

## محاسبه تابع عضویت $p$ -مقدار فازی در نرم‌افزار R

محبوبه سادات مدنی، محمدرضا ربیعی، عباس پرچی

دانشگاه صنعتی شاهرود، دانشکده ریاضی، بخش آمار

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

تاریخ پذیرش: ۱۴۰۰/۲/۶

تاریخ دریافت: ۱۳۹۹/۱۱/۱۴

نوع مقاله: علمی-پژوهشی

### 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 در نظر 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 در نظر 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.

۱۳۹۹ (انجمن سیستم‌های فازی ایران)

## 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 در نظر 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 اولین بار فیلیزموسر و ویرتل [۱] مساله‌ی آزمون فرضیه‌ها و  $p$ -مقدار فازی را در آزمون فرضیه‌ها وقتی که 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

$$F(\mathbb{R}) = \{\tilde{A} | \tilde{A} : \mathbb{R} \rightarrow [0, 1]\}$$

مجموعه تمام مجموعه‌های فازی روی  $\mathbb{R}$  باشد. همچنین فرض کنید  $F_c(\mathbb{R})$  مجموعه تمامی مجموعه‌های فازی نرمال، محدب و پیوسته روی  $\mathbb{R}$  باشد. تکیه‌گاه  $\tilde{A} \in F(\mathbb{R})$  مجموعه غیرفازی  $\text{supp}(\tilde{A}) = \{x | \tilde{A}(x) > 0\}$  و همچنین  $\delta$ -برش  $\tilde{A} \in F(\mathbb{R})$  مجموعه غیرفازی  $\tilde{A}_\delta = \{x | \tilde{A}(x) \geq \delta\}$  به ازای  $\delta \in (0, 1]$  است.

**تعریف ۱.۲.** هر فرضیه به صورت « $\theta$  تابع  $H$  است :  $\tilde{H}$ » فرضیه فازی نامیده می‌شود به طوری که  $\tilde{H} : \theta \rightarrow (0, 1]$ ، عبارت « $\theta$  تابع  $H$  است» دلالت بر آن دارد که  $\theta$ ، با درجه  $H(\theta)$  متعلق به زیرمجموعه فازی  $\tilde{H}$  از فضای پارامتر  $\Theta$  است. به عبارتی هر ادعا در خصوص پارامتر مجهول جامعه که به وسیله یک مجموعه فازی بیان گردد، یک فرضیه فازی است.

توجه کنید که فرضیه معمولی « $H : \theta \in \Theta$ » یک فرضیه فازی با تابع عضویت  $H_\theta = I_\theta$  است. که در آن  $I_\theta$  تابع نشانگر فضای پارامتر می‌باشد.

### مسئله 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  $\tilde{x}_1, \dots, \tilde{x}_n \in F_c(\mathbb{R})$  برای این نمونه

تصادفی گزارش یا ثبت شده باشد. براساس اصل گسترش می توان تابع عضویت آماره آزمون  $t = h(\tilde{x}_1, \dots, \tilde{x}_n)$ ، را بدست آورد که آن را با نماد  $t(t)$  نشان می دهیم. مساله اصلی مورد بحث در این مقاله آزمون فرضیه های فازی:

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

بر اساس مشاهدات فازی  $\tilde{x}_1, \dots, \tilde{x}_n$  است. که آن را به اختصار «آزمون فرضیه در محیط فازی» می نامیم. هدف اصلی در این مقاله حل این مساله براساس  $p$ -مقدار فازی به کمک بسته ی نرم افزاری Fuzzy.p.value است.

**تعریف ۲.۲.** سطح معنی داری فازی، مجموعه ای فازی است که بر بازه  $(0, 1)$  تعریف می شود. در حالت معمولی می توان سطح معنی داری غیر فازی را با تابع نشانگر  $I_{\{\alpha\}}$  نشان داد به طوری که  $\alpha \in (0, 1)$ .

**تعریف ۳.۲.** در مساله آزمون فرضیه در محیط فازی  $D(P > S)$  درجه پذیرش فرضیه صفر و  $D(S > P) = 1 - D(P > S)$  درجه رد فرضیه صفر نامیده می شود که در آن

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

$$\Delta_{AB} = \int_{a_{A\delta}^+ > a_{B\delta}^-} (a_{A\delta}^+ - a_{B\delta}^-) + \int_{a_{A\delta}^- > a_{B\delta}^+} (a_{A\delta}^- - a_{B\delta}^+) \quad (2)$$

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

در بخش بعدی مقاله، به حل مساله آزمون فرضیه در محیط فازی به کمک بسته ی نرم افزاری Fuzzy.p.value که قابلیت نصب بر روی نرم افزار آماری R را دارد، می پردازیم.

