سیزدهمین کنفرانس
آمار ایران
۲-۴ شهریور ۱۳۹۵

آمار توصیفی بر اساس اطلاعات مستخرج از یک جامعه فازی

عباس پرچمی [۱]، سید محمود طاهری [۲]

[۱] دانشگاه شهید باهنر کرمان، دانشکده ریاضی و کامپیوتر، بخش آمار

[۲] دانشگاه تهران، دانشکده فنی

چکیده: در بسیاری از مسائل کاربردی، هدف، بررسی یک جامعه‌ی فازی (مانند جامعه‌ی جوانان و یا جامعه‌ی مستمندان یک شهر) بر اساس نمونه‌ای از آن جامعه است. رویکرد متداول آمار کلاسیک در مواجهه با این مسائل، صورت‌بندی جامعه‌ی فازی موردنظر بوسیله‌ی زیرمجموعه‌ای از جامعه‌ی آماری است (مثلاً در آمار کلاسیک افرادی را که سن آنها بین ۱۸ تا ۳۰ سال است به عنوان جامعه جوانان در نظر می‌گیرند). ضمن انتقاد به این رویکرد، در این مقاله روشی جدید و منصفانه‌تر برای بررسی چنین مسائل کاربردی پیشنهاد شده است. در این روش، ابتدا جامعه‌ی فازی مورد نظر به کمک یک تابع عضویت مشخص می‌شود. سپس، درجات عضویت تک‌تک داده‌های نمونه به جامعه‌ی فازی در بررسی جامعه‌ی فازی مشارکت داده می‌شوند. برای نمونه در محاسبه‌ی میانگین قد جوانان، منصفانه‌تر آن است که افراد جوان‌تر تاثیرگذاری بیشتری نسبت به سایر افراد داشته باشند. در این مقاله مفاهیم اصلی آمار توصیفی کلاسیک، شامل میانگین، واریانس، انحراف معیار، کوواریانس و ضریب همبستگی برای چنین مسائلی تعمیم و مورد بازنگری قرار گرفته‌اند. در این تعمیم، درجه‌ی عضویت هر یک از داده‌ها به جامعه‌ی فازی موردنظر، میزان تاثیرگذاری آنها را در محاسبات آماری تعیین می‌کند. برای تشریح بهتر، مطالب همراه با چند مثال عددی مطرح شده است.

واژه‌های کلیدی: جامعه فازی، تابع عضویت، شاخص‌های مرکزی، شاخص‌های پراکندگی.

کد موضوع‌بندی ریاضی (۲۰۱۰): 62A86, 26E50.

[۱] عباس پرچمی ؛ parchami@uk.ac.ir

۱ انواع جامعه‌ها و نمونه‌های آماری

در بررسی‌های آماری، با دو مجموعه‌ی اساسی/کلیدی روبرو هستیم: مجموعه (جامعه‌ی) هدف و (مجموعه‌ی) نمونه. در آمار کلاسیک فرض می‌شود که هر دوی این مجموعه‌ها (یعنی جامعه و نمونه) مشخص و دقیق هستند، بدین معنی که عضویت یا عدم عضویتِ هر فرد/شی در جامعه و نمونه، واضح و روشن است. ولی در عمل با دو وضعیت زیر نیز روبرو هستیم:

۱) مواردی که اعضای نمونه، مشخص و دقیق نیستند (و یا مشاهدات نمونه، دقیق گزارش نمی‌شوند). برای مثال هنگامی که می‌خواهیم شوری خاک را در نقاط مختلف یک منطقه اندازگیری کنیم، ممکن است نتایج حاصله همراه با ابهام باشد (محمدی و طاهری (۲۰۰۴)). یا هنگامی که می‌خواهیم میزان درد را در نمونه‌ای از افراد بیمار بررسی کنیم، پاسخ‌های آنها معمولاً به صورت «درد زیاد»، «درد متوسط»، «درد قابل تحمل»، «درد کم» و ... است (نامداری و همکاران (۲۰۱۴)).

