پانزدهمین سمینار احتمال
و فرآیندهای تصادفی
۸ و ۹ شهریور ۱۴۰۴
دانشگاه کردستان

15th Seminar On Probability and Stochastic Processes
August 30-31, 2025
University of Kurdistan

انجمن آمار ایران

مدل‌بندی تصادفات جاده‌ای استان خراسان جنوبی با استفاده از مدل‌های اتورگرسیو صحیح مقدار

زهره نخعی زاده [۱] ، سارا جمهوری[۲]

[۱،۲] گروه آمار، دانشگاه بیرجند

چکیده: در علم آمار، به دنباله‌ای از مشاهدات که به طور معمول در فواصل زمانی منظم ثبت می‌شوند، "سری زمانی" گفته می‌شود. بسیاری از متغیرها، مانند تعداد تصادفات، قربانیان جرایم و سایر موارد مشابه، ماهیتی گسسته دارند. بنابراین، تحلیل و پیش‌بینی این نوع داده‌ها با استفاده از مدل‌های سری زمانی شمارشی، مناسب‌تر از دیگر روش‌ها است. در این مقاله، از مدل‌های سری زمانی شمارشی برای بررسی داده‌های تصادفات روزانه برون‌شهری در شهرستان‌های استان خراسان جنوبی استفاده شده است. هدف از این پژوهش، شناسایی مدل مناسبی است که بهترین برازش را به داده‌ها داشته باشد. پس از مقایسه مدل‌های مختلف، مدلی که بهترین تطابق را با داده‌های موجود نشان می‌دهد، انتخاب شده و بر اساس آن پیش‌بینی‌های لازم انجام می‌شود.

واژه‌های کلیدی: پیش‌بینی، عملگر نازک‌کننده، سری زمانی شمارشی.

کد موضوع‌بندی ریاضی (۲۰۲۰): 62M10.

۱ مقدمه

تصادفات جاده‌ای یکی از مهم‌ترین و چالش‌برانگیزترین مسائل در حوزه حمل و نقل مدرن محسوب می‌شوند. این تصادفات بر اثر عوامل مختلفی از جمله افزایش جمعیت، رشد تقاضای سفر و بنابراین استفاده بیشتر از وسیله‌های نقلیه موتوری رخ می‌دهند. در نتیجه، ایمنی جاده‌ها و خیابان‌ها اهمیت ویژه‌ای پیدا می‌کند، چرا که کاهش تصادفات و جلوگیری از وقوع آن‌ها هدف اصلی در ارتقاء ایمنی حمل و نقل است.

در بسیاری از کشورها، از جمله ایران، تصادفات جاده‌ای به عنوان یکی از اصلی‌ترین دلایل مرگ‌ومیر و جراحات تشخیص داده شده است. در نتیجه توجه به این امر یکی از اندیشه‌های اصلی در برنامه‌ریزی‌های حمل و نقل و سیاست‌های ایمنی است. بر اساس آمارهای

[۱] سخنران، znakhaeezadeh72@gmail.com

موجود، ایران یکی از کشورهایی است که با بالاترین نرخ مرگ‌ومیر ناشی از تصادفات جاده‌ای مواجه است. تحلیل آمارهای تصادفات می‌تواند به مدیران و کارشناسان این حوزه کمک کند تا روندهای تصادفات را بهتر درک کرده و اقدامات موثری برای کاهش آن‌ها انجام دهند.

برای پیش‌بینی تعداد تصادفات در مناطق خاص و در بازه‌های زمانی مختلف، از مدل‌های پیش‌بینی و سری‌های زمانی استفاده می‌شود. در سال‌های اخیر، تکنیک‌های رگرسیونی که بر فرض استقلال داده‌ها تکیه دارند، کمتر مورد استفاده قرار می‌گیرند، چرا که محدودیت‌هایی در مدل‌سازی سری‌های زمانی پیچیده دارند. در این راستا، محققان در قالب مدل‌های سری زمانی پیوسته مقدار، همچون مدل‌های $ARMA$، $ARIMA$ و $SARIMA$، داده‌های مربوط به تصادفات را مورد مطالعه قرار داده‌اند و نتایج مفیدی در حوزه پیش‌بینی و تحلیل روندهای تصادفات ارائه داده‌اند. مطالعات متعددی در قلمرو مدل‌های سری زمانی انجام شده است که از این میان می‌توان به یوسف‌زاده چابک و همکاران (۲۰۱۶)، آگیمنگ و همکاران (۲۰۲۳)، نصیری و همکاران (۲۰۲۳) و ولوگدین و همکاران (۲۰۲۴) و از میان پژوهش‌های داخلی به توکلی و رحیم‌اف (۱۳۹۰)، عسگری و همکاران (۱۳۹۴) و امیدی و همکاران (۱۳۹۶) اشاره کرد.

