

## آمار توصیفی بر اساس داده‌های دقیق از یک جامعه فازی

عباس پرچمی\* و سید محمود طاهری

چکیده. در بسیاری از مسائل کاربردی هدف، بررسی یک جامعه فازی (مانند جامعه جوانان و یا جامعه مستمندان یک شهر) بر اساس نمونه‌ای از آن جامعه است. رویکرد متداول آمار کلاسیک در مواجهه با این مسائل، صورت‌بندی جامعه فازی موردنظر به عنوان زیرمجموعه‌ای از جامعه 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 می‌شود. در این روش، ابتدا جامعه فازی مورد نظر به وسیله یک تابع عضویت مشخص می‌شود. سپس، درجات عضویت تک تک داده‌های نمونه در جامعه فازی هنگام بررسی جامعه مشارکت داده می‌شوند. برای مثال، در محاسبه میانگین قد جوانان یک جامعه، منصفانه‌تر آن است که افراد جوان‌تر تاثیر بیشتری نسبت به سایر افراد داشته باشند. در این مقاله مفاهیم اصلی آمار توصیفی کلاسیک، شامل میانگین، واریانس، انحراف معیار، کواریانس و ضریب همبستگی برای چنین مسائلی تعمیم داده می‌شوند. در این تعمیم، درجه عضویت هر یک از داده‌ها به جامعه فازی موردنظر، میزان تاثیرگذاری آنها را در محاسبات آماری تعیین می‌کند. چندین مثال عددی نیز برای تشریح بهتر مطالب ارائه شده است.

### ۱. انواع جامعه‌ها و نمونه‌های آماری

در بررسی‌های آماری، با دو مجموعه‌ی اساسی/کلیدی روبرو هستیم: مجموعه یا 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 عضویت هر فرد/شیء 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، یا اینکه مشاهدات نمونه، دقیق گزارش نمی‌شوند را نمونه فازی یا نمونه نادقیق می‌گوییم. برای مثال هنگامی که می‌خواهیم شوری خاک را در نقاط مختلف یک منطقه اندازه‌گیری کنیم، ممکن است نتایج حاصل همراه با ابهام باشد [۳]. 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 [۴].

عبارات و کلمات کلیدی. جامعه فازی، تابع عضویت، شاخص‌های مرکزی، شاخص‌های پراکندگی، نمودار پراکندگی.

\* نویسنده مسئول

دبیر تخصصی رابط: حمید پزشک

تاریخ دریافت: ۱۳۹۶/۰۶/۱۶ تاریخ پذیرش: ۱۳۹۸/۰۶/۱۷

DOI: <http://dx.doi.org/10.22108/msci.2019.106316.1245>

۲) مواردی که جامعه‌ی هدف به‌طور دقیق و واضح تعریف نشده است، در این حالت با یک جامعه‌ی هدف نادقیق روبرو هستیم. برای مثال اگر بخواهیم میزان رطوبت هوا در روزهای بارانی را بررسی کنیم، جامعه‌ی هدف «روزهای بارانی» است. در اینجا بدیهی است که روزهای مختلف با مقادیر مختلف بارندگی به یک اندازه مد نظر نیستند بلکه روزهای با بارندگی بیشتر، عضویت بالاتری در مجموعه‌ی «روزهای بارانی» دارند و برعکس. به‌عنوان مثالی دیگر، اگر هدف بررسی وضع درآمد «جوانان کرمانی» باشد، آنگاه جامعه‌ی هدف یعنی «جوانان کرمانی» یک مجموعه‌ی دقیق و خوش‌تعریف نیست.

به بیان روشن، مجموعه‌ای فازی از افراد یا اشیاء را که یک یا چند ویژگی آنها مورد بررسی است، یک جامعه‌ی فازی می‌نامیم. هر جامعه‌ی فازی به‌وسیله‌ی تابع عضویت مربوطه، به‌طور یکتا، مشخص می‌شود.