۲) مواردی که جامعه‌ی هدف به طور دقیق و واضح تعریف نشده است، بلکه با یک جامعه‌ی هدف نادقیق روبرو هستیم. برای مثال اگر بخواهیم «بررسی میزان رطوبت هوا در روزهای بارانی» انجام دهیم، جامعه‌ی هدف «روزهای بارانی» است. در اینجا بدیهی است که روزهای مختلف با مقادیر مختلفِ بارندگی به یک اندازه مد نظر نیستند بلکه روزهای با بارندگی بیشتر، عضویت بالاتری در مجموعه‌ی «روزهای بارانی» دارند و بالعکس. به عنوان مثالی دیگر، اگر هدف بررسی وضع درآمد «جوانان کرمانی» باشد، آنگاه جامعه‌ی هدف یعنی «جوانان کرمانی» یک مجموعه‌ی دقیق و خوش تعریف نیست.

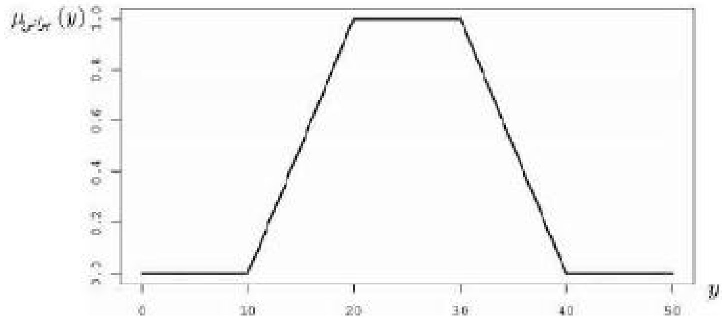
با توجه به دو حالت بالا، یعنی حالتی که مشاهدات نمونه کمیت‌های فازی هستند و حالتی که جامعه‌ی هدف یک جامعه‌ی فازی است، می‌توانیم در مجموع چهار حالت زیر را برای وضعیت یک بررسی آماری درنظر بگیریم:

حالت ۱ (جامعه دقیق - مشاهدات نمونه دقیق): برای مثال، ممکن است بخواهیم متوسط قند خون زنان ۱۵–۳۵ سال بوشهری را برآورد کنیم. در این حالت جامعه‌ی زنان ۱۵–۳۵ سال بوشهری کاملاً مشخص است و لذا پس از اخذ یک نمونه‌ی تصادفی از این جامعه می‌توان متوسط قند خون جامعه‌ی هدف را بر اساس مشاهدات نمونه‌ای دقیق، برآورد کرد.

حالت ۲ (جامعه دقیق - مشاهدات نمونه نادقیق): برای مثال، ممکن است قصد برآورد متوسط درآمد ماهانه تاکسی‌رانان بوشهری را داشته باشیم. گر چه در این حالت نیز جامعه‌ی تاکسی‌رانان بوشهری مشخص است اما واضح است که درآمد ماهانه‌ی یک تاکسی‌ران، معمولاً به صورت تقریبی بیان می‌شود (مثلاً به صورت ۲ میلیون تومان، حدود ۲ تا ۲/۵ میلیون تومان). لذا در این حالت معمولاً داده‌ها به صورت اعداد مثلثی فازی یا اعداد ذوزنقه‌ای فازی ثبت می‌شوند.

حالت ۳ (جامعه نادقیق - مشاهدات نمونه دقیق): فرض کنید می‌خواهیم متوسط درآمد کارمندان جوان بوشهری را محاسبه کنیم. از آنجا که «جوان» یک مفهوم فازی و نادقیق است لذا افراد جامعه‌ی جوانان بوشهری به طور دقیق مشخص نیست. در واقع هر بوشهری با درجه‌ای بین صفر و یک متعلق به مجموعه‌ی فازی جوانان بوشهری می‌باشد. در این حالت، ابتدا جامعه‌ی فازی جوانان بوشهری را به کمک یک تابع عضویت، مثلاً تابع عضویت زیر تعیین و تعریف می‌کنیم (شکل ۱ را ببینید)

$$\mu_{\text{جوان}}(y) = \begin{cases} \frac{y-10}{10} & \text{اگر} \quad 10 \leq y < 20 \\ 1 & \text{اگر} \quad 20 \leq y < 30 \\ \frac{40-y}{10} & \text{اگر} \quad 30 \leq y < 40 \\ 0 & \text{جاهای دیگر} \end{cases} \qquad (1.1)$$



