مدلسازی و آینده‌نگری مینیمم و ماکسیمم درجهٔ حرارت ماهانهٔ شهر اصفهان

منوچهر خردمندنیا*

چکیده

در این مقاله برای مینیمم و ماکسیمم درجهٔ حرارت ماهانهٔ شهر اصفهان، یک مدل آنالیزکواریانس تعیین می‌کنیم. پس از بررسی مدل از لحاظ درستی فرضها، با استفاده از آن مقادیر آینده را تا سال ۱۳۸۰ پیش‌بینی می‌کنیم. مدل تعیین شده شامل یک متغیر توضیحی رسته‌ای دو سطحی با سطوح مینیمم (۱) و ماکسیمم (۲)، و تعداد معدودی متغیرهای توضیحی سینوسی ـ کسینوسی است. برای تعیین مؤلفه‌های سینوسی‌ـکسینوسی معنی‌دار، با استفاده از دوره‌نگار، ترتیب اهمیت همسازها را در هر یک از دو سطح تعیین می‌کنیم. سپس با جستجو در بین چند همساز مهم اول، چرخه‌های معنی‌دار پنهان را پیدا می‌کنیم. الگوریتم موفقی که در این مقاله برای مدلسازی ارائه شده است برای هر سری دو سطحی تناوبی مشابه دیگری کاربرد دارد.

۱ تحلیل دوره‌نگار

یک سری زمانی $\{Z_t\}_{t=1}^{n} = \{Z_1, Z_2, \dots, Z_n\}$ را می‌توان به شکل زیر نوشت

$$Z_t = a_0 + h_1(t) + h_2(t) + \cdots + h_q(t) \qquad (1)$$

که در آن

$$h_i(t) = a_i C_{it} + b_i S_{it} \qquad (2)$$

همساز iـاُم نامیده می‌شود. در رابطهٔ اخیر

$$C_{it} = \mathrm{Cos}(2\pi f_i t) \quad , \quad S_{it} = \mathrm{Sin}(2\pi f_i t) \qquad (3)$$

که در آن $f_i = i/n$ فراوانی همساز iـاُم نامیده می‌شود. تعداد مشاهدات در یک چرخه کامل از همساز iـاُم را دورهٔ تناوب همساز iـاُم می‌نامند که برابر $n/i = 1/f_i$ است. اگر n فرد باشد، آنگاه $q = (n-1)/2$ و ضرایب a_0، a_i، و b_i که ضرایب فوریه نامیده می‌شوند از روابط زیر به دست می‌آیند:

$$a_0 = \frac{1}{n}\sum_{t=1}^{n} Z_t$$

$$\begin{cases} a_i = \frac{2}{n}\sum_{t=1}^{n} Z_t C_{it} \\ b_i = \frac{2}{n}\sum_{t=1}^{n} Z_t S_{it} \end{cases} \qquad i = 1, 2, \dots, q \quad (4)$$

اگر n زوج باشد، آنگاه $q = n/2$ و به‌استثنای دو ضریب a_q و b_q، بقیه ضرایب از روابط فوق به دست می‌آیند. در حالتی که n زوج است دو ضریب a_q و b_q از روابط زیر تعیین می‌شوند:

*) دکتر منوچهر خردمندنیا، گروه آمار دانشگاه اصفهان

$$a_q = \frac{1}{n}\sum_{t=1}^{n}(-1)^t Z_t \quad , \quad b_q = 0. \qquad (5)$$

اگر یک سری زمانی از جمع یک مؤلفهٔ تصادفی و یک مؤلفهٔ تناوبی تشکیل شده باشد، تحت شرایطی کلی که معمولاً برقرار است، جمع تعداد معدودی از q همساز ممکن، تقریب خوبی برای توصیف آن سری تلقی می‌شود. به عنوان مثال فرض کنید که همساز kـاُم به تنهایی تقریب خوبی برای سری $\{Z_t\}_{t=1}^{n}$ باشد. در این صورت می‌توان نوشت

$$Z_t = a_0 + a_k \mathrm{Cos}(2\pi\frac{k}{n}t) + b_k \mathrm{Sin}(2\pi\frac{k}{n}t) + e_t \qquad (6)$$

که در آن e_t مجموع $q-1$ همساز دیگر است. رابطهٔ (۶) را یک مدل فوریه گوییم. مدل فوریه در واقع یک مدل رگرسیون است. به سهولت می‌توان نشان داد که برآورد کمترین توانهای دوّم a_0، a_k، و b_k از روابطی که در (۴) و (۵) داده شده به دست می‌آیند. رابطهٔ (۱) را نمایش فوریهٔ یک سری زمانی گویند که آن را به صورت زیر می‌توان نوشت:

$$Z_t = a_0 + \sum_{i=1}^{q}(a_i \mathrm{Cos}2\pi\frac{i}{n}t + b_i \mathrm{Sin}2\pi\frac{i}{n}t). \qquad (7)$$