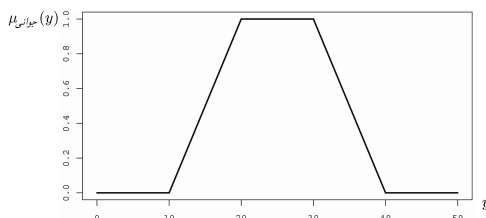
با توجه به دو حالت بالا، یعنی حالتی که مشاهدات نمونه کمیت‌های فازی هستند و حالتی که جامعه‌ی هدف یک جامعه‌ی فازی است، می‌توانیم در مجموع چهار حالت زیر را برای وضعیت یک بررسی آماری در نظر بگیریم:

حالت ۱ (جامعه دقیق - 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 - نمونه نادقیق): در این حالت، گرچه اعضای جامعه کاملاً مشخص هستند، ولی مشاهدات نمونه‌ای، نادقیق‌اند. برای مثال، ممکن است بخواهیم متوسط درآمد ماهانه تاکسی‌رانان بوشهری را برآورد کنیم. گرچه جامعه‌ی تاکسی‌رانان بوشهری کاملاً مشخص است اما درآمد ماهانه‌ی یک تاکسی‌ران، معمولاً به‌صورت تقریبی بیان می‌شود (مثلاً به‌صورت ۲ میلیون تومان، حدود ۲ تا ۲/۵ میلیون تومان). در این حالت معمولاً داده‌ها به‌صورت اعداد مثلی فازی یا اعداد دوزنقه‌ای فازی ثبت می‌شوند.

حالت ۳ (جامعه نادقیق - نمونه دقیق): فرض کنید می‌خواهیم متوسط درآمد کارمندان جوان بوشهری را محاسبه کنیم. از آنجا که «جوان» یک مفهوم فازی و نادقیق است لذا اعضای جامعه‌ی جوانان بوشهری به‌طور 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):

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



شکل ۱: نمودار تابع عضویت «جوانی»

پس از اخذ نمونه‌ی تصادفی از افراد این شهر (به حجم  $n$ )، اطلاعات مربوط به نمونه را به صورت  $(x_1, \mu_1), \dots, (x_n, \mu_n)$  ثبت می‌کنیم که در آن  $x_i$  درآمد فرد  $i$ ام و  $\mu_i$  درجه‌ی عضویت او 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 آماری برای توصیف این نوع داده‌ها بحث خواهیم نمود.

**حالت ۴ (جامعه نادقیق - نمونه نادقیق):** این حالت پیچیده‌تری نسبت به سه حالت قبل است. 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 ([۹] و [۱] را ببینید). اما تاکنون روشی برای انجام محاسبات حالت‌های ۳ و ۴ که در آنها داده‌ها از جوامع 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. معرفی کواریانس و ضریب همبستگی، و همچنین روش رسم نمودار پراکنش وزنی برای این نوع داده‌ها در بخش ۳ مطرح شده‌اند. بخش ۴ شامل نتیجه‌گیری و اشاره‌ای به برخی از کارهای مرتبط است.

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

تحلیل‌های آماری در بسیاری از مسائل کاربردی مبتنی بر سرشماری/نمونه‌گیری از یک جامعه‌ی فازی است. در روش‌های رایج، پس از مدل‌بندی جامعه‌ی فازی به کمک یک رده یا یک بازه‌ی غیرفازی، جمع‌آوری و بررسی داده‌ها انجام می‌شود. 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 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 شورپسند 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 بارانی در یک شهر؛ در حوزه اقتصاد: بررسی الگوی مصرف در خانوارهای با درآمد متوسط.

