

بررسی آماری زوایای جهت وزش باد

سید مرتضی نجیبی^۱، موسی گل علی زاده^۲

چکیده

مشکلات جدید در علوم کاربردی، که بخشی از آن توسط آمار حل می‌شود، نگرش‌های نوینی را بوجود آورده است. بطور مثال در علوم مثل هواشناسی، زیست‌شناسی و زمین‌شناسی داده‌هایی وجود دارند که ماهیت غیرخطی و یا عبارتی دیگر ناقلیدسی دارند. شاخه جدید آمار دایره‌ای به بررسی و مطالعه پدیده‌هایی می‌پردازد که بصورت زاویه و جهت می‌باشند. مقاله حاضر به تجزیه و تحلیل داده‌های جهت باد ایستگاه سنج از دید آمار دایره‌ای می‌پردازد. جهت ایجاد انگیزه برای محققین علاقه‌مند به آمار دایره‌ای تحلیل توصیفی داده‌ها انجام شده و نمودارهای مناسب رسم خواهد شد. سپس، برازش داده‌ها به یکی از توزیع‌های دایره‌ای مورد ارزیابی قرار گرفته و پارامترهای موجود برآورد می‌گردد. امید است نتایج نهایی تحلیل، راهی را در پیش‌روی علاقه‌مندان به حوزه آمار غیرخطی قرار دهد.

واژه های کلیدی: متغیر جهت‌دار، توزیع‌های غیراقلیدسی، آمار دایره‌ای، توزیع ون‌میسز، نمودار دایره‌ای سه بعدی و برازش مدل.

۱- مقدمه

روش‌های موجود آمار خطی به تجزیه و تحلیل آنها پرداخت. اینگونه مشاهدات به داده‌های غیر خطی معروفند [۶]. شاید کسانی باشند که با نادیده گرفتن این خاصیت مهم داده‌ها به مدل‌بندی و تحلیل مشاهدات، در فضای اقلیدسی، بپردازند. اما در ادامه مقاله نشان داده خواهد شد که نتایج حاصل از این رویکرد منطقی و قابل تفسیر نخواهد بود و چه بسا منجر به نتایج غیر عادی شود.

یکی از ساده‌ترین نوع متغیر که مقادیرش در فضای غیراقلیدسی قرار دارد، متغیر جهت دار یا دایره‌ای^۳ می‌باشد. این نامگذاری بدین دلیل است که مشاهدات این متغیر دارای دور یا جهت خاصی در صفحه می‌باشد و بدلیل داشتن جهت یا دور می‌توانند با یک تبدیل مناسب بر روی دایره واحد قرار گیرند. مشاهداتی که از این نوع متغیرها ناشی می‌شوند در

روش‌های بیشمار در آمار وجود دارند که به محققین آمار و علوم دیگر کمک می‌کنند به تجزیه و تحلیل مناسب داده‌های مورد مطالعه بپردازند. تقریباً در تمامی این روش‌ها فرض بر این است که مشاهدات اولیه داده‌های خطی می‌باشند و یا عبارتی دیگر فضای نمونه‌ای مشاهدات زیرمجموعه‌ای از یک فضای اقلیدسی است. نرم‌افزارهای آماری بسیاری وجود دارند که قابلیت‌های بالایی را برای تحلیل این نوع مشاهدات در اختیار کاربر قرار می‌دهند. اما مثال‌های بیشماری نیز وجود دارد که مشاهدات آماری مورد مطالعه، بطور بالقوه، در یک فضای غیر خطی قرار دارند و لذا نمی‌توان با استفاده از

^۱ دانشگاه تربیت مدرس تهران

^۲ دانشگاه شهید بهشتی تهران

^۳ Directional or Circular

نمودارها و آماره‌های توصیفی داده‌های مشاهده شده می‌باشد. نمودارهای جدید بسیاری برای داده‌های دایره‌ای در چند سال اخیر معرفی گردیده است که نمایش تغییرات اینگونه مشاهدات را واضح‌تر می‌نماید [۱۵و۱۱].

مقاله حاضر به پنج بخش زیر تقسیم‌بندی شده است. جهت معرفی و آشنایی بیشتر با متغیر دایره‌ای، طریقه نمایش این متغیرها در یک فضای غیر اقلیدسی در بخش دوم ارائه می‌شود. بعلاوه به کمک مفهوم فاصله در فضای غیراقلیدسی مربوط به متغیر دایره‌ای، نحوه بدست آوردن چند آماره توصیفی و رسم مشاهدات نیز تشریح می‌شود. در ادامه با استفاده از داده‌های مربوط به زاویه جهت باد در ایستگاه سنندج با طول جغرافیایی ۴۷ درجه و عرض جغرافیایی ۳۵ درجه که بصورت ماهیانه در طی سال‌های ۲۰۰۰ الی ۲۰۰۶ جمع‌آوری شده است، مقایسه معیارهای معمولی (اقلیدسی) با معیارهای آمار دایره‌ای انجام می‌شود. مقدمه‌ای از توزیع‌های آماری در بخش سوم آورده شده و خواص یک توزیع مهم مورد مطالعه قرار می‌گیرد. برازش یک مدل مناسب آماری به داده‌های باد سنندج در بخش چهارم آمده است. پایان بخش مقاله حاضر نتیجه‌گیری از تحلیل آماری انجام‌شده روی داده‌های باد و توصیف تحقیقات آتی این حوزه می‌باشد.