شکل ۱: نمودار تابع عضویت جوانی

بعد از اخذ نمونه‌ی تصادفی از افراد این شهر (به حجم n)، اطلاعات مربوط به این نمونه را به صورت $(x_1, \mu_1), \ldots, (x_n, \mu_n)$ ثبت می‌کنیم که در آن x_i درآمد فرد iام و μ_i درجه‌ی عضویت فرد iام در مجموعه‌ی فازی جوانان بوشهری است. اما میانگین این نمونه‌ی تصادفی و سایر شاخص‌های آماری چگونه محاسبه می‌گردند؟ در این مقاله، چنین داده‌هایی را «داده‌های مستخرج از یک جامعه‌ی فازی» می‌نامیم و در بخش‌های دیگر مقاله درباره‌ی روش‌های آمار توصیفی آنها بحث خواهیم کرد.

حالت ۴ (جامعه نادقیق - مشاهدات نمونه نادقیق): حالت پیچیده‌تری نسبت به سه حالت قبل وجود دارد. فرض کنید می‌خواهیم متوسط درآمد تاکسی‌رانان جوان بوشهری را محاسبه کنیم. در این حالت، درآمد هر راننده‌ی تاکسی یک عدد فازی و مبهم است و از سوی دیگر جامعه‌ی تاکسی‌رانان جوان بوشهری که از آن نمونه‌گیری می‌کنیم نیز یک مجموعه فازی است.

می‌توان محاسبات مربوط به حالت ۱ را بر پایه‌ی آمار کلاسیک انجام داد. از طرفی محاسبات مربوط به حالت ۲، با استفاده از حساب اعداد فازی و یا با استفاده از روش‌های فازی‌زدایی انجام می‌شود (فیتل (۲۰۱۱) و طاهری (۱۳۸۵) را ببینید). اما تا کنون روشی برای انجام محاسبات حالت‌های ۳ و ۴ که در آنها داده‌ها از جوامعی مبهم و فازی گرفته شده‌اند مطرح نشده است و لذا روش‌های آماری برای داده‌های مستخرج از جوامع فازی نیاز به تعمیم دارند. در این مقاله قصد داریم تا به بررسی و محاسبه برخی از شاخص‌های آماری بر اساس داده‌های مستخرج از یک جامعه‌ی فازی، که در حالت ۳ مطرح شد بپردازیم. از ترکیب ایده‌ی مطرح شده برای حالت ۳ با راهکارهای موجود برای حالت ۲، شاید بتوان روش‌هایی نوین برای صورت‌بندی و محاسبات آماری حالت ۴ در آینده مطرح نمود.

در ادامه، در بخش ۲ برخی از شاخص‌های مرکزی و پراکندگی بر اساس داده‌های دقیق مستخرج از جامعه‌ی فازی تعریف شده است. کوواریانس و ضریب همبستگی برای این نوع داده‌ها در بخش ۳ معرفی شده‌اند. بخش ۴ مقاله شامل نتیجه‌گیری و اشاره‌ای به برخی از کارهای مرتبط است.