به دلیل ماهیت گسسته و غیرمنفی داده‌های تصادفات، مدل‌های سنتی و پیوسته مقدار سری‌های زمانی برای تحلیل این داده‌ها مناسب نیستند. از این‌رو، توسعه مدل‌هایی که به طور خاص برای این نوع داده‌ها طراحی شده‌اند، ضروری است. یکی از ساده‌ترین و در عین حال پرکاربردترین مدل‌ها برای داده‌های شمارشی، مدل $INAR(۱)$ معرفی‌شده توسط مک‌کنزی (۱۹۸۵) است. این مدل نسخه‌ای صحیح‌مقدار از مدل خودرگرسیو مرتبه اول $AR(۱)$، است و با وجود ساختار ساده، از انعطاف‌پذیری بالایی برخوردار است. مدل $INAR(۱)$ نه تنها امکان تحلیل فرآیندهای شمارشی ایستا را فراهم می‌کند، بلکه مبنایی برای توسعه مدل‌های پیشرفته‌تر نیز محسوب می‌شود. همچنین، مفاهیم مهمی مانند برآورد پارامترها، تحلیل مانده‌ها و استنباط آماری را می‌توان به سادگی در قالب این مدل پیاده‌سازی کرد. اساس ساخت مدل‌های $INAR(۱)$ مبتنی بر عملگر نازک‌کننده[1] است. با در نظر گرفتن عملگرهای نازک کننده متفاوت، مدل‌های سری زمانی متنوعی به‌دست می‌آیند. در پژوهش حاضر از سری‌های زمانی شمارشی برای مدل‌سازی داده‌های تصادفات جاده‌ای در سطح استان خراسان جنوبی بهره گرفته شده است.

۲ مدل‌های شمارشی

در این بخش، خانواده مدل‌های سری زمانی شمارشی $INAR(۱)$ و خواص آن مورد بررسی قرار گرفته است. ابتدا مفهوم عملگر نازک‌کننده دوجمله‌ای[2] بیان می‌شود.

تعریف ۱.۲. فرض کنید X یک متغیر تصادفی گسسته و $\{Z_i, i = ۱, ۲, \ldots\}$ دنباله‌ای از متغیرهای تصادفی مستقل و هم‌توزیع با توزیع مشترک برنولی و احتمال موفقیت $\alpha \in (۰, ۱)$ باشند. در این‌صورت توزیع شرطی متغیر تصادفی $\alpha \circ X = \sum_{i=۱}^{X} Z_i$ به شرط X دوجمله‌ای $Bin(X, \alpha)$ است. عملگر $\circ$ عملگر نازک‌کننده دوجمله‌ای نامیده می‌شود.

ال اوش و الزید (۱۹۸۷) مدل $INAR(۱)$ را با استفاده از عملگر دوجمله‌ای به صورت زیر معرفی کردند.

[1] Thinning operator
[2] Binomial thinning

تعریف ۲.۲. فرآیند تصادفی شمارشی $\{X_t, t \in \mathcal{T}\}$ یک فرآیند $INAR(۱)$ نامیده می‌شود هرگاه

$$X_t = \alpha \circ X_{t-۱} + \varepsilon_t, \quad t \in \mathbb{Z}, \quad \alpha \in (۰, ۱), \tag{۱.۲}$$

که در آن جملات نوآوری[۳] $\{\varepsilon_t; t \in \mathcal{T}\}$ دنباله‌ای از متغیرهای تصادفی نامنفی، صحیح مقدار، مستقل و هم‌توزیع با میانگین μ_ε و واریانس $\sigma^۲_\varepsilon$ است. همچنین فرض می‌شود که متغیرهای تصادفی موجود در عملگر نازک‌کننده α و ε_t از $\{X_t; t \in \mathcal{T}\}$ مستقل هستند.

