 \text{ است } : \tilde{H}_1$$

بر اساس مشاهدات فازی $\tilde{x}_1, \ldots, \tilde{x}_n$ است. که آن را به اختصار «آزمون فرضیه در محیط فازی» می‌نامیم. هدف اصلی در این مقاله حل این مساله براساس p-مقدار فازی به‌کمک بسته‌ی نرم‌افزاری Fuzzy.p.value است.

تعریف ۲.۲. سطح معنی‌داری فازی، مجموعه‌ای فازی است که بر بازه $(0, 1)$ تعریف می‌شود. در حالت معمولی می‌توان سطح معنی‌داری غیرفازی را با تابع نشانگر $I_{\{\alpha\}}$ نشان داد به‌طوری که $\alpha \in (0, 1)$.

تعریف ۳.۲. در مساله آزمون فرضیه در محیط فازی $D(\mathrm{P} > \mathrm{S})$ درجه پذیرش فرضیه صفر و $D(\mathrm{S} > \mathrm{P}) = 1 - D(\mathrm{P} > \mathrm{S})$ درجه رد فرضیه صفر نامیده می‌شود که در آن

$$D(A > B) = \frac{\Delta_{AB}}{\Delta_{AB} + \Delta_{BA}}, \qquad A, B \in F_c(\mathbb{R}), \tag{1}$$

$$\Delta_{AB} = \int_{a^+_{A_\delta} > a^-_{B_\delta}} (a^+_{A_\delta} - a^-_{B_\delta}) + \int_{a^-_{A_\delta} > a^+_{B_\delta}} (a^-_{A_\delta} - a^+_{B_\delta}) \tag{2}$$

و همچنین $a^+_{A_\delta} = \sup\{x; x \in A_\delta\}$، $a^-_{A_\delta} = \inf\{x; x \in A_\delta\}$ و $\delta \in (0, 1]$.

در بخش بعدی مقاله، به حل مساله آزمون فرضیه در محیط فازی به‌کمک بسته‌ی نرم‌افزاری Fuzzy.p.value که قابلیت نصب بر روی نرم‌افزار آماری R را دارد، می‌پردازیم.

۳ بسته نرم‌افزاری Fuzzy.p.value

عنوان اصلی پکیج Fuzzy.p.value به‌صورت Copmputing Fuzzy p-value است که نسخه 1.0 روی CRAN بارگزاری شده است. در این بخش پس از بررسی چندین روش برای معرفی مجموعه‌های فازی در بسته Fuzzy.p.value برخی توابع اصلی این بسته با چند مثال ارائه شده است.

۱.۳ معرفی چند مجموعه‌ی فازی در بسته نرم‌افزاری

مجموعه فازی $\boldsymbol{T} \in \boldsymbol{F}(\mathbb{R})$ عدد فازی مثلثی گفته می‌شود که تابع عضویت آن به شکل زیر است:

$$\boldsymbol{T(a,b,c)(x)} = \begin{cases} \frac{x-a}{b-a} & \text{اگر} \quad a < x \leq b \\ \frac{x-c}{b-c} & \text{اگر} \quad b < x \leq c \\ ۰ & \quad \text{o.w.} \end{cases}$$

که در آن $\boldsymbol{a,b,c} \in \mathbb{R}$ و $\boldsymbol{a \leq b \leq c}$. این عدد فازی مثلثی $\boldsymbol{T(a,b,c)}$ به آسانی در بسته Fuzzy.p.value با دستور $\boldsymbol{T(a,b,c)}$ قابل معرفی است. به‌طور مشابه دستور $\boldsymbol{Tr(a,b,c,d)}$ برای معرفی عدد فازی ذوزنقه‌ای $\boldsymbol{Tr(a,b,c,d)}$ با تابع عضویت زیر استفاده می‌شود:

$$\boldsymbol{Tr(a,b,c,d)(x)} = \begin{cases} \frac{x-a}{b-a} & \text{اگر} \quad a < x \leq b \\ ۱ & \text{اگر} \quad b < x \leq c \\ \frac{x-d}{c-d} & \text{اگر} \quad c < x \leq d \\ ۰ & \quad \text{o.w.} \end{cases}$$