۲- آمار دایره‌ای

متغیرهای جهت‌دار بسته به اینکه در چه بعدی تعریف می‌شوند، انواع متفاوتی دارند. اگر متغیر جهت‌دار مقادیرش را بر روی صفحه اختیار کند، می‌توان آن را بر روی یک دایره نمایش داد. بدین معنی که با تبدیل مقیاس تاثير دور یا جهت موجود در داده‌ها را بر روی دایره‌ای با شعاع واحد نشان داد. سپس متغیر مورد مطالعه زاویه (θ) خواهد بود که دارای دامنه تغییرات بین صفر تا 2π رادیان (صفر تا 360 درجه) است. اما ممکن است متغیری دارای دور یا جهت باشد ولی طول مقیاس اندازه‌گیری آن در یک دور، بیشتر یا کمتر از محیط دایره باشد. برای تحلیل این متغیرها نیاز است که تبدیل مقیاس بگونه‌ای انجام گیرد که داده‌ها بر روی دایره به شعاع واحد قرار گیرند. یکی از انواع این نوع مشاهدات،

علمی مانند زیست‌شناسی، جغرافیا، زمین‌شناسی و علوم اجتماعی فراوان دیده می‌شوند. این مشاهدات عموماً نشان دهنده جهت، دور یا چرخش زمان در یک بازه مشخص می‌باشند. بعنوان مثال زاویه جهت وزش باد یک متغیر دایره‌ای است زیرا می‌توان با در نظر گرفتن جهت عقربه‌های ساعت هر جهت جغرافیایی را به عنوان یک زاویه بر روی دایره‌ای به شعاع واحد در نظر گرفت [۸].

بدلیل ماهیت فضای نااقلیدسی می‌توان تصور کرد که خصوصیات و ویژگی‌های متغیر دایره‌ای و خطی تفاوت‌های ویژه‌ای دارند. یکی از خصوصیات که متغیرهای جهت‌دار یا دایره‌ای دارند این است که ابتدا و انتهای برد اندازه‌گیری آنها بر همدیگر منطبق می‌باشند. مثلاً زوایای 5° شرقی و 365° شرقی، وقتی بر روی دایره واحد رسم شوند، بر همدیگر منطبق خواهند شد (زاویه θ شرقی بدین معنی است که مقدار θ از شرق زاویه دارد). وجود چنین مشخصه‌هایی در متغیر دایره‌ای موجب می‌شود که محاسبه معیارهای آماری به کمک روش‌های مرسوم آمار خطی غیر منطقی به نظر برسد. به عنوان مثال میانگین حسابی سه جهت 10° و 15° و 350° زاویه 125° شرقی جغرافیا است، واقع در جهت شمال غربی، است. درحالی‌که در آمار خطی انتظار داریم معیار مرکزی جایی در بین مشاهدات باشد. واضح است که در این مثال میانگین حسابی این حقیقت را نشان نمی‌دهد. حال اگر 350° را با 10° - شرقی که معادل آن است عوض کنیم، میانگین حاصل برابر 5° شرق خواهد شد که منطقی‌تر می‌باشد. عدم کارایی معیارهای مرسوم دیگر آمار خطی در ارائه تصویر مناسبی از شاخص‌های داده‌های دایره‌ای را می‌توان با این مثال و یا مثال‌های دیگر مورد مطالعه قرار داد.

از آنجایی که زاویه یک متغیر تصادفی تعریف شده در فضای غیراقلیدسی دایره می‌باشد، بررسی آماری آن تنها به کمک آمار جهت‌دار ممکن است [۱۳و۶]. مسائلی از قبیل اربیی برآوردها، توزیع‌های آماری مناسب و آزمون فرض در آمار دایره‌ای مورد توجه محققین قرار گرفته است [۲، ۷ و ۱۰]. یکی از بخش‌های اصلی در پیش‌بینی وضعیت هوا استفاده از

علاقه‌مند می‌تواند منابع [۶ و ۸] را برای این منظور مطالعه نماید.