ال اوش و الزید (۱۹۸۷) نشان دادند که فرآیند$INAR(۱)$، یک زنجیر مارکوف همگن با احتمال انتقال یک مرحله‌ای به صورت زیر است

$$\begin{aligned} p_{k|l} &:= P(X_t = k | X_{t-۱} = l) \\ &= \sum_{j=۰}^{min\{k,l\}} \binom{l}{j} \alpha^j (۱-\alpha)^{l-j} P(\varepsilon_t = k - j). \end{aligned} \tag{۲.۲}$$

توجه ۳.۲. با انتخاب توزیع‌های مختلفی برای جملات نوآوری، مدل‌های $INAR(۱)$ مختلفی حاصل می‌شود. یکی از مدل‌های پرکاربرد، حالتی است که در آن توزیع حاشیه‌ای X_t پواسون با میانگین $\mu = \frac{\lambda}{(۱-\alpha)}$ است و $\{\varepsilon_t; t \in \mathcal{T}\}$ دنباله‌ای از متغیرهای تصادفی پواسون با میانگین و واریانس λ می‌باشد.

توجه ۴.۲. با وجود اینکه توزیع پواسون یکی از مهم‌ترین توزیع‌های آماری مرتبط با فرآیندهای شمارشی است، اما استفاده از آن همیشه در مدل‌های سری زمانی شمارشی مناسب نیست. زیرا در این توزیع میانگین و واریانس با یکدیگر برابرند و این خاصیت عملاً در سری‌های زمانی مشاهده نمی‌شود. به همین علت، مدل‌های دیگری مورد مطالعه قرار گرفته است. اگر در (۱.۲) جملات نوآوری از توزیع دوجمله‌ای منفی $NB(n, \pi)$ پیروی کنند، مدل $NBINAR(۱)$ حاصل می‌شود.

توجه ۵.۲. در برخی مدل‌های $INAR(۱)$ فرض می‌شود که ضریب نازک‌کننده متغیری تصادفی است. در اینصورت مدل به صورت $X_t = \alpha_\phi \circ X_{t-۱} + \varepsilon_t$ نوشته می‌شود. جو (۱۹۹۶) با در نظر گرفتن $\phi = \frac{۱}{(n+۱)}$ و توزیع بتا $Beta(\frac{۱-\phi}{\phi}\alpha, \frac{۱-\phi}{\phi}(۱-\alpha))$ برای α_ϕ وقتی جملات نوآوری دارای توزیع $NB(n(۱-\alpha), \pi)$ باشند، مدل $RCNBINAR(۱)$ را معرفی کرد. در این مدل، توزیع حاشیه‌ای X_t دوجمله‌ای منفی $NB(n, \pi)$ خواهد بود.

۱.۲ برآورد پارامترهای مدل

روش‌های متعددی مانند روش گشتاورها، برآورد حداقل مربعات شرطی و برآورد ماکزیمم درستنمایی برای تخمین پارامترهای مدل وجود دارد. برای انجام برآورد پارامترها بر پایه روش درستنمایی، می‌توان از احتمال‌های شرطی و خاصیت مارکوفی سری زمانی $INAR(۱)$ بهره برد. تابع لگاریتم درستنمایی در مدل $INAR(۱)$ با بردار پارامتر $\boldsymbol{\theta}$ عبارت است از

$$\ell(\boldsymbol{\theta}) = \log p_{x_۱}(\boldsymbol{\theta}) + \sum_{t=۲}^{T} \log p_{x_t|x_{t-۱}}(\boldsymbol{\theta}), \tag{۳.۲}$$

که در آن احتمالات انتقال با توجه به رابطه (۲.۲) محاسبه می‌شوند. پارامتر $\boldsymbol{\theta}$ در مدل‌های مختلف $INAR(۱)$ متفاوت است. به عنوان مثال در مدل پواسون $INAR(۱)$ که در ملاحظه ۳.۲ معرفی شد $\boldsymbol{\theta} = (\alpha, \lambda)$ و در مدل $NBINAR(۱)$ معرفی شده در ملاحظه ۴.۲، $\boldsymbol{\theta} = (\alpha, n, \pi)$ است. در این صورت برآوردگر ماکسیمم درستنمایی پارامتر $\boldsymbol{\theta}$ به صورت $\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\text{argmax}} \ell(\boldsymbol{\theta})$ می‌باشد.