در آمار کلاسیک همه مسائل بالا با در نظر گرفتن یک مجموعه/رده/بازه‌ی غیرفازی برای جامعه‌ی فازی بررسی می‌شوند. بدین ترتیب از ابهام موجود در جامعه‌ی فازی چشم‌پوشی می‌شود و لذا بخشی مفید از اطلاعات نادیده گرفته می‌شود. اما مناسب‌تر و طبیعی‌تر، این است که، ابتدا جامعه‌ی فازی مورد نظر را به وسیله‌ی یک تابع عضویت به‌طور دقیق مدل‌سازی کنیم. در این صورت می‌توان درجه‌ی عضویت هر داده در این جامعه‌ی فازی را مشخص نمود. در چنین حالتی منطقی است که روش‌های آمار کلاسیک را به‌گونه‌ای 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 به گونه‌ای صورت گیرد که داده‌های با درجات عضویت بزرگ‌تر، نقشی پررنگ‌تر و تاثیرگذارتر در محاسبات، نسبت به داده‌های با درجات عضویت کوچک‌تر، داشته باشند.

برای پرداختن به این موضوع، فرض کنید بر اساس نمونه‌ای تصادفی از جامعه‌ای فازی، داده‌های  $(\underline{x}, \underline{\mu}) = ((x_1, \mu_1), \dots, (x_n, \mu_n))$  ثبت شده‌اند که در آنها  $x_i$  داده‌ی مربوط به فرد  $i$ ام و  $\mu_i$  میزان عضویت فرد  $i$ ام به جامعه‌ی فازی موردنظر است. 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  $(x_1, \mu_1), \dots, (x_n, \mu_n)$ ، مقادیر

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

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

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

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 تغییرات نامیده می‌شوند.

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

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

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

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

$$(۶.۲) \quad g = \frac{m_3}{s^3}$$

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

اگر به جای تابع عضویت مجموعه‌ی فازی  $\mu$  در تعریف ۱.۲ از تابع نشانگر جامعه (یعنی  $I_{\{\Omega\}}$ ) استفاده شود، آنگاه بدیهی است که  $\sum_{i=1}^n \mu_i = n$  و در نتیجه شاخص‌های (۱.۲) تا (۷.۲) به ترتیب معادل میانگین، واریانس، متوسط انحرافات، ضریب چولگی و ضریب کشیدگی داده‌های  $x_1, \dots, x_n$  در آمار کلاسیک می‌شوند.

**قضیه ۳.۲.**  $s^2 = \overline{x^2} - \bar{x}^2$ .

**اثبات.**

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

$$\begin{aligned}
 &= \frac{1}{\sum_{i=1}^n \mu_i} \sum_{i=1}^n (\mu_i x_i^2 + \mu_i \bar{x}^2 - 2\bar{x} \mu_i x_i) \\
 &= \frac{\sum_{i=1}^n \mu_i x_i^2}{\sum_{i=1}^n \mu_i} + \frac{\bar{x}^2 \sum_{i=1}^n \mu_i}{\sum_{i=1}^n \mu_i} - 2\bar{x} \frac{\sum_{i=1}^n \mu_i x_i}{\sum_{i=1}^n \mu_i} \\
 &= \frac{\sum_{i=1}^n \mu_i x_i^2}{\sum_{i=1}^n \mu_i} + \bar{x}^2 - 2\bar{x}^2 \\
 &= x^2 - \bar{x}^2.
 \end{aligned} \tag{۸.۲}$$

□

مثال ۴.۲. می‌خواهیم متوسط درآمد کارمندان جوان بوشهری را محاسبه کنیم که مفهوم «جوانی» بوسیله‌ی تابع عضویت (۱.۱) داده شده است. همچنین واریانس، انحراف معیار و ضریب تغییرات درآمد کارمندان جوان بوشهری را محاسبه می‌کنیم. میزان درآمد و سن ده کارمند بوشهری در جدول ۱ درج شده است. درجه‌ی جوانی هر کارمند نیز بر اساس تابع عضویت (۱.۱) محاسبه و در این جدول نوشته شده است.