### ۳ بسته نرم‌افزاری Fuzzy.p.value

عنوان اصلی پکیج Fuzzy.p.value به صورت 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  $T \in F(\mathbb{R})$  عدد فازی مثلثی گفته می‌شود که تابع عضویت آن به شکل زیر است:

$$T(a, b, c)(x) = \begin{cases} \frac{x-a}{b-a} & \text{اگر } a < x \leq b \\ \frac{x-c}{b-c} & \text{اگر } b < x \leq c \\ 0 & \text{o.w.} \end{cases}$$

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

$$Tr(a, b, c, d)(x) = \begin{cases} \frac{x-a}{b-a} & \text{اگر } a < x \leq b \\ 1 & \text{اگر } b < x \leq c \\ \frac{x-d}{c-d} & \text{اگر } c < x \leq d \\ 0 & \text{o.w.} \end{cases}$$

علاوه بر این ممکن است تصمیم بگیریم که توابع عضویت خطی برای مفاهیم «تقریباً کوچکتر از» و «تقریباً بزرگتر از» در قالب فرضیه‌های فازی در نظر 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  $B(a, b)$  به سهولت قابل فراخوانی و کاربرد هستند:

$$S(a, b)(x) = \begin{cases} 1 & x \leq a \\ \frac{x-b}{a-b} & a < x \leq b \\ 0 & x > b \end{cases}$$

$$B(a, b)(x) = \begin{cases} 1 & b \leq x \\ \frac{x-a}{b-a} & a \leq x < b \\ 0 & x < a, \end{cases}$$

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:

$$p\_value.norm(kind, H_0, H_1, t, s.d, n, sig)$$

که در آن  $H_0$  و  $H_1$  به ترتیب فرضیه‌های صفر و جایگزین،  $t$  مقادیر مشاهده شده آماره آزمون (میانگین)، s.d. انحراف معیار توزیع نرمال،  $n$  حجم نمونه و  $sig$  سطح معنی‌داری در آزمون مورد نظر است. در ضمن 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  $H_1$  از نوع نامساوی باشد،  $kind = 0$ ، اگر فرضیه  $H_1$  از نوع کوچکتی باشد،  $kind = 1$  و در صورتی که  $H_1$  از نوع بزرگتری باشد،  $kind = 2$  را در نظر بگیرید.

## ۴ مثال‌های عددی

مثال ۰.۱۰۴. طول عمر لامپ‌های تولیدی یک کارخانه (بر حسب ساعت) دارای توزیع نرمال با میانگین مجهول  $\mu$  و انحراف معیار  $\sigma = 120$  است. برپایه مشاهدات نمونه‌ای تصادفی به حجم ۳۶ لامپ مقدار  $\bar{x} = 1327$  ثبت شده است. قصد داریم فرضیه‌های فازی زیر را در سطح معنی‌داری  $S = T(0, 0.15, 0.3)$  آزمون می‌کنیم

$$\begin{cases} \tilde{H}_0: \mu \text{ تقریباً } 1300 \text{ است} \\ \tilde{H}_1: \mu \text{ تقریباً بزرگتر از } 1300 \text{ است} \end{cases}$$

که در آن‌ها  $\tilde{H}_0$  و  $\tilde{H}_1$  به ترتیب توابع عضویت

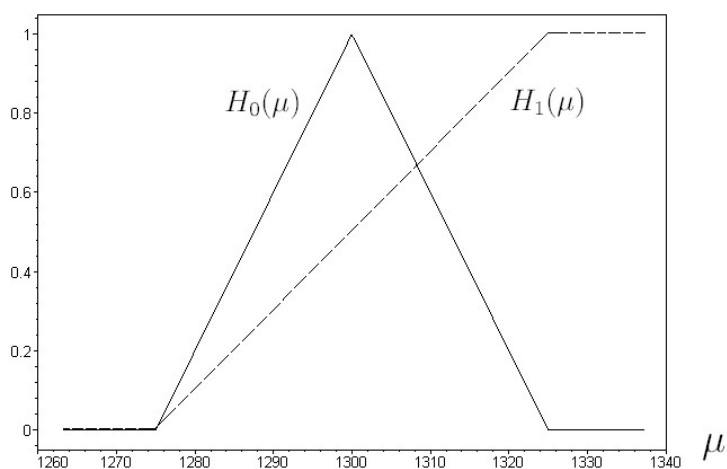
$$T(1275, 1300, 1325) \in F_T(\mathbb{R})$$

و

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

صورت‌بندی شده است (شکل ۱).