[۳] Innovations

۲.۲ تشخیص و کفایت مدل

در این بخش، ابزارهای مورد نیاز برای تشخیص مناسب‌ترین مدل برای برازش به داده‌ها معرفی می‌شوند. برای تشخیص ساختار وابستگی سری، نمودار خودهمبستگی جزئی نمونه‌ای $(SPACF)$رسم می‌شود. اگر $\hat{\rho}_{par}(k)$ در تاخیر اول مخالف صفر و در بقیه تاخیرها به سمت صفر میل کند، مدل $INAR(۱)$ برای برازش به داده‌ها مناسب خواهد بود. برای تعیین مدل $INAR(۱)$ مناسب برای داده‌ها، می‌توان از شاخص $\hat{I} = \frac{S^۲}{\bar{X}}$ استفاده کرد که در آن $\bar{X} = \frac{۱}{T}\sum_{t=۱}^{T} X_t$ و $S^۲ = \frac{۱}{T}\sum_{t=۱} T(X_t - \bar{X})^۲$. اگر مدل پواسون برای برازش به داده‌ها مناسب باشد، در اینصورت باید مقدار این آماره تقریبا برابر یک باشد. همچنین این آماره به طور مجانبی از توزیع نرمال با

$$E(\hat{I}) = ۱ - \frac{۱}{T}\frac{۱+\alpha}{۱-\alpha}, \quad V[\hat{I}] \approx \frac{۲}{T}\frac{۱+\alpha^۲}{۱-\alpha^۲},$$

پیروی می‌کند (شوئر و ویب ، ۲۰۱۴). اگر چندین مدل برای برازش به داده‌ها مناسب باشند، از معیارهایی مانند آکائیک و معیار اطلاع بیزی که به صورت زیر هستند

$$AIC = -۲\ell(\boldsymbol{\theta}) + ۲m, \quad BIC = -۲\ell(\boldsymbol{\theta}) + ۲\log T,$$

برای شناسایی مدل مناسب‌تر استفاده می‌شود (m تعداد پارامترهای مجهول مدل و T تعداد مشاهدات است). بعد از شناسایی بهترین مدل، لازم است که کفایت مدل بررسی شود. محاسبه باقی‌مانده‌های استاندارد شده پیرسون اولین رویکرد در کفایت مدل است. این مقادیر طبق رابطه زیر حاصل می‌شود.

$$e_t = \frac{x_t - E(X_t|x_{t-۱})}{\sqrt{V(X_t|x_{t-۱})}}, \quad t = ۲, ۳, \ldots, T.$$

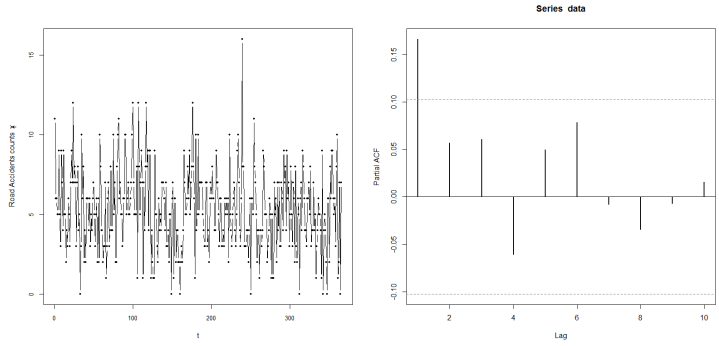
برای یک مدل مناسب انتظار می‌رود که این باقی‌مانده‌ها دارای میانگین و واریانس به ترتیب برابر صفر و یک بوده و با یکدیگر همبستگی نداشته باشد. واریانس بزرگتر (کوچکتر) از یک نشان دهنده وجود بیش‌پراکندگی (کم‌پراکندگی) نسبت به مدل در نظر گرفته شده است (هاروی و فرناندز ، ۱۹۸۹).

