مجله‌ی بررسی‌های آمار رسمی ایران
سال ۳۱، شماره‌ی ۱، بهار و تابستان ۱۳۹۹، صص ۱ – ۲۰
DOR: 20.1001.1.25385798.1399.31.1.6.0

برازش مدل‌های سری‌زمانی شمارشی روی تعداد مراجعین به مراکز ترک اعتیاد در شهرستان سمنان

امید کریمی*، فاطمه حسینی و سپیده تشرفی

دانشگاه سمنان

چکیده. داده‌های شمارشی برحسب زمان در خیلی از زمینه‌های کاربردی مشاهده می‌شوند. خیلی از پژوهشگران برای تحلیل این داده‌ها از الگوهای سری‌زمانی استفاده می‌کنند. در این مقاله، مدل‌های خطی سری‌زمانی شمارشی پواسنی و دوجمله‌ای منفی روی این نوع از داده‌ها با حضور متغیرهای توضیحی مورد مطالعه قرار می‌گیرند. تحلیل درست‌نمایی و ارزیابی مدل سری‌زمانی شمارشی براساس مدل‌های خطی تعمیم‌یافته ارائه و بررسی می‌شود و همچنین داده‌های تعداد معتادین مراجعه کننده به مراکز ترک اعتیاد شهرستان سمنان براساس مدل‌ها و روش‌های بیان شده مورد تحلیل قرار می‌گیرد. با توجه به نتایج حاصل از ملاک‌های انتخاب مدل، مدل سری‌زمانی شمارشی دوجمله‌ای منفی برازش مناسبی برای این داده‌ها است و براساس آن پیش‌بینی ماهانه برای سال ۹۸ صورت گرفت.

واژگان کلیدی: مدل شکنندگی، مدل خطرهای متناسب کاکس، داده بقا، سانسور، مدت زمان بیکاری.

۱- مقدمه

داده‌های سری‌زمانی شمارشی، مانند تعداد ماهانه یک بیماری خاص، تعداد روزهای بارانی در هفته، تعداد معتادین، تعداد مراجعات و غیره در بسیاری از زمینه‌های کاربردی علوم پزشکی، علوم اجتمایی و فیزیکی مشاهده می‌شوند. یک مدل مناسب برای تحلیل داده‌های سری‌زمانی شمارشی، استفاده از مدل رگرسیون پواسن تحت

* نویسنده‌ی عهده‌دار مکاتبات
دریافت: ۱۳۹۹/۵/۸، پذیرش: ۱۴۰۰/۷/۱۴.

مدل‌های خطی تعمیم‌یافته (GLM)[1] است [۸]. فوکیانوس [۵] مدل رگرسیون پواسنی برای داده‌های سری‌زمانی شمارشی را مورد مطالعه قرار داد و روش ماکسیمم درستنمایی برای براورد پارامترهای مدل به‌کار گرفت.
فرلند و همکاران [۴] مانایی مدل‌های اتورگرسیو شرطی با مقادیر صحیح را مورد تحلیل قرار دادند. کریستو و فوکیانوس [۱] مدل‌های خطی تعمیم‌یافته را با تابع پیوند لگ‌خطی بررسی کردند ولیبوشیک و همکاران [۷] اثرات مداخله‌ای را در مدل‌های سری‌زمانی دوجمله‌ای منفی در نظر گرفتند. مدل‌های سری زمانی شمارشی دارای مشاهدات صحیح نامنفی هستند و باید همبستگی بین آن‌ها به درستی شناسایی شود. یک روش مناسب و منعطف، به‌کارگیری روش GLM نلدر و ودربرن [۹]، برای مدل کردن مشاهدات به صورت شرطی روی اطلاعات گذشته می‌باشد. این روش با انتخاب یک توزیع مناسب برای داده‌های شمارشی و یک تابع پیوند مناسب، پیاده‌سازی می‌شود [۴]، [۲]. در این مقاله به تحلیل مدل‌های سری‌زمانی شمارشی پرداخته می‌شود، سپس این مدل‌ها روی داده‌های ترک اعتیاد شهرستان سمنان پیاده‌سازی می‌شوند.
ساختار مقاله به این صورت است که در بخش ۲ مدل‌های سری‌زمانی شمارشی معرفی و مورد مطالعه قرار می‌گیرند. همچنین جزئیات مدل‌های خطی و لگ‌خطی سری‌زمانی شمارشی بیان می‌شود. در بخش ۳ داده‌های واقعی، تعداد معتادین مراجعه کننده به مراکز ترک اعتیاد شهرستان سمنان براساس مدل‌های ارائه شده مورد بررسی قرار می‌گیرد. در انتها بحث و نتیجه‌گیری ارائه می‌شود.

۲- مدل‌های سری‌زمانی شمارشی

۲-۱- مدل‌سازی رگرسیون پواسون

توزیع پواسون معمولاً برای مدل‌بندی نرخ رویدادهای تصادفی که در برخی از فاصله‌های زمانی ثابت رخ می‌دهند، استفاده می‌شود. اگر فرض شود λ نشان دهنده نرخ ورود است؛ متغیر تصادفی Y نشان دهنده نرخ ورودی‌ها در یک فاصله زمانی ثابت است و دارای توزیع پواسون با تابع چگالی احتمال زیر است:

[DOR: 20.1001.1.25385798.1399.31.1.6.0]

$$\mathrm{P[Y=y]}=\frac{\exp(-\lambda)\,\lambda^{y}}{\mathrm{y!}},\mathrm{y}=۰,۱,۲$$

میانگین و واریانس Y هر دو برابر λ است. در اکثر مسائل کاربردی، داده‌های شمارشی معمولاً با برخی اطلاعات متغیر کمکی مشاهده می‌شود [۸]. به‌طور کلی فرض کنید که $X_۱,\cdots,X_p$، p متغیر رگرسیونی مشاهده شده همراه با متغیر پاسخ شمارشیY با توزیع پواسون است. در آن صورت یک مدل رگرسیونی برای ارتباط بین میانگین متغیر پاسخ با متغیرهای کمکی به‌صورت $\lambda=\beta_۰+\sum_{i=۱}^{p}\beta_i X_i$ است که در آن پارامتر λ به‌خاطر این‌که میانگین توزیع پواسن است، باید مثبت باشد. یک انتخاب ساده برای مدل‌سازی رگرسیون داده‌های شمارشی، مدل لگ‌خطی به‌صورت $\log(\lambda)=\beta_0+\sum_{i=1}^{p}\beta_i X_i$ است. هر دو مدل متعلق به کلاس مدل‌های خطی تعمیم یافته‌اند که توسط نلدر و ودربرن [۹] معرفی شده است. در تعریف مدل‌های سری‌زمانی شمارشی از مدل اتورگرسیو مرتبه یک AR(۱) به‌صورت

$$Y_t=\beta_۱Y_{t-۱}+\epsilon_t, \qquad (۱)$$

استفاده می‌شود. که در آن $\left|\beta_۱\right|<۱$ و $\{\epsilon_t\}$ دنباله متغیرهای تصادفی نرمال با میانگین صفر و واریانس $\sigma^۲$ (دنباله نوفه سفید) است. این یک مدل استاندارد برای تجزیه و تحلیل سری‌های زمانی است که مقدار فرآیند در زمان t بسته به مقدار فرایند در زمان $t-۱$ به همراه یک خطای تصادفی است [۱۰].

