

تحلیل تابعی دما و بارندگی در ایران با استفاده از مؤلفه‌های اصلی تابعی

نورالله تازیکه میاندره و سید محمدابراهیم حسینی نسب*

دانشگاه تربیت مدرس

چکیده. وقتی که مشاهدات، توابع پیوسته‌ای از یک متغیر (معمولاً زمان) هستند؛ استفاده از روش‌های متداول در آمار کلاسیک برای تحلیل آن‌ها ممکن است به چالش کشیده شود. بنا بر این برای تحلیل آن‌ها انطباق‌هایی در تئوری‌های موجود، لازم به نظر می‌رسد. چنین اقدامی به رویکرد جدیدی در آمار منتهی می‌شود که به آن تحلیل داده‌های تابعی (FDA) گویند. با توجه به این‌که در این رویکرد، مشاهدات به فضاهای تابعی با بُعد بی‌نهایت متعلق می‌باشند، برای تحلیل آن‌ها ناگزیر به کاهش بُعد هستیم. همانند یک مطالعه‌ی چندمتغیره، یکی از روش‌های متداول برای این کار استفاده از تحلیل مؤلفه‌های اصلی (PCA) می‌باشد. اما استفاده از PCA برای داده‌های تابعی نیز منوط به انطباق‌هایی در تئوری‌های مربوط و ارائه‌ی آن‌ها برای این‌گونه داده‌ها می‌باشد.

در این مقاله ابتدا تعاریف و مفاهیم لازم برای بحث PCA در حالت تابعی ارائه می‌شود. سپس به تحلیل داده‌های دما و مقدار بارندگی در ایران و استخراج الگوهای پراکندگی و همبستگی بین این دو پدیده و نیز تفسیر آن‌ها می‌پردازیم. داده‌های موجود که از ایستگاه‌های مختلف هواشناسی در سراسر ایران جمع‌آوری شده‌اند، به صورت میانگین ماهانه‌ی دما و مقدار بارندگی در طول سال ۲۰۰۵ بوده‌اند. به دلیل ماهیت تابعی دما و مقدار بارندگی (تابعی از زمان)، اولین گام در تحلیل آن‌ها به روش تابعی، برازش منحنی‌های مناسب به این داده‌های گسسته‌ی اولیه می‌باشد که به آن ثبت نام داده‌ها گویند. از طرف دیگر، مشاهدات اولیه ممکن است با خطا اندازه‌گیری شوند، از این رو لازم است تا نخست با استفاده

* نویسنده‌ی عهده‌دار مکاتبات

از یکی از روش‌های هموارسازی به ثبت نام داده‌ها پرداخته، سپس با به‌کارگیری PCA اقدام به تحلیل آن‌ها نمود. در پایان فواصل اطمینان بوت‌استرپی برای مقادیر و توابع ویژه برای داده‌های دما و بارندگی به دست آورده شده است.

واژگان کلیدی. تحلیل داده‌های تابعی؛ تحلیل مؤلفه‌های اصلی تابعی؛ ثبت نام داده‌ها؛ فاصله‌ی اطمینان بوت‌استرپی.

۱ مقدمه

وقتی مشاهدات دارای ماهیت تابعی باشند، روش‌های متداول در تحلیل چندمتغیره برای تحلیل آن‌ها ممکن است به چالش کشیده شوند (رمسی، ۱۹۸۲). هرچند که داده‌هایی از قبیل داده‌های روزانه یا ماهانه‌ی مالی یا هواشناسی که در طول یک سال مشاهده می‌شوند می‌توانند همانند یک مسئله‌ی چندمتغیره به‌صورت بردارهایی به طول ۱۲ یا ۳۶۵ در نظر گرفته و تحلیل شوند، چالش‌هایی که در تفسیر نتایج به‌واسطه‌ی تغییرپذیری بسیار پایین بین مؤلفه‌ها و نیز از دست دادن توان آماری برای استنباط در مورد پارامترها حاصل می‌شود، محققان را بر آن می‌دارد که یک منحنی هموار به این ۱۲ یا ۳۶۵ نقطه برازش دهند و هر مجموعه داده‌ی سالانه را به‌صورت یک تابع در نظر بگیرند و تحلیل کنند. چنین اقدامی به رویکرد جدیدی در آمار منتهی می‌شود که تحلیل داده‌های تابعی^۱ (FDA) نامیده می‌شود (رمسی و سیلورمن، ۲۰۰۵). در این گرایش، داده‌ها در ماهیت واقعی خود، یعنی توابع پیوسته، بررسی می‌شوند و اثر زمان در کل بازه‌ی زمانی و نه لزوماً در زمان‌های گسسته یا فواصل زمانی برابر، مورد توجه می‌باشد. این موضوع نیز یکی از مزیت‌های دیگر این روش را نسبت به روش‌های متداول در آمار کلاسیک نشان می‌دهد.

داده‌های چندمتغیره بر اساس ماتریس مشاهدات تحلیل می‌شوند که در آن، مؤلفه‌ی x_{ij} نشان‌دهنده‌ی مقدار متغیر j ام برای آزمودنی i ام می‌باشد. اما در FDA، مقدار متغیر پاسخ برای آزمودنی i ام به‌صورت $x_i(t)$ ، $t \in I = [0, T]$ ، نمایش داده می‌شود. با وجود این تفاوت، با اقتباس از نگاشت‌های ماتریسی در حالت چندمتغیره، نگاشت‌های تابعی برای مشاهداتی با بُعد بی‌نهایت قابل تعریف هستند (رمسی، ۱۹۸۲). در عمل، پدیده‌های تابعی معمولاً در زمان‌های گسسته، $t_1 = 0 < t_2 < \dots < t_n = T$ ، مشاهده می‌شوند. در نتیجه، نخستین گام در FDA تبدیل این مشاهدات گسسته به قالب «تابعی» می‌باشد به طوری که هر تابع را بتوان برای تمام مقادیر t در بازه‌ی $[0, T]$ مورد ارزیابی قرار داد. این تبدیل با استفاده از روش‌های درون‌یابی یا روش‌های هموارسازی امکان‌پذیر است. شایان ذکر است که روش‌های هموارسازی اغلب مبتنی بر به‌کارگیری یک مجموعه از توابع پایه‌ی مناسب می‌باشند. از طرف دیگر چون داده‌های مشاهده‌شده در اکثر مواقع با خطا اندازه‌گیری می‌شوند، در هنگام برازش مدل از روش‌های هموارسازی

استفاده می‌شود و در آن‌ها سعی بر این است که خطا تا حد امکان پالایش شود. آن‌گاه در رویکرد FDA، هر تابع برازنده شده به عنوان یک مشاهده در نظر گرفته می‌شود. در ادامه، بعد از تعریف آماره‌های توصیفی تابعی در بخش ۲، تعاریف و مفاهیم در بحث تحلیل مؤلفه‌های اصلی تابعی^۲ (FPCA) ارائه می‌شوند. ساختار و نحوه‌ی انتخاب داده‌های دما و مقدار بارندگی در بخش ۴ بررسی خواهد شد. در بخش ۵، تحلیل تابعی برای نمونه‌ی منتخب، بررسی و نتایج حاصل بیان می‌شود.

۲ آماره‌های توصیفی برای داده‌های تابعی

در این بخش، نخست آماره‌های توصیفی در FDA را معرفی و سپس این آماره‌ها را برای مجموعه داده‌های دما و میزان بارندگی محاسبه و نتایج را ارائه می‌کنیم.

۲٫۱ میانگین و واریانس تابعی

میانگین یک مجموعه از توابع را می‌توان با استفاده از میانگین نقطه‌به‌نقطه‌ی تمامی آزمودنی‌ها به دست آورد. از این رو، تابع میانگین به صورت متوسط تکرارهای عرضی توابع، قابل تعریف است:

$$(۱) \quad \bar{x}(t) = \frac{1}{n} \sum_{i=1}^n x_i(t).$$

در عمل، معمولاً با استفاده از روش‌های هموارسازی مناسب، از جمله اسپلاین‌ها، باید مشاهدات اولیه را به صورت منحنی‌ها ثبت نام (registration) و سپس تابع میانگین را محاسبه نمود. همچنین تابع واریانس نمونه‌ای به صورت

$$(۲) \quad \text{var}_x(t) = \frac{1}{n-1} \sum_{i=1}^n [x_i(t) - \bar{x}(t)]^2,$$

تعریف می‌شود. در واقع این تابع، مقدار واریانس را در هر مقدار t محاسبه می‌کند.