با در نظر گرفتن (۷) به عنوان یک مدل رگرسیون که در آن $e_t = 0$، به سهولت می‌توان نشان داد که برآورد کمترین توانهای دوّم a_0، a_iها، و b_i ها از روابط (۴) و (۵) به دست می‌آیند. توجه کنید که اگر q زوج باشد، آنگاه $q = n/2$ و $b_q = 0$ و ضرایب فوریه عبارت‌اند از a_0، a_1، a_2، $\ldots$، a_q، و b_1، b_2، $\ldots$، b_{q-1} که تعداد آنها n است و اگر q فرد باشد، آنگاه $q = (n-1)/2$ و ضرایب فوریه عبارت‌اند از a_0، a_1، a_2، $\ldots$، a_q، b_1، b_2، $\ldots$، b_q. ملاحظه می‌شود که در (۷) چه n زوج و چه n فرد باشد، تعداد ضرایب برابر n یعنی برابر طول بردار پاسخ است.

تعیین تعداد معدودی از همسازها که از جمع آنها تقریب خوبی برای یک سری زمانی به دست آید مدلسازی فوریه نامیده می‌شود. یک ابزار مؤثر و مفید در مدلسازی فوریه، دوره‌نگار است. دوره‌نگار شامل q مقدار $I(f_1)$، $I(f_2)$، $\ldots$، $I(f_q)$ است. اگر n فرد باشد، آنگاه

$$I(f_i) = \frac{n}{2}(a_i^2 + b_i^2), \qquad i = 1, 2, \ldots, q. \qquad (8)$$

اگر n زوج باشد، آنگاه برای $i \in \{1, 2, \ldots, q-1\}$ مقادیر $I(f_i)$ از رابطهٔ فوق به دست می‌آیند ولی برای $i = q$ داریم $I(f_q) = na_q^2$، که در آن a_iها و b_iها ضرایب فوریه هستند. نموداری که مقادیر $I(f_i)$ را در مقابل i نشان دهد دوره‌نگار، نامیده می‌شود. با استفاده از دوره‌نگار، ترتیب اهمیت همسازها را می‌توان تعیین کرد. مقصود از اهمیت یک همساز، سهمی است که آن همساز از کل تغییرپذیری دارد. با در نظر گرفتن (۷) به عنوان یک الگوی رگرسیون که در آن $e_t = 0$، می‌توان نشان داد که در یک جدول آنالیز واریانس، $I(f_i)$ در واقع مجموع توان دوم متناظر با همساز iـاُم است (به [۱] نگاه کنید). فرض کنید که

$$I(f_k) = \max\{I(f_1), I(f_2), \ldots, I(f_q)\}. \qquad (9)$$

در این صورت اگر بخواهیم امکان تقریب زدن سری زمانی $\{Z_t\}$ را فقط با یک همساز بررسی کنیم، مدل (۶) یک مدل آزمایشی مناسب است. اگر بخواهیم امکان تقریب زدن سری را با دو همساز بررسی کنیم مدل

$$Z_t = a_0 + h_k(t) + h_\ell(t) + e_t \qquad (10)$$

یک مدل آزمایشی مناسب است، که در آن برای تعیین ℓ ابتدا $I(f_\ell)$ را تعیین می‌کنیم. $I(f_\ell)$ ماکسیمم $I(f_i)$ ها بدون در نظر گرفتن $I(f_k)$ است.

هر چند نمایش فوریهٔ یک سری زمانی مقوله‌ای صرفاً ریاضی است ولی تعیین آن تعداد معدود از همسازها که تقریب خوبی به دست می‌دهند یک مقوله آماری است. به زبان آماری مقصود از مدلسازی فوریه تعیین همسازهایی است که نقش معنی‌داری در تغییرپذیری سری دارند. هر روش معقولی که برای تعیین همسازهای معنی‌دار به کار برده شود، مستلزم جستجو در بین چند همساز مهم اول است. با استفاده از دوره‌نگار به سهولت می‌توان اهمیت همسازها و در نتیجه چند همساز مهم اول را تعیین کرد. یک روش ساده و مؤثر در الگوسازی فوریه به این ترتیب است که همسازها را به ترتیب اهمیت یکی‌یکی به مدل اضافه کنیم و پارامترهایی را که تفاوت معنی‌دار با صفر ندارند، حذف کنیم. عمل افزودن همسازهای جدید را تا جایی ادامه می‌دهیم که به همسازی برسیم که هر دو مؤلفه کسینوسی و سینوسی آن نقش معنی‌داری در تغییرپذیری سری نداشته باشند. در اینجا عمل افزودن همساز جدید را متوقف نموده مدل آخری را به عنوان یک مدل آزمایشی مناسب کاندید می‌کنیم. بالاخره مدل آزمایشی را از لحاظ درستی فرضها مورد بررسیهای همه‌جانبه قرار داده در صورت لزوم تعدیلهایی در آن انجام می‌دهیم تا به مدلی برسیم که از مرحله بررسی درستی فرضها نیز موفق بیرون آید. در آزمونهای فرض برابری ضرایب فوریه با صفر، هرچه که سطح معنی‌دار بودن را کوچکتر بگیریم مدل کوچکتری به عنوان مدل نهایی آزمایشی تعیین خواهد شد. تجارب مدلسازی فوریه حاکی از آن است که برای سریهایی که دارای یک روند آشکار تناوبی هستند، سطح معنی‌دار $\alpha = 0{,}01$ معمولاً منجر به مدل معقولی می‌شود که به طور متعادلی اصل کمترین توانهای دوّم و اصل امساک را ارضاء نموده از مرحله بررسی درستی فرضها نیز با موفقیت عبور می‌کند. برای توضیحات نظری بیشتر راجع به تحلیل دوره‌نگار و مدلسازی فوریه به خردمندنیا (۱۳۷۶) مراجعه کنید. در این مرجع همچنین یک برنامه مینی‌تب برای محاسبات مربوط به دوره‌نگار ارائه شده و چگونگی استفاده از دستور GLM در مینی‌تب برای برآورد پارامترهای یک مدل خطی توضیح داده شده است.