۲-۱-۱- مدل‌های خطی برای سری‌های زمانی شمارشی

فرض کنید که $\{Y_t\}$ یک سری‌زمانی شمارشی را نشان می‌دهد. مدل اتورگرسیو پواسون برای فرایند پاسخ $\{Y_t\}$ به‌صورت

$$Y_t\mid\mathcal{F}_{t-۱}^{y,\lambda}\sim\mathrm{Poisson}(\lambda_t),\quad \lambda_t=\beta_۰+\beta_۱Y_{t-۱},\ t\geq ۱, \qquad (۲)$$

[DOR: 20.1001.1.25385798.1399.31.1.6.0]

تعریف می‌شود که در آن $\mathcal{F}_t^{y,\lambda} = \sigma\left(Y_s, \lambda_0, s \le t\right)$ سیگما میدان تولید شده به‌وسیله $\{Y_0, \dots, Y_t, \lambda_0\}$ است. β_1 ، β_0 پارامترهای نامنفی و $\{\lambda_t\}$ به‌عنوان میانگین فرآیند $\{Y_t\}$ با توجه به گذشته آن است. مقادیر مثبت β_0 و β_1 تضمین می‌کند که $\lambda_t > 0$ است، به‌خاطر این‌که Y_t یک عدد صحیح نامنفی است. با توجه به این نکات، واضح است که مدل (۲) با مولفه تصادفی توزیع پواسون و مؤلفه سیستماتیک $\lambda_t = \beta_0 + \beta_1 Y_{t-1}$ و تابع ربط همانی در چارچوب مدل‌های خطی تعمیم‌یافته قرار می‌گیرد. مدل (۲) همان پویایی مدل (۱) را نشان می‌دهد، بنابراین داریم:

$$Y_t = \lambda_t + (Y_t - \lambda_t) = \beta_0 + \beta_1 Y_{t-1} + \epsilon_t, \quad t \ge 1, \tag{۳}$$

که در آن $\epsilon_t = Y_t - \lambda_t$ نوفه سفید است، چون ناهمبسته با میانگین صفر و واریانس ثابت است. البته با فرض مانایی Y_t به راحتی می‌توان نشان داد:

$$E(\epsilon_t) = E(Y_t - \lambda_t) = E\big(E(Y_t - \lambda_t \mid \mathcal{F}_t)\big) = 0,$$

$$Var(\epsilon_t) = Var\big(E(\epsilon_t \mid \mathcal{F}_t)\big) + E\big(Var(\epsilon_t \mid \mathcal{F}_t)\big) = E(\lambda_t) = E(Y_t),$$

$$Cov(\epsilon_t, \epsilon_{t+k}) = E(\epsilon_t \epsilon_{t+k}) = E\left(\epsilon_t E\left(\epsilon_{t+k} \,\middle|\, \mathcal{F}_{t+k-1}\right)\right) = 0.$$

با توجه به معادله (۳) و با فرض مانایی داریم:

$$Var(\epsilon_t) = E(Y_t) = \frac{\beta_0}{1 - \beta_1},$$

بنابراین باید پارامتر $0 < \beta_1 < 1$ باشد.

حالت کلی مدل (۲) با مرتبه‌ی (p, q) به‌صورت زیر تعریف می‌شود:

$$\begin{aligned} Y_t | \mathcal{F}_{t-1}^{y,\lambda} &\sim Poisson(\lambda_t), \\ \lambda_t &= \beta_0 + \sum_{k=1}^{p} \beta_k Y_{t-k} + \sum_{\ell=1}^{q} \alpha_\ell \lambda_{t-1}, \quad t \ge \max(p, q), \end{aligned} \tag{۴}$$

[DOR: 20.1001.1.25385798.1399.31.1.6.0]

که در آن مانایی مرتبه دوم با شرط $0 < \sum_{k=1}^{p} \beta_k + \sum_{l=1}^{q} \alpha_l < 1$ فراهم می‌شود. برای اطلاعات بیشتر به فرلند و همکاران [۴] مراجعه شود. مدل (۲) به مدل واریانس ناهمسانی شرطی اتورگرسیو صحیح مقدار NGARCH[۲] معروف است. با در نظر گرفتن تابع پیوند کانونی $\nu_t = \log(\lambda_t)$ خانواده مدل‌های لگ‌خطی اتورگرسیو به‌صورت

$$Y_t \mid \mathcal{F}_{t-1}^{y,\nu} \sim \text{Poisson}(\lambda_t),$$

$$\nu_t = \beta_0 + \sum_{k=1}^{p} \beta_k \log\left(Y_{t-k} + 1\right) + \sum_{\ell=1}^{q} \alpha_\ell \nu_{t-\ell} \qquad (۵)$$

حاصل می‌شود که در آن $\mathcal{F}_t^{y,\nu}$ سیگما میدان تولید شده از $\{Y_0, \ldots, Y_t, \nu_0\}$ است. به‌طور کلی، پارامترهای $\beta_0, \beta_k, \alpha_\ell$ می‌توانند مثبت یا منفی باشند. اما باید شرایط خاصی روی پارامترها مشابه مدل قبلی در نظر گرفت تا یک سری زمانی مانا به‌دست آید، برای جزئیات بیشتر در مورد آن به [۱] مراجعه شود. با در نظر گرفتن بردار پارامترهای مدل

$$\boldsymbol{\theta} = \left(\beta_0, \beta_1, \ldots, \beta_p, \alpha_1, \ldots, \alpha_q\right)',$$

لگاریتم تابع درست‌نمایی مدل بدون مقدار ثابت را می‌توان به‌صورت

$$l(\boldsymbol{\theta}) = \sum_{t=1}^{n} \left(Y_t \log\left(\lambda_t(\boldsymbol{\theta})\right) - \lambda_t(\boldsymbol{\theta})\right),$$

نوشت که دارای شکل بسته‌ای برای محاسبه براورد پارامترها نیست و باید از الگوریتم‌های عددی برای به‌دست آوردن براورد ماکسیمم درست‌نمایی استفاده کرد. در عمل گاهی اوقات با متغیرهای کمکی که روی متغیر پاسخ تأثیرگذارند روبه‌رو می‌شویم. در آن صورت مدل (۵) به‌صورت زیر بیان می‌شود:

$$\nu_t = \beta_0 + \sum_{k=1}^{p} \beta_k \log\left(Y_{t-k} + 1\right) + \sum_{\ell=1}^{q} \alpha_\ell \nu_{t-\ell} + \boldsymbol{\eta}^T \boldsymbol{X_t},$$

[DOR: 20.1001.1.25385798.1399.31.1.6.0]