۲ شاخص‌های تمرکز و پراکندگی بر اساس داده‌های دقیق از جامعه فازی

تحلیل بسیاری از مسائل کاربردی مبتنی بر سرشماری/نمونه‌گیری از یک جامعه‌ی فازی می‌باشد که عملاً پس از مدل‌بندی جامعه‌ی فازی به کمک یک رده یا یک بازه‌ی غیرفازی، جمع‌آوری داده‌ها از آن بازه امکان‌پذیر است. برای نمونه در ادامه به چند مورد از این نوع مسائل، که طیف وسیعی از مسائل کاربردی را در علوم مختلف به خود اختصاص می‌دهند اشاره می‌کنیم. در حوزه پزشکی: بررسی پوکی استخوان در زنان مسن، بررسی کلسترول (چربی خون) میانسالان، تحقیق آماری در مورد فشار خون افراد چاق، بررسی و کنترل زردی در نوزادان کم‌وزن؛ در حوزه کشاورزی: بررسی گیاهان مقاوم به تنش شوری/خشکی، میزان رشد گیاهان شوررپسند در یک اقلیم خاص، میزان

محصول درختان کهنسال یک باغ؛ در حوزه صنایع؛ بررسی کالاهای باکیفیت یک خط تولید؛ در حوزه هواشناسی؛ بررسی آلودگی ناشی از ماشین‌های دودزا، میانگین تراکم ابرهای باران‌زا، متوسط دمای روزهای برفی در یک شهر؛ در حوزه اقتصاد: بررسی الگوی مصرف در میان خانوارهای با درآمد متوسط.

همان گونه که مطرح شد، در آمار کلاسیک تمامی مسائل فوق با در نظر گرفتن یک مجموعه/رده/بازه‌ی غیرفازی برای جامعه‌ی فازی بررسی می‌شوند. اما مناسب‌تر و طبیعی‌تر این است که ابتدا جامعه‌ی فازی مورد نظر را بوسیله‌ی یک تابع عضویت به طور دقیق مدل‌سازی کنیم. در این صورت می‌توان تک‌تک درجات عضویت داده‌ها را به این جامعه‌ی فازی مشخص نمود. در چنین حالتی منطقی است که روش‌های آمار کلاسیک را بگونه‌ای مورد بازنگری و تعمیم قرار دهیم که نقش و میزان تاثیرگذاری تک‌تک داده‌ها در محاسبات آماری بر اساس میزان عضویت آن‌ها به جامعه‌ی فازی مشخص و تعیین گردد. به عبارت دیگر، تعمیم مورد نظر باید بگونه‌ای صورت گیرد که داده‌های با درجات عضویتِ بزرگتر، نقشی پررنگ‌تر و تاثیرگذارتر در محاسبات، نسبت به داده‌های با درجات عضویتِ کوچک‌تر داشته باشند.

برای پرداختن به این موضوع، فرض کنید بر اساس نمونه‌ای تصادفی از جامعه‌ای فازی، داده‌های $(\underline{x},\underline{\mu}) = ((x_1,\mu_1),\ldots,(x_n,\mu_n))$ ثبت شده‌اند که در آنها x_i داده‌ی مربوط به فرد iام و μ_i میزان عضویت فرد iام به جامعه‌ی فازی موردنظر می‌باشد. چنین داده‌هایی را که از جوامع فازی گرفته می‌شوند، به اختصار «داده‌های دقیق مستخرج از یک جامعه‌ی فازی» می‌نامیم.

تعریف ۱.۲. *بر اساس داده‌های $(x_1,\mu_1),\ldots,(x_n,\mu_n)$، مقادیر*

$$\bar{x} = \frac{\sum_{i=1}^{n}\mu_i x_i}{\sum_{i=1}^{n}\mu_i}, \qquad (۱.۲)$$

$$s^2 = \frac{1}{\sum_{i=1}^{n}\mu_i}\sum_{i=1}^{n}\mu_i\,[x_i-\bar{x}]^2, \qquad (۲.۲)$$

$$A.D. = \frac{1}{\sum_{i=1}^{n}\mu_i}\sum_{i=1}^{n}\mu_i|x_i-\bar{x}| \qquad (۳.۲)$$

به ترتیب میانگین، واریانس و متوسط انحرافات داده‌ها نامیده می‌شوند. همچنین $s=\sqrt{s^2}$ و $C.V.=\frac{s}{\bar{x}}$، به ترتیب، انحراف استاندارد (انحراف معیار) و ضریب پراکندگی (ضریب تغییرات) نامیده می‌شوند.