جدول ۱: اطلاعات مربوط به ده کارمند بوشهری در مثال ۴.۲

| $i$                                      | 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   | ۴۳   |
| $\mu_i$ (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، انحراف معیار و ضریب تغییرات داده‌های  $(x_1, \mu_1), \dots, (x_{10}, \mu_{10})$  به صورت زیر به دست می‌آیند

$$\begin{aligned}
 \bar{x} &= \frac{\sum_{i=1}^{10} \mu_i x_i}{\sum_{i=1}^{10} \mu_i} \\
 &= \frac{(0/6 \times 1/52) + (0 \times 1/60) + \dots + (0 \times 1/27)}{0/6 + 0 + \dots + 0} \\
 &= \frac{3/99}{3/3} = 1/21,
 \end{aligned}$$

$$\begin{aligned}
 s^2 &= \overline{x^2} - \bar{x}^2 \\
 &= \frac{\sum_{i=1}^{10} \mu_i x_i^2}{\sum_{i=1}^{10} \mu_i} - \bar{x}^2 \\
 &= \frac{(0/6 \times 1/52^2) + (0 \times 1/60^2) + \dots + (0 \times 1/27^2)}{0/6 + 0 + \dots + 0} - 1/21^2 \\
 &= 0/11.
 \end{aligned}$$

همچنین  $C.V. = \frac{0/34}{1/21} = 0/28$ ،  $s = \sqrt{0/11} = 0/34$  و  $g = 0/14$  و ضرایب چولگی و کشیدگی به ترتیب برابر ۱۴ و  $k = -1/85$  هستند.

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

گاهی علاقه‌مندیم 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 درمانگاه‌ها افزایش می‌یابد یا خیر. یا اینکه می‌خواهیم بدانیم آیا زیاد شدن غیبت دانشجویان 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، مربوط به دو متغیر  $X$  و  $Y$  مبتنی بر زوج داده‌های دقیق از جامعه‌ی فازی  $(x_1, y_1, \mu_1), \dots, (x_n, y_n, \mu_n)$  به‌ترتیب به صورت زیر تعریف می‌شوند که در آنها مقادیر  $\bar{x}$ ،  $\bar{y}$ ،  $s_x$  و  $s_y$  در تعریف ۱.۲ معرفی شده‌اند:

$$(۱.۳) \quad s_{xy} = \frac{1}{\sum_{i=1}^n \mu_i} \sum_{i=1}^n \mu_i (x_i - \bar{x})(y_i - \bar{y})$$

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

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

**اثبات.**

$$\begin{aligned} s_{xy} &= \frac{1}{\sum_{i=1}^n \mu_i} \sum_{i=1}^n \mu_i (x_i - \bar{x})(y_i - \bar{y}) \\ &= \frac{1}{\sum_{i=1}^n \mu_i} \sum_{i=1}^n [\mu_i x_i y_i - \mu_i x_i \bar{y} - \mu_i \bar{x} y_i + \mu_i \bar{x} \bar{y}] \\ &= \frac{\sum_{i=1}^n \mu_i x_i y_i}{\sum_{i=1}^n \mu_i} - \bar{y} \frac{\sum_{i=1}^n \mu_i x_i}{\sum_{i=1}^n \mu_i} - \bar{x} \frac{\sum_{i=1}^n \mu_i y_i}{\sum_{i=1}^n \mu_i} + \bar{x} \bar{y} \\ &= \frac{1}{\sum_{i=1}^n \mu_i} \sum_{i=1}^n \mu_i x_i y_i - \bar{y} \bar{x} - \bar{x} \bar{y} + \bar{x} \bar{y} \\ &= \overline{xy} - \bar{x} \bar{y}. \end{aligned} \quad (۳.۳)$$

□

علاوه بر تعریف ۱.۳، می‌توان ضریب همبستگی پیرسون را پس از استانداردسازی داده‌ها به کمک قضیه زیر محاسبه نمود که محاسبات کمتری، مخصوصاً برای اعداد بزرگ، نیاز دارد.