۲ مدلسازی سری دو سطحی مورد بررسی

مینیمم مطلق و ماکسیمم مطلق درجهٔ حرارت ماهانهٔ شهر اصفهان از فروردین ۱۳۶۵ لغایت اسفند ۱۳۷۴ در جدول (۱) ملاحظه می‌شود. این داده‌ها با استفاده از سالنامه‌های آماری که همه ساله توسط مرکز آمار ایران منتشر می‌شود به دست آمده است. ارقام مربوط به بهمن و اسفند ۷۲ در سالنامه مربوطه گزارش نشده است. در جدول (۱) موارد گزارش نشده را با علامت * مشخص نموده‌ایم. با نگاهی اجمالی به جدول (۱) ملاحظه می‌کنیم که بسیاری از داده‌های مربوط به دو سال ۷۳ و ۷۴ با داده‌های سالهای قبل ناهمخوانی دارند. در بررسیهایی که به عمل آمد معلوم شد که از فروردین ۷۳ مسئولیت اندازه‌گیری درجه‌حرارت هوای اصفهان به ایستگاه هواشناسی دیگری که از لحاظ شرایط محیطی و ارتفاع، تفاوتهای قابل توجهی با ایستگاه هواشناسی قبلی دارد واگذار شده است. با توجه به این موارد تصمیم گرفتیم که مدلسازی را بر اساس داده‌های مربوط به هشت سال ۶۵-۷۲ قرار دهیم. در تجزیه و تحلیلهای مقدماتی داده‌های هشت سالهٔ مذکور سه مورد مینیمم بهمن ۷۰، ماکسیمم بهمن ۷۰ و ماکسیمم اسفند ۷۰ به عنوان داده‌هایی پرت یا مشکوک تلقی شدند. داده‌های پرت و گزارش نشده را به عنوان داده‌های مفقود تلقی نموده با دستور GLM مینی‌تب و براساس مدل (۱۱) آنها را برآورد کردیم. به این ترتیب مقادیر مینیمم بهمن ۷۰، مینیمم بهمن ۷۲، مینیمم اسفند ۷۲، ماکسیمم بهمن ۷۰، ماکسیمم اسفند ۷۰، ماکسیمم بهمن ۷۲ و ماکسیمم اسفند ۷۲ به ترتیب برابر ۶/۲-، ۶/۲-، ۲/۵-، ۱۶/۵، ۲۰/۳، ۱۶/۵ و ۲۰/۳ برآورد گردید. این هفت عدد را به جای هفت مورد فوق‌الذکر قرار داده فرآیند مدلسازی را براساس سری حاصله قرار دادیم. دستور GLM مینی‌تب از جمله دستوراتی است که اجازه می‌دهد تعدادی از داده‌ها مفقود باشند. در مینی‌تب یک داده مفقود با علامت * مشخص می‌شود. با توجه به ویژگیهای آشکار، سری مورد بررسی (نمودار ۱) که دو سطحی است و دارای یک دوره تناوب آشکار ۱۲ است، مدل آنالیز کوواریانس (۱۱) یک مدل آزمایشی اولیه معقول به نظر می‌رسد.

$$y_{jt} = \mu + \lambda_j + a_8 \mathrm{Cos}(2\pi \frac{8}{96} t) + b_8 \mathrm{Sin}(2\pi \frac{8}{96} t) + e_{jt} \qquad (11)$$