علاوه براین ممکن است تصمیم بگیریم که توابع عضویت خطی برای مفاهیم «تقریباً کوچکتر از» و «تقریباً بزرگتر از» در قالب فرضیه‌های فازی درنظر بگیریم. برای این کار از توابع S و B استفاده می‌کنیم که به ترتیب دارای توابع عضویت زیر هستند و در بسته‌ی نرم‌افزاری Fuzzy.p.value با دستورات $S(a,b)$ و $B(a,b)$ به‌سهولت قابل فراخوانی و کاربرد هستند:

$$S(a,b)(x) = \begin{cases} ۱ & x \le a \\ \frac{x-b}{a-b} & a < x \le b \\ ۰ & x > b \end{cases}$$

$$B(a,b)(x) = \begin{cases} ۱ & b \le x \\ \frac{x-a}{b-a} & a \le x < b \\ ۰ & x < a, \end{cases}$$

نکته ۱. اعداد فازی دیگر و پیچیده‌تر مانند اعداد فازی LR، اعداد فازی قطعه قطعه شده‌ی خطی و اعداد فازی توانی نیز با بسته‌ی نرم‌افزاری FuzzyNumbers در دسترس هستند که برای جزئیات بیشتر می‌توان به پرچمی [۲، ۳]، مراجعه کرد.

۲.۳ محاسبه p-مقدار فازی

در بسته‌ی نرم‌افزاری Fuzzy.p.value، محاسبه آزمون فرضیه‌ها در محیط فازی و رسم تابع عضویت p-مقدار فازی را می‌توان به کمک دو تابع اصلی p-value.norm و p-value.pois انجام داد. تابع اول زمانی اعمال می‌شود که آماره آزمون دارای توزیع نرمال باشد و تابع دوم زمانی کاربرد دارد که آماره آزمون از توزیع پواسون تبعیت کند. اگرچه این بسته اکنون تنها محدود به توزیع نرمال و پواسون، به‌عنوان دو نامزد برای کلاس همه‌ی توزیع‌های پیوسته و گسسته است، اما ایده‌ی آن می‌تواند برای سایر توزیع‌های گسسته و پیوسته نیز تعمیم یابد. برای مثال شکل استفاده از تابع

p-value.norm به‌صورت زیر است:

$$\mathbf{p\_value.norm}(\mathbf{kind}, \boldsymbol{H}_{\circ}, \boldsymbol{H}_{1}, \boldsymbol{t}, \mathbf{s.d}, \boldsymbol{n}, \boldsymbol{sig})$$

که در آن $\boldsymbol{H}_{\circ}$ و $\boldsymbol{H}_{1}$ به ترتیب فرضیه‌های صفر و جایگزین، $\boldsymbol{t}$ مقادیر مشاهده شده آماره آزمون (میانگین)، $\mathbf{s.d}$ انحراف معیار توزیع نرمال، $\boldsymbol{n}$ حجم نمونه و $\boldsymbol{sig}$ سطح معنی‌داری در آزمون مورد نظر است. درضمن $\mathbf{kind}$ نوع آزمون فرضیه‌هاست و یکی از اعداد ۰، ۱ و ۲ را می‌تواند اختیار کند. که با توجه به صورت فرضیه‌ی جایگزین قابل تشخیص و تعیین است. اگر فرضیه $\boldsymbol{H}_{1}$ از نوع نامساوی باشد، $\mathbf{kind} = \circ$، اگر فرضیه $\boldsymbol{H}_{1}$ از نوع کوچکتری باشد، $\mathbf{kind} = 1$ و درصورتی که $\boldsymbol{H}_{1}$ از نوع بزرگتری باشد، $\mathbf{kind} = 2$ را درنظر بگیرید.