تعریف ۲.۲. *الف) گشتاورهای مرتبه‌ی سوم و چهارم برای داده‌های دقیق از یک جامعه‌ی فازی به ترتیب برابرند با*

$$m_3 = \frac{1}{\sum_{i=1}^{n}\mu_i}\sum_{i=1}^{n}\mu_i\,[x_i-\bar{x}]^3 \qquad (۴.۲)$$

$$m_4 = \frac{1}{\sum_{i=1}^{n}\mu_i}\sum_{i=1}^{n}\mu_i\,[x_i-\bar{x}]^4 \qquad (۵.۲)$$

ب) بر اساس بند (الف)، ضریب چولگی و ضریب کشیدگی به ترتیب به صورت زیر تعریف می‌شوند

$$g = \frac{m_3}{s^3} \qquad (۶.۲)$$

$$k = \frac{m_4}{s^4} - 3 \qquad (۷.۲)$$

توجه ۳.۲. *اگر به جای تابع عضویت مجموعه‌ی فازی μ در تعریف ۱.۲ از تابع نشانگر جامعه (یعنی $I_{\{x\}}$) استفاده شود، آنگاه بدیهی است که $\sum_{i=1}^{n} \mu_i = n$ و درنتیجه شاخص‌های (۲-۱-۲-۷) به ترتیب معادل میانگین، واریانس، متوسط انحرافات، ضریب چولگی و ضریب کشیدگی داده‌های $x_1, \dots, x_n$ در آمار کلاسیک می‌شوند.*

قضیه ۴.۲. $s^2 = \overline{x^2} - \bar{x}^2$.

مثال ۵.۲. در ادامه‌ی حالت ۳ از بخش ۱، به منظور محاسبه‌ی برخی از شاخص‌ها برای کارمندان جوان بوشهری، میزان درآمد و سن برای ده کارمند بوشهری در جدول ۱ درج شده است. همچنین، درجه‌ی جوانی هر کارمند بر اساس تابع عضویت (۱.۱) محاسبه و در جدول ۱ درج شده است.

جدول ۱: اطلاعات مربوط به ده کارمند بوشهری در مثال ۵.۲

i	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
x_i (درآمد ماهانه بر حسب میلیون تومان)	۱/۵۲	۱/۶۰	۱/۳۴	۱/۶۲	۲/۳۸	۰/۹۵	۰/۸۲	۲/۲۳	۱/۰۴	۱/۲۷
y_i (سن بر حسب سال)	۳۴	۴۹	۵۵	۳۱	۴۰	۲۶	۳۳	۵۲	۳۹	۴۲
μ_i (درجه عضویت در جامعه فازی جوانی)	۰/۶	۰	۰	۰/۹	۰	۱	۰/۷	۰	۰/۱	۰

طبق تعریف ۱.۲، میانگین، واریانس، انحراف‌معیار و ضریب پراکندگی داده‌های $(x_1, \mu_1), \dots, (x_{10}, \mu_{10})$ به صورت زیر به دست می‌آیند

$$\bar{x} = \frac{\sum_{i=1}^{۱۰} \mu_i x_i}{\sum_{i=1}^{۱۰} \mu_i} = \frac{(۰/۶ \times ۱/۵۲) + (۰ \times ۱/۶۰) + \cdots + (۰ \times ۱/۲۷)}{۰/۶ + ۰ + \cdots + ۰} = \frac{۳/۹۹}{۳/۳} = ۱/۲۱,$$

$$s^2 = \overline{x^2} - \bar{x}^2 = \frac{\sum_{i=1}^{۱۰} \mu_i x_i^2}{\sum_{i=1}^{۱۰} \mu_i} - \bar{x}^2 = \frac{(۰/۶ \times ۱/۵۲^2) + (۰ \times ۱/۶۰^2) + \cdots + (۰ \times ۱/۲۷^2)}{۰/۶ + ۰ + \cdots + ۰} - ۱/۲۱^2 = ۰/۱۱.$$

همچنین $s = \sqrt{۰/۱۱} = ۰/۳۴$، $C.V. = \frac{۰/۳۴}{۱/۲۱} = ۰/۲۸$ و ضرایب چولگی و کشیدگی به ترتیب برابر $g = ۰/۱۴$ و $k = -۱/۸۵$ هستند.