که در آن X_t بردار متغیر کمکی r بعدی $X_t = \left(X_{t,1}, \dots, X_{t,r}\right)^T$ می‌باشد و $\boldsymbol{\eta} = \left(\eta_1, \dots, \eta_r\right)^T$ بردار پارامتر ضرائب متغیرهای کمکی است. برای نحوه برازش و پیش‌بینی در این نوع مدل‌ها به [۴] مراجعه شود. یک حالت دیگر برای پاسخ‌های شمارشی استفاده از توزیع دوجمله‌ای منفی است. در توزیع دوجمله‌ای‌منفی، واریانس شرطی می‌تواند از میانگین λ_t بزرگتر باشد که به‌عنوان یک جایگزین برای وقتی که داده‌ها بیش پراکنش دارند، معرفی می‌شود. کریستو و فوکیانوس [۱] مدل دوجمله‌ای منفی را به‌صورت

$$Y_t \mid \mathcal{F}_{t-1} \sim NegBin(\lambda_t, \phi)$$

بیان کردند، که بر حسب پارامتر میانگین λ_tو پارامتر پراکندگی $\phi \in \left(0, \infty\right)$ است. تابع جرم احتمال توزیع دوجمله‌ای منفی به‌صورت

$$P\left(Y_t = y \middle| \mathcal{F}_{t-1}\right) = \frac{\Gamma(\phi + y)}{\Gamma\left(y+1\right)\Gamma(\phi)}\left(\frac{\phi}{\phi+\lambda_t}\right)^{\phi}\left(\frac{\lambda_t}{\phi+\lambda_t}\right)^{y},$$

$$y = 0, 1, \dots$$

تعریف می‌شود که در آن $Var(Y_t|\mathcal{F}) = \lambda_t + \lambda_t^2/\phi$ است. یعنی، واریانس شرطی با توان دوم λ_t افزایش می‌یابد. توزیع پواسون یک حالت خاص دوجمله‌ای منفی است که در آن$\phi \to \infty$ میل می‌کند. در توزیع دوجمله‌ای منفی λ_t مشابه توزیع دوجمله‌ای مدل‌بندی می‌شود. یکی از مسائل مهم در تحلیل داده‌های واقعی این است که کدام مدل برای این داده‌ها مناسب است. بنابراین ملاک‌هایی برای ارزیابی مدل‌های ارائه شده نیاز است که در ادامه بیان می‌شود و از این ملاک‌ها برای تشخیص مدل مناسب در تحلیل داده‌های ترک اعتیاد شهرستان سمنان استفاده شده است.

۲-۲- ارزیابی مدل

ابزارهایی که برای مدل‌های خطی تعمیم‌یافته و نیز برای سری‌زمانی توسعه یافته‌اند، می‌توانند برای ارزیابی برازش مدل و عملکرد پیشگویی آن مورد استفاده قرار گیرند. در کلاس سری‌های‌زمانی شمارشی و مدل‌های خطی تعمیم‌یافته باید ارزیابی پیشگوی خطی و انتخاب تابع پیوند برای توزیع شرطی، انجام شود. معیارهای ارائه شده در این

[DOR: 20.1001.1.25385798.1399.31.1.6.0]

بخش، انتخاب مدل مناسب برای یک مجموعه داده را فراهم می‌کند. این معیارها در بخش بعدی روی مجموعه داده واقعی تعداد معتادین مراجعه کننده به مراکز ترک اعتیاد شهرستان سمنان برای انتخاب مدل مناسب استفاده می‌شوند.

با در نظر گرفتن مقادیر برازش برای هر مدل به‌صورت $\hat{\lambda}_t = \lambda_t(\hat{\theta})$، باقیمانده مدل نیز به‌صورت $r_t = y_t - \hat{\lambda}_t$ به‌دست می‌آیند. همانند مدل‌های خطی این باقیمانده‌ها برای ارزیابی مدل برازش شده به‌کار می‌روند.

تابع خودهمبستگی تجربی مانده‌ها برای تشخیص همبستگی‌های سریالی استفاده می‌شوند که توسط مدل برازش شده بیان نشده‌اند. شکل این مانده‌ها در برابر زمان می‌تواند تغییرات فرآیند تولید داده‌ها روی زمان را آشکار سازد. به‌علاوه، شکل توان دوم باقیمانده‌ها r_t^2 در برابر مقادیر برازش شده $\hat{\lambda}_t$ رابطه‌ی میانگین و واریانس را نمایش می‌دهد. اگر نقاط، اطراف تابع همانی یا همان خط نیمساز قرار گیرند، توزیع پواسون مناسب است و اگر یک رابطه‌ی درجه دو مشاهده شود، توزیع دوجمله‌ای منفی مناسب است، [۱۱].

فرض کنید $P_t(y) = P\left(Y_t \le y \middle| \mathcal{F}_{t-1}\right)$ تابع توزیع تجمعی، $p_t(y) = P\left(Y_t = y \middle| \mathcal{F}_{t-1}\right)$ تابع احتمال برای $y \in N_0$ و $v_t = \sqrt{Var\left(Y_t \middle| \mathcal{F}_{t-1}\right)}$ انحراف استاندارد توزیع پیشگویی باشد که یا دارای توزیع پواسون با میانگین $\widehat{\lambda_t}$ و یا دارای توزیع دوجمله‌ای منفی با میانگین $\widehat{\lambda_t}$ و ضریب بیش‌پراکنش $\widehat{\sigma^2} = \frac{1}{\hat{\phi}}$ است.

یک معیار ارزیابی کالیبدن احتمالی[3] برای توزیع پیشگو، تبدیل انتگرال احتمال[4] (PIT) است که اگر توزیع پیشگو صحیح باشد؛ دارای توزیع یکنواخت است. برای داده‌های شمارشی چادو و همکاران [۲] معیار PIT را برای مقدار مشاهده شد y_t و توزیع پیشگویی $P_t(y)$ به‌صورت

$$F_t(u|y) = \begin{cases} 0 & u \le P_t\left(y-1\right) \\ \frac{u - P_t(y-1)}{P_t(y) - P_t(y-1)} & P_t\left(y-1\right) < u < P_t(y) \\ 1 & u \ge P_t(y) \end{cases} \qquad (۶)$$

تعریف کرده‌اند که میانگین PIT به‌صورت زیر محاسبه می‌شود:

[DOR: 20.1001.1.25385798.1399.31.1.6.0]

$$\overline{F}(u) = \frac{1}{n}\sum_{t=1}^{n} F_t(u|y_t)\,, \qquad 0 \le u \le 1.$$

برای بررسی این‌که آیا $\overline{F}(u)$، تابع توزیع تجمعی یک توزیع یکنواخت است، چادو و همکاران [۲] ترسیم یک شکل ستونی با H ستون را پیشنهاد کردند که ستون h، دارای ارتفاع

$$f_h = \overline{F}(h/H) - \overline{F}\left(\left(h-1\right)/H\right),\ h = 1, \dots, H$$