۲-۲- آمار توصیفی در داده‌های جهت‌دار

نمایش اطلاعات جمع‌آوری شده به کمک آمارها و نمودارهای توصیفی، اولین گام در تجزیه و تحلیل آماری یک تحقیق می‌باشد. در ادامه بعد از بیان روش‌های کلی محاسبه شاخصهای مرکزی و پراکندگی داده‌های جهت‌دار دایره‌ای، چند آماره توصیفی را برای برای مشاهدات زوایای جهت باد ایستگاه سنج بدست می‌آوریم. همچنین نمایش داده‌های مورد نظر را بصورت نمودارهای دایره‌ای ارائه خواهیم کرد.

۲-۲-۱- شاخص‌های تمرکز و پراکندگی

میانگین:

یکی از تکنیک‌های جالبی که برای محاسبه میانگین داده‌های جهت‌دار دایره‌ای (μ) مورد استفاده قرار می‌گیرد بکارگیری توابع \sin و \cos کمان موجود در دایره می‌باشد. در این حالت میانگین عبارتست از:

مشاهداتی است که متغیر مربوط به آن بصورت محوری دارای دور می‌باشد و به مشاهدات محوری ۱ معروفند. بعنوان مثال متغیری که دامنه تغییراتش زاویه بین صفر تا 180° درجه است و زاویه 30° درجه معادل زاویه 210° درجه در نظر گرفته شود، یک متغیر محوری است. این نوع از متغیرها را می‌توان هم باتغییر مقیاس اندازه‌گیری بر روی دایره واحد تعریف کرد و هم با استفاده از روش‌های تحلیل داده‌های محوری به بررسی آنها پرداخت [۳ و ۸].

۲-۱- فواصل نقاط دایره‌ای

همانگونه که بحث شد معیارهای مرسوم آمار خطی بطور مستقیم قابل استفاده برای متغیرهای جهت‌دار روی دایره نیستند. میانگین حسابی در آمار خطی کمیتی است که فاصله اقلیدسی بین مجموعه داده‌ها و یک مقدار ثابت را مینیمم می‌نماید. حال اگر شخصی بخواهد میانگین دایره‌ای مجموعه‌ای از داده‌های دایره‌ای را بیابد شاید راه حل ساده بررسی تعریف فاصله برای متغیر دایره‌ای باشد. قابل ذکر است که واریانس نیز کمیتی برآمده از مفهوم فاصله می‌باشد. لذا جهت روشن شدن میانگین دایره‌ای و مباحث آماری دیگری که در پی آن می‌آید، به بررسی انواع طول فاصله‌ها روی دایره واحد می‌پردازیم.

دو نقطه $x = (x_1, y_1)$ و $y = (x_2, y_2)$ را بر روی دایره واحد در نظر بگیرید. اگر این دو نقطه را بصورت $x = [\cos(\theta_1), \sin(\theta_1)]$ و $y = [\cos(\theta_2), \sin(\theta_2)]$ نمایش دهیم، آنگاه فاصله اقلیدسی بین آنها بصورت زیر خواهد بود.

$$\|x - y\| = \sqrt{2[1 - \cos(\theta_1 - \theta_2)]} \quad (1)$$

اما به این دلیل که هر دو بردار x و y دارای اندازه واحد می‌باشند ($\|x\| = \|y\| = 1$) این فاصله یک فاصله غیر اقلیدسی بر روی دایره می‌باشد. انواع دیگری از طول فاصله بین نقاط روی دایره موجود می‌باشد که بدلیل جلوگیری از گستردگی حجم مقاله از ارائه آنها خودداری می‌گردد. خواننده

$$\mu = \arctan^* \left(\frac{S}{C} \right) = \begin{cases} \arctan\left(\frac{S}{C}\right) & S \geq 0, C > 0 \\ \frac{\pi}{2} & S > 0, C = 0 \\ \arctan\left(\frac{S}{C}\right) + \pi & C < 0 \\ \arctan\left(\frac{S}{C}\right) + 2\pi & S < 0, C \geq 0 \\ \text{indefinite} & S = 0, C = 0 \end{cases} \quad (2)$$

ادامه نمایش نموداری نیز عدم تناسب آن را نمایان خواهد کرد.

۲-۲-۲- نمودارها

در آمار دایره‌ای برای نمایش مشاهدات از آمارها و نمودارهای متفاوتی استفاده می‌شود. این آمارها و نمودارها به دو بخش تقسیم می‌شوند:

- ۱- آنهایی که هم در آمار دایره‌ای و هم در آمار خطی استفاده می‌شود مانند هیستوگرام.
- ۲- آنهایی که مخصوص آمار دایره‌ای می‌باشد مانند نمودار گل سرخی^۲.