که درآن $\sum_{j=1}^{2} \lambda_j = 0$، $(j = 1, 2)$، $(t = 1, 2, \ldots, 96)$. پارامتر μ، یک میانگین عمومی است، پارامتر $\lambda_1 = -\lambda_2$ برای به حساب آوردن دو سطح مینیمم و ماکسیمم است و همساز هشتم یعنی $h_8(t) = a_8 C_{8t} + b_8 S_{8t}$ برای به حساب آوردن فراوانی آشکار $\frac{8}{96} = \frac{1}{12}$ است. در (۱۱) به جای λ_j می‌توانیم عبارت $(-1)^j \lambda$ را قرار دهیم. واضح است که $\sum_{j=1}^{2} (-1)^j \lambda = 0$. با استفاده از دستور GLM مینی‌تب، مدل برازش شده به صورت زیر به دست می‌آید

$$y_{jt} = \underset{(0)}{16.58} + (-1)^j \underset{(0)}{11.36} - \underset{(0)}{7.67} \mathrm{Cos}(2\pi \frac{8}{96} t) + \underset{(0)}{9.61} \mathrm{Sin}(2\pi \frac{8}{96} t) + e_{jt} \qquad (12)$$

که در آن $\hat{\sigma}^2 = 4$، $R^2 = 0.9825$ و $R_a^2 = 0.9822$. مقادیر R^2 و R_a^2 هر یک به نحوی میزان موفقیت مدل را نشان می‌دهند (نگاه کنید به [۳]). اعدادی که در داخل پرانتز، زیر برآورد پارامترها نوشته شده است، مقادیر احتمال برای آزمون فرض برابری ضریب فوریه با صفر است. با توجه به اینکه در هر چهار مورد مقدار احتمال برابر صفر است، هر یک از فرضهای صفر $\mu = 0$، $\lambda = 0$، $a_8 = 0$ و $b_8 = 0$ با قاطعیت رد می‌شود.

اکنون امکان بهبود مدل (۱۲) را از جهات مختلف مورد بررسی قرار می‌دهیم. یکی از جهاتی که می‌بایست بررسی شود، بررسی معنی‌دار بودن اثر متقابل بین متغیر توضیحی رسته‌ای و متغیرهای توضیحی غیررسته‌ای است. به عبارت دیگر بررسی این فرض که ضریب C_{8t} در دو سطح مینیمم و ماکسیمم تفاوت معنی‌دار دارند و نیز بررسی این فرض که ضریب S_{8t} در دو سطح مینیمم و ماکسیمم به طور معنی‌داری متفاوت‌اند. مدل زیر فرض می‌کند که اثرات متقابل وجود دارند.

$$y_{jt} = \mu + (-1)^j \lambda + (a_8 + (-1)^j a) C_{8t} + (b_8 + (-1)^j b) S_{8t} + e_{jt}. \qquad (13)$$

مدل برازش شده به صورت زیر به دست می‌آید:

$$y_{jt} = \underset{(0)}{16.58} + (-1)^j \underset{(0)}{11.36} - (\underset{(0)}{7.67} + (-1)^j \underset{(0.158)}{0.28}) C_{8t} + (\underset{(0)}{9.61} \dot{-} (-1)^j \underset{(0.903)}{0.024}) S_{8t} + e_{jt}. \qquad (14)$$

مقدار احتمال برای آزمون فرضهای صفر عدم وجود اثر متقابل یعنی فرضهای $a = 0$ و $b = 0$ به ترتیب ۰/۱۵۸ و ۰/۹۰۳ به دست می‌آید. لذا فرض عدم وجود اثر متقابل پذیرفته می‌شود. جهت دیگری که برای بهبود مدل (۱۲) می‌بایست مورد بررسی قرارگیرد یافتن چرخه‌های معنی‌دار پنهان دیگری است که محتملاً در باقیمانده‌های مدل (۱۲) وجود دارند. در نمودارهای (۳) و (۴)، دوره‌نگار باقیمانده‌های مدل (۱۲) به تفکیک دو سطح مینیمم و ماکسیمم ملاحظه می‌شوند. براساس این نمودارها پنج همساز مهم اول مربوط به باقیمانده‌های سطح مینیمم به ترتیب اهمیت عبارت‌اند از: هفتم، شانزدهم، بیست ونهم، بیست و یکم و نهم؛ و در مورد سطح ماکسیمم به ترتیب اهمیت عبارت‌اند از: هفتم، شانزدهم، بیست و هفتم، بیست و یکم و دهم. ملاحظه می‌شود که دو همساز هفتم و شانزدهم، بعد از همساز هشتم مهمترین همسازهایی هستند که احتمال دارد حضور