۲٫۲ تابع کوواریانس و همبستگی

در تحلیل چندمتغیره، میزان همبستگی بین هر جفت از متغیرها به وسیله‌ی مقدار کوواریانس نشان داده می‌شود. مشابه با آن، چنین عملی در حالت تابعی با ثابت نگه داشتن دو مقدار زمانی، یعنی t_1 و t_2 و از طریق محاسبه‌ی

$$\text{cov}_X(t_1, t_2) = \frac{1}{n-1} \sum_{i=1}^n \{x_i(t_1) - \bar{x}(t_1)\} \{x_i(t_2) - \bar{x}(t_2)\}$$

امکان‌پذیر می‌باشد. با توجه به رابطه‌ی فوق، تابع همبستگی متناظر برابر است با

$$(3) \quad \text{corr}_X(t_1, t_2) = \frac{\text{cov}_X(t_1, t_2)}{\sqrt{\text{var}_X(t_1)\text{var}_X(t_2)}}.$$

همان‌طور که مشاهده می‌شود فرمول (۳) نیز با فرمول ضریب همبستگی برای هر جفت از داده‌های چندمتغیره مشابهت دارند. اما بر اساس آنچه در بالا دیده شد، در FDA برای بررسی تغییرات به جای ماتریس‌ها از توابع استفاده می‌شود. از این رو، می‌توان برای نمایش این ناحیه، از روش ترسیمی و از نمودارهای خطوط تراز (contour plots) یا سه‌بعدی پرسپکتیو (perspective plots) بهره برد.

۲٫۳ توابع کوواریانس و همبستگی متقابل

در بررسی پدیده‌های تابعی ممکن است دو یا چند متغیر تحت بررسی قرار گیرند. مثلاً در تحلیل داده‌های هواشناسی ممکن است متغیرهای دما و مقدار بارندگی مورد مطالعه قرار گیرند. در این گونه موارد، امکان بررسی ساختار همبستگی متغیرها با استفاده از توابع کوواریانس یا همبستگی متقابل وجود دارد. اگر مشاهدات ما به صورت زوج‌های $(x_i(t_1), y_i(t_2))$ باشند، تابع کوواریانس متقابل (cross-covariance function) این دو متغیر به صورت

$$\text{cov}_{X,Y}(t_1, t_2) = \frac{1}{n-1} \sum_{i=1}^n \{x_i(t_1) - \bar{x}(t_1)\} \{y_i(t_2) - \bar{y}(t_2)\},$$

و تابع همبستگی متقابل (cross-correlation function) آن‌ها به شکل

$$\text{corr}_{X,Y}(t_1, t_2) = \frac{\text{cov}_{X,Y}(t_1, t_2)}{\sqrt{\text{var}_X(t_1)\text{var}_Y(t_2)}},$$

تعریف می‌شود. از این رو، امکان نمایش بصری چنین توابعی نیز با استفاده از نمودارهای خطوط تراز یا سه‌بعدی وجود دارد.

۳ تحلیل مؤلفه‌های اصلی تابعی

استفاده از PCA به عنوان رویکردی استاندارد در بررسی ساختار داده‌ها و الگوهای پراکندگی موجود در داده‌های چندمتغیره، بسیار رایج است. در واقع، PCA با تجزیه‌ی ماتریس واریانس-کوواریانس در پی یافتن جهت‌هایی در فضای مشاهدات است که در آن راستا بیش‌ترین تغییرات وجود دارد. در این رویکرد برای هر مؤلفه‌ی اصلی، یک بردار وزنی که جهت تغییرات آن متناظر با مؤلفه‌ی اصلی است نمایش داده می‌شود (جولیفه، ۲۰۰۲). در حالت تابعی، به دلیل آن‌که مشاهدات به صورت تابع هستند، هر مؤلفه‌ی اصلی از طریق تابع وزنی $\psi(t)$ مشخص و روی بازه‌ی زمانی متناظر با داده‌های تابعی تعریف می‌شود. فرض کنید داده‌های تابعی $x_1(t), \dots, x_n(t)$ موجود باشند. آنگاه امتیازات متناظر با هر مؤلفه‌ی اصلی برای آزمودنی i ام به صورت زیر محاسبه می‌شود:

$$z_i = \int \psi(t)x_i(t)dt.$$

هدف در FPCA، یافتن یک تابع وزنی مانند $\psi_1(t)$ می‌باشد که $\frac{1}{n} \sum_{i=1}^n (z_{i1} - \bar{z})^2$ واریانس امتیازات مؤلفه‌های اصلی اول (z_{i1}) را نسبت به قید

$$(4) \quad \int \psi(t)^2 dt = 1$$

ماکسیمم می‌کند، که در آن، z_{i1} امتیاز مؤلفه‌ی اصلی اول برای مشاهده‌ی i ام می‌باشد. مؤلفه‌های اصلی دیگر نیز مشابه با مؤلفه‌ی اصلی اول به دست می‌آیند. به عنوان مثال، تابع مؤلفه‌ی اصلی دوم $\psi_2(t)$ طوری تعریف می‌شود که واریانس امتیازات مؤلفه‌ی اصلی را نسبت به قید (۴) ماکسیمم کند و علاوه بر آن در شرط زیر نیز صادق باشد:

$$\int \psi_2(t)\psi_1(t)dt = 0.$$

در حالت کلی، برای j امین مؤلفه داریم:

$$\int \psi_j(t)\psi_1(t)dt = \int \psi_j(t)\psi_2(t)dt = \dots = \int \psi_j(t)\psi_{j-1}(t)dt = 0,$$

که متعامد بودن توابع مؤلفه‌ی اصلی را تضمین می‌کند.

۳/۱ نظریه برای تحلیل مؤلفه‌های اصلی تابعی

به‌طور مشابه با تحلیل چندمتغیره، از تحلیل مؤلفه‌های اصلی به‌طور وسیع در مطالعه‌ی داده‌های تابعی استفاده می‌شود؛ زیرا با استفاده از آن، مسئله‌ای که دارای بُعد نامتناهی است به‌صورت مسئله‌ای با بُعد متناهی قابل تحلیل می‌شود. برای اطلاعات بیشتر در این زمینه، برای مثال می‌توان به فصل ۸ از رمسی و سیلورمن (۲۰۰۵)، هال و حسینی نسب (۲۰۰۶ و ۲۰۰۷) و چندین مسئله‌ی واقعی که به‌وسیله‌ی رمسی و سیلورمن (۲۰۰۲) حل شده‌اند رجوع کرد.

۳/۱/۱ هسته برای تحلیل مؤلفه‌های اصلی

فرض کنید $X(t)$ یک تابع تصادفی (فرایند تصادفی) باشد که روی فاصله‌ی $I = [a, b]$ تعریف شده است و در شرط $\int_I E[X^2(t)] dt < \infty$ صدق می‌کند. اگر $\mu(\cdot) = E[X(\cdot)]$ باشد، برای هر $u, v \in I$ تعریف می‌کنیم:

$$K(u, v) \equiv \text{cov}\{X(u), X(v)\} = E\{(X(u) - \mu(u))(X(v) - \mu(v))\}.$$

همچنین فرض کنید مجموعه‌ی توابع حقیقی مقدار روی I را که توان دوم آن‌ها انتگرال‌پذیر است با $L_2(I)$ نشان دهیم. در این صورت، تبدیل K که هر $\psi \in L_2(I)$ را به $K\psi$ می‌برد به‌صورت زیر تعریف می‌شود:

$$(K\psi)(x) = \int K(x, y) \psi(y) dy.$$

K یک عملگر فشرده و هیلبرت‌اشمیت روی $L_2(I)$ است و آن را عملگر کوواریانس تابع تصادفی $X(t)$ می‌نامیم. عملگر کوواریانس K ، متقارن و نیمه‌معین مثبت است و نیز هسته‌ی آن، یعنی $K(\cdot, \cdot)$ قطری‌شدنی است. به عبارت دیگر،

$$(5) \quad K(u, v) = \sum_{j=1}^{\infty} \theta_j \psi_j(u) \psi_j(v),$$

که در آن، θ_j ها ویژه‌مقدارهای عملگر K ، و ψ_j ها توابع ویژه‌ی متناظر با آن‌ها هستند. به‌خصوص می‌توان گفت که $K\psi_j = \theta_j \psi_j$.