یکی از مهمترین نمودارهایی که در داده‌های جهت‌دار استفاده می‌شود، هیستوگرام است. بدلیل ماهیت داده‌های جهتی استفاده از هیستوگرام‌های خطی مفید نبوده و لازم است که نمودار هیستوگرام بر روی دایره رسم شود. قابل ذکر است که با اعمال محدودیت بازه‌ای، مثلاً $[0, \pi]$ ، می‌توان داده‌های جهتی دایره‌ای را بعنوان داده‌های اقلیدسی در نظر گرفت و نمودارهای هیستوگرام را در صفحه مختصات دکارتی رسم کرد. همچنین برای نمایش فراوانی یا فراوانی نسبی داده‌های جهتی دایره‌ای از نمودارهای مختلفی می‌توان کمک گرفت

[۳ و ۶]، از آن جمله می‌توان به هیستوگرام دایره‌ای، نمودار گل سرخی و نمودار پیکانی نمونه‌ای اشاره کرد. قابل ذکر است که از میان نمودارهای موجود نمودار گل سرخی یکی از اصلی‌ترین نمودارهای جغرافیای اقلیم می‌باشد که در محافل

که در آن $S = \sum_{i=1}^n \sin(\theta_i)$ و $C = \sum_{i=1}^n \cos(\theta_i)$ می‌باشد [۶].

طول بردار میانگین (R) را اصطلاحاً برآیند میانگین^۱ می‌گویند که بصورت $R = \sqrt{C^2 + S^2}$ تعریف می‌شود.
واریانس:

واریانس داده‌های جهت‌دار دایره‌ای (σ^2) نیز مانند واریانس در داده‌های خطی تعبیر می‌شود، اما فرمول محاسبه آن تا حدودی متفاوت می‌باشد. ثابت می‌شود که واریانس در این حالت با برآیند میانگین رابطه دارد. به این ترتیب که $\sigma^2 = 1 - \bar{R}$ که در آن $\bar{R} = \frac{1}{n} R$ می‌باشد [۸].

انحراف معیار:

در داده‌های جهت‌دار مانند داده‌های خطی جذر واریانس را انحراف معیار نمی‌نامند و بجای آن مقدار $S = \sqrt{-2 \ln(\bar{R})}$ را انحراف معیار معرفی می‌نمایند.

با توجه به شاخص‌های معرفی شده فوق می‌توان آنها را برای زوایای باد ایستگاه سنندج به آسانی محاسبه نمود. میانگین، واریانس، انحراف معیار و برآیند میانگین ۸۴ داده میانگین ماهیانه جهت وزش باد عبارتند از: $\mu = -2/63$ رادیان

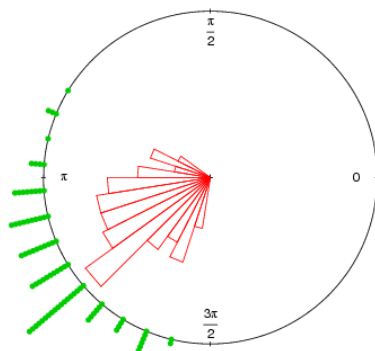
($-150/7617^\circ$)، $\sigma^2 = 0/0802$ ، $\bar{R} = 0/9197$. اگر مشاهدات فوق مشاهدات خطی در نظر گرفته شود میانگین و واریانس بترتیب برابر $\mu_1 = 3/64$ رادیان (معادل $280/97^\circ$) و $\sigma_1^2 = 0/1689$ خواهند شد. علاوه بر ضعف شهود میانگین در عدم وجودش در قسمت مرکزی مشاهدات، در

²Rose diagram

¹ Resultant mean

جغرافیایی ایران به نام نمودار گلباد معروف است. نمودار گلباد در بدست آوردن جهت باد غالب در منطقه دارای باد بکار می‌رود. این امر به کمک نمایش تصویری صورت می‌پذیرد. نکته قابل تأمل نقش تجربه و تخصص در عوض بکارگیری ابزارهای آماری در یافتن باد غالب است. نمودار گل سرخی در نرم‌افزار آماری R در صفحه مختصات دکارتی رسم می‌شود. با اینحال شخص می‌تواند آنرا به محیط برنامه‌نویسی و گرافیکی دیگری مانند Matlab منتقل نماید و نمودار سه‌بعدی واضح‌تری را رسم کند. هیستوگرام و نمودار گل سرخی خروجی نرم‌افزار (R) داده‌های ماهیانه زاویه جهت وزش باد در ایستگاه سنندج در شکل (۱) رسم شده است. از نمودار واضح است که باد غالب در منطقه سنندج در جهت

جنوب غربی است. علاوه بر آن بیشترین جهت باد در 150° - شرقی (جنوب غربی) است. محققین جغرافیا و هواشناسی همواره علاقمندند که جهت وزش باد را در آینده پیش‌بینی کنند. برای این منظور لازم است توزیع آماری مربوط به جهت وزش باد را یافته و به کمک آن داده‌ها را بیشتر ارزیابی کرد. پس از آن با در نظر گرفتن مدل رگرسیونی مناسب به پیش‌بینی جهت وزش باد پرداخت. اما هدف نویسندگان مقاله پیش‌بینی نبوده بلکه قصد یافتن توزیع مناسب دایره‌ای برای مجموعه داده‌ی موجود می‌باشد. در بخش بعد به این مسئله توجه شده است.