**قضیه ۳.۳.** ضریب همبستگی پیرسون برابر است با

$$(۴.۳) \quad r = \frac{\sum_{i=1}^n z_{x_i} z_{y_i}}{\sum_{i=1}^n \mu_i}$$

که در آن  $z_{y_i} = \frac{y_i - \bar{y}}{s_y}$  و  $z_{x_i} = \frac{x_i - \bar{x}}{s_x}$  استاندارد شده‌ی  $x_i$  و  $y_i$  می‌باشند و آماره‌های  $\bar{x}$ ،  $\bar{y}$ ،  $s_x$  و  $s_y$  در تعریف ۱.۲ معرفی شده‌اند.

## اثبات.

$$\begin{aligned}
 r &= \frac{s_{xy}}{s_x s_y} \\
 &= \frac{\sum_{i=1}^n \mu_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n \mu_i} \\
 &= \frac{s_x s_y}{\sum_{i=1}^n \mu_i} \left( \frac{x_i - \bar{x}}{s_x} \right) \left( \frac{y_i - \bar{y}}{s_y} \right) \\
 &= \frac{\sum_{i=1}^n \mu_i z_{x_i} z_{y_i}}{\sum_{i=1}^n \mu_i}. \quad (5.3)
 \end{aligned}$$

□

مثبت، منفی و صفر بودن کواریانس مشاهدات مربوط به دو متغیر  $X$  و  $Y$  (یا به طور معادل ضریب همبستگی پیرسون)، به ترتیب به معنای رابطه‌ی خطی مستقیم، معکوس و عدم وجود رابطه‌ی خطی (ناهمبستگی) بین دو متغیر  $X$  و  $Y$  است. اگر به جای تابع عضویت مجموعه‌ی فازی  $\mu$  در تعریف ۱.۳ از تابع نشانگر جامعه (یعنی  $I_{\{\Omega\}}$ ) استفاده شود، آنگاه  $\sum_{i=1}^n \mu_i = n$  و در نتیجه  $s_{xy}$  و  $r$  به ترتیب معادل کواریانس و ضریب همبستگی میان داده‌های  $x_1, \dots, x_n$  و  $y_1, \dots, y_n$  در آمار کلاسیک می‌شوند.

قضیه ۴.۳.  $-1 \leq r \leq 1$ .

اثبات. طبق نامساوی کوشی-شوارتز<sup>۱</sup> [۲] داریم

$$\left( \sum_{i=1}^n a_i b_i \right)^2 \leq \left( \sum_{i=1}^n a_i^2 \right) \left( \sum_{i=1}^n b_i^2 \right)$$

بنابراین، با جایگذاری  $a_i = \sqrt{\mu_i}(x_i - \bar{x})$  و  $b_i = \sqrt{\mu_i}(y_i - \bar{y})$  می‌توان نوشت

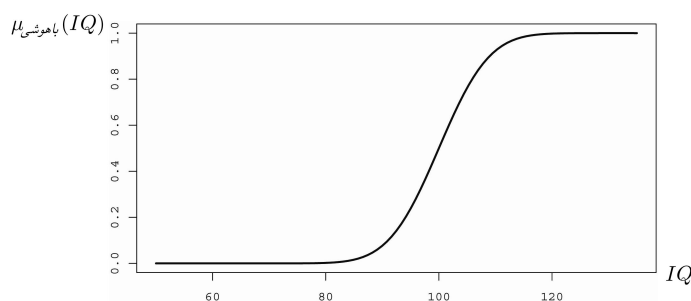
$$\begin{aligned}
 &\left( \sum_{i=1}^n \mu_i (x_i - \bar{x})(y_i - \bar{y}) \right)^2 \leq \left( \sum_{i=1}^n \mu_i (x_i - \bar{x})^2 \right) \left( \sum_{i=1}^n \mu_i (y_i - \bar{y})^2 \right) \\
 \Rightarrow &\left\{ \left| \sum_{i=1}^n \mu_i (x_i - \bar{x})(y_i - \bar{y}) \right| \right\}^{\frac{1}{2}} \leq \left\{ \left( \sum_{i=1}^n \mu_i (x_i - \bar{x})^2 \right) \left( \sum_{i=1}^n \mu_i (y_i - \bar{y})^2 \right) \right\}^{\frac{1}{2}} \\
 \Rightarrow &\frac{\left| \sum_{i=1}^n \mu_i (x_i - \bar{x})(y_i - \bar{y}) \right|}{\sqrt{\sum_{i=1}^n \mu_i (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n \mu_i (y_i - \bar{y})^2}} \leq 1 \\
 \Rightarrow &\left| \frac{\sum_{i=1}^n \mu_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n \mu_i (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n \mu_i (y_i - \bar{y})^2}} \right| \leq 1 \\
 \Rightarrow &\left| \frac{s_{xy}}{s_x s_y} \right| \leq 1 \\
 \Rightarrow &-1 \leq r \leq 1.
 \end{aligned}$$