فرض می‌کنیم که θ_j ها در (۵) به‌صورت $\theta_1 \geq \theta_2 \geq \dots \geq 0$ مرتب می‌شوند. اگر θ_j ها مثبت باشند، ψ_j ها (توابع ویژه) یک پایه‌ی متعامد کامل در $L_2(I)$ تشکیل می‌دهند.

۳/۱/۲ بسط کارهونن-لوی

تابع تصادفی $X(t)$ دارای بسط کارهونن-لوی

$$(۶) \quad X(t) - \mu(t) = \sum_{j=1}^{\infty} \xi_j \psi_j(t),$$

است که در آن $\xi_j = \int_I (X(t) - \mu(t)) \psi_j(t) dt$ نشان‌دهنده‌ی زامین امتیاز $X(t)$ می‌باشد. سری (۶) در $L_2(I)$ همگرا است و متغیرهای ξ_1, ξ_2, \dots ناهمبسته‌اند؛ زیرا

$$E(\xi_j \xi_k) = \int_I \int_I \psi_j(u) K(u, v) \psi_k(v) du dv = \theta_j \delta_{jk},$$

که در آن، δ_{jk} دلتای کرونه‌کر است. علاوه بر آن، $E(\xi_j) = 0$ ، $E(\xi_j^2) = \theta_j$ ، و

$$\int_I E[(X(t) - \mu(t))^2] dt = \sum_{j=1}^{\infty} \theta_j < \infty.$$

۳/۱/۳ برآوردهای تجربی

اگر $X_1(t), \dots, X_n(t)$ یک نمونه به اندازه‌ی n باشد، برآورد تجربی $K(u, v)$ به صورت زیر تعریف می‌شود:

$$\widehat{K}(u, v) = \frac{1}{n} \sum_{i=1}^n \{X_i(u) - \bar{X}(u)\} \{X_i(v) - \bar{X}(v)\},$$

که در آن $\bar{X}(\cdot)$ معرف تابع میانگین نمونه‌ی فوق است. مشابه با عملگر K ، عملگر \widehat{K} که به صورت

$$(\widehat{K}\widehat{\psi})(x) = \int \widehat{K}(x, y) \widehat{\psi}(y) dy,$$

تعریف می‌شود، یک عملگر فشرده و هیلبرت اشمیت روی $L_2(I)$ است و نیز

$$\widehat{K}(u, v) = \sum_{j=1}^{\infty} \hat{\theta}_j \widehat{\psi}_j(u) \widehat{\psi}_j(v),$$

که در آن $\hat{\theta}_j$ ها ویژه‌مقدارهای عملگر \widehat{K} هستند و $\widehat{\psi}_j$ ها توابع ویژه‌ی متناظر با آن‌ها. همچنین عملگر \widehat{K} متقارن و نیمه‌معین مثبت است و متغیرهای $\hat{\theta}_j$ به صورت $\hat{\theta}_1 \geq \hat{\theta}_2 \geq \dots \geq 0$ مرتب شده‌اند. به طور

مشابه با (۶)،

$$X_i(t) - \bar{X}(t) = \sum_{j=1}^{\infty} \hat{\xi}_{ij} \hat{\psi}_j(t),$$

که در آن $\hat{\xi}_{ij} = \int_I (X_i - \bar{X})(t) \hat{\psi}_j(t) dt$ عبارت است از زامین امتیاز تجربی $X_i(t)$. علاوه بر آن،

$$\frac{1}{n} \sum_{i=1}^n \hat{\xi}_{ij} \hat{\xi}_{ik} = \int_I \int_I \hat{\psi}_j(u) \hat{K}(u, v) \hat{\psi}_k(v) du dv = \hat{\theta}_j \delta_{jk}.$$

۳/۲ هموارسازی در FPCA

فرایند به دست آوردن توابع مؤلفه‌ی اصلی هموار، مشابه با این فرایند برای یافتن توابع مؤلفه‌ی اصلی، بدون استفاده از هموارسازی می‌باشد، با این تفاوت که در هنگام ماکسیم کردن واریانس، باید از نرم دیگری به جای نرم به کار گرفته شده در طرف چپ رابطه‌ی (۴) استفاده نمود (پتسولی و سیلورمن، ۱۹۹۳).

برای به دست آوردن FPCA هموار، نه تنها اندازه‌ی ξ ، بلکه ناهمواری (roughness) آن نیز باید کنترل شود. با این هدف، شرط (۴) جایگزین قیدی می‌شود که ناهمواری را نیز به خوبی تضمین می‌کند. بنا بر این، برای یافتن اولین مؤلفه‌ی اصلی تابعی $\psi_1(t)$ (FPC) هموار، تابع $\psi_1(t)$ طوری تعیین می‌شود که واریانس امتیازات مؤلفه‌ی اصلی را تحت قید

$$(۷) \quad \int \psi(t)^2 dt + \alpha \int \psi''(t)^2 dt = 1$$

ماکسیم کند. شایان ذکر است که پارامتر $\alpha \geq 0$ در رابطه‌ی فوق، مقدار «همواری» را کنترل می‌کند؛ از این رو، به آن «پارامتر همواری» گفته می‌شود.

در ضمن، تاوان ناهمواری (roughness penalty)، غیر از قیده‌های اضافی، برای FPC‌های هموار شده‌ی دوم و سوم و مراتب بالاتر ضمیمه می‌شود. حال برای یافتن دومین مؤلفه‌ی اصلی تابعی، تابع $\psi_2(t)$ به گونه‌ای اختیار می‌شود که واریانس امتیازات مؤلفه‌ی اصلی را نسبت به قید (۷) و

$$\int \psi_2(t) \psi_1(t) dt + \alpha \int \psi_2''(t) \psi_1''(t) dt = 0$$

ماکسیم کند. این روند تا به دست آوردن تمامی مؤلفه‌های اصلی تابعی ادامه دارد.

۴ داده‌های دما و مقدار بارندگی

دو پدیده‌ی بارندگی و دما ماهیت تابعی دارند، زیرا مقادیر بارندگی و دما در هر منطقه به زمان بستگی دارند، و در طول زمان‌های مختلف در حال تغییر می‌باشند؛ از این رو، استفاده از FDA برای تحلیل آن‌ها توجیه‌پذیر است. داده‌هایی که در این مقاله مورد تحلیل قرار گرفته‌اند از داده‌هایی که سازمان هواشناسی ایران در سال ۲۰۰۵، از ۱۷۶ ایستگاه سینوپتیکی در نقاط مختلف کشور جمع‌آوری کرده است انتخاب شده‌اند. در این مجموعه داده، برای هر ایستگاه، ۱۲ عدد که همان میانگین ماهانه‌ی دما و مقدار بارندگی در آن منطقه می‌باشند گزارش شده است.