جدول (۱)مینیمم مطلق و ماکسیمم مطلق درجهٔ حرارت ماهانهٔ شهر اصفهان ۷۴ـ۶۵

مینیمم

ماه / سال	۶۵	۶۶	۶۷	۶۸	۶۹	۷۰	۷۱	۷۲	۷۳	۷۴
۱	۶٫۶	۲٫۶	۳٫۰	۱٫۰	-۲٫۱	۵٫۰	-۰٫۸	-۱٫۶	۱٫۴	-۲٫۴
۲	۸٫۶	۱۱٫۵	۱۱٫۶	۸٫۰	۹٫۲	۱۰٫۰	۸٫۰	۸٫۴	۵٫۸	۶٫۶
۳	۱۶٫۲	۱۳٫۴	۱۵٫۴	۱۶٫۰	۱۵٫۰	۱۰٫۰	۱۳٫۶	۱۵٫۲	۱۱٫۰	۱۰٫۴
۴	۱۹٫۴	۱۷٫۴	۱۷٫۴	۱۸٫۴	۱۹٫۶	۱۸٫۶	۱۹٫۲	۱۷٫۸	۱۴٫۵	۱۴٫۴
۵	۱۵٫۶	۱۸٫۸	۱۸٫۰	۱۹٫۰	۲۰٫۰	۱۹٫۲	۱۶٫۸	۱۸٫۶	۱۴٫۰	۱۴٫۶
۶	۱۲٫۰	۱۳٫۵	۱۱٫۴	۱۱٫۰	۱۲٫۰	۱۵٫۰	۱۰٫۶	۱۳٫۲	۴٫۴	۹٫۶
۷	۶٫۰	۴٫۰	۹٫۰	۷٫۰	۷٫۰	۸٫۰	۸٫۰	۵٫۴	۱٫۶	۱٫۶
۸	-۱٫۴	-۰٫۳	-۳٫۴	۱٫۰	۲٫۰	۱٫۸	۲٫۲	۰٫۰	-۲٫۵	-۱٫۰
۹	-۵٫۶	-۵٫۶	-۸٫۶	-۲٫۴	-۱٫۲	-۴٫۴	-۳٫۶	-۳٫۴	-۱۳٫۶	-۱۰٫۰
۱۰	-۴٫۶	-۶٫۵	-۶٫۶	-۷٫۲	-۸٫۶	-۸٫۰	-۷٫۲	-۴٫۲	-۹٫۶	-۱۹٫۴
۱۱	-۲٫۰	-۴٫۹	-۷٫۶	-۴٫۰	-۷٫۶	-۱۳٫۲	-۶٫۸	*	-۸٫۴	-۱۱٫۴
۱۲	-۰٫۶	-۰٫۸	-۳٫۰	-۲٫۱	-۲٫۰	-۶٫۰	-۱٫۲	*	-۸٫۰	-۳٫۴

ماکسیمم

ماه / سال	۶۵	۶۶	۶۷	۶۸	۶۹	۷۰	۷۱	۷۲	۷۳	۷۴
۱	۲۴٫۴	۲۸٫۵	۲۷٫۸	۲۷٫۴	۲۹٫۶	۲۶٫۰	۲۶٫۰	۲۸٫۰	۲۶٫۴	۲۵٫۶
۲	۳۲٫۰	۳۳٫۰	۳۲٫۴	۳۵٫۸	۳۳٫۶	۳۱٫۰	۲۹٫۴	۳۱٫۶	۳۰٫۲	۳۰٫۴
۳	۳۹٫۰	۳۶٫۲	۳۳٫۷	۳۹٫۰	۳۸٫۰	۳۷٫۰	۳۹٫۸	۳۷٫۶	۳۹٫۰	۳۸٫۰
۴	۳۸٫۰	۳۹٫۰	۴۰٫۰	۴۰٫۶	۳۸٫۸	۴۰٫۴	۳۹٫۲	۳۹٫۰	۴۱٫۰	۴۱٫۴
۵	۳۶٫۵	۳۷٫۴	۳۸٫۵	۳۷٫۲	۳۹٫۰	۳۹٫۰	۳۷٫۸	۳۹٫۲	۴۱٫۶	۳۸٫۶
۶	۳۲٫۳	۳۶٫۰	۳۶٫۰	۳۵٫۲	۳۷٫۵	۳۷٫۰	۳۶٫۲	۳۵٫۶	۳۷٫۰	۳۸٫۶
۷	۳۰٫۶	۲۹٫۶	۲۷٫۴	۳۱٫۲	۳۲٫۵	۳۳٫۸	۳۲٫۰	۳۱٫۵	۲۸٫۸	۳۳٫۰
۸	۲۲٫۴	۲۴٫۰	۲۳٫۰	۲۰٫۰	۲۶٫۰	۲۴٫۶	۲۴٫۶	۲۳٫۰	۲۴٫۴	۲۴٫۴
۹	۱۵٫۰	۲۱٫۰	۱۷٫۰	۱۸٫۴	۲۰٫۰	۱۶٫۲	۱۹٫۴	۱۸٫۲	۱۷٫۲	۱۸٫۸
۱۰	۱۸٫۰	۱۴٫۰	۱۲٫۰	۱۳٫۰	۱۲٫۰	۱۲٫۵	۱۱٫۸	۱۹٫۶	۱۶٫۶	۱۲٫۴
۱۱	۱۹٫۵	۱۷٫۲	۱۸٫۰	۱۸٫۰	۱۶٫۰	۱۱٫۵	۱۵٫۶	*	۱۷٫۰	۲۲٫۰
۱۲	۱۹٫۸	۲۳٫۰	۲۱٫۴	۲۳٫۸	۲۰٫۶	۱۱٫۶	۱۹٫۰	*	۱۹٫۰	۱۷٫۸