۳ کواریانس و ضریب همبستگی بر اساس داده‌های دقیق از جامعه فازی

برخی مواقع علاقمندیم تا نوع رابطه‌ی خطی بین دو متغیر را مشخص و اندازه‌ی قوت آن را تعیین کنیم. مثلاً می‌خواهیم بدانیم که آیا با زیاد شدن میزان آلودگی در شهرهای بزرگ ایران، مراجعه به درمانگاه افزایش می‌یابد یا خیر. یا اینکه می‌خواهیم بدانیم آیا زیاد شدن غیبت دانشجویانِ ممتاز یک کلاس، باعث کاهش و یا افزایش نمره‌ی آنها می‌شود. برای پرداختن به این موضوع، در این بخش قصد داریم تا کواریانس و ضریب همبستگی پیرسون را برای زوج داده‌های $(\underline{x}, \underline{y}, \underline{\mu}) = ((x_1, y_1, \mu_1), \dots, (x_n, y_n, \mu_n))$ تعمیم دهیم. یادآور می‌شویم که این دو شاخص برای اندازه‌گیری همبستگی و میزان رابطه‌ی خطی بین دو متغیر کاربرد دارند.

تعریف ۱.۳. *کواریانس و ضریب همبستگی پیرسون مشاهدات مربوط به دو متغیر X و Y مبتنی بر زوج داده‌های دقیق از جامعه‌ی فازی $(x_1, y_1, \mu_1), \dots, (x_n, y_n, \mu_n)$ بترتیب به صورت زیر تعریف می‌شوند که در آنها مقادیر $\bar{x}$، $\bar{y}$، s_x و s_y در تعریف ۱.۲*

معرفی شده‌اند

$$s_{xy} = \frac{۱}{\sum_{i=۱}^{n} \mu_i} \sum_{i=۱}^{n} \mu_i (x_i - \bar{x})(y_i - \bar{y}) \qquad (۱.۳)$$

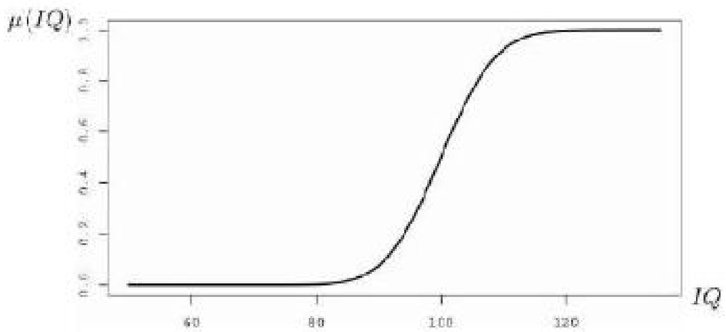
$$r = \frac{s_{xy}}{s_x s_y} \qquad (۲.۳)$$

قضیه ۲.۳. $s_{xy} = \overline{xy} - \bar{x}\bar{y}$.

توجه ۳.۳. اگر به جای تابع عضویت مجموعه‌ی فازی μ در تعریف ۱.۳ از تابع نشانگر جامعه (یعنی $I_{\{\Omega\}}$) استفاده شود، آنگاه بدیهی است که $\sum_{i=۱}^{n} \mu_i = n$ و درنتیجه s_{xy} و r به ترتیب معادل کوواریانس و ضریب همبستگی میان داده‌های $x_۱, \dots, x_n$ و داده‌های $y_۱, \dots, y_n$ درآمار کلاسیک می‌شوند.