است. مقدار H به‌طور پیش‌فرض ۱۰ انتخاب می‌شود. شکل U حاکی از کم پراکنش توزیع پیش‌گو است درحالی‌که شکل U وارونه نشان دهنده‌ی بیش‌پراکنش است. کالبیدن حاشیه‌ای[5] به‌صورت اختلاف میانگین تابع توزیع تجربی مشاهدات به ازای هر $y \in R$ به‌صورت زیر تعریف می‌شود:

$$\frac{1}{n}\sum_{t=1}^{n} P_t(y) - \frac{1}{n}\sum_{t=1}^{n} \mathbb{I}(y_t \le y). \qquad (۷)$$

در عمل کالبیدن حاشیه‌ای برای مقادیر y در حوزه مقادیر مشاهدات رسم می‌شود. اگر مدل مناسب باشد، توزیع حاشیه‌ای پیش‌بینی‌ها، شبیه توزیع حاشیه‌ای مشاهدات است و(۷) باید نزدیک صفر باشد. انحراف عمده از صفر نشان‌دهنده نامناسب بودن مدل است. نایتینگ و همکاران [۶] نشان دادند که ارزیابی مدل توسط شکل ستونی PIT یا شکل کالبیدن حاشیه‌ای یک شرط لازم و نه کافی برای ایده‌ال بودن پیش‌گویی است. آن‌ها با مدلی موافق هستند که حداکثر شفافیت را در میان تمام مدل‌های پیشنهادی دارد. شفافیت، تمرکز توزیع پیش‌بینی است و به‌وسیله پهنای بازه‌های پیش‌بینی اندازه‌گیری می‌شود. یک ارزیابی هم‌زمان کالبیدن و شفافیت که در یک امتیاز عددی خلاصه شود، با قوانین امتیازدهی مناسب قابل انجام است، [۶]. تعدادی از قوانین امتیازدهی در جدول ۱ ارائه شده‌اند. هر یک از قوانین امتیازدهی ویژگی‌های مختلف توزیع پیش‌بینی و فاصله آن با داده‌های مشاهده شده را اندازه‌گیری می‌کنند. به‌جز میانگین توان دوم خطای نرمال شده، مدل با پایین‌ترین امتیاز ترجیح داده می‌شود. یعنی مدلی که پایین‌ترین مقدار را از معیارهای امتیازدهی جدول ۱ اختیار کند مناسب است. امتیاز میانگین توان دوم خطا تنها موردی است که به توزیع بستگی ندارد و با عنوان میانگین توان دوم خطای پیشگویی شناخته می‌شود. میانگین

[DOR: 20.1001.1.25385798.1399.31.1.6.0]

توان دوم خطاهای نرمال شده، واریانس مانده‌های پیرسون را اندازه‌گیری می‌کند. اگر مدل مناسب باشد خطای نرمال شده NSE (جدول ۱) نزدیک به یک است. امتیازدهی دیوید-سباستینی نوع دیگری با یک جمله‌ی اضافی برای جبران بیش‌برآورد انحراف استاندارد است.

جدول ۱- قوانین امتیازدهی برای انتخاب مدل مناسب

معیارهای امتیازدهی	
میانگین توان دوم خطاها	$MSE = \frac{1}{n}\sum_{t=1}^{n} r_t^2$
میانگین توان دوم خطاهای نرمال شده	$NSE = \frac{1}{n}\sum_{t=1}^{n}\left(\frac{r_t}{v_t}\right)^2$
امتیازدهی دیوید-سباستینی	$DS = \frac{1}{n}\sum_{t=1}^{n}\left(\left(\frac{r_t}{v_t}\right)^2 + \log(p_t(y_t))\right)$
امتیازدهی لگاریتمی	$LOG = \frac{1\sum_{t=1}^{n}\log(p_t(y_t))}{n}$
امتیازدهی درجه دوم	$QUD = -\frac{1}{n}\sum_{t=1}^{n}\left(2p_t(y_t) - \sum_{y=0}^{\infty} p_t^2(y)\right)$
امتیازدهی کروی	$SPH = -\frac{1}{n}\sum_{t=1}^{n}\left(p_t(y_t)/\sum_{y=0}^{\infty} p_t^2(y)\right)$
امتیازدهی رتبه احتمال	$RP = \frac{1}{n}\sum_{t=1}^{n}\left(\sum_{y=0}^{\infty}\left(P_t(y) - I(y_t \le y)\right)\right)$

ابزارهای دیگر انتخاب مدل، معیار اطلاع آکائیکه (AIC) و معیار اطلاع بیزی (BIC) هستند. مدل با کمترین مقدار معیار اطلاع مناسب است. معیارهای اطلاع به‌صورت

$$AIC = -2\tilde{\ell}\left(\hat{\theta}, \widehat{\sigma^2}\right) + 2df$$

و

$$BIC = -2\tilde{\ell}\left(\hat{\theta}, \widehat{\sigma^2}\right) + \log(n_{eff})\, df$$

[DOR: 20.1001.1.25385798.1399.31.1.6.0]

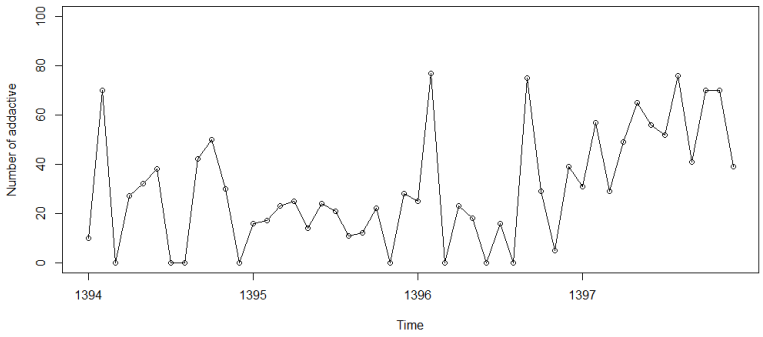
محاسبه می‌شوند که در آن $\tilde{\ell}\left(\hat{\theta}, \widehat{\sigma^2}\right) = \sum_{t=1}^{n} \log(p_t(y_t))$، df تعداد کل پارامترها (شامل ضریب پراکندگی) و n_{eff} تعداد مشاهدات موثر هستند. به‌طور کلی BIC نسبت به AIC مدل‌های خلاصه‌تری را انتخاب می‌کند. در ادامه روش‌ها و مدل‌های ارائه شده روی داده‌های ترک اعتیاد شهرستان سمنان مورد تحلیل قرار می‌گیرد.