شکل ۱: هیستوگرام دایره‌ای و نمودار گل سرخی داده‌های ماهیانه زاویه جهت وزش باد در ایستگاه سنندج. با توجه به شکل میانگین جهت وزش باد غالب جنوب غربی می‌باشد.

توزیع‌های متفاوت تعریف شده بر روی دایره، بسته به اینکه چگونه تولید می‌شوند، به بخش‌های متفاوتی تقسیم می‌شوند. به عنوان مثال توزیع‌هایی که از پیچاندن یک توزیع یک متغیره یا دو متغیره بر روی محیط دایره بدست می‌آیند به توزیع‌های پیچشی معروفند. توزیع نرمال پیچشی یا ون میسز^۱ و توزیع کوشی پیچشی^۲ از نوعی هستند که از پیچیدن توزیع‌های یک متغیره نرمال و کوشی بر روی دایره بوجود می‌آیند. همچنین بعضی توزیع‌های دایره را می‌توان با انتگرال گیری متغیرهای چند متغیری اقلیدسی بدست آورد. بعنوان

۳- توزیع های دایره ای

برای متغیر تصادفی دایره‌ای X ، با فضای نمونه‌ای S $\chi \in [0, 2\pi]$ ، تابع توزیع $F_X(x)$ علاوه بر خواص مرسوم بایستی دارای ویژگی زیر نیز باشد.

$$F_X(x + 2\pi) - F_X(x) = 1, -\infty < x < \infty$$

این رابطه عنوان می‌کند که احتمال طول کمان 2π برابر یک باشد. برای هر تابع توزیع متغیر دایره‌ای، تابع چگالی متناظر به کمک روابط آمار خطی بدست می‌آید. در حالت پیوسته تابع چگالی احتمال این چنین متغیرهایی به ازای هر عدد صحیح k در خاصیت $f(\theta) = f(\theta + 2k\pi)$ صدق می‌کنند.

¹ Wrapped Normal(WN) or Von Mises(VM) distribution

² Wrapped Cauchy(WC) distribution

توزیع از نقطه نظر تئوری حذف خواهد شد و خواننده علاقمند به [۸] ارجاع داده می شود. جهت آشنایی مقدماتی نحوه رفتار توزیع ون میسر در مقابل مقادیر متفاوت κ در شکل (۲) نمایش داده شده است. از شکل (۲) واضح است که برای مقادیر ثابت μ ، افزایش κ منجر به گرایش توزیع ون میسر بصورت زنگی شکل حول میانگین می شود.

مثال توزیع نرمال دوخمی^۱ را می توان از توزیع نرمال دو متغیره بدست آورد [۸].

در بعضی موارد شباهت های جالبی بین توابع چگالی تعریف شده روی خط و دایره ای با شعاع واحد وجود دارد. بطور مثال توزیع یکنواخت برای متغیر جهتی دایره ای که دارای بازه تغییرات $[\alpha, \beta]$ به قسمی که

$$0 \leq \alpha < \beta \leq \alpha + 2\pi$$

می باشد عبارتست از $\frac{\beta - \alpha}{2\pi}$.

این تابع چگالی سنگ بنای محک داده های دایره ای در بررسی یکنواختی داده ها می باشد. توزیع ون میسر نیز یکی از انواع توزیع های دایره ای می باشد که مشابه چگالی نرمال در فضای اقلیدسی است. توزیع ون میسر را می توان از شرطی کردن توزیع نرمال دو متغیره نسبت به شعاع واحد بدست آورد. نام دیگر این توزیع، توزیع نرمال دایره ای می باشد و تابع چگالی آن بصورت زیر تعریف می شود،

$$g(\theta, \mu, \kappa) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa \cos(\theta - \mu)}, \quad 0 \leq \theta \leq 2\pi. \quad (۳)$$

که در آن $I_0(\kappa)$ ، تابع بسل از مرتبه صفر بوده که بصورت زیر نوشته می شود [۱]:

$$I_0(\kappa) = \frac{1}{2\pi} \int_0^{2\pi} e^{\kappa \cos(\theta)} d\theta = \sum_{r=0}^{\infty} \left(\frac{\kappa}{2}\right)^{2r} \left(\frac{1}{r!}\right)^2. \quad (۴)$$