به دلیل همبستگی مکانی بین ایستگاه‌ها (معمولاً میزان بارندگی و دما که از ایستگاه‌های مجاور جمع‌آوری می‌شوند به هم وابسته‌اند)، بررسی مشاهدات مربوط به هر ایستگاه در قالب تابعی مجاز نیست؛ چرا که فرض استقلال بین داده‌ها در این حالت نقض می‌شود. این مشکل را می‌توان با در نظر گرفتن مشخصه‌های عرض جغرافیایی و ارتفاع مربوط به ایستگاه‌ها به‌طور نسبی رفع کرد؛ زیرا کارشناسان هواشناسی این عوامل را مؤثرترین عوامل در وابستگی مکانی بین ایستگاه‌ها معرفی می‌کنند. به‌عنوان مثال، در ۳۴ درجه‌ی عرض جغرافیایی که از رشته‌کوه زاگرس عبور می‌کند، مقدار دما و بارندگی در دو طرف این رشته‌کوه را نمی‌توان وابسته به هم دانست (تأثیر مشخصه‌ی عرض جغرافیایی). این موضوع با این واقعیت که در دو طرف این رشته‌کوه، دو ناحیه‌ی متفاوت معتدل و گرم و خشک قرار می‌گیرند نیز سازگار است. به دلیل آن‌که شهرهای دارای ارتفاع زیاد معمولاً هوای سردتر و در عین حال باثبات‌تری نسبت به شهرهای دارای ارتفاع کم دارند، در نظر گرفتن معیار ارتفاع می‌تواند منجر به انتخاب ایستگاه‌هایی در هر طرف این رشته‌کوه شود که مقادیر دما و بارندگی آن‌ها به هم وابسته نباشند؛ یعنی از نظر ارتفاع، تفاوت زیادی با هم دارند (تأثیر عامل ارتفاع در آب و هوا). شایان ذکر است که در عمل، امکان تأثیرگذاری عوامل دیگری بر بعضی از ایستگاه‌های نزدیک به هم یا دارای ارتفاع تقریباً یکسان وجود دارد به‌طوری که شرایط جوی کاملاً متفاوتی را برای این ایستگاه‌ها پدید می‌آورد. بنا بر این، در هنگام بررسی دو شرط فوق، باید دیگر شرایط تأثیرگذار نیز بررسی شوند تا چنین ایستگاه‌هایی کنار گذاشته نشوند. از این رو، برای حصول نتایج بهتر، از مقایسه‌ی بصری نیز بهره جسته‌ایم. بر اساس مطالب فوق و نظر کارشناسان هواشناسی، در نهایت، ۶۶ ایستگاه انتخاب شد و داده‌های دما و بارندگی مربوط به آن‌ها مورد تحلیل قرار گرفت.

همان‌طور که در مقدمه ذکر شد، در ابتدا باید منحنی‌های مناسب، طوری که مجموعه داده‌ها برازش شوند که خطای اندازه‌گیری تا حد ممکن پالایش شود. به دلیل ماهیت دوره‌ای (تناوبی) پدیده‌های دما و بارندگی، برای برازش منحنی به مجموعه داده‌ها از بسط سری‌های فوریه استفاده نموده‌ایم. در نهایت، استفاده از یک بسط فوریه با ۱۳ جمله، برازش مناسبی به داده‌ها ایجاد نمود؛ یعنی برای هر ایستگاه، برای برازش

منحنی $x(\cdot)$ ضرایب $\beta_1, \beta_2, \dots, \beta_r$ در سری

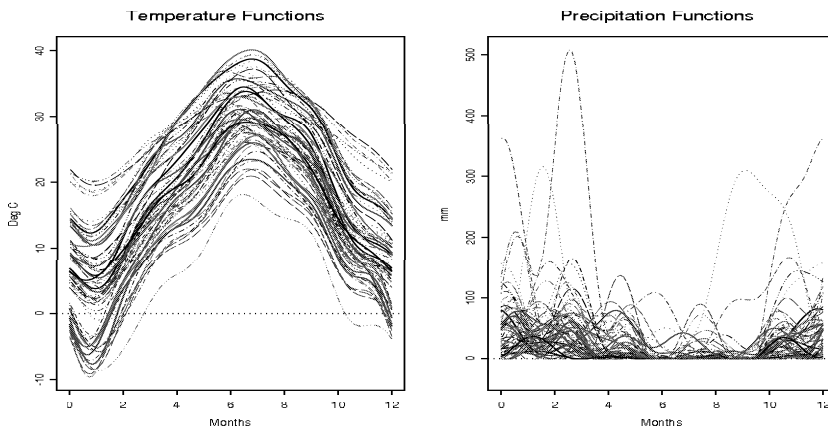
$$x(t) \approx \beta_0 + \beta_1 \sin\left(\frac{\pi}{6}t\right) + \beta_2 \cos\left(\frac{\pi}{6}t\right) + \beta_3 \sin\left(\frac{\pi}{3}t\right) + \beta_4 \cos\left(\frac{\pi}{3}t\right) \\ + \dots + \beta_{1r} \sin(2\pi t) + \beta_{1r} \cos(2\pi t)$$

برآورد شده است. نمودار داده‌های مورد بررسی در شکل ۱ آورده شده است.

۵ تحلیل داده‌های دما و مقدار بارندگی

در ادامه، تحلیل اولیه‌ی داده‌های تابعی دما و بارندگی با استفاده از FDA مورد بحث قرار می‌گیرد. برای این منظور، در ابتدا آماره‌های توصیفی به دست می‌آیند و سپس با استفاده از تحلیل مؤلفه‌ی اصلی تابعی، ساختار همبستگی داده‌ها مورد بررسی قرار می‌گیرد. شایان ذکر است که تحلیل جامع‌تر را می‌توان با بررسی FPCA چندمتغیره‌ی فواصل بوت‌استریپی دنبال کرد، که در این جا به این موارد پرداخته نخواهد شد.

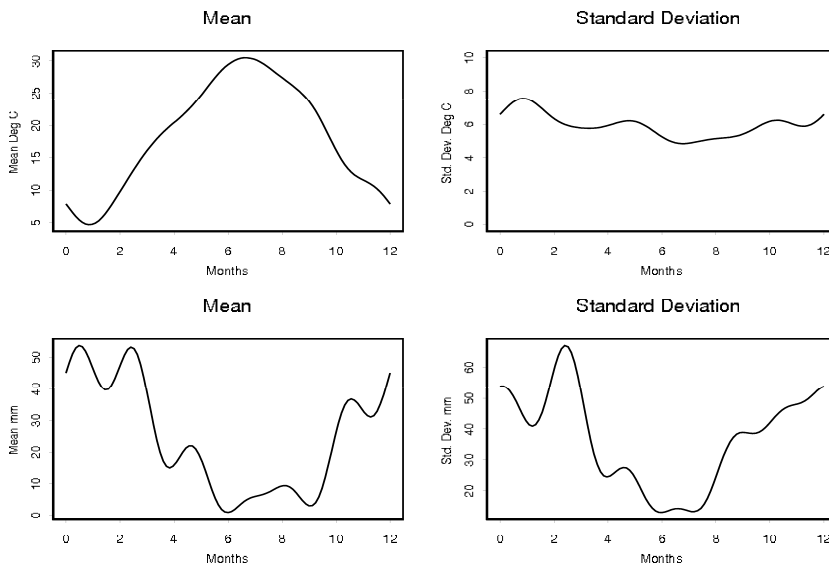
شکل ۲ توابع میانگین و انحراف معیار را برای داده‌های دما و مقدار بارندگی نشان می‌دهد. بر اساس نمودار میانگین که از رابطه‌ی (۱) به دست آمده است، مشخصه‌های کلی زیر برای پدیده‌ی دما قابل بیان می‌باشند: به‌طور متوسط در طول سال، دمای تقریباً زیر ۵ درجه و بیش‌تر از ۳۰ درجه رخ نمی‌دهد و دامنه‌ی تغییرات آن حول ۲۵ درجه‌ی ساتی‌گراد است. در نیمه‌ی دوم زمستان تا اوایل بهار، معمولاً در ابتدا اُفت و



شکل ۱. برازش منحنی به داده‌های دما و میزان بارندگی مربوط به ۶۶ ایستگاه هواشناسی در ایران با استفاده از سری‌های فوریه.

سپس افزایش دما را شاهد هستیم و به‌طور متوسط، دمای ماکسیمم در بین ماه‌های ژوئن و ژوئیه (تقریباً نیمه‌ی خرداد تا نیمه‌ی امرداد) رخ می‌دهد. بیش‌ترین تغییرات دما بر مبنای تابع انحراف معیار که از جذر فرمول (۲) به دست آمده است، در اواخر پاییز و نیمه‌ی اول زمستان رخ می‌دهد. به‌طور دقیق‌تر، بیش‌ترین تغییرات در ماه ژانویه (تقریباً نیمه‌ی دی تا نیمه‌ی بهمن) اتفاق می‌افتد. این رخداد بر اساس منحنی‌های دما در شکل ۱ نیز توجیه‌پذیر است؛ زیرا در این زمان خاص، اُفت زیاد دما را برای چند ایستگاه نسبت به ایستگاه‌های دیگر، و در نتیجه دامنه‌ی بزرگ‌تر تغییرات را برای نمونه‌های تابعی شاهد هستیم. از طرف دیگر، کم‌ترین تغییرات در ماه‌های ژوئن و ژوئیه رخ می‌دهد؛ زیرا برای اکثر ایستگاه‌ها در این بازه‌ی زمانی بیش‌ترین دما اتفاق می‌افتد و در ضمن، دمای این مکان‌ها از ۲۲ تا ۳۰ درجه‌ی سانتی‌گراد می‌باشد. در واقع، بیش‌ترین تطابق دمایی در این ماه‌ها (کاهش دامنه‌ی پراکندگی برای دما) مشاهده می‌شود. برای بقیه‌ی زمان‌ها تغییرات تقریباً یکسان و حول ۶ درجه‌ی سانتی‌گراد می‌باشد.