۳- داده‌های ترک اعتیاد شهرستان سمنان

اطلاعات و داده‌های این بخش با همکاری اداره بهزیستی استان سمنان گردآوری شده است. داده‌ها با توجه به مراجعین افراد معتاد به یکی از مواد مخدر به مراکز ترک اعتیاد در شهرستان سمنان طی مدت چهار سال از ماه فروردین سال ۱۳۹۴ تا اسفند ماه سال ۱۳۹۸ به‌صورت ماهانه جمع‌آوری شده‌اند. مراکز ترک اعتیاد به دو صورت سرپایی و اقامتی می‌باشد که در این مقاله آمار مراکز سرپایی شهرستان سمنان مورد بررسی قرار گرفته است. برای در نظر گرفتن اثر زمان، مدل رگرسیونی روی مشاهدات قبلی لحاظ شده است. اثر فصلی آن با رگرسیون بر λ_{12} یعنی ۱۲ واحد زمانی گذشته (یک سال) تعیین شده‌اند. عوامل بی‌شماری در اعتیاد تاثیرگذار می‌باشد و به همین نسبت برخی عامل‌ها نیز مانع و یا انگیزه‌ای برای ترک اعتیاد است. گروه معتادین مورد مطالعه زنان و مردان می‌باشند. عوامل تاثیرگذار به این شرح هستند: سن، وضعیت شغلی، میزان تحصیلات، وضعیت تأهل، سن شروع مصرف مواد، تعداد دفعاتی که اقدام به ترک کرده است. در مورد وضعیت شغلی، دو گروه شاغل و بیکار در نظر گرفته شده است. میزان تحصیلات گروه مورد بررسی بازه‌ی بی‌سواد تا دکتری را در برگرفته که آن در دو طبقه‌ی دارای سواد زیر دیپلم و بالای دیپلم گروه‌بندی شده‌اند. همچنین افراد ممکن است، زندگی مشترک را تجربه کرده باشند و بعد از آن به هر دلیلی طلاق گرفته و یا جدا زندگی کنند. در این بررسی تمامی افرادی که مجرد بوده و یا به هر دلیلی جدا از همسرشان زندگی می‌کنند در یک گروه قرار گرفته‌اند و وضعیت تاهل افراد معتاد در دو صورت متاهل و غیرمتاهل لحاظ شده‌اند. داده‌ها با روش‌های ارائه شده مورد تحلیل و بررسی قرار گرفته‌اند و برای انجام این‌کار از نرم‌افزار R بسته‌ی tscount استفاده شده است. تعداد معتادین به مراکز ترک اعتیاد در یک ماه به

[DOR: 20.1001.1.25385798.1399.31.1.6.0]

عنوان متغیر پاسخ در نظر گرفته شده است. شکل سری زمانی تعداد معتادین مراجعه کننده جهت ترک در یک بازه‌ی چهار ساله در شکل ۱ رسم شده است. ابتدا مدل‌های پواسونی و دوجمله‌ای منفی با توجه به متغیرهای توضیحی برازش شدند، سپس با ملاک‌های معرفی شده در بخش ۲ بهترین مدل برای برازش به داده‌ها انتخاب می‌شود. نتایج حاصل از برازش مدل پواسونی در جدول ۲ خلاصه شده‌اند.



شکل ۱- تعداد معتادین مراجعه‌کننده جهت ترک اعتیاد

جدول ۲- برآورد ضرایب مدل پواسون

پارامترهای مدل	براورد	خطای استاندارد	فاصله اطمینان ۹۵ درصد
$\beta_{\circ}$ (عرض از مبدأ)	۰٫۲۳	۰٫۲۶	$(۰٫۳-, ۰٫۷۵)$
$(Y_t - ۱)\beta_{۱}$	-۰٫۰۰۷	۰٫۰۲	$(-۰٫۰۰۵, ۰٫۰۴)$
$(\lambda_{t-۱})\beta_{۱۲}$	۰٫۰۰۷	۰٫۰۲	$(-۰٫۰۳, ۰٫۰۴)$
سن $(X_{۱})$	۰٫۰۵	۰٫۰۰۸	$(۰٫۰۴, ۰٫۰۷)$
وضعیت تحصیلات $(X_{۲})$	۰٫۰۲	۰٫۰۰۶	$(۰٫۰۱, ۰٫۰۴)$
وضعیت تأهل $(X_{۳})$	-۰٫۰۲	۰٫۰۰۶	$(-۰٫۰۳, -۰٫۰۰۶)$
تعداد فرزندان $(X_{۴})$	۰٫۱۹	۰٫۱۲	$(-۰٫۰۵, ۰٫۴۳۳)$
بعد خانوار $(X_{۵})$	۰٫۰۴	۰٫۰۵	$(-۰٫۰۵, ۰٫۰۱۳)$
وضعیت شغلی $(X_{۶})$	۰٫۰۲	۰٫۰۰۵	$(۰٫۰۱, ۰٫۰۰۳)$
وضعیت مصرف سیگار $(X_{۷})$	-۰٫۰۰۱	۰٫۰۰۴	$(-۰٫۰۱, ۰٫۰۰۷)$

[DOR: 20.1001.1.25385798.1399.31.1.6.0]

بنابراین مدل حاصل برابراست با:

$$M_۱: log(\lambda_t) = ۰/۲۳ - ۰/۰۰۷Y_{t-۱} + ۰/۰۰۷\lambda_{t-۱۲} + ۰/۰۵X_{۱t} + ۰/۰۲X_{۲t} - ۰/۰۲X_{۳t} + ۰/۱۹X_{۴t} + ۰/۰۴X_{۵t} + ۰/۰۲X_{۶t} - ۰/۰۰۱X_{۷t}, \qquad t = ۱, \dots, ۴۸.$$

که در آن متغیرهای توضیحی $X_۱, \cdots, X_۷$ در جدول ۲ مشخص شده‌اند. با توجه به فواصل اطمینان ۹۵ درصد جدول ۲ متغیرهای سن، میزان تحصیلات، وضعیت تاهل و وضعیت شغلی تأثیر معنی‌داری روی تعداد مراجعات معتادین به مراکز ترک اعتیاد شهرستان سمنان دارند. نتایج برازش مدل دوجمله‌ای منفی در جدول ۳ خلاصه شده است. معیارهای AIC و BIC در این مدل به‌ترتیب برابر با ۲۴۴/۰۸۱ و ۲۶۲/۷۹ است که در آن متغیرهای وضعیت شغلی و مصرف سیگار معنی‌دار شده‌اند.

جدول ۳- برآورد ضرایب مدل دوجمله‌ای منفی

پارامترهای مدل	براورد	خطای استاندارد	فاصله اطمینان ۹۵ درصد
$\beta_۰$	$۳/۴۲ \times ۱۰^{-۷}$	$۳/۹۳ \times ۱۰^{-۴}$	$(-۷/۷ \times ۱۰^{-۴}, -۷/۷ \times ۱۰^{-۴})$
$\beta_۱$	$۱/۰۱ \times ۱۰^{-۱۱}$	$۸/۲۷ \times ۱۰^{-۶}$	$(۱/۶۲ \times ۱۰^{-۵}, ۱/۶۲ \times ۱۰^{-۵})$
$\beta_{۱۲}$	$۲/۸۸ \times ۱۰^{-۱۰}$	$۱/۷۳ \times ۱۰^{-۵}$	$(-۳/۳۹ \times ۱۰^{-۵}, ۳/۳۹ \times ۱۰^{-۵})$
سن	$۶/۸۹ \times ۱۰^{-۳}$	۰/۱۳۷	$(-۹/۰۸ \times ۱۰^{-۴}, ۰/۶۶)$
میزان تحصیلات	۰/۳۳	۰/۱۷	(−۰/۰۸۴۴, ۰/۵۱)
وضعیت تأهل	۰/۲۱	۰/۱۵	(−۰/۰۸۴۴, ۰/۵۱)
تعداد فرزندان	۰/۴۷	۱/۶۱	(−۲/۶۹, ۳/۶۳)
بعد خانوار	$۱/۲۸ \times ۱۰^{-۶}$	۱/۴۱	(−۲/۷۶, ۲/۷۶)
وضعیت شغلی	۰/۴۴	۰/۱۵	(۰/۱۴, ۰/۷۳)
وضعیت مصرف سیگار	۰/۳۱	۰/۱۳	(۰/۰۴۹, ۰/۵۷)
$\sigma^۲$	۰/۳۹		