در رابطه (۳) پارامتر μ ، میانگین توزیع و ثابت κ ، پارامتر تمرکز^۲ می باشد. در این توزیع اندازه برآیند میانگین برابر

$$R = A(\kappa) = \frac{I_1(\kappa)}{I_0(\kappa)}$$

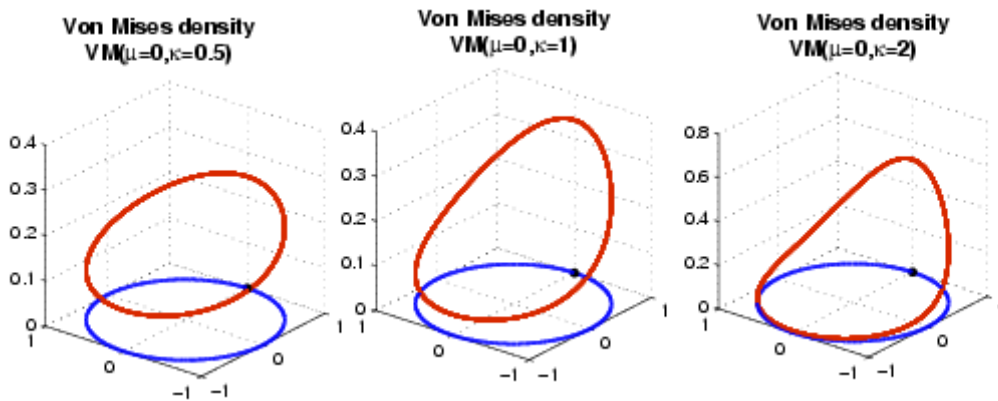
ماردیا و جاب بررسی شده است [۷].

می توان توزیع ون میسر را برای κ های بزرگ با توزیع نرمال تقریب زد. بدین صورت که اگر $\theta \sim VM(\mu, \kappa)$ باشد

آنگاه توزیع حدی $\frac{1}{\kappa^2}(\theta - \mu)$ وقتی که κ به سمت بی نهایت میل می کند، توزیع نرمال استاندارد است. توزیع ون میسر رفتارها و ویژگی هایی شبیه توزیع نرمال دارد. بدلیل جلوگیری از گستردگی مقاله حاضر بررسی رفتارهای این

^۱ Offset Normal(ON) distribution

^۲ Concentration parameter



شکل ۲: توابع چگالی ون میسز با مقادیر مختلف پارامترهای توزیع. نمودار تابع چگالی ون میسز با میانگین صفر و معیار تمرکز، از راست به چپ، ۱، ۲ و ۰/۵ را در صفحه سه بعدی نشان می دهد. نقطه برجسته سیاه رنگ روی دایره میانگین توزیع می باشد. مشاهده می شود با افزایش κ از پراکندگی داده ها حول میانگین کاسته می شود.

$$W_n^2 = \int_0^{2\pi} \left[(F_n - F) - \int_0^{2\pi} (F_n - F) dF \right] dF \quad (5)$$

هر چه این آماره بزرگتر باشد، تایید فرضیه H_0 با احتمال بیشتری مورد شک قرار می گیرد. واتسون توزیع مجانبی این آماره را با استفاده از تقریب ذیل محاسبه کرد.

$$\lim_{x \rightarrow \infty} P[W_n^2 \geq w] = 2 \sum_{m=1}^{\infty} (-1)^{m-1} e^{-2m^2 \pi^2 w} \quad (6)$$

برازش توزیع های متفاوت دایره ای به کمک آزمون واتسون براحتی امکان پذیر است. از نمودار شکل (۱) پیش بینی می شود که توزیع ون میسز بتواند برازش خوبی برای داده های باد در ایستگاه سنندج باشد. لذا در ادامه نحوه برازش این توزیع را به داده های باد در ایستگاه سنندج مورد بررسی قرار می دهیم. با در نظر گرفتن خطای آزمون ۰/۰۵ آزمون می کنیم که " آیا توزیع ون میسز به داده های سنندج می برازد؟ " با توجه به جدول (۱) قادر به رد فرض صفر نمی باشیم. لذا پیروی تغییرات داده های جهت باد ایستگاه سنندج از توزیع ون میسز را تایید می کنیم.

۴- مدل بندی داده های جهت باد

همانگونه که در آمار ناپارامتری برای داده های خطی می توان با استفاده از آزمون کرامر، که بر اساس تابع توزیع تجربی است، آزمون نیکویی برازش انجام داد، در داده های دایره ای نیز می توان این آماره را بگونه ای اصلاح کرد که نسبت به مبدا تغییرات پایا باشد [۳]. واتسون آماره دیگری را بر اساس این آماره معرفی کرد که به آماره واتسون معروف است و ثابت کرد که این آماره نسبت به دوران و انتقال پایا می باشد [۶].

فرض کنید x_1, \dots, x_n یک نمونه تصادفی n تایی از توزیع $F_X(x)$ باشند. فرضیه صفر آزمون نیکویی برازش را بصورت $H_0: F_X(x) = F_X(x_0)$ در نظر بگیرید که در آن توزیع F_0 ، توزیعی پیوسته و یک به یک باشد. حال اگر نمونه تصادفی را بصورت صعودی مرتب کرده $(x_{(1)} \leq \dots \leq x_{(n)})$ و تابع توزیع تجربی $F_n(x)$ را محاسبه کنیم، آنگاه مقدار آماره واتسون بر اساس توزیع F_n و F به صورت زیر بدست می آید.