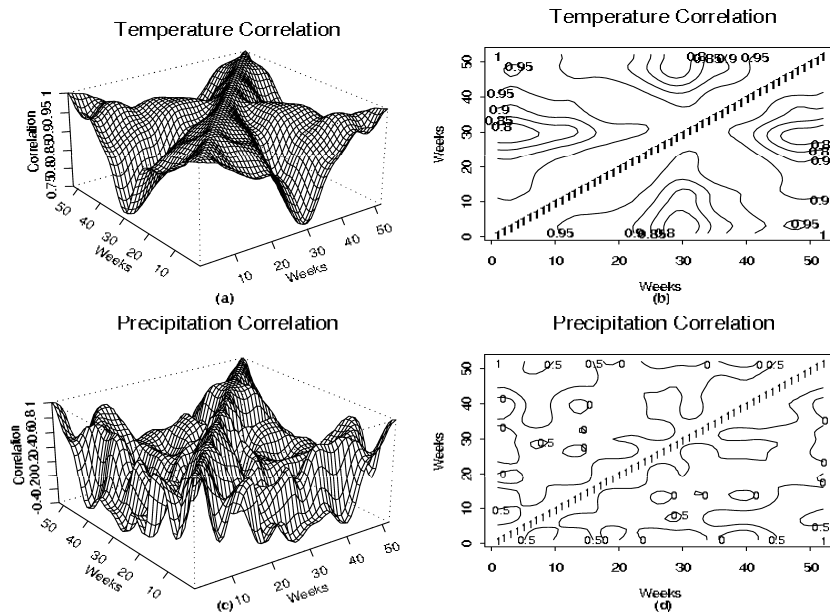
تابع میانگین و انحراف معیار برای داده‌های مقدار بارندگی در دو نمودار پایینی شکل ۲ ترسیم شده است. با توجه به تابع میانگین، به‌طور متوسط کم‌ترین مقدار بارندگی بین ماه‌های ژوئن و سپتامبر (اواخر بهار تا اواخر تابستان) رخ می‌دهد. از طرف دیگر، بیش‌ترین میزان بارندگی را می‌توان در زمستان و اوایل بهار مشاهده نمود. با نگاه دقیق‌تر به این نمودار، مشاهده می‌شود که به‌طور متوسط در ماه‌های ژانویه و



شکل ۲. توابع میانگین و انحراف معیار برای داده‌های دما و میزان بارندگی. دو نمودار بالایی توابع میانگین و انحراف معیار را برای پدیده‌ی دما نشان می‌دهند و دو نمودار پایینی، این توابع را برای میزان بارندگی نشان می‌دهند.

مارس (تقریباً نیمه‌ی دی تا اوایل بهار) بیش‌ترین میزان بارندگی، و در ماه‌های ژوئن و سپتامبر (تقریباً نیمه‌ی خرداد تا نیمه‌ی شهریور) کم‌ترین میزان بارندگی نسبت به بقیه‌ی ماه‌های سال رخ داده است. بیش‌ترین تغییرات برای پدیده‌ی بارندگی در ماه مارس (تقریباً نیمه‌ی اسفند تا نیمه‌ی فروردین) مشاهده می‌شود. این رخداد، مصادف با روزهایی است که در مناطق شمالی بیش‌ترین مقدار بارندگی سال به وقوع می‌پیوندد و از سوی دیگر، مناطق جنوبی و جنوب شرقی روزهای بدون بارندگی را سپری می‌کنند. کم‌ترین تغییرات برای ماه‌های ژوئن و ژوئیه رخ می‌دهد؛ یعنی در روزهایی که بارندگی تقریباً کمی در تمامی نقاط کشور رخ می‌دهد. از ماه ژوئیه تا ماه دسامبر (تقریباً نیمه‌ی خرداد تا نیمه‌ی دی) به‌طور نسبی، افزایش تغییرات در میزان بارندگی را شاهد هستیم.

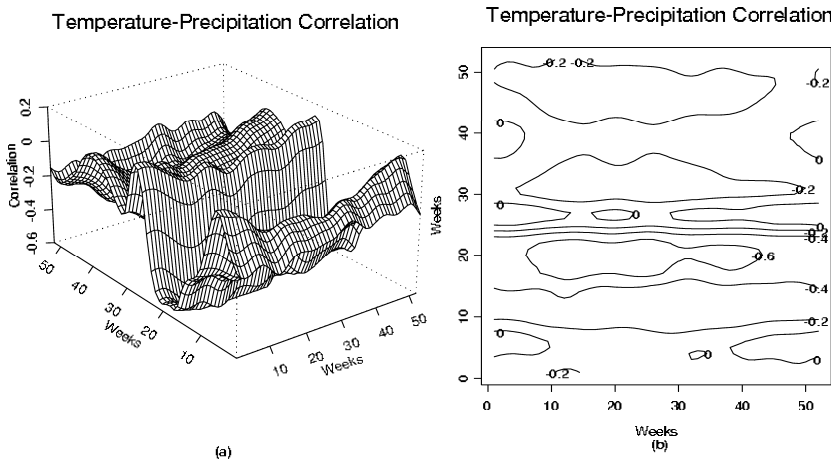
برای بررسی دقیق‌تر ساختار همبستگی داده‌های دما و میزان بارندگی، می‌توان از نمودارهای سطوح تراز یا سه‌بعدی استفاده نمود. به‌عنوان مثال، نمودارهای (a) و (b) در شکل ۳، مربوط به داده‌های دما می‌باشند. در این نمودارها، محورهای x و y متناظر با تقریباً ۵۲ هفته‌ی موجود در یک سال مقیاس‌بندی شده‌اند. بر اساس این نمودارها، با گذشت زمان، بعد از هفته‌ی سی‌ام، حد اکثر تا ۱۵ هفته، همبستگی بین



شکل ۳. توابع همبستگی و نمودارهای خطوط تراز برای دو پدیده‌ی دما و میزان بارندگی.

دما در این هفته‌ها با دما در اوایل فصل بهار از ۸° به ۹۵° افزایش می‌یابد. همچنین این روند به‌طور متقارن با شروع از هفته‌ی سی‌ام تا هفته‌ی پانزدهم نیز ادامه دارد. از طرفی همبستگی بین هفته‌ی اول و هفته‌ی پنجاه و یکم سال میلادی (اواخر دی با اوایل دی) نیز بیش‌ترین مقدار ممکن را دارد. در دو نمودار (c) و (d) در شکل ۳، ساختار همبستگی برای داده‌های میزان بارندگی نشان داده شده است. بر مبنای این نمودارها، همبستگی بین هفته‌های ۲۵ تا ۴۲ (تقریباً ژوئن تا اکتبر) با هفته‌های ۱۰ تا ۱۳ (تقریباً مارس تا آوریل) حدوداً برابر با صفر می‌باشد. این موضوع، اظهار نظر در مورد میزان بارش در فصل پاییز بر اساس اطلاعاتی در مورد میزان بارندگی در اواخر فصل بهار را مشکل می‌کند. از طرف دیگر، همبستگی زیادی بین هفته‌های آذر و دی وجود دارد. در واقع، حاکم شدن جریان هوای سرد بر اکثر نقاط کشور در این زمان‌ها، افزایش بارندگی را نسبت به بقیه‌ی سال نتیجه می‌دهد. برای همین هفته‌ها، یعنی اوایل زمستان، مقدار همبستگی مثبت اندکی با فصل تابستان مشاهده می‌شود.