معنی‌داری در هر دو سطح مینیمم و ماکسیمم داشته باشند. در صورتی که به منظور بررسی امکان بهبود مدل (۱۲)، همساز هفتم را بر مدل بیفزاییم مدل زیر به دست می‌آید:

$$y_{jt} = \mu + (-1)^j \lambda + a_{۸} C_{۸t} + b_{۸} S_{۸t} + a_{۷} C_{۷t} + b_{۷} S_{۷t} + e_{jt}. \qquad (۱۵)$$

مدل برازش شده به صورت زیر است:

$$y_{jt} = \underset{(۰)}{۱۶٫۵۸} + (-۱)^j \underset{(۰)}{۱۱٫۳۶} - \underset{(۰)}{۷٫۶۷C_{۸t}} + \underset{(۰)}{۹٫۶۱S_{۸t}} + \underset{(۰٫۰۱۴)}{۰٫۴۶۳C_{۷t}} - \underset{(۰)}{۰٫۷۶۴S_{۷t}} + e_{jt}$$

که در آن $\hat{\sigma}^۲ = ۳٫۳$. ملاحظه می‌شود که همه متغیرهای توضیحی موجود در این مدل به استثنای $C_{۷t}$ معنی‌دار هستند. فرض $a_{۷} = ۰$ در سطح ۰٫۰۱ پذیرفته می‌شود، لذا $C_{۷t}$ را از مدل حذف می‌کنیم.

در بررسی همسازهای دیگر ملاحظه نمودیم که فقط مؤلفهٔ کسینوسی همساز بیست و یکم در سطح ۰٫۰۱ معنی‌دار است و مؤلفه‌های همسازهای دیگر در سطح ۰٫۰۱ معنی‌دار نیستند. با افزودن مؤلفهٔ کسینوسی همساز بیست و یکم، مدل برازش شده نهایی به صورت زیر به دست می‌آید:

$$y_{jt} = \underset{(۰)}{۱۶٫۵۸} + (-۱)^j \underset{(۰)}{۱۱٫۳۶} - \underset{(۰)}{۷٫۶۷C_{۸t}} + \underset{(۰)}{۹٫۶۱S_{۸t}} - \underset{(۰)}{۰٫۷۶۴S_{۷t}} - \underset{(۰٫۰۰۳)}{۰٫۵۶\,C_{۲۱t}} + e_{jt}, \qquad (۱۶)$$

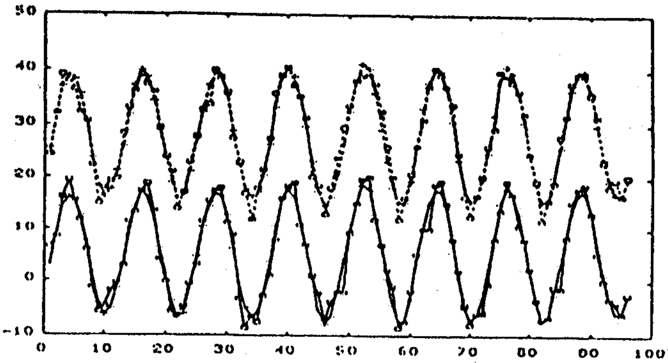
که در آن $\hat{\sigma}^۲ = ۳٫۳$ ، $R^۲ = ۰٫۹۸۴۷$ و $R_a^۲ = ۰٫۹۸۴۳$.

در مدل (۱۳) ملاحظه کردیم که فرض عدم وجود اثر متقابل بین متغیر توضیحی رسته‌ای و هر یک از متغیرهای توضیحی غیر رسته‌ای قویاً پذیرفته شد. محاسبات مربوط به مدلهای دیگر و از جمله مدل نهایی (۱۷) نیز نشان می‌دهد که در همه موارد بین متغیر توضیحی رسته‌ای و هر یک از متغیرهای توضیحی غیررسته‌ای اثر متقابل وجود ندارد. موجود نبودن اثرات متقابل فوق‌الذکر در واقع انعکاس همبستگی مثبت و تقریباً کامل دو سری مینیمم و ماکسیمم می‌باشد.