قضیه ۴.۳. $-۱ \leq r \leq ۱$.

مثال ۵.۳. به منظور بررسی نوع و شدت همبستگی میان نمرات فارسی و ریاضی دانش‌آموزانِ تیزهوش سال چهارم ابتدایی، ۲۰ دانش‌آموز سال چهارمی به تصادف انتخاب شده‌اند. نمرات فارسی، ریاضی و نیز ضریب هوشی دانش‌آموزان (در یک آزمونِ هوش) در جدول ۲ ثبت شده است. همچنین، میزان عضویت هر دانش‌آموز به مجموعه‌ی افراد تیزهوش (درجه‌ی تیزهوشی هر دانش‌آموز)، بر اساس تابع عضویت $\mu(IQ) = \Phi\left(\frac{IQ-۱۰۰}{۷}\right)$ محاسبه و در جدول ۲ درج شده است که در آن Φ تابع توزیع تجمعی نرمال استاندارد است (شکل ۲ را ببینید).



شکل ۲: نمودار تابع عضویت تیزهوشی در مثال ۵.۳

اگر چه میانگین و انحراف‌معیار نمرات فارسی این ۲۰ دانش‌آموز بترتیب ۱۸/۴۱ و ۱/۶۸ است، اما توجه داشته باشید که طبق تعریف ۱.۲، میانگین و انحراف‌معیار نمرات فارسیِ دانش‌آموزان تیزهوش بر اساس مجموعه‌ی داده‌های دقیق مستخرج از جامعه‌ی فازی که به صورت $(x_۱, \mu_۱), \dots, (x_{۲۰}, \mu_{۲۰})$ است، اینگونه به دست می‌آیند

$$\bar{x} = \frac{\sum_{i=۱}^{۲۰} \mu_i x_i}{\sum_{i=۱}^{۲۰} \mu_i} = \frac{(۰/۹۶ \times ۱۹) + (۱ \times ۱۸/۷۵) + \cdots + (۰/۲۸ \times ۲۰)}{۰/۹۶ + ۱ + \cdots + ۰/۲۸} = \frac{۱۸۹/۷۶}{۱۰/۰۲} = ۱۸/۹۴,$$

$$s = \sqrt{\frac{\sum_{i=۱}^{۲۰} \mu_i x_i^۲}{\sum_{i=۱}^{۲۰} \mu_i} - \bar{x}^۲} = \sqrt{\frac{(۰/۹۶ \times ۱۹^۲) + (۱ \times ۱۸/۷۵^۲) + \cdots + (۰/۲۸ \times ۲۰^۲)}{۰/۹۶ + ۱ + \cdots + ۰/۲۸} - \bar{x}^۲} = ۱/۴۴.$$

جدول ۱۲ اطلاعات مربوط به ۲۰ دانش‌آموز مثال ۵.۳ که در آن μ_i درجه عضویت دانش‌آموز iام به جامعه فازی تیزهوشان است

i	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
x_i (نمره فارسی)	۱۹	۱۸/۷۵	۱۷	۱۹/۲۵	۱۵/۵	۲۰	۱۸/۲۵	۱۹/۲۵	۲۰	۱۸
y_i (نمره ریاضی)	۱۹/۲۵	۱۸/۷۵	۱۶	۱۴/۲۵	۱۶	۲۰	۱۸/۲۵	۱۹/۲۵	۲۰	۱۸
$\mu_i = \mu(IQ_i)$	۰/۹۶	۱	۰/۱۰	۰/۰۸	۰	۰/۹۹	۰/۰۴	۰/۸۴	۱	۰/۰۴
i	۱۱	۱۲	۱۳	۱۴	۱۵	۱۶	۱۷	۱۸	۱۹	۲۰
x_i (نمره فارسی)	۱۷	۱۹/۵	۲۰	۲۰	۱۴	۱۷	۱۹	۱۹/۷۵	۱۷	۲۰
y_i (نمره ریاضی)	۱۶	۲۰	۱۷	۱۸/۵	۱۷/۲۵	۱۴	۱۶/۲۵	۱۹	۱۷	۱۸
$\mu_i = \mu(IQ_i)$	۰/۴۴	۱	۰/۹۲	۰	۰/۴۴	۰/۱۶	۰/۰۶	۱	۰/۸۰	۰/۲۸