<sup>1</sup>Cauchy-Schwarz inequality

□

هر چه  $r$  به  $+1$  نزدیکتر باشد، قوت همبستگی (رابطه) خطی مستقیم بیشتر است و هر چه به  $-1$  نزدیکتر باشد، قوت همبستگی (رابطه) خطی معکوس بیشتر است.

**مثال ۵.۳.** به منظور بررسی نوع و قوت همبستگی میان نمرات فارسی و ریاضی دانش‌آموزان تیزهوش سال چهارم ابتدایی، ۲۰ دانش‌آموز سال چهارمی به تصادف انتخاب شده‌اند. نمرات فارسی و ریاضی دانش‌آموزان در جدول ۲ ثبت شده‌اند. پس از سنجش ضریب هوشی دانش‌آموزان در یک آزمون هوش، میزان عضویت هر دانش‌آموز به مجموعه‌ی افراد تیزهوش (درجه‌ی تیزهوشی هر دانش‌آموز)، بر اساس تابع عضویت  $\mu(IQ) = \Phi\left(\frac{IQ-100}{15}\right)$  محاسبه و در جدول ۲ درج شده است که در آن  $\Phi$  تابع توزیع تجمعی نرمال استاندارد است (شکل ۲ را ببینید).



شکل ۲: نمودار تابع عضویت «تیزهوشی» در مثال ۵.۳

اگر چه میانگین و انحراف معیار نمرات فارسی این ۲۰ دانش‌آموز، به ترتیب  $18/41$  و  $1/68$  است، اما توجه داشته باشید که طبق تعریف ۱.۲، میانگین و انحراف معیار نمرات فارسی دانش‌آموزان تیزهوش بر اساس مجموعه‌ی داده‌های دقیق مستخرج از جامعه‌ی فازی که به صورت  $(x_1, \mu_1), \dots, (x_{20}, \mu_{20})$  است، این‌گونه به دست می‌آیند

$$\bar{x} = \frac{\sum_{i=1}^{20} \mu_i x_i}{\sum_{i=1}^{20} \mu_i} = \frac{(0/96 \times 19) + (1 \times 18/75) + \dots + (0/28 \times 20)}{0/96 + 1 + \dots + 0/28} = \frac{189/76}{10/02} = 18/94,$$

$$s = \sqrt{\frac{\sum_{i=1}^{20} \mu_i x_i^2}{\sum_{i=1}^{20} \mu_i} - \bar{x}^2} = \sqrt{\frac{(0/96 \times 19^2) + (1 \times 18/75^2) + \dots + (0/28 \times 20^2)}{0/96 + 1 + \dots + 0/28} - \bar{x}^2} = 1/44.$$