۳ پیش‌بینی تصادفات جاده‌ای

در این پژوهش، ابتدا آمار و ارقام مربوط به تصادفات جاده‌ای [۴] در سه سال متوالی ۱۴۰۱، ۱۴۰۲ و ۱۴۰۳ گزارش شده است. این گزارش نشان‌دهنده وضعیت تصادفات، میزان تلفات، محل وقوع و زمان‌بندی آن‌ها است که به منظور شناسایی روندها و ارائه راهکارهای بهبود امنیت جاده‌ای اهمیت فراوان دارد.

در سال ۱۴۰۱، تعداد کل تصادفات ثبت‌شده به ۲۰۵۲ مورد رسید که از این تعداد، ۲۵۸ مورد مرگ و میر ناشی از تصادف گزارش شده است. همچنین، حدود ۳۴ درصد از این تصادفات در جاده‌های روستایی، ۶۰ درصد در جاده‌های اصلی و ۶ درصد در جاده‌های فرعی رخ داده است. در این سال، ۶۴ درصد تصادفات در طول روز و ۳۶ درصد در شب اتفاق افتاده است، که نشان‌دهنده اهمیت زمان‌بندی و دقت در رانندگی در ساعات مختلف شبانه‌روز است. در سال ۱۴۰۲، تعداد تصادفات کمی کاهش یافته و به ۱۹۸۶ مورد رسید. در این سال، تعداد ۲۴۸ فوتی به ثبت رسیده که نسبت به سال قبل کمی افزایش یافته است. همچنین ۵۷ درصد تصادفات در جاده‌های اصلی، ۳۶ درصد در جاده‌های روستایی و ۷ درصد در جاده‌های فرعی رخ داده است. الگوی زمانی همچنان برتری تصادفات

[۴] داده‌ها از پلیس راه استان خراسان جنوبی اخذ شده است.



شکل ۱: تعداد تصادفات روزانه جاده‌ای استان خراسان جنوبی و نمودار ACF داده‌ها

در روز را نشان می‌دهد (۶۳ درصد در روز و ۳۷ درصد در شب). در سال ۱۴۰۳، تعداد تصادفات دوباره کمی کاهش یافته و به ۱۹۵۳ مورد رسیده است. تعداد فوتی‌ها نیز به ۲۰۰ تن کاهش یافته است، که نشان‌دهنده نسبتی پایدار است. در این سال، توزیع جغرافیایی تصادفات نسبت به سال‌های قبل تقریباً ثابت است (۵۷ درصد در جاده‌های اصلی، ۳۵ درصد در جاده‌های روستایی و ۸ درصد در جاده‌های فرعی). همچنین، الگوهای زمانی همچنان بر تصادفات روز تأکید دارند (۶۴ درصد در طول روز و ۳۶ درصد در شب).