جدول ۱: مقدار آماره آزمون و سطح معناداری مربوط به آزمون نیکویی برازش واتسون

p-value	مقدار w^2
۰/۱۸۷	۴/۳۱۸۸

ون میسز، که به خوبی به داده‌های جهت باد سنندج برازش شد، می‌توان به ارزیابی بیشتر داده‌ها پرداخت. بطور مثال با رسم نمودار P-P می‌توانیم بررسی کنیم که " آیا توزیع تجربی حاصل از داده‌ها و چگالی تئوری ون میسز با همدیگر مطابقت دارند یا خیر؟". نمودار سمت چپ در شکل (۳) تطبیق این دو توزیع را بخوبی نشان می‌دهد. علاوه بر این تابع توزیع تجمعی تجربی داده‌ها و تابع توزیع تجمعی ون میسز این تطبیق را هم تایید می‌نماید.

بیشتر نمودارهای موجود در صفحه مختصات دکارتی هستند. برای درک بیشتر تطبیق داده‌ها به توزیع ون میسز نمودار فراوانی نسبی در صفحه مختصات سه‌بعدی رسم و سپس توزیع ون میسز نیز در آن صفحه ترسیم شد. شکل این نمودار در شکل (۴) آمده است.

حال با تایید پیروی داده‌ها از توزیع ون میسز، می‌توان پارامترهای توزیع را نیز برآورد کرد. بست و فیشر روشی را برای محاسبه برآورد ماکزیمم درست‌نمایی پارامترهای توزیع ون میسز ارائه نمودند [۲]. اگر $\theta_1, \dots, \theta_n$ یک نمونه تصادفی از توزیع ون میسز با تابع چگالی (۳) باشد، به قسمی که هر دو پارامتر مجهول باشند برآورد ماکزیمم درست‌نمایی میانگین

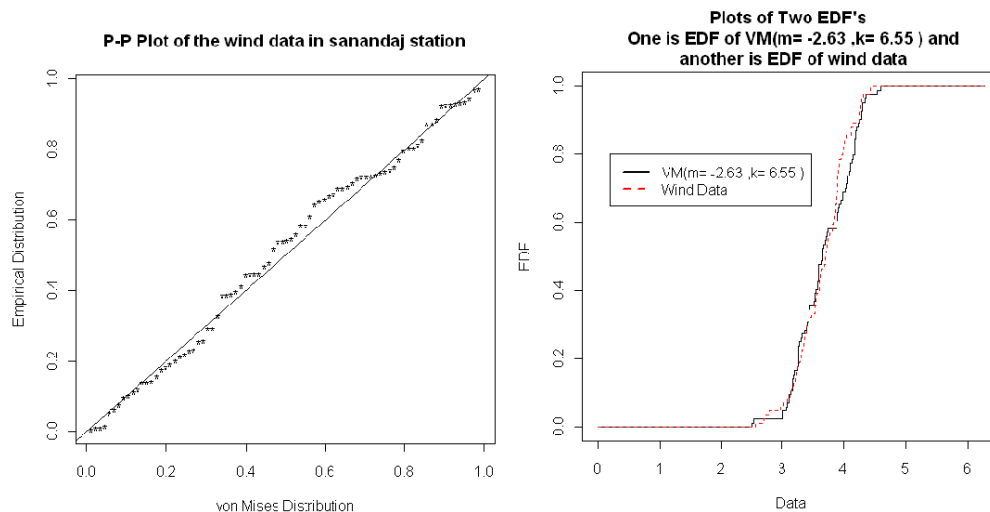
توزیع برابر $\hat{\mu} = \text{arctan}^* \left(\frac{S}{C} \right)$ خواهد بود که در آن تابع arctan^* در رابطه (۲) آمده است. برآورد ماکزیمم درست‌نمایی معیار تمرکز نیز از حل معادله

$$A(\hat{\kappa}) = \frac{I_1(\hat{\kappa})}{I_0(\hat{\kappa})} = \frac{R}{n}$$

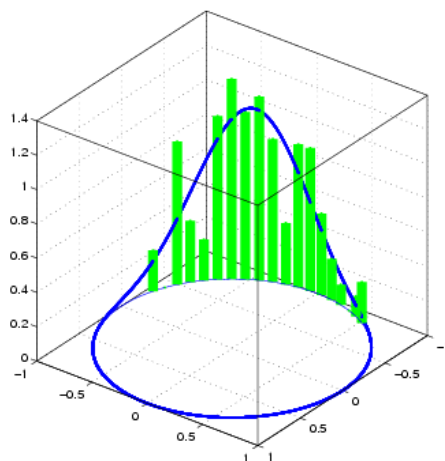
بدست می‌آید که در آن همان برآیند

$$R = \sqrt{C^2 + S^2} = \sum_{i=1}^n \cos(\theta_i - \hat{\mu})$$

میانگین می‌باشد. با در دست داشتن برآورد پارامترهای توزیع



شکل ۳: راست: توزیع تجربی داده‌های ماهیانه وزش باد در ایستگاه سنندج (خط‌چین) و توزیع ون میسز (خط ممتد). پارامترهای توزیع از روش ماکزیمم رست‌نمایی برآورد شده است. چپ: نمودار P-P داده‌های ماهیانه جهت باد در ایستگاه سنندج. برازش مناسب داده‌ها به توزیع ون میسز نمایان است.