مدل برازش شده به‌صورت

[DOR: 20.1001.1.25385798.1399.31.1.6.0]

$$M_۲:\ \lambda_t = ۳/۴۲ \times ۱۰^{-۷} + ۱/۰۱ \times ۱۰^{-۱۱} Y_{t-۱} + ۲/۸۸ \times ۱۰^{-۱۰} \lambda_{t-۱۲} + ۶/۸۹ \times ۱۰^{-۳} X_{۱t} + ۰/۳۳ X_{۲t} + ۰/۲۱ X_{۳t} + ۰/۴۶ X_{۴t} + ۰/۱۳ X_{۵t} + ۰/۴۴ X_{۶t} - ۰/۳۱ X_{۷t}, \quad t = ۱, \dots, ۴۸, \tag{۸}$$

است. از آنجایی که متغیر وضعیت شغلی در هر دو مدل معنی‌دار بود، مدل‌های پواسونی و دوجمله‌ای منفی فقط برای این متغیر هریک به‌صورت جداگانه به داده‌ها برازش شدند. نتایج برازش مدل پواسونی با متغیر وضعیت شغلی در جدول ۴ خلاصه شده است.

جدول ۴- برآورد ضرایب مدل پواسونی با متغیر وضعیت شغلی

پارامترهای مدل	براورد	خطای استاندارد	فاصله اطمینان ۹۵ درصد
$\beta_۰$	۲/۳۳	۰/۰۹۷	$(۲/۱۴۵, ۲/۵۲۷)$
$\beta_۱$	۰/۰۳۲	۰/۰۱۹	$(-۰/۰۰۵, ۰/۰۶۹)$
$\beta_{۱۲}$	-۰/۰۲	۰/۰۲۱	$(-۰/۰۶۴, ۰/۰۲۱)$
وضعیت شغلی	۰/۰۳۲	۰/۰۰۱	$(۰/۰۳۰, ۰/۰۳۵)$

مدل پواسونی با متغیر وضعیت شغلی به‌صورت

$$M_۳:\ log(\lambda_t) = ۲/۳۳ + ۰/۰۳۲ Y_{t-۱} - ۰/۰۲ \lambda_{t-۱۲} + ۰/۰۳۲ X_{۶t}, \quad t = ۱, \dots, ۴۸,$$

است. جدول ۵ نتایج برازش مدل دوجمله‌ای منفی با متغیر وضعیت شغلی را نشان می‌دهد.

جدول ۵- برآورد ضرایب دوجمله‌ای منفی با متغیر وضعیت شغلی

پارامترهای مدل	براورد	خطای استاندارد	فاصله اطمینان ۹۵ درصد
$\beta_۰$	۷×۱۰^{-۸}	$۲/۲۹ \times ۱۰^{-۴}$	$(۲/۱۴۵, ۲/۵۲۷)$
$\beta_۱$	$۵/۹ \times ۱۰^{-۱۱}$	$۵/۴۲ \times ۱۰^{-۶}$	$(-۱۰۶ \times ۱۰^{-۵}, ۱/۰۶ \times ۱۰^{-۵})$
$\beta_{۱۲}$	۰/۰۱۳۵	۰/۰۱۲۹	$(-۰/۰۱۲, ۰/۰۳۹)$
وضعیت شغلی	۱/۲۴	۷/۲۳	$(۱/۱, ۱/۳۸)$
$\sigma^۲$	۰/۰۷۲		

[DOR: 20.1001.1.25385798.1399.31.1.6.0]

همچنین مدل برازش شده به‌صورت

$$M_۴: \quad \lambda_t = ۷ \times ۱۰^{-۸} + ۵/۹ \times ۱۰^{-۱۱} Y_{t-۱} + ۰/۰۱۳۵\lambda_{t-۱۲} + ۱/۲۴X_{۶t}, \quad t = ۱, \dots, ۴۸$$

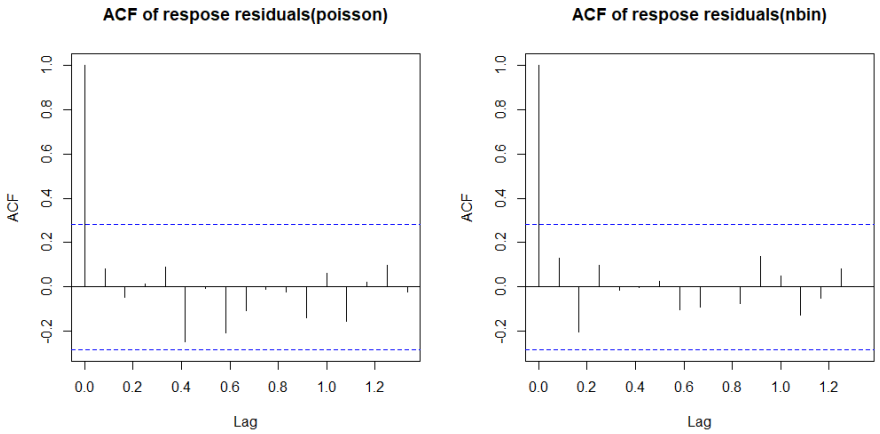
به‌دست می‌آید. نتایج معیارهای AIC، BIC و MSE به‌دست‌آمده از برازش مدل‌های ذکر شده در جدول ۶ خلاصه شده است.

جدول ۶- نتایج معیارهای BIC، AIC و MSE به‌دست آمده از برازش مدل‌ها

مدل‌ها	AIC	BIC	MSE
$M_۱$	۳۱۱/۴۹	۳۳۰/۲۰	۳۴/۷۸
$M_۲$	۲۴۴/۰۸۱	۲۶۲/۷۹	۱۱/۵۵
$M_۳$	۳۱۱/۴۹	۳۳۰/۲۰	۸۸/۶۴
$M_۴$	۳۰۱/۶۶	۳۱۱/۰۲۵	۳۳/۱۹