۴ مثال‌های عددی

مثال ۱.۴. طول عمر لامپ‌های تولیدی یک کارخانه (بر حسب ساعت) دارای توزیع نرمال با میانگین مجهول $\boldsymbol{\mu}$ و انحراف معیار $\boldsymbol{\sigma} = 120$ است. برپایه مشاهدات نمونه‌ای تصادفی به حجم ۳۶ لامپ مقدار $\bar{\boldsymbol{x}} = 1327$ ثبت شده است. قصد داریم فرضیه‌های فازی زیر را در سطح معنی‌داری $\boldsymbol{S} = \boldsymbol{T}(\circ, \circ/15, \circ/3)$ آزمون می‌کنیم

$$\begin{cases} \tilde{\boldsymbol{H}}_{\circ} : \text{است ۱۳۰۰ تقریبا } \boldsymbol{\mu} \\ \tilde{\boldsymbol{H}}_{1} : \text{است ۱۳۰۰ تقریبابزرگتر از } \boldsymbol{\mu} \end{cases}$$

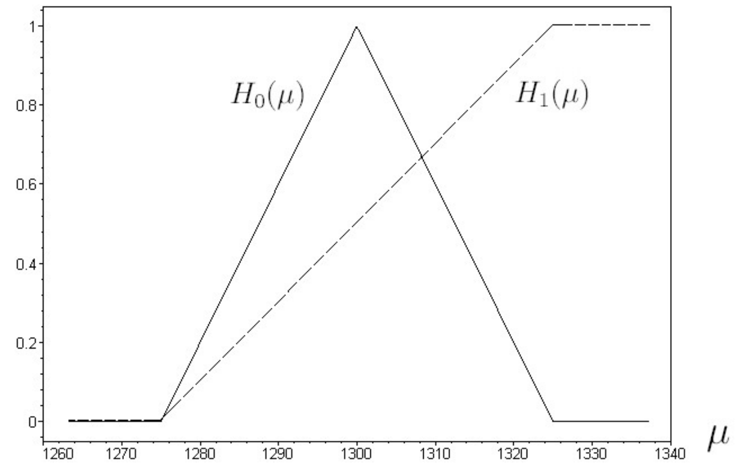
که در آن‌ها $\tilde{\boldsymbol{H}}_{\circ}$ و $\tilde{\boldsymbol{H}}_{1}$ به‌ترتیب توابع عضویت

$$\boldsymbol{T}(1275, 1300, 1325) \in \boldsymbol{F_T}(\mathbb{R})$$

و

$$\boldsymbol{B}(1275, 1325) \in \boldsymbol{F_B}(\mathbb{R})$$

صورت‌بندی شده است (شکل ۱).

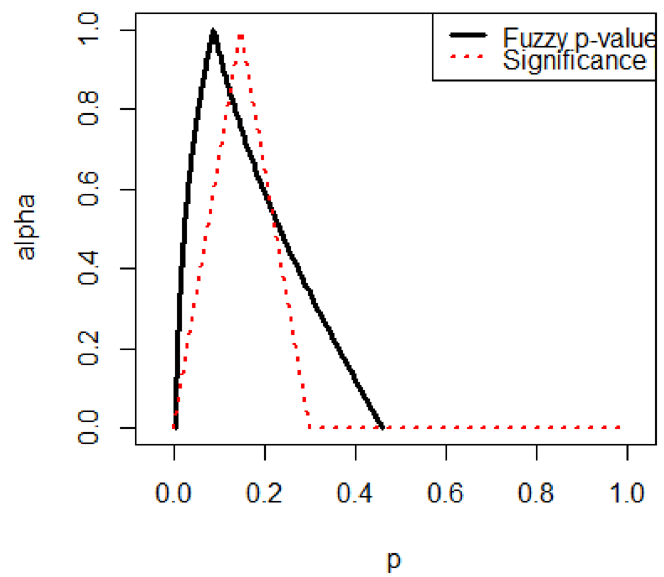


شکل ۱: توابع عضویت فرضیه‌های فازی در مثال ۱.۴

باتوجه به نوع فرضیه $\tilde{H}_۱$، ناحیه بحرانی آزمون از نوع بزرگتری بوده و لذا $\mathbf{kind} = ۲$ و آماره آزمون $\frac{\sum_{i=۱}^{۳۶} X_i}{۳۶} = \bar{X} = T(X_۱, \ldots, X_n)$ است. پس از نصب و فراخوانی بسته‌ی نرم‌افزاری Fuzzy.p.value، برای رسم تابع عضویت p-مقدار فازی و نیز انجام آزمون فرضیه‌های فازی در سطح معنی‌داری $S = T(۰, ۰/۱۵, ۰/۳)$ از کدهای زیر در نرم‌افزار R استفاده می‌کنیم:

```
H0 = T(1275, 1300, 1325)
H1 = B(1275, 1325)
t = T(1327, 1327, 1327) # i.e.   crisp t = 1327
sig = T(0.0, 0.15, 0.3)
p_value.norm(kind = 2, H0, H1, t, s.d = 120, n = 36, sig)
```