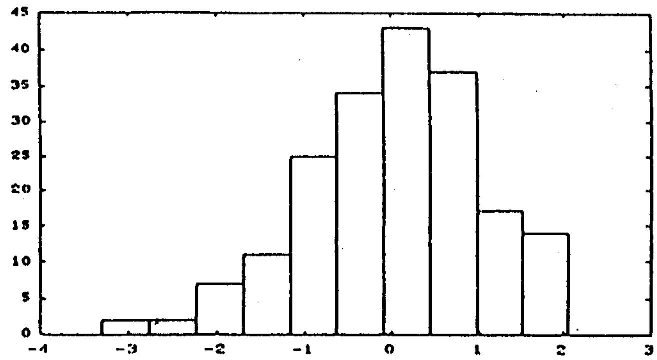
۳ بررسی درستی تشخیص مدل و آینده‌نگری

در نمودار (۱) سری دو سطحی مشاهده شده و مقادیر برازش شده بر اساس مدل (۱۷) مقایسه شده‌اند. این نمودار حاکی از انطباقی عالی است. در نمودار (۲) بافت‌نگار باقیمانده‌های مدل (۱۷) ملاحظه می‌شود. بررسیها و آزمونهای استاندارد مربوط به نرمال و ناهمبسته (و در نتیجه مستقل) بودن باقیمانده‌ها و ثابت بودن واریانس آنها حاکی از قابل قبول بودن این فرضهاست. به عبارت دیگر فرض $e_{jt} \sim \mathrm{iidN}(o, \sigma^2)$ در مورد باقیمانده‌های مدل (۱۷) قابل قبول است. توجه کنید که بر اساس نتایج بخش قبل، حذف هر یک از پارامترهای مدل (۱۷) موجب می‌شود که این مدل به‌طور معنی‌داری بدتر شود. از طرف دیگر افزودن هر همساز دیگری مدل را به‌طور معنی‌داری بهتر نمی‌کند. بنابراین سری دو سطحی مشاهده شده را می‌توان تحققی از مدل (۱۷) دانست. با در نظر گرفتن قسمت سیستماتیک این مدل به عنوان مکانیزم مولد مقادیر آینده، مینیمم و ماکسیمم درجهٔ حرارت ماهانهٔ شهر اصفهان از فروردین ۷۶ لغایت اسفند ۸۰ در جدول (۲) پیش‌بینی شده است.

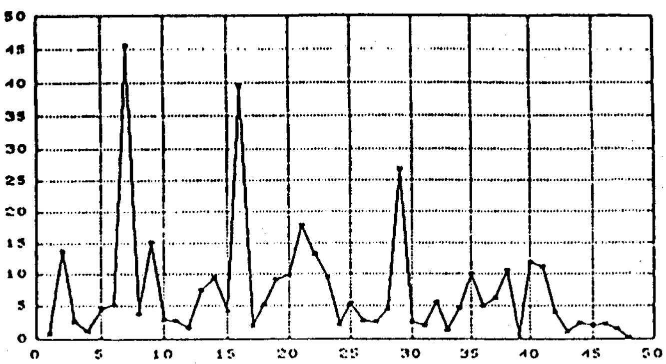
نمودار (۱) سری دو سطحی مشاهده شده (∘) و سری دو سطحی برازش شده (+)



نمودار (۲) بافت نگار باقیمانده‌های مدل (۱۷)



نمودار (۳) دوره‌نگار باقیمانده‌های مدل (۱۲) مربوط به سطح مینیمم



جدول (۲) پیش‌بینی مینیمم مطلق و ماکسیمم مطلق درجهٔ حرارت ماهانهٔ شهر اصفهان ۷۶-۸۰

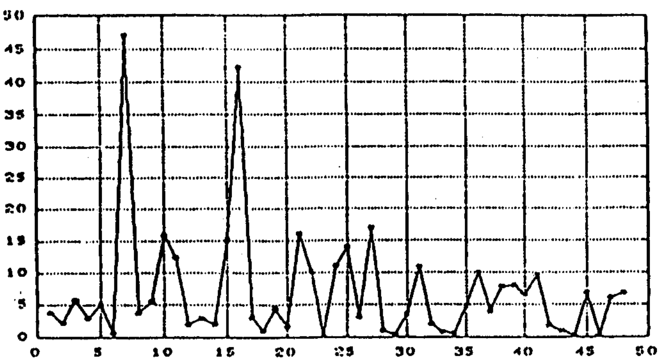
مینیمم

سال	۷۶	۷۷	۷۸	۷۹	۸۰
ماه ۱	۳٫۶	۳٫۸	۳٫۴	۲٫۱	۳٫۱
۲	۱۰٫۶	۹٫۷	۱۰٫۳	۹٫۰	۸٫۷
۳	۱۶٫۰	۱۵٫۲	۱۵٫۱	۱۵٫۱	۱۳٫۶
۴	۱۷٫۷	۱۸٫۵	۱۷٫۴	۱۷٫۹	۱۶٫۹
۵	۱۶٫۱	۱۷٫۷	۱۷٫۳	۱۶٫۸	۱۷٫۱
۶	۱۲٫۳	۱۲٫۹	۱۴٫۱	۱۳٫۰	۱۳٫۳
۷	۶٫۹	۶٫۴	۷٫۸	۷٫۹	۷٫۱
۸	۰٫۳	۰٫۳	۰٫۵	۱٫۹	۱٫۰
۹	-۵٫۴	-۴٫۴	-۵٫۰	-۳٫۸	-۳٫۳
۱۰	-۸٫۰	-۷٫۴	-۷٫۱	-۷٫۳	-۵٫۸
۱۱	-۶٫۴	-۷٫۴	-۶٫۳	-۶٫۷	-۵٫۹
۱۲	-۱٫۸	-۳٫۳	-۳٫۲	-۲٫۵	-۳٫۰