شکل ۴: نمودار فراوانی نسبی داده‌های وزش باد در ایستگاه سنندج و تابع چگالی ون میسز. پارامترهای تابع چگالی به کمک داده‌های باد بصورت $\mu = -2/63$ و $\kappa = 6/54$ برآورد شده است.

۵- نتیجه گیری

گرفته شده و انتظار می‌رود نقطه شروعی برای تحقیقات بیشتر در حوزه آمار غیرخطی باشد.

متغیرهای بیشمار در مسائل کاربردی وجود دارند که ماهیت غیراقلیدسی دارند. از آن جمله متغیرهایی هستند که دارای دور یا گردش بوده و بخوبی با آمار دایره‌ای تجزیه و تحلیل می‌شوند. نادیده گرفتن ماهیت دایره‌ای چنین متغیرهایی و اعمال تحلیل‌های مرسوم آمار خطی منجر به نتایج نامناسبی خواهد شد.

تقدیر و تشکر

داده‌های موجود در این مقاله توسط آقای اسدا... خورانی در اختیار نگارندگان قرار گرفت. بدین وسیله از ایشان کمال تشکر و قدردانی را داریم. همچنین از کمیته علمی و برگزارکنندگان دهمین کنفرانس آمار ایران در فراهم کردن زمینه‌ای جهت ارائه این مقاله سپاسگزاریم.

در این مقاله با معرفی آمار دایره‌ای و بررسی چند مفهوم آماری، داده‌های جهت باد ایستگاه سنندج مورد مطالعه قرار گرفت. پیروی داده‌ها از توزیع ون میسز مورد تایید قرار گرفت و پارامترهای توزیع بر اساس مشاهدات موجود برآورد شد. این نتایج راهی را برای استنباط بیشتر در آینده فراروی علاقمندان قرار می‌دهد.

منابع

- [1]. Abramowitz, M. and Stegun, I. A. (1965). Handbook of mathematical Functions. Dover Pub., New York.
- [2]. Best, D. and Fisher N. (1981). The bias of the maximum likelihood estimators of the von Mises-Fisher concentration parameters. Communications in Statistics – Simulation and Computation, 10(5), 493-502.
- [3]. Fisher, N., Lewis, T. and Embleton, B. (1993). Statistical analysis of spherical Data. Cambridge University Press, Cambridge.
- [4]. Hassan, S. F. and Hussin, A. G. and Zubairi, Y. Z. (2009). Analysis of Malaysian wind Direction Data Using ORIANA. Modern Applied Science, 3(3), 115-119.

اگرچه از نقطه نظر تئوری پیشرفتهای خوبی در زمینه آمار دایره‌ای صورت گرفته اما بسته‌های نرم‌افزاری کمتری در این خصوص وجود دارد. در نرم‌افزار آماری R [۹] بسته کتابخانه‌ای Circular بخشی از نیازمندی‌های محققین آمار دایره‌ای را برآورده می‌سازد. نرم‌افزاری بنام ORIANA نیز اخیراً توسعه یافته و برای داده‌های جهت باد در مالزی مورد استفاده قرار گرفت [۴]. مطالب ارائه شده در این مقاله به عنوان انگیزه‌ای برای محققین در حوزه‌های جدید آمار در نظر

- [5]. Hussain, F., Zubairi, Y. Z. and Hussin, A. G. (2007). An alternative diagrammatical representation of wind data. World Scientific and Engineering Academy and Society (WSEAS). 292-294.
- [6]. Jammalamadaka, S. R. and Sengupta, A. (2001). Topics in circular statistics, World Scientific Publishing Co., Inc, Singapore.
- [7]. Khatri, C. G. and Mardia, K. V. (1977). The von Mises-Fisher Matrix distribution in orientation statistics. Journal of the Royal Statistical Society, Ser. B, 39, 95-106.
- [8]. Mardia, K. V. and Jupp, P. E. (2000). Directional statistics, Wiley, Chichester.
- [9]. R Development Core Team (2009). R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- [10]. Watson, G.S. (1962). Goodness-of-fit tests on a circle. II., Biometrika, 49, 57-63.
- [11]. Zubairi, Y., Hussain, F. and Hussin, A. (2008). An Alternative Analysis of Two Circular Variables via Graphical Representation. Computer and Information Science, 1(4), 3-8.