بر اساس این آمار، می‌توان نتیجه گرفت که تعداد تصادفات جاده‌ای در سه سال مورد بررسی کاهش نسبی داشته است، اما میزان تلفات در سال ۱۴۰۲ کمی بیشتر شده است که نیازمند توجه و اقدامات بیشتری در زمینه ایمنی جاده‌ها و آموزش رانندگان است. همچنین، سهم قابل توجه تصادفات در جاده‌های روستایی و در ساعات روز نشان می‌دهد که برنامه‌ریزی برای ارتقاء زیرساخت‌ها و اقدامات مراقبتی در این فضاها اهمیت دارد. به طور کلی، تمرکز بر کاهش تصادفات در جاده‌های روستایی و کنترل سرعت و رعایت قوانین در زمان‌های مختلف شبانه‌روز می‌تواند نقش مؤثری در بهبود وضعیت ایمنی جاده‌ای کشور ایفا کند. در این بخش مجموعه داده‌های تصادفات روزانه جاده‌ای در استان خراسان جنوبی در سال ۱۴۰۳ مورد بررسی قرار می‌گیرد ($T = ۳۶۵$). شکل (a) ۱، نمودار سری زمانی تعداد تصادفات را نشان می‌دهد. مشاهدات در فاصله صفر تا ۱۶ تصادف در روز قرار دارند. روند یا تغییرات فصلی قابل توجهی در داده‌ها مشاهده نمی‌شود. نوسانات بالا و پایین موجود در این نمودار بیان‌کننده سطح متوسطی از همبستگی است. این موضوع از نمودار ACF نمونه‌ای نیز مشهود است. با توجه به مقدار $\hat{\rho}_{par}(۱) = ۰/۱۶۶۲$ ملاحظه می‌شود که این مقدار به‌طور قابل توجهی از صفر بزرگتر است. بنابراین، مدلی شبیه به $AR(۱)$ برای توصیف مشاهدات مناسب است. به وضوح، مقدار میانگین $\bar{x} \approx ۵/۳۲۰۵$ کمتر از واریانس $s^۲ \approx ۶/۶۵۷۹$ است. لذا، در داده‌ها پراکندگی زیادی وجود دارد. از طرف دیگر، مقدار احتمال برآورد شده در نقطه صفر برابر $p_۰ \approx ۰/۰۲۱۹۱$ است که بزرگتر از مقدار متناظر آن از توزیع پواسون، یعنی $exp(-\bar{x}) \approx ۰/۰۰۴۸$ می‌باشد. بنابراین، برازش یک مدل $INAR(۱)$ به داده‌ها، قابل قبول به نظر می‌رسد. برای انجام آزمون تشخیص بیش‌پراکندگی در مدل $INAR(۱)$، فرضیه صفر اینکه داده‌ها از مدل $INAR(۱)$ با نواوری‌های پواسون [۵] پیروی می‌کنند، آزمون می‌شود. تحت فرضیه صفر، $E(\hat{I}) = ۰/۹۹۶۱$ و $V(\hat{I}) = ۰/۰۷۰۰$ است. با توجه به مقدار $p - value = ۰/۰۰۰۴$ مشخص می‌شود که پراکندگی بیش از حد و احتمالا قابل توجه است. بنابراین مدل $RCNBINAR(۱)$ برای برازش به داده‌ها مناسب‌تر است. این مدل با مدل پواسون $INAR(۱)$ از نظر معیارهای AIC و BIC مقایسه می‌شود. برآورد میانگین و شاخص پراکندگی عوامل نوآوری‌ها به ترتیب $\hat{I} = \hat{I}(۱ + \hat{\alpha}) - \hat{\alpha} \approx ۱/۲۸۹۱$ و $\bar{\mu}_\varepsilon = \bar{x}(۱ - \hat{\alpha}) \approx ۴/۴۳۵۸$ است. از بررسی مقادیر AIC و BIC موجود در جدول

[۵] $PoissonINAR(۱)$

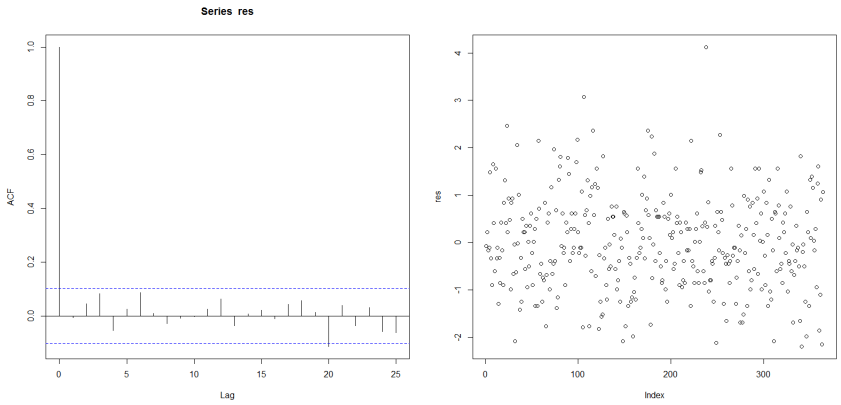
جدول ۱: آمار توصیفی داده‌های تصادفات در شهرهای استان خراسان جنوبی در سال ۱۴۰۳

متغیر	میانگین	میانه	انحراف معیار	مینم	ماکزیمم	چولگی	کشیدگی
تصادفات سال ۱۴۰۳	۵٫۳۲۱	۵	۲٫۵۸۰	۰	۱۶	۰٫۳۴۹	۳٫۳۱۹