ماکسیمم

سال	۷۶	۷۷	۷۸	۷۹	۸۰
ماه ۱	۲۶٫۳	۲۶٫۵	۲۶٫۱	۲۴٫۸	۲۵٫۸
۲	۳۳٫۳	۳۲٫۵	۳۳٫۰	۳۱٫۷	۳۱٫۴
۳	۳۸٫۷	۳۷٫۹	۳۷٫۸	۳۷٫۸	۳۶٫۳
۴	۴۰٫۴	۴۱٫۲	۴۰٫۱	۴۰٫۶	۳۹٫۷
۵	۳۸٫۸	۴۰٫۴	۴۰٫۰	۳۹٫۵	۳۹٫۸
۶	۳۵٫۱	۳۵٫۶	۳۶٫۸	۳۵٫۷	۳۶٫۱
۷	۲۹٫۶	۲۹٫۱	۳۰٫۵	۳۰٫۶	۲۹٫۸
۸	۲۳٫۱	۲۳٫۰	۲۳٫۲	۲۴٫۶	۲۳٫۷
۹	۱۷٫۲	۱۸٫۲	۱۷٫۷	۱۸٫۸	۱۹٫۳
۱۰	۱۴٫۶	۱۵٫۲	۱۵٫۵	۱۵٫۳	۱۶٫۹
۱۱	۱۶٫۲	۱۵٫۳	۱۶٫۳	۱۵٫۹	۱۶٫۷
۱۲	۲۰٫۸	۱۹٫۳	۱۹٫۵	۲۰٫۱	۱۹٫۷

نمودار (۴) دوره‌نگار باقیمانده‌های مدل (۱۲) مربوط به سطح ماکسیمم



۴ تقدیر و تشکر

از پیشنهادهای داوران محترم که برای بهبود مقاله مفید بوده‌اند سپاسگزارم. از آقای دکتر هوشنگ طالبی که با پیشنهاد معقولتری برای برآورد داده‌های مفقود موجب بهترشدن مدل نهایی شدند تشکر می‌کنم. انگیزه اولیهٔ تهیهٔ مقالهٔ حاضر در جریان راهنمایی پروژهٔ خانم فیروزه روزبه، فارغ‌التحصیل رشتهٔ آمار دانشگاه اصفهان، به وجود آمد. زحمات ایشان قابل تقدیر است.

مراجع

[۱] باکس و جنکینز، تحلیل سریهای زمانی: پیش‌بینی و کنترل، ترجمه دکتر محمدرضا مشکانی ۱۳۷۱، انتشارات دانشگاه شهید بهشتی.

[۲] خردمندنیا، م (۱۳۷۶) الگوسازی فوریه و آینده‌نگری درجهٔ حرارت ماهانهٔ شهر اصفهان، طرح پژوهشی مصوب ۷۶۰۲۱۵ ، معاونت پژوهشی دانشگاه اصفهان.

[3] Draper, N. R. and Smith, H. (1981) *Applied Regression Analysis.* 2nd ed. Willey. New York.

برآورد تعداد اشتباهات کتاب

غلامحسین شاهکار*

جرج پولیا* ریاضیدان معاصر در ماهنامهٔ ریاضی، شمارهٔ ژانویهٔ سال ۱۹۷۶ این مسأله را به صورت زیر مطرح و به آن جواب می‌گوید.

مسأله- قبل از چاپ کتابی دو نمونه‌خوان کتاب را خوانده و هر یک مستقلاً m و n اشتباه در آن یافته‌اند. اگر تعداد اشتباهاتی که دو نمونه خوان مشترکاً پیدا کرده‌اند k باشد، در این صورت تعداد اشتباهات پیدا نشدهٔ کتاب را چقدر برآورد می‌کنید؟

حل- فرض کنید احتمال یافتن اشتباه توسط نمونه‌خوان اول p، توسط نمونه‌خوان دوم q، و این دو پیشامد مستقل باشند. همچنین فرض کنید کل تعداد اشتباهات موجود در کتاب M باشد. میانگین تعداد اشتباهاتی که به وسیلهٔ نمونه خوانهای اول، دوم، و هر دو پیدا می‌شوند به ترتیب برابرند با Mp، Mq، و Mpq. اگر تعداد اشتباهات پیدا شده را تقریباً برابر با میانگین آنها بگیریم، در این صورت داریم:

$$Mp = n, \qquad Mq = m, \qquad Mpq = k$$

و از اینجا به دست می‌آوریم:

$$M = \frac{(Mp)(Mq)}{Mpq} = \frac{mn}{k}.$$

در نتیجه برآورد تعداد اشتباهات پیدا نشدهٔ کتاب عبارت است از:

$$M - (m+n-k) = \frac{mn}{k} - (m+n-k) = \frac{(m-k)(n-k)}{k}$$

*) غلامحسین شاهکار، گروه آمار، دانشگاه فردوسی مشهد

*) Gorge Polya (1888-1985), American Mathematical Monthly.