باتوجه به نوع فرضیه  $\tilde{H}_1$ ، ناحیه بحرانی آزمون از نوع بزرگتری بوده و لذا  $kind = 2$  و آماره آزمون  $T(X_1, \dots, X_n) = \bar{X} = \frac{\sum_{i=1}^n X_i}{36}$  پس از نصب و فراخوانی بسته‌ی



شکل ۱: توابع عضویت فرضیه‌های فازی در مثال ۱.۴

نرم‌افزاری Fuzzy.p.value، برای رسم تابع عضویت  $p$ -مقدار فازی و نیز انجام آزمون فرضیه‌های فازی در سطح معنی‌داری  $S = T(0, 0.15, 0.3)$  از کدهای زیر در نرم‌افزار R استفاده می‌کنیم:

$$H_0 = T(1275, 1300, 1325)$$

$$H_1 = B(1275, 1325)$$

$$t = T(1327, 1327, 1327) \# \text{i.e. crisp } t = 1327$$

$$\text{sig} = T(0.0, 0.15, 0.3)$$

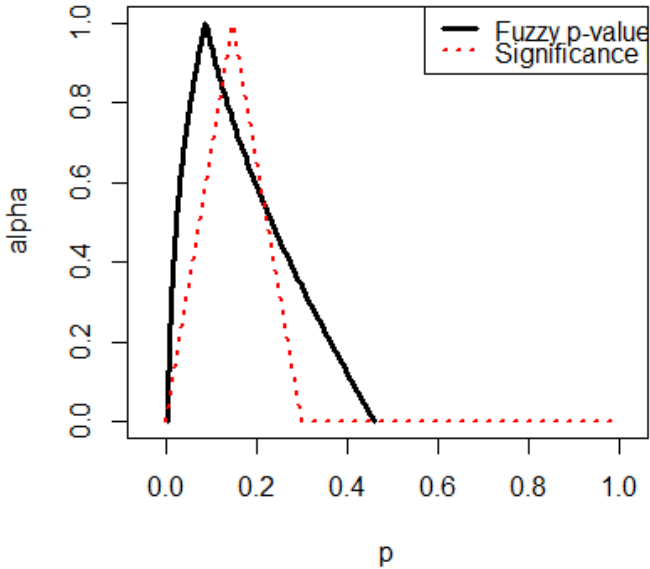
$$\text{p\_value.norm}(\text{kind} = 2, H_0, H_1, t, \text{s.d} = 120, n = 36, \text{sig})$$

در نهایت، پس از محاسبه  $\Delta_{PS} = 0.17$  و  $\Delta_{SP} = 0.19$  به وسیله بسته Fuzzy.p.value،

فرضیه فازی  $H_1$ ، در سطح معنی‌داری  $S$  با درجه  $D(S > P) = 0.53$  پذیرفته و شکل ۲ رسم می‌شود.

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 پواسون با میانگین مجهول  $\lambda$  است. به منظور بررسی کاهش احتمالی تعداد حوادث، مدیریت





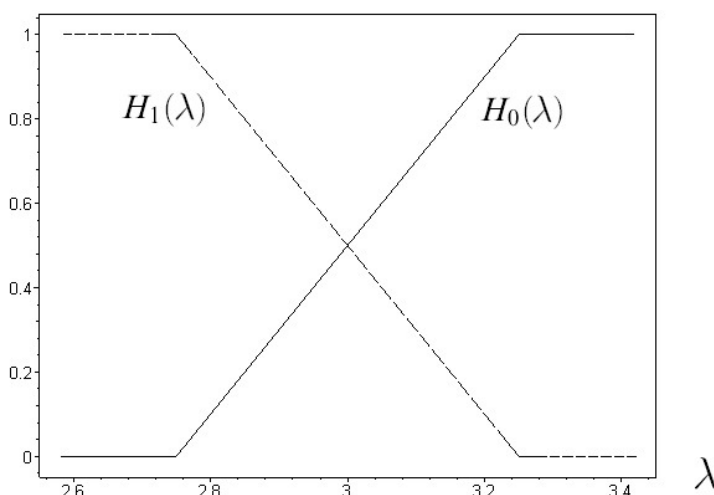
شکل ۲: توابع عضویت  $p$ -مقدار فازی و سطح معنی داری فازی در مثال ۱.۴