دو نمودار (a) و (b) در شکل ۴، همبستگی بین دما و میزان بارندگی را آشکار می‌سازد. محورهای افقی و عمودی در این نمودارها به ترتیب، مقدار بارندگی و دما را بر حسب زمان (هفته) بیان می‌کنند. میزان بارندگی در هفته‌های ۱۰ تا ۴۲ (تقریباً ژانویه تا اکتبر) همبستگی منفی نسبتاً زیادی ($-۰/۶$) را با دما برای هفته‌های ۱۷ تا ۲۲ (آوریل تا نیمه‌ی ژوئن) نسبت به بقیه‌ی سال نشان می‌دهد. این پدیده به دلیل افزایش دما و کاهش بارندگی برای هفته‌های میانی سال رخ می‌دهد. این همبستگی منفی با افزایش فاصله از این بازه‌ی زمانی (هفته‌های ۱۷ تا ۲۲) کاهش می‌یابد. در ضمن، همبستگی مقدار دما در هفته‌های ۲۶ و ۲۷ (اواخر تیر و اوایل مرداد) با مقدار بارندگی، برای تمامی سال، تقریباً برابر با صفر می‌باشد.

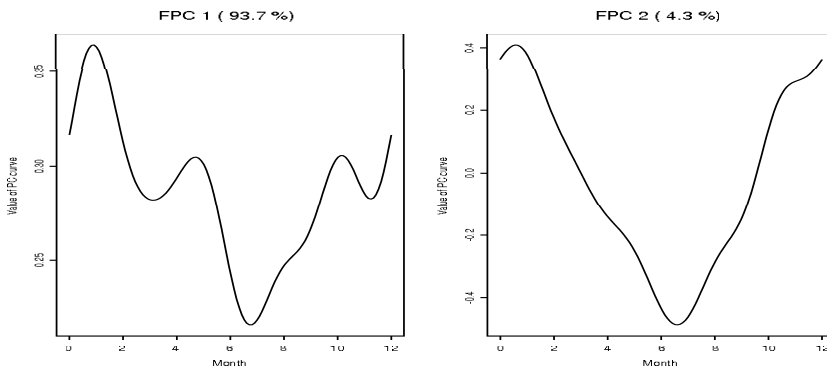


شکل ۴. تابع همبستگی متقابل میزان بارندگی-دما و نمودار خطوط تراز مربوط به آن.

در ادامه با استفاده از تحلیل مؤلفه‌های اصلی تابعی، ساختار تغییرات دما بررسی و الگوهای موجود در آن استخراج می‌شود. به‌طور مشابه، این رویکرد برای داده‌های بارندگی نیز قابل بررسی است، که در این جا به آن پرداخته نمی‌شود.

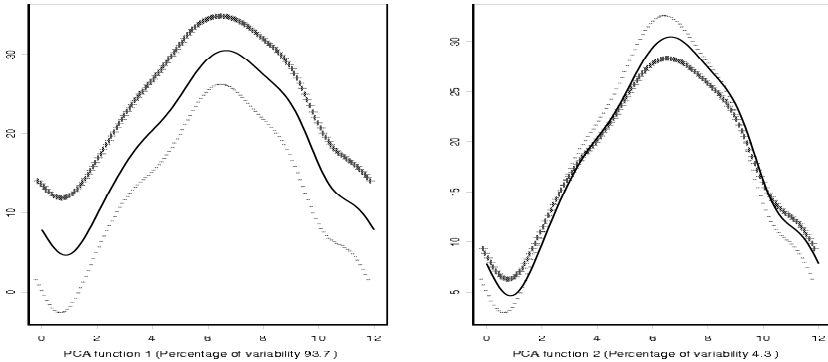
بعد از به دست آوردن مؤلفه‌های اصلی تابعی، مانند PCA در حالت چندمتغیره، انتخاب تعداد مناسبی از آن‌ها مد نظر خواهد بود. به دلیل آن‌که انتخاب دو مؤلفه‌ی اصلی اول، ۹۸ درصد از تغییرات اولیه را حفظ می‌کند، انتخاب این دو مؤلفه که به ترتیب ۹۳/۷ و ۴/۳ درصد از تغییرات کل را بیان می‌کنند، مناسب خواهد بود.

نمودار ۵ دو مؤلفه‌ی اول را نشان می‌دهد. همان‌طور که ملاحظه می‌شود، اولین مؤلفه درصد بالایی از تغییرات را به خود اختصاص می‌دهد. به‌طور کلی، وزنی که برای فصل زمستان قرار داده شده، ۱/۵ برابر وزنی است که برای فصل تابستان در نظر گرفته شده است. در این مرحله برای این مؤلفه اراهه‌ی یک تفسیر بهتر امکان‌پذیر نمی‌باشد؛ از این رو، باید از روش‌های دیگر برای تفسیر آن استفاده نمود. یک روش مناسب برای تفسیر اولین FPC، ترسیم نمودارهای تابع میانگین، همراه با توابعی است که با اضافه کردن یا کم کردن ضریب مناسبی از تابع وزنی $\psi_1(t)$ به دست می‌آیند (رمسی و سیلورمن، ۲۰۰۵). شکل ۶ چنین نموداری را نشان می‌دهد. با توجه به این شکل، اولین منحنی مؤلفه‌ی اصلی نشان می‌دهد که اکثر تغییرپذیری در داده‌ها (تقریباً ۹۴ درصد) از تفاوت‌هایی ناشی می‌شود که بین دما در تابستان و زمستان وجود دارد. سه ایستگاه جزیره‌ی سیری، جزیره‌ی کیش و بندر لنگه که دمای این فصول برای آن‌ها بالا می‌باشد، بیش‌ترین امتیاز را به خود اختصاص داده‌اند. از سوی دیگر، سه ایستگاه با کم‌ترین امتیاز که عبارت‌اند از ایستگاه‌های پل فیروزکوه، خلخال و سراب، در طول سال، هوای نسبتاً سردی دارند (شکل ۷ را ببینید).

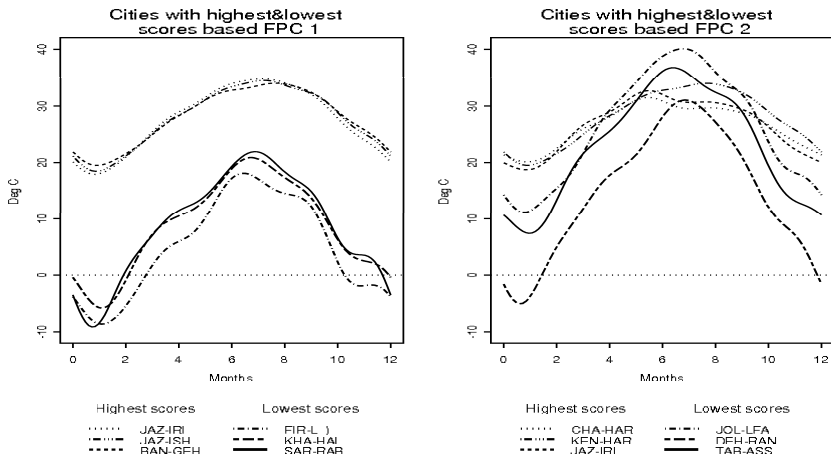


شکل ۵. دو FPC اول هموار شده برای داده‌های دما که به ترتیب ۹۳/۷ و ۴/۳ درصد از تغییرات را بیان می‌کنند.

اما نمودار تابع وزنی $\psi_2(t)$ یک منحنی هموارتر را نسبت به منحنی $\psi_1(t)$ نشان می‌دهد. در واقع، این نوع از تغییرات فقط ۳/۴ درصد از تغییرات اولیه را شامل می‌شود و تغییر پذیری از متوسط تفاوت بین



شکل ۶. دو FPC اول هموار شده برای داده‌های دما. در هر نمودار، منحنی پیوسته تابع میانگین را نشان می‌دهد و دو منحنی که با علامت - و + ترسیم شده‌اند، به ترتیب، منحنی‌هایی را نشان می‌دهند که از جمع زدن ضرایب منفی و مثبت از هر FPC با تابع میانگین به دست آمده‌اند.