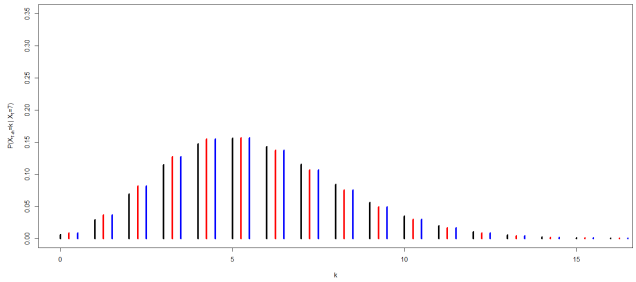
جدول ۲: برآوردهای درستنمایی ماکسیمم پارامترها و AIC و BIC در مدل‌های مختلف

$Model$	Parameter			AIC	BIC
	۱	۲	۳		
$i.i.d.Poisson$	۵٫۳۲۰			۱۷۳۵٫۱۰۷	۱۷۳۹٫۰۰۷
μ	۰٫۱۲۰				
$Poisson.INAR(۱)$	۴٫۶۰۴	۰٫۱۳۵		۱۷۲۶٫۹۱۶	۱۷۳۴٫۷۱۶
(λ, α)	(۰٫۲۴۵)	(۰٫۰۴۱)			
$i.i.d.NB$	۱۹٫۷۲۶	۰٫۷۸۵		۱۷۲۶٫۸۵۱	۱۷۳۴٫۶۵۱
(n, π)	(۷٫۲۴۵)	(۰٫۰۶۱)			
$NBINAR(۱)$	۱۵٫۳۴۰	۰٫۷۷۳	۰٫۱۵۵	۱۷۱۹٫۳۲۹	۱۷۳۱٫۰۲۸
(n, π, α)	(۶٫۳۷۸)	(۰٫۰۶۹)	(۰٫۰۴۹)		
$RCNBINAR(۱)$	۲۱٫۴۵۹	۰٫۰۸۰۱	۰٫۱۵۸	۱۷۱۹٫۲۳۰	۱۷۳۱٫۰۳۰
(n, π, α)	(۹٫۳۱۶)	(۰٫۸۰۱)	(۰٫۱۵۸)		

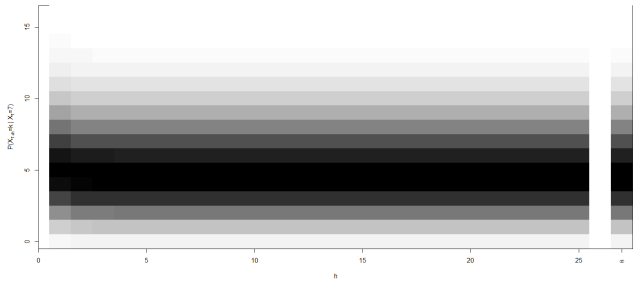
۲ مشخص می‌شود که مدل‌های $INAR(۱)$ بهتر از مدل $i.i.d$ عمل می‌کنند. مقدار برآورد شده α به طور قابل توجهی از صفر بزرگتر است. از مقایسه مدل‌های $INAR(۱)$ برحسب معیارهای AIC و BIC، ملاحظه می‌شود که مدل $RCNBINAR(۱)$ برای برازش به داده‌ها مناسب‌تر است. برای بررسی کفایت این مدل، نمودارهای پراکنش و ACF نمونه‌ای باقیمانده‌های پیرسون در شکل ۲ رسم شده‌اند. با توجه به این شکل، هیچگونه همبستگی معنی‌داری بین باقیمانده‌ها مشاهده نمی‌شود. برای این مدل، توزیع‌های پیش‌بینی h–گام روبه جلو به شرط مقدار آخرین مشاهده ($x_{۳۶۵} = ۷$)، با استفاده از رابطه (۲.۲) به دست می‌آیند. نمودار این توزیع‌ها در شکل ۳ و ۴ رسم شده است. با توجه به شکل ۳ پیش‌بینی مبتنی بر میانه توزیع به ازای سه مقدار $h = ۱$ (مشکی)، $h = ۵$ (قرمز) و $h = ۱۰$ (آبی) در نقطه $k = ۵$ اتفاق می‌افتد. شکل ۴ نیز مقدار احتمال بیشتر را روی مقادیر ۴، ۵ و ۶ نشان می‌دهد. لذا انتظار می‌رود که در روزهای آتی، به طور متوسط ۵ تصادف در روز اتفاق بیافتد.