به طور مشابه، میانگین و انحراف معیار نمرات ریاضی دانش‌آموزان تیزهوش بر اساس داده‌های  $(y_1, \mu_1), \dots, (y_{20}, \mu_{20})$  به ترتیب برابر  $18/49$  و  $1/42$  است. به دست می‌آیند. همچنین، کواریانس نمرات فارسی و ریاضی دانش‌آموزان تیزهوش بر اساس مقادیر  $(x_1, y_1, \mu_1), \dots, (x_{20}, y_{20}, \mu_{20})$  به صورت زیر به دست می‌آید

$$\begin{aligned} s_{xy} &= \frac{1}{\sum_{i=1}^{20} \mu_i} \sum_{i=1}^{20} \mu_i x_i y_i - \bar{x} \bar{y} \\ &= \frac{(0/96 \times 19 \times 19/25) + (1 \times 18/75 \times 18/75) + \dots + (0/28 \times 20 \times 18)}{0/96 + 1 + \dots + 0/28} - \bar{x} \bar{y} = 1/11 \end{aligned}$$

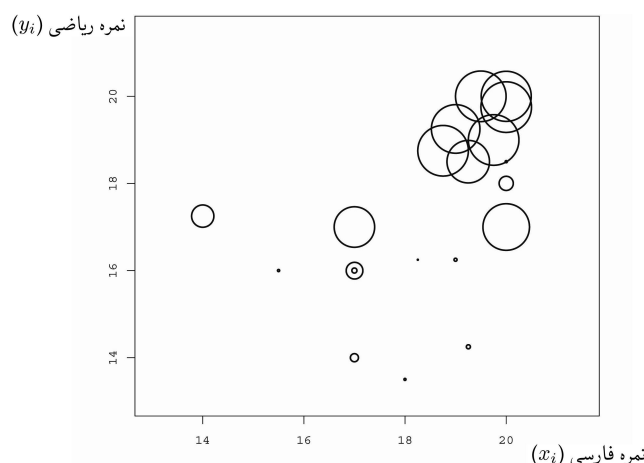
جدول ۲: اطلاعات مربوط به ۲۰ دانش‌آموز مثال ۵.۳ که در آن  $\mu_i$  درجه عضویت دانش‌آموز  $i$  در جامعه فازی تیزهوشان است

| $i$                 | 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 | ۰/۰۲  | 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   |
| $\mu_i = \mu(IQ_i)$ | ۰/۳۳  | ۱     | ۰/۹۲ | ۰     | ۰/۴۴  | ۰/۱۶ | ۰/۰۶  | ۱     | ۰/۸۰ | ۰/۲۸ |

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  $r = \frac{s_{xy}}{s_x s_y} = \frac{1/11}{1/44 \times 1/42} = 0/54$  به دست می‌آید.

می‌توان برای تحلیل شهودی همبستگی زوج داده‌های دقیق  $(x_1, y_1, \mu_1), \dots, (x_n, y_n, \mu_n)$  که از یک جامعه‌ی فازی گرفته شده‌اند، از نمودار پراکنش وزنی<sup>۲</sup> یا نمودار نقطه‌ای وزنی استفاده کرد. روش ترسیم این نمودار همانند نمودار پراکنش معمولی است با این تفاوت که تمامی نقاط نمودار پراکنش معمولی از نظر مساحت یکسان هستند اما مساحت نقاط (دایره‌ها) در نمودار پراکنش وزنی متناسب با درجه‌ی عضویت آنها در جامعه‌ی فازی است. به عبارت دقیق‌تر، نمودار پراکنش وزنی برای زوج داده‌های دقیق  $(x_1, y_1, \mu_1), \dots, (x_n, y_n, \mu_n)$ ، متشکل از  $n$  دایره در صفحه‌ی دوبعدی است که در آن  $(x_i, y_i)$  مرکز و  $\mu_i$  شعاع (و یا نسبتی از شعاع) دایره  $i$ ام، به ازای  $i = 1, \dots, n$  می‌باشد. شکل ۳ یک نمونه از نمودار پراکنش وزنی برای زوج داده‌های حقیقی مقدار فازی  $(x_1, y_1, \mu_1), \dots, (x_{20}, y_{20}, \mu_{20})$  از مثال ۵.۳ را نشان می‌دهد که در آن قطر هر دایره برابر درجه‌ی عضویت دانش‌آموز مورد نظر در جامعه‌ی فازی دانش‌آموزان تیزهوش سال چهارمی است.