شکل ۷. در هر یک از دو نمودار، منحنی دما مربوط به ایستگاه‌های دارای بالاترین و پایین‌ترین امتیاز متناظر با هر یک از دو مؤلفه‌ی اول ترسیم شده است. اسم هر سه ایستگاه با بالاترین و پایین‌ترین امتیاز، با درج سه حرف اول و آخر نام آن در پایین هر نمودار آورده شده است.

دما در زمستان و دما در تابستان را منعکس می‌کند. به عبارت دیگر، این مؤلفه به مناطقی که اختلاف دمای زمستان و تابستان آن‌ها کوچک است امتیازهای مثبت بزرگ اختصاص می‌دهد. به عنوان مثال، ایستگاه‌های چابهار، کنارک چابهار و جزیره سیری بیش‌ترین امتیازات را با تابع وزنی $\psi_2(t)$ می‌گیرند. دمای این موقعیت‌ها در طول سال، بیش‌تر از 20° درجه و کم‌تر از 32° درجه سانتیگراد می‌باشد. از طرف دیگر، این مؤلفه برای مناطقی که اختلاف دمای آن‌ها در زمستان و تابستان زیاد است، امتیازهای منفی بزرگ در نظر می‌گیرد، که ایستگاه‌های جلفا و دهلران و طبس از آن جمله‌اند (شکل ۷). از این رو، مؤلفه‌ی دوم را می‌توان به‌عنوان یک اندازه از یکنواختی دما وقتی که از زمستان به طرف تابستان در حرکت هستیم در نظر گرفت.

۵/۱ فواصل اطمینان بوت‌استرپی برای مقادیر و توابع ویژه

تشکیل فواصل اطمینان انفرادی و همزمان برای مقادیر و توابع ویژه در حالت «تابعی»، در انجام دادن استنباط‌های آماری در مورد این پارامترها به ما کمک می‌کند.

فرض کنید $\chi = \{X_1, X_2, \dots, X_n\}$ یک نمونه به اندازه‌ی n از توابع تصادفی باشد. بازنمونه‌گیری $\chi^* = \{X_1^*, X_2^*, \dots, X_n^*\}$ را که به‌طور مستقل و با جایگذاری از نمونه‌ی χ انتخاب می‌شود در نظر بگیرید. برای این بازنمونه، برآوردهای بوت‌استرپی $\hat{\theta}_j^*$ و $\hat{\psi}_j^*$ را محاسبه می‌کنیم. حال، توزیع‌های غیر شرطی $\theta_j - \hat{\theta}_j$ و $\psi_j - \hat{\psi}_j$ را به‌ترتیب با توزیع‌های شرطی $\theta_j^* - \hat{\theta}_j^*$ و $\psi_j^* - \hat{\psi}_j^*$ به‌شرط χ تقریب می‌زنیم.

یک فاصله‌ی اطمینان بوت‌استرپی متقارن برای هر θ_j به‌طور انفرادی به‌صورت زیر تشکیل می‌شود. در ابتدا $\hat{z}_{1-\alpha}$ را با استفاده از توزیع شرطی $|\hat{\theta}_j^* - \hat{\theta}_j|$ به‌شرط χ طوری پیدا می‌کنیم که داشته باشیم

$$P(|\hat{\theta}_j^* - \hat{\theta}_j| \leq \hat{z}_{1-\alpha} | \chi) = 1 - \alpha.$$

سپس با استفاده از آن، فاصله‌ی اطمینان $(1 - \alpha)100^\circ$ درصدی برای θ_j به‌صورت $(\hat{\theta}_j - \hat{z}_{1-\alpha}, \hat{\theta}_j + \hat{z}_{1-\alpha})$ می‌باشد. این فاصله دارای خطای پوشانشی از مرتبه‌ی $O(n^{-1})$ می‌باشد (هال و حسینی‌نسب، ۲۰۰۶). شایان ذکر است برای تشکیل یک فاصله‌ی اطمینان بوت‌استرپی با دم‌های برابر، چندک‌های توزیع $\hat{\theta}_j^* - \hat{\theta}_j$ به‌شرط χ باید به‌گونه‌ای انتخاب شوند که در دو دم توزیع، احتمال‌های مساوی $\frac{\alpha}{2}$ تشکیل شود.

نوار اطمینان بوت‌استرپی برای هر $\psi_j(t)$ که همزمان برای هر $t \in I$ صادق باشد، به‌صورت

جدول ۱. فواصل اطمینان بوت‌استرپ منفرد و دوگانه برای پنج ویژه‌مقدار اول داده‌های دمای ایران در سال ۲۰۰۵. سطح پوشانش اسمی $1 - \alpha = 0.95$ بوده است.

نوع روش	فاصله‌ی اطمینان با دم‌های برابر		فاصله‌ی اطمینان متقارن	
	منفرد	دوگانه	منفرد	دوگانه
θ_1	(۲۹۴٫۶، ۵۰۶٫۴)	(۲۹۵٫۸، ۵۲۸٫۳)	(۲۹۵٫۲، ۵۰۷٫۲)	(۲۸۶٫۱، ۵۱۶٫۴)
θ_2	(۱۱٫۷۴، ۲۵٫۸۶)	(۱۲٫۱۵، ۲۸٫۹۵)	(۱۱٫۲۷، ۲۵٫۶۱)	(۹٫۶۴۹، ۲۷٫۲۳)
θ_3	(۲٫۵۲۲، ۴٫۲۰۳)	(۲٫۵۲۶، ۴٫۳۷۰)	(۲٫۶۲۵، ۴٫۳۰۶)	(۲٫۶۶۵، ۴٫۲۶۶)
θ_4	(۲٫۱۸۷، ۳٫۸۳۷)	(۱٫۹۹۲، ۴٫۳۲۸)	(۱٫۷۴۳، ۳٫۶۸۹)	(۱٫۳۹۲، ۴٫۰۳۹)
θ_5	(۰٫۶۱۵، ۰٫۹۷۷)	(۰٫۶۰۹، ۱٫۰۳۹)	(۰٫۶۰۳، ۰٫۹۶۸)	(۰٫۵۹۱، ۰٫۹۸۱)

زیر تشکیل می‌شود: در ابتدا چندک مرتبه‌ی $(1 - \alpha)$ ام، یعنی $\hat{z}_{1-\alpha}$ ، به‌گونه‌ای تعیین می‌شود که داشته باشیم

$$P \left\{ \sup_{t \in \mathcal{I}} |\hat{\psi}_j^* - \hat{\psi}_j| \leq \hat{z}_{1-\alpha} | \mathcal{X} \right\} = 1 - \alpha.$$

با استفاده از آن، ناحیه‌ی $\{(t, u) : t \in \mathcal{I} \text{ \& } |\hat{\psi}(t) - u| \leq \hat{z}_{1-\alpha}\}$ یک ناحیه‌ی اطمینان در سطح اسمی $(1 - \alpha) 100\%$ درصد برای منحنی $\hat{\psi}_j(\cdot)$ می‌باشد (هال و حسینی‌نسب، ۲۰۰۶). آن‌ها همچنین ثابت کردند که خطای پوشانش این فاصله‌ی اطمینان از مرتبه‌ی $O(n^{-1})$ است. برای تشکیل یک فاصله‌ی اطمینان هم‌زمان (برای همه‌ی z ها) برای ویژه‌مقدارها، در ابتدا $\hat{\Delta}_{1-\alpha}$ را به‌گونه‌ای تعیین می‌کنیم که داشته باشیم $P(\hat{\Delta}^* \leq \hat{\Delta}_{1-\alpha} | \mathcal{X}) = 1 - \alpha$ ، که در آن $\hat{\Delta}^* = \|\hat{K}^* - \hat{K}\| = \int (\hat{K}^* - \hat{K})^2$ ، دوگانه‌ی توان دوم تفاضل \hat{K}^* و \hat{K} روی فاصله‌ی \mathcal{I} می‌باشد. هال و حسینی‌نسب (۲۰۰۶) نشان دادند که

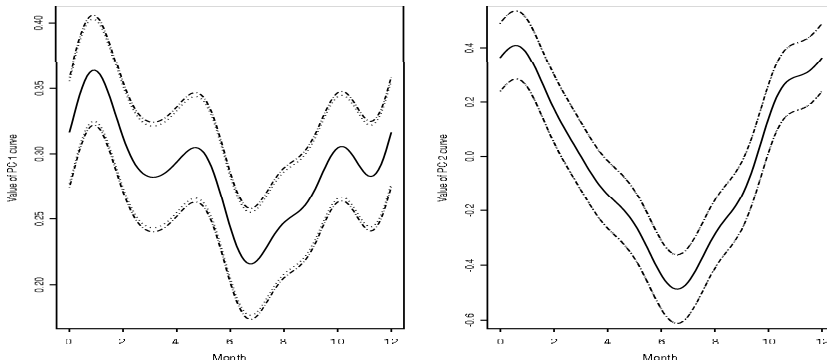
$$P(\sup_{j \geq 1} |\hat{\theta}_j - \theta_j| \leq \hat{\Delta}_{1-\alpha}) = 1 - \alpha + O(n^{-1}),$$

که در آن $O(n^{-1})$ خطای پوشانش می‌باشد. شایان ذکر است که استفاده از روش‌های بوت‌استرپ دوگانه در تمام فواصل اطمینان فوق باعث می‌شود تا خطای پوشانش از مرتبه‌ی $O(n^{-1})$ به $O(n^{-2})$ کاهش یابد.