کارخانه قصد دارد فرضیه‌های

$$\begin{cases} \tilde{H}_0: \lambda \text{ تقریباً بیشتر از } 3 \text{ است} \\ \tilde{H}_1: \lambda \text{ تقریباً کمتر از } 3 \text{ است} \end{cases}$$

را آزمون کنند که در آن‌ها  $\tilde{H}_0$  و  $\tilde{H}_1$  به ترتیب با توابع عضویت  $B(275, 325) \in F_B(\mathbb{R})$  و  $S(275, 325) \in F_S(\mathbb{R})$  صورت‌بندی شده‌اند (شکل ۳).

اگر متغیرهای تصادفی  $X_1, \dots, X_{12}$  تعداد حوادث رخ داده در ماه‌های سال باشند، آنگاه آماره  $T = \sum_{i=1}^{12} X_i$  دارای توزیع پواسون با پارامتر مجهول ۱۲۸ است. براساس اطلاعات ثبت شده در یک سال گذشته و باتوجه به تعریف مبهمی که برای رخداد یک حادثه در این کارخانه



شکل ۳: توابع عضویت فرضیه‌های فازی در مثال ۲.۴

در نظر گرفته شده است، تعداد تقریباً ۲۷ حادثه در این کارخانه، پس از نصب دستگاه جدید، با تابع عضویت  $T(24, 27, 30)$  رخ داده است. با توجه به فرم فرضیه جایگزین  $kind = 1$  در نظر می‌گیریم و برای آزمون فرضیه‌های فازی بالا در سطح معنی‌داری  $S = T(0, 0.05, 0.1)$  تنها کافیست کدهای زیر را در R استفاده کنیم (شکل ۴ را ببینید).

$$H_0 = B(12 \times 2.75, 12 \times 3.25)$$

$$H_1 = S(12 \times 2.75, 12 \times 3.25)$$

$$t = T(24, 27, 30)$$

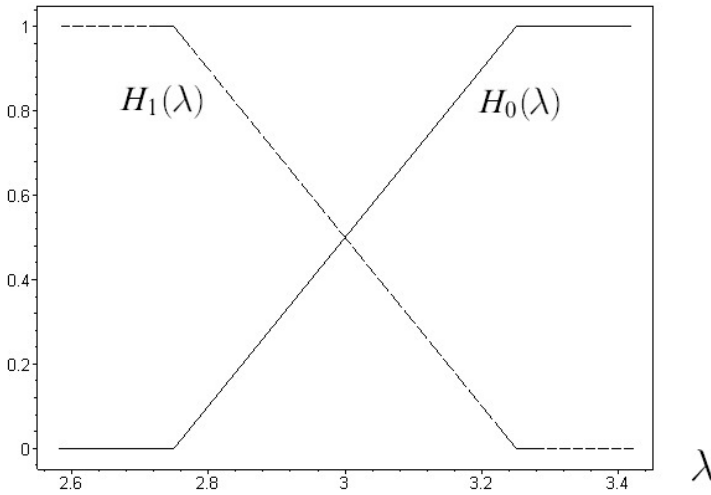
$$sig = T(0, 0.05, 0.1)$$

$$p\_value.pois(kind = 1, H_0, H_1, t, n = 12, sig)$$

در نهایت، پس از محاسبه  $\Delta_{PS} = 0.59$  و  $\Delta_{SP} = 0.22$  به وسیله‌ی بسته

$D(P > S) = Fuzzy.p.value$ ، فرضیه فازی  $H_0$ ، در سطح معنی‌داری  $S$  با درجه  $0.72$

پذیرفته و شکل ۴ رسم می‌شود.



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

## ۵ نتیجه

در حل مسأله‌ی آزمون فرضیه‌ها با داده‌ها یا فرضیه‌های فازی با رویکرد  $p$ -مقدار، ابهام فرضیه‌ها یا مشاهدات به  $p$ -مقدار منتقل می‌شود. به منظور 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 Numbers on 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.