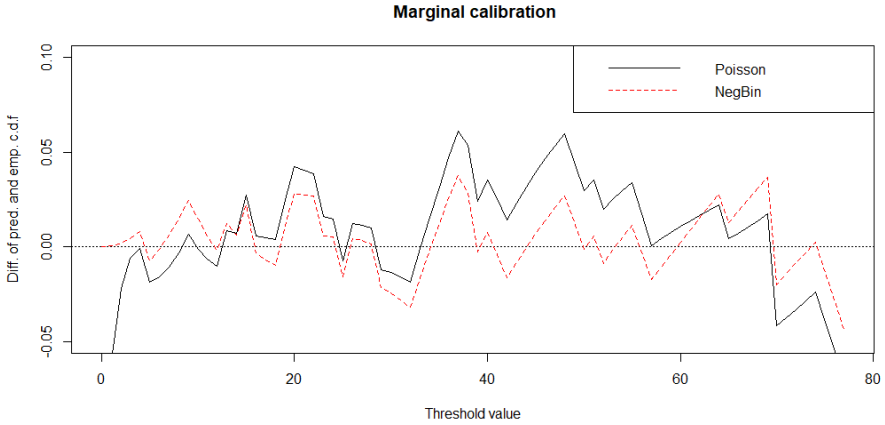
با توجه به جدول ۶ همه معیارها مدل دوجمله‌ای منفی $M_۲$ را تأیید می‌کنند، که این مدل برازش بهتری در مقایسه‌ی مدل پواسونی دارد. برای اطمینان بیشتر، از معیارهای دیگر نیز استفاده شده است که در ادامه بررسی می‌شوند. شکل ۲ مانده پاسخ‌ها را برای هر دو مدل پواسونی و دوجمله‌ای منفی نشان می‌دهد. با توجه به تابع خودهمبستگی، هیچ خودهمبستگی یا اثر فصلی مشاهده نمی‌شود.

[DOR: 20.1001.1.25385798.1399.31.1.6.0]



شکل ۲- شکل خودهمبستگی باقیمانده‌های دو مدل پواسنی (M_1) و دوجمله‌ای منفی (M_2)

شکل کالیبدن حاشیه‌ای در شکل ۳ برای دو مدل رسم شده‌اند، با توجه به فرمول ۶ هرچه این مقادیر به صفر نزدیکتر باشند، برازش بهتر را بیان می‌کنند. در اکثر مواقع شکل دوجمله‌ای منفی به خط صفر نزدیکتر است.

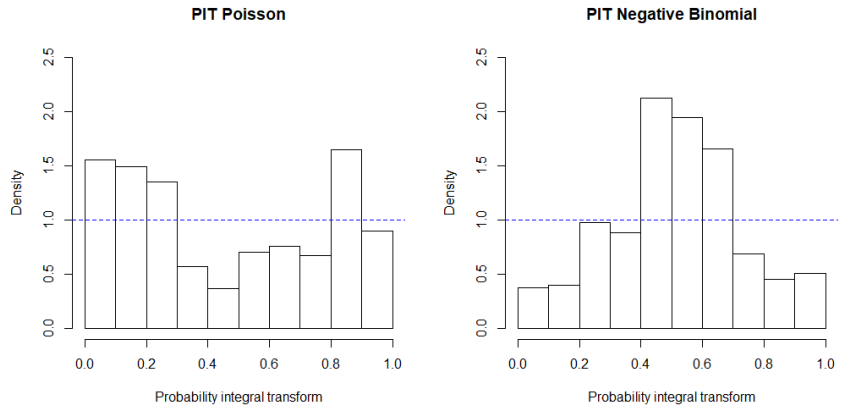


شکل ۳- شکل کالیبدن حاشیه‌ای دو مدل پواسنی (M_1) و دوجمله‌ای منفی (M_2)

شکل ۴ مقادیر PIT برای هر دو برازش مدل را نشان می‌دهد. همان‌طور که مشاهده می‌شود، شکل پواسونی U شکل است و پراکنش کمتر را نشان می‌دهد. برخلاف آن

[DOR: 20.1001.1.25385798.1399.31.1.6.0]

شکل ستونی دوجمله‌ای منفی شکل U برعکس است و بیش‌پراکنش را نشان می‌دهد.



شکل ۴- شکل شکلPIT برای دو مدل پواسنی و دوجمله‌ای منفی

به‌عنوان آخرین ابزار مقایسه و ارزیابی مدل، قوانین امتیازدهی را محاسبه کرده‌ایم که نتایج در جدول ۷ آورده شده است.

جدول ۷- مقایسه مدل پواسونی و دوجمله‌ای منفی با قوانین امتیازدهی

معیارهای	دوجمله‌ای منفی	پواسون
LOG	۲٫۳۳۴	۳٫۰۳۶
QUD	-۰٫۲۳۳	-۰٫۰۸۲
SPH	-۰٫۴۰۲	-۰٫۲۵۸
RP	۱٫۷۳۹	۳٫۲۰۷
DS	۰٫۶۴۹	۴٫۴۴۲
NSE	۰٫۲۸۰	۱٫۵۵۴
MSE	۱۱٫۵۵۵	۳۴٫۷۸۲

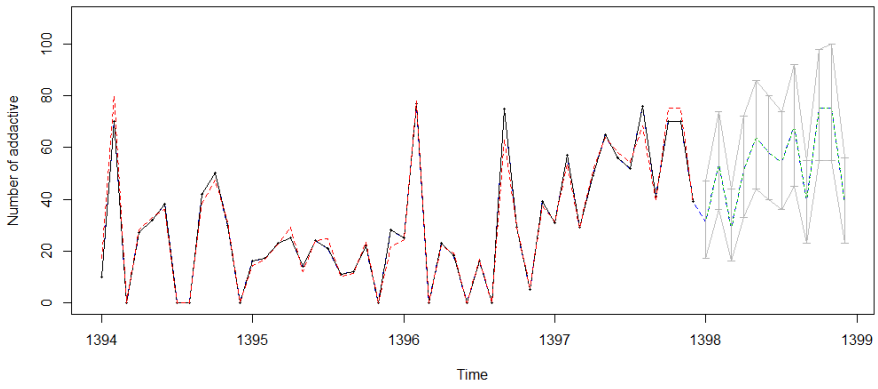
با توجه به نتایج امتیازدهی چون مقادیر آن برای مدل دوجمله‌ای منفی کمتر است پس برازش بهتری نسبت به مدل پواسونی دارد. تفسیر نتایج ضرائب رگرسیونی از برازش مدل دوجمله‌ای منفی در جدول ۸ به این صورت است که ضریب β_1 مربوط به رگرسیون روی مشاهدات گذشته پاسخ با تاخیر یک واحد زمان و برآورد β_{12} رگرسیون روی مقدار میانگین شرطی ۱۲ واحد گذشته در زمان است. برآورد ضریب