در نهایت، پس از محاسبه $\Delta_{\mathbf{PS}} = ۰/۱۷$ و $\Delta_{\mathbf{SP}} = ۰/۱۹$ به‌وسیله بسته Fuzzy.p.value، فرضیه فازی $H_۱$، درسطح معنی‌داری $\mathbf{S}$ با درجه $D(\mathbf{S} > \mathbf{P}) = ۰/۵۳$، پذیرفته و شکل ۲ رسم می‌شود.

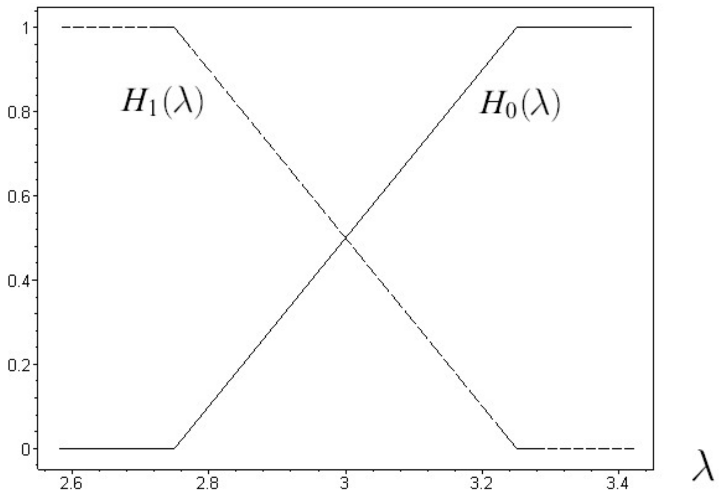


شکل ۲: توابع عضویت p-مقدار فازی و سطح معنی‌داری فازی در مثال ۱.۴

مثال ۲.۴. مدیر یک کارخانه جهت ارتقای سطح ایمنی کارگران، دستگاه جدیدی نصب و راه‌اندازی کرده است. فرض کنید تعداد حوادثی که در هر ماه در این کارخانه رخ می‌دهد دارای توزیع پوآسون با میانگین مجهول λ است. به‌منظور بررسی کاهش احتمالی تعداد حوادث، مدیریت کارخانه قصد دارد فرضیه‌های

$$\begin{cases} \tilde{H}_0 : \text{است ۳ از تقریبابیشتر } \lambda \\ \tilde{H}_1 : \text{است ۳ از تقریبا کمتر } \lambda \end{cases}$$

را آزمون کند که در آن‌ها $\tilde{H}_0$ و $\tilde{H}_1$ به‌ترتیب با توابع عضویت $B(۲/۷۵, ۳/۲۵) \in F_B(\mathbb{R})$ و $S(۲/۷۵, ۳/۲۵) \in F_S(\mathbb{R})$ صورت‌بندی شده‌اند (شکل ۳).