جدول ۱ فواصل اطمینان بوت‌استرپی متقارن و با دم‌های برابر را در سطح ۹۵ درصد برای پنج ویژه‌مقدار اول داده‌های دما نشان می‌دهد که با روش‌های بوت‌استرپ منفرد (single) و دوگانه (double) تشکیل شده‌اند. تمامی فواصل اطمینان منفرد با استفاده از $B = 4000$ بازنمونه‌گیری به دست آمده‌اند. همان‌طور که مشاهده می‌شود، فواصل اطمینان متقارن برای θ_j ها که با استفاده از روش‌های بوت‌استرپی منفرد به دست آمده‌اند، تفاوت چندانی با فواصل دارای دم‌های برابر ندارند. فاصله‌ی اطمینان دارای دم‌های برابر

برای ویژه مقدار اول که از روش بوت استرپ منفرد به دست آمده است به صورت $(۲۹۴/۶, ۵۰۶/۴)$ می باشد. این فاصله برای ویژه مقدار دوم برابر با $(۱۱/۷۴, ۲۵/۸۶)$ است. علاوه بر آن با برآورد $\Delta_{0.95} = ۱۰۸$ می توان فواصل اطمینان همزمان برای ویژه مقدارها را تشکیل داد.

همچنین برای تشکیل یک فاصله اطمینان بوت استرپ دوگانه، هر باز نمونه $C = ۵۰^\circ$ بار باز نمونه گیری شده است. این فواصل نسبت به فواصل اطمینانی که از روش بوت استرپ منفرد به دست آمده اند دارای خطای پوشانش کمتری می باشند. در این حالت، فاصله های اطمینان به دست آمده برای θ_1 ها در دو حالت دارای دم های برابر و متقارن، متفاوت اند. برای مثال، اگر از فواصل اطمینان دارای دم های برابر استفاده شود، این فاصله برای θ_1 به صورت $(۲۹۵/۸, ۵۲۸/۳)$ می باشد؛ در حالی که این فاصله برای θ_1 در حالت متقارن به صورت $(۲۸۶/۱, ۵۱۶/۴)$ می باشد. همچنین فاصله اطمینان دارای دم های برابر برای θ_2 به صورت $(۱۲/۱۵, ۲۸/۹۵)$ است، اما این فاصله در حالت متقارن $(۹/۶۴۹, ۲۷/۲۳)$ می باشد. نوارهای اطمینان برای دو تابع ویژه اول در شکل ۸ رسم شده اند. نزدیکی بیش از حد این دو نوار، نشان دهنده آن است که نوارهای به دست آمده از روش بوت استرپ منفرد، به خوبی نوارهای به دست آمده از روش بوت استرپ دوگانه می باشند.



شکل ۸. دو FPC اول هموار شده برای داده های دما و نوارهای اطمینان بوت استرپی متناظر با آنها. در هر نمودار، منحنی پیوسته تابع ویژه را نشان می دهد و خطوط نقطه چین و شکسته به ترتیب، فواصل اطمینانی را نشان می دهند که با استفاده از روش های بوت استرپ منفرد و دوگانه به دست آمده اند.

۶ بحث و نتیجه‌گیری

با توجه به این‌که داده‌های مربوط به دو پدیده‌ی دما و مقدار بارندگی، که دارای ماهیت تابعی هستند، به صورت گسسته (روزانه یا ماهانه) گزارش می‌شوند، اولین گام در تحلیل این داده‌ها، برازش منحنی‌های مناسب به داده‌های اولیه می‌باشد. با توجه به ماهیت دوره‌ای دما و بارندگی، داده‌های گسسته‌ی اولیه را با استفاده از بسط سری‌های فوریه از مرتبه‌ی ۱۳ به توابعی پیوسته از زمان تبدیل کردیم. سپس با استفاده از آماره‌های توصیفی، ویژگی‌های کلی این دو پدیده استخراج شد. ساختار تابع همبستگی برای داده‌های دما ترسیم گردید و تحلیل‌ها و تفسیرهای مربوط به آن ارائه شد. در مرحله‌ی بعد، با استفاده از FPCA، الگوهای مهم پراکنندگی برای این پدیده استخراج گردید. به دلیل آن‌که دو مؤلفه‌ی اول، تقریباً بیش از ۹۸ درصد از تغییرات اولیه را بیان می‌کنند، فقط این دو مؤلفه مورد توجه قرار گرفتند. FPC اول نشان می‌دهد که اکثر تغییرپذیری (تقریباً ۹۴ درصد) در داده‌ها از تفاوت‌هایی ناشی می‌شود که بین دما در تابستان و زمستان وجود دارد و FPC دوم که ۴/۳ درصد از تغییر را حفظ می‌کند، منعکس‌کننده‌ی این است که این الگو از متوسط تفاوت بین دما در زمستان و دما در تابستان ناشی می‌شود. در انتها نیز فواصل اطمینان بوت‌استرپی منفرد و دوگانه را برای این دو FPC و برای پنج ویژه‌مقدار اول تشکیل دادیم.

سیاس‌گذاری

نویسندگان مقاله از داوران محترم و از آقای دکتر رنجبر و سرکار خانم تقی‌پور در سازمان هواشناسی کشور به دلیل مشاوره و راهنمایی در انتخاب ایستگاه‌های مورد بررسی تشکر و قدردانی می‌نمایند.

توضیحات

۱. Functional Data Analysis
۲. Functional Principal Components Analysis
۳. Functional Principal Component

مرجعه‌ها

- Hall, P.; Hosseini-nasab, M. (2006). On properties of functional principal components analysis. *J. R. Stat. Soc. Ser. B* **68**, 109-126.
- Hall, P.; Hosseini-nasab, M. (2007). Theory for high-order bounds in functional principal components analysis. Accepted by Mathematical Proceedings of Cambridge Philosophical Society.
- Jolliffe, I.T. (2002). *Principal Component Analysis, 2nd edn*. Springer, New York.
- Pezzulli, S.; Silverman, B.W. (1993). Some properties of smoothed principal components analysis for functional data. *Comput. Statist* **8**, 1-16.
- Ramsay, J.O. (1982). When the data are functions. *Psychometrika* **47**, 379-96.
- Ramsay, J.O.; Silverman, B.W. (2002). *Applied Functional Data Analysis: Methods and Case Studies*. Springer, New York.
- Ramsay, J.O.; Silverman, B.W. (2005). *Functional Data Analysis, 2nd edn*. Springer, New York.

سید محمدابراهیم حسینی نسب

گروه آمار، دانشکده‌ی علوم پایه،
دانشگاه تربیت مدرس،
پل نصر، بزرگراه جلال آل احمد،
تهران، ایران.

پایان‌نگار: m.hosseininasab@modares.ac.ir

نورالله تازیکه میاندره

گروه آمار، دانشکده‌ی علوم پایه،
دانشگاه تربیت مدرس،
پل نصر، بزرگراه جلال آل احمد،
تهران، ایران.

پایان‌نگار: tazikeh_nm@modares.ac.ir