بیش‌پراکنش σ^2 به پارامتر پراکندگی ϕ توزیع دوجمله‌ای منفی مربوط است و این رابطه $\phi = \frac{1}{\sigma^2}$ است. بر این اساس، مدل برازش شده برای Y_t در طول زمان t به‌صورت $Y_t \mid \mathcal{F}_{t-1} \sim NegBin(\lambda_t, \phi = 13/07)$ با رابطه‌ی (۸) می‌باشد. خطاهای استاندارد برآورد پارامتر رگرسیون و فاصله اطمینان پراکندگی در جدول ۳ بر اساس تقریب نرمال داده شده در (۶) است. برای ضریب پیش پراکنش اضافی σ^2 در توزیع دوجمله‌ای منفی، هیچ تحلیل تقریبی برای خطای استاندارد در دسترس نیست. گاها مشکل برآورد برای پارامتر پراکندگی مدل‌ها اتفاق می‌افتد و ممکن است پراکندگی σ^2 کوچک باشد. مثلاً در این‌جا مدل برازش شده نزدیک مدل پوآسون است و به همین علت مقدار σ^2 کوچک است اما این مسئله عدم قطعیت انتخاب مدل را نشان نمی‌دهد. ضریب برآورد β_1 مربوط به مرتبه‌ی اول اتورگرسیو بسیار کوچک است و حتی پایین‌تر از اندازه‌ی خطای تقریبی آن است که نشان می‌دهد وابستگی قابل توجهی به تعداد رگرسیون‌ها در ماه‌های گذشته وجود ندارد. یک اثر فصلی با ضریب همبستگی مرتبه دوازدهم β_{12} مشاهده می‌شود. با توجه به مدل برازش شده براساس داده‌ها تا اسفند ۱۳۹۷، می‌توان تعداد معتادین مراجعه کننده در سال ۱۳۹۸ پیش‌بینی کرد. نتایج پیش‌بینی در جدول ۸ خلاصه شده‌اند.

جدول ۸- پیش‌بینی تعداد مراجعین در سال ۱۳۹۸ براساس مدل دوجمله‌ای منفی

	فروردین	اردیبهشت	خرداد	تیر
۱۳۹۸	۳۱	۵۴	۲۹	۵۱
	مرداد	شهریور	مهر	آبان
۱۳۹۸	۶۴	۵۸	۵۴	۶۹
	آذر	دی	بهمن	اسفند
۱۳۹۸	۳۹	۷۵	۷۵	۳۹

[DOR: 20.1001.1.25385798.1399.31.1.6.0]



شکل ۵- مقادیر برازش شده و مقادیر پیش‌بینی شده بر اساس مدل دوجمله‌ای منفی

نمایش گرافیکی پیش‌بینی‌ها در شکل ۵ آمده است. شکل سری زمانی داده‌ها، برازش و پیش‌بینی نیز در شکل ۵ رسم شده است. خطوط شکسته مقادیر برازش شده و خط چین‌ها مقادیر پیش‌بینی همراه با فاصله پیشگویی ۹۵ درصد می‌باشد.
در نهایت باتوجه به نتایج مشاهده می‌شود که از متغیرهای کمکی در نظر گرفته شده، وضعیت شغلی بر تعداد مراجعین ترک اعتیاد معنی‌دار است. در این میان مدل دوجمله‌ای منفی بهترین برازش را به این داده‌ها دارد.

۴- بحث و نتیجه‌گیری

در این مقاله مدل‌های سری‌زمانی شمارشی بر اساس فرایند پواسن ارائه و مورد مطالعه قرار گرفتند. یک مدل کلی اتورگرسیو تعمیم یافته بیان و معیارهای ارزیابی مدل نیز ارائه شد. مدل‌های سری‌زمانی پواسنی و دوجمله‌ای منفی روی داده‌های ترک اعتیاد شهرستان سمنان به‌کار گرفته شد. همچنین عوامل تاثرگذار بر تعداد معتادین مراجعه کنند به مراکز ترک اعتیاد مانند سن، وضعیت شغلی و غیره به‌صورت متغیرهای توضیحی در این مدل‌ها مورد تحلیل قرار گرفتند. در میان آن‌ها متغیر وضعیت شغلی که حاکی از سطح رفاه و خطر از دست دادن شغل معتادین می‌باشد، با اطمینان ۹۵ درصد در هر دو مدل پواسنی و دوجمله‌ای تاثیر معنی‌داری بر تعداد مراجعین به مراکز

[DOR: 20.1001.1.25385798.1399.31.1.6.0]

ترک اعتیاد دارد. همه معیارهای پیشنهادی انتخاب مدل، مدل دوجمله‌ای منفی را برای تحلیل این داده‌ها پیشنهاد کردند. در انتها پیش‌بینی تعداد مراجعات براساس مدل انتخابی صورت گرفت که افزایش تعداد مراجعین را نسبت به سال‌های گذشته نشان می‌داد.

در این مقاله رهیافت درستنمایی برای براورد پارامترهای مدل به کار گرفته شد، به عنوان پیشنهاد می‌توان از رهیافت بیزی سلسله مراتبی و یا رهیافت بیز تقریبی با وجود پیچیدگی مدل‌ها برای بالا بردن سرعت محاسبات استفاده کرد.

توضیحات

1. Generalized Linear Model
2. Integer-Valued Autoregressive conditional heteroskedasticity
3. Probabilistic calibration
4. Probability integral transform
5. Marginal calibration

مرجع‌ها

[1] Christou, V., and Fokianos, K. (2014). Quasi-Likelihood Inference for Negative Binomial Time Series Models. *Journal of Time Series Analysis*, **35**, 55–78.

[2] Czado, C., Gneiting, T., and Held, L. (2009). Predictive Model Assessment for Count Data. *Biometrics*, **65**, 1254–1261.

[3] Fahrmeir, L., and Tutz, G. (2001). *Multivariate Statistical Modelling Based on Generalized Linear Models*. Springer-Verlag, New York.

[4] Ferland, R., Latour, A., and Oraichi, D. (2006). Integer-Valued GARCH Processes. *J. Time Ser. Anal.*, **27**, 249–329.

[5] Fokianos, K. (2012). Count time series models. In: Handbook of Statistics, **30**, 315–347.

[DOR: 20.1001.1.25385798.1399.31.1.6.0]

[6] Gneiting, T., Balabdaui, F., and Raftery, A.E. (2007). Probabilistic Forecasts, Calibration and Sharpness. *Journal of the Royal Statistical Society B*, **69**, 243–268.

[7] Liboschik, T., Kerschke, P., Fokianos, K., and Fried, R. (2016). Modelling Interventions in INGARCH Processes. *International Journal of Computer Mathematics*, **93**, 640–657.

[8] McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models*. Second ed. Chapman & Hall, London.

[9] Nelder, J.A., and Wedderburn, R.W.M. (1972). Generalized Linear Models. *J. R. Stat. Soc. Ser. A,* **135**, 370–384.

[10] Priestley, M. B. (1981). *Spectral Analysis and Time Series*. Academic Press, London.

[11] Ver, H.J., and Boveng, P. (2007). Quasi-Poisson vs. Negative Binomial Regression: How Should We Model Overdispersed Count Data?. *Ecology*, **88**, 2766–2772.

امید کریمی
دکتری آمار
سمنان،دانشگا سمنان، گروه آمار.
رایانشانی: omid.karimi@semnan.ac.ir

فاطمه حسینی
دکتری آمار
سمنان،دانشگا سمنان، گروه آمار.
رایانشانی: fatemeh.hoseini@semnan.ac.ir

سپیده تشرفی
فوق‌لیسانس آمار
سمنان،دانشگا سمنان، گروه آمار.
رایانشانی: s.tasharofy@gmail.com

[DOR: 20.1001.1.25385798.1399.31.1.6.0]