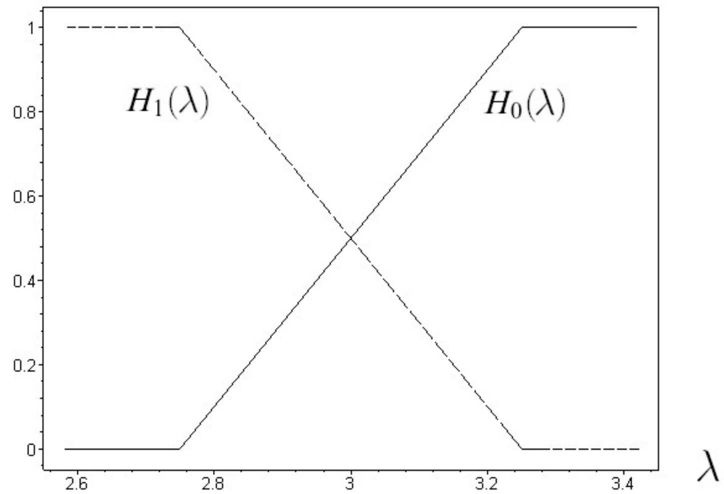


شکل ۳: توابع عضویت فرضیه‌های فازی در مثال ۲.۴

اگر متغیرهای تصادفی $X_{۱},\ldots,X_{۱۲}$ تعداد حوادث رخ داده در ماه‌های سال باشند، آن‌گاه آماره $T=\sum_{i=۱}^{۱۲} X_i$ دارای توزیع پوآسون با پارامتر مجهول ۱۲λ است. براساس اطلاعات ثبت شده در یک سال گذشته و باتوجه به تعریف مبهمی که برای رخداد یک حادثه در این کارخانه درنظر گرفته شده است، تعداد تقریباً ۲۷ حادثه در این کارخانه، پس از نصب دستگاه جدید، با تابع عضویت $T(۲۴,۲۷,۳۰)$ رخ‌داده است. باتوجه به فرم فرضیه جایگزین $\mathbf{kind}=۱$ درنظر می‌گیریم و برای آزمون فرضیه‌های فازی بالا در سطح معنی‌داری $S=T(۰,۰/۰۵,۰/۱)$ تنها کافیست کدهای زیر را در R استفاده کنیم (شکل ۴ را ببینید).

```
H0 = B(12 × 2.75, 12 × 3.25)
H1 = S(12 × 2.75, 12 × 3.25)
t = T(24, 27, 30)
sig = T(0, 0.05, 0.1)
p_value.pois(kind = 1, H0, H1, t, n = 12, sig)
```

درنهایت، پس از محاسبه $\Delta_{\mathbf{PS}}=۰/۰۵۹$ و $\Delta_{\mathbf{SP}}=۰/۰۲۲$ به‌وسیله‌ی بسته Fuzzy.p.value، فرضیه فازی $H_۰$، در سطح معنی‌داری $\mathbf{S}$ با درجه $D(\mathbf{P}>\mathbf{S})=۰/۷۲$



شکل ۴: توابع عضویت p-مقدار فازی و سطح معنی‌داری فازی در مثال ۲.۴

پذیرفته و شکل ۴ رسم می‌شود.

۵ نتیجه

در حل مساله‌ی آزمون فرضیه‌ها با داده‌ها یا فرضیه‌های فازی با رویکرد p-مقدار، ابهام فرضیه‌ها یا مشاهدات به p-مقدار منتقل می‌شود. به‌منظور پرهیز از دشواری‌های محاسباتی در رویکرد p-مقدار فازی، پس از نصب بسته نرم‌افزاری Fuzzy.p.value بر روی نرم‌افزار R، کاربران می‌توانند از این بسته برای رسم توابع عضویت p-مقدار فازی و سطح معنی‌داری فازی و نیز تصمیم‌گیری و انجام آزمون فرضیه‌ها (در محیط نایقینی) استفاده نمایند.

مراجع

[1] P. Filzmoser and R. Viertl, Testing hypotheses with fuzzy data: the fuzzy p-value, *Metrika.*, **59** (2004), 21-29.

[2] A. Parchami, Calculator for Fuzzy Numbers, *Complex & Intelligent Systems*.,**5(3)**, (2019), 331-342.

[3] A. Parchami, FuzzyNumbers.Ext.2: Apply Two Fuzzy Numberson a Monotone Function. R package version 3.2. https://CRAN.Rproject.org/package=FuzzyNumbers.Ext.2, (2017).

[4] A. Parchami, Fuzzy.p.value: Computing Fuzzy p-Value. R package version 1.0. https://CRAN.R-project.org/package=Fuzzy.p.value, (2016).

[5] A. Parchami, SM. Taheri and M. Mashinchi, Fuzzy p-value in testing fuzzy hypotheses with crisp data, *Statistical Papers*., **51** (2010), 209-226.

[6] A. Parchami, SM. Taheri and M. Mashinchi, Testing fuzzy hypotheses based on vague observations: a p-value approach, *Statistical Papers*., **53** (2012), 469-484.