<sup>۲</sup>weighted scatter plot



شکل ۳: نمودار پراکنش وزنی در مثال ۵.۳

#### ۴. بحث و نتیجه‌گیری

تاکنون طیف نسبتاً وسیعی از روش‌های آمار توصیفی و آمار استنباطی برای حالتی که جامعه دقیق و مشاهدات نمونه نادقیق هستند (حالت ۲)، تعمیم و گسترش یافته است که علاقمندان می‌توانند برای نمونه به [۹] مراجعه نمایند. همچنین، ایده‌ی فراوانی موزون بر حسب میزان عضویت در [۶]، [۷] و [۸] معرفی و کاربردهایی از آن در برخی مسائل احتمال و آمار بررسی شده است. ولی تا جایی که نویسندگان اطلاع دارند، تا کنون مطالعه‌ای مبتنی بر ایده‌ی «داده‌های دقیق از یک جامعه‌ی فازی» معرفی نشده است.

در این مقاله رویکردی جدید برای 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 در جامعه‌ی فازی مورد نظر است، وارد محاسبات می‌شود. گرچه 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 در جامعه‌ی فازی، نقش و میزان تاثیرگذاری آن داده را در محاسبات آماری تعیین می‌کند.

#### مراجع

[۱] س. م. طاهری، غیرفازی سازی، ششمین کنفرانس سیستم‌های فازی ایران، شیراز، ۱۳۸۵، ۲۸۸-۲۷۹.

[2] Z. Lin and Z. Bai, *Probability Inequalities*, Science Press Beijing and Springer-Verlag, Berlin, 2011.

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

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

- [5] A. Parchami, *Weighted. Desc. Stat: Weighted Descriptive Statistics*, R package version 1.0, 2016. <https://cran.r-project.org/package=Weighted.Desc.Stat>
- [6] S. M. Taheri and G.R. Hesamian, Goodman–Kruskal measure of association for fuzzy–categorized variables, *Kybernetika*, **47** (2011) 110-122.
- [7] S. M. Taheri, G. Hesamian and R. Viertl, Contingency table with fuzzy categories, *Proc. of the 11th International Conference on Intelligent Technologies*, Bangkok, Thailand (2010) 100-104.
- [8] S. M. Taheri, G. Hesamian and R. Viertl, Contingency tables with fuzzy information. *Communications in Statistics-Theory and Methods*, **45** (2016) 5906-5917.
- [9] R. Viertl, *Statistical Methods for Fuzzy Data*, John Wiley & Sons, New York, 2011.

#### عباس پرچمی

دانشگاه شهید باهنر کرمان، دانشکده ریاضی و کامپیوتر، بخش آمار  
parchami@uk.ac.ir

عباس پرچمی متولد اردیبهشت ۱۳۵۶ در شهر مشهد است. وی در سال‌های ۱۳۸۰، ۱۳۸۳ و ۱۳۹۴، به ترتیب دوره‌های کارشناسی آمار (دانشگاه فردوسی مشهد)، کارشناسی ارشد آمار (دانشگاه شهید باهنر کرمان) و دکترای آمار (دانشگاه فردوسی مشهد) را به پایان رساند. او هم‌اکنون عضو هیات علمی دانشگاه شهید باهنر کرمان است.



#### سید محمود طاهری

تهران، دانشگاه تهران، دانشکده فنی  
sm\_taheri@ut.ac.ir

سید محمود طاهری متولد دی ۱۳۴۳ در شهر تهران است. وی در سال‌های ۱۳۶۷، ۱۳۷۰ و ۱۳۷۸، به ترتیب، دوره‌های کارشناسی و کارشناسی ارشد آمار را در دانشگاه فردوسی مشهد و دکترای آمار را در دانشگاه شیراز به پایان رساند. وی حدود ۱۵ سال عضو هیات علمی دانشگاه صنعتی اصفهان بوده و هم‌اکنون عضو هیات علمی دانشگاه تهران است.