به طور مشابه، میانگین و انحراف‌معیار نمرات ریاضی دانش‌آموزان تیزهوش بر اساس داده‌های $(y_1, \mu_1), \ldots, (y_{20}, \mu_{20})$ به ترتیب برابر $\bar{y} = ۱۸/۴۹$ و $s_y = ۱/۴۲$ بدست می‌آیند. همچنین با توجه به تعریف ۱.۳ و قضیه ۲.۳، کوواریانس نمرات فارسی و ریاضی دانش‌آموزان تیزهوش بر اساس مقادیر $(x_1, y_1, \mu_1), \ldots, (x_{20}, y_{20}, \mu_{20})$ به صورت زیر بدست می‌آید

$$\begin{aligned} s_{xy} &= \frac{۱}{\sum_{i=۱}^{۲۰} \mu_i} \sum_{i=۱}^{۲۰} \mu_i x_i y_i - \bar{x}\bar{y} \\ &= \frac{(۰/۹۶ \times ۱۹ \times ۱۹/۲۵) + (۱ \times ۱۸/۷۵ \times ۱۸/۷۵) + \cdots + (۰/۲۸ \times ۲۰ \times ۱۸)}{۰/۹۶ + ۱ + \cdots + ۰/۲۸} - \bar{x}\bar{y} = ۱/۱۱ \end{aligned}$$

که نشان‌دهنده‌ی یک رابطه‌ی مستقیم میان نمرات فارسی و ریاضی دانش‌آموزان تیزهوش سال چهارمی در این مثال است. ضریب همبستگی میان نمرات فارسی و ریاضی دانش‌آموزان تیزهوش سال چهارمی نیز برابر $r = \frac{s_{xy}}{s_x s_y} = \frac{۱/۱۱}{۱/۴۴ \times ۱/۴۲} = ۰/۵۴$ به دست می‌آید.

بحث و نتیجه‌گیری

در این مقاله رویکردی جدید به روش محاسبه‌ی شاخص‌های اصلی آماری، مبتنی بر داده‌های دقیق مستخرج از یک جامعه‌ی فازی مطرح شد. در این رویکرد، هر داده‌ی خام با ضریبی که برابر با درجه‌ی عضویت آن داده به مجموعه‌ی فازی مورد نظر است، وارد محاسبات می‌شود. گر چه این مقاله تنها به توضیح مبانی این رویکرد و برخی از مباحث آمار توصیفی پرداخته است، اما این رویکرد جدید می‌تواند آغاز شاخه‌ی جدیدی از آمار، به نام آمار وزنی یا آمار موزون را فراهم آورد که در آن میزان عضویت هر یک از داده‌ها به جامعه‌ی فازی، نقش و میزان تاثیرگذاری آن داده را در محاسبات آماری تعیین می‌کند.

مراجع

طاهری، س.م. (۱۳۸۵). فیفازی سازی، ششمین کنفرانس سیستمهای فازی ایران، شیراز، ۲۰-۲۲ شهریور ۱۳۸۵، ۲۷۹-۲۸۸.

Mohammadi, J. and S. M. Taheri (2004), Pedomodels fitting with fuzzy least squares regression, *Iranian Journal of Fuzzy Systems* 1, 45-61.

Namdari, M., Yoon, J.H., Abadi, A.R., Taheri, S.M. and Choi, S.H. (2014), Fuzzy logistic regression with least absolute deviations estimators. *Soft Computing* 19, 909-917.

Viertl, R. (2011), *Statistical Methods for Fuzzy Data*. John Wiley & Sons, New York.