شکل ۲: نمودارهای پراکنش و ACF نمونه‌ای باقیمانده‌های پیرسون



شکل ۳: هیستوگرام توزیع پیش‌بینی h گام به جلو برای $h = ۱, ۵, ۱۰$



شکل ۴: توزیع پیش‌بینی h گام به جلو برای $h = ۱, \ldots, ۲۵, \infty$

بحث و نتیجه‌گیری

در این مقاله، با استفاده از مدل‌های سری زمانی شمارشی، تعداد تصادفات رانندگی جاده‌ای در سطح استان خراسان جنوبی، مورد مطالعه قرار گرفت. پس از تحلیل، مشخص گردید که مدل $RCNBINAR(۱)$ برای برازش به داده‌ها، مناسب‌تر از سایر مدل‌های رقیب است.

مراجع

امیدی، نبی.، خلیلی، کرم.، امیدی، محمدرضا. و جعفری اسکندری، میثم. (۱۳۹۶). ارائه مدل اقتصادسنجی پیش بینی کشته شدگان و مصدومان تصادفات جاده ای، مطالعات مدیریت ترافیک، شماره ۴۴، ۲۱-۴۴.

توکلی شیرازی، نیما و رحیم اف، کامران. (۱۳۹۰). مدلسازی سری زمانی مجروحین و متوفیات حوادث تصادفات جاده ای ایران با رویکردهای MA، AR، $ARMA$، دومین کنفرانس ملی تصادفات جاده ای، سوانح ریلی و هوایی، زنجان.

عسگری، حشمت اله.، امیدی، محمد رضا.، امیدی، نبی. و مفتاحی، هادی. (۱۳۹۴). پیش بینی بروز تخلفات از نوع سرعت غیرمجاز در آزادراه‌ها، مطالعات مدیریت ترافیک. شماره ۳۸، ۲۵-۳۸.

Agyemang, E. F., Mensah, J. A., Ocran, E., Opoku, E. Nortey, E. N. N. (2023). Time series based road traffic

accidents forecasting via SARIMA and Facebook Prophet model with potential changepoints. *Heliyon*, 9, e22544.

Al-Osh, M. A., Alzaid, A. A. (1987). First-order integer-valued autoregressive (INAR(1)) process. *Journal of Time Series Analysis*, 8(3), 261-275.

Freeland, R. K. (1998). Statistical analysis of discrete time series with application to the analysis of workers' compensation claims data (Doctoral dissertation, University of British Columbia).

Harvey, A.C., Fernandes, C. (1989). Time series models for count or qualitative observations. Journal of Business &*Economic Statistics*, 7(4), 407–417.

Joe, H. (1996). Time series models with univariate margins in the convolution-closed infinitely divisible class. *Journal of Applied Probability*, 33(3), 664–677.

McKenzie, E. (1985). Some simple models for discrete variate time series. *Water Resources Bulletin*, 21(4), 645–650.

Nasiri, H., Mohammadpour, S. I., Dahaghin, M. (2023). Forecasting time trend of road traffic crashes in Iran using the macro-scale trafiic flow characteristics. *Heliyon*, 9(3), e14481.

Schweer, S., Weiß, C.H. (2014). Compound Poisson INAR(1) processes: stochastic properties and testing for overdispersion. Computational Statistics and Data Analysis, 77, 267–284.

Vologdin, S., Kasatkina, E., Kasatkin, A. (2024). Analyzing and forecasting road traffic accidents and their consequences: a case study of the Udmurt republic. *E3S Web of Conferences*, 471, 06006.

Yousefzadeh-Chabok, S., Ranjbar-Taklimie, F., Malekpouri, R., Razzaghi, A. (2016). A Time Series Model for Assessing the Trend and Forecasting the Road Traffic Accident Mortality. *Arch Trauma Res*, 5(3), e36570.

Modeling Road Traffic Accidents in South Khorasan Province Using Integer-Valued Autoregressive Models

Nakhaeezadeh, Z.[1], **Jomhoori, S.**[2]

[1,2]Department of Statistics, University of Birjand, Iran.

Abstract: In statistics, a sequence of observations that are typically recorded at regular time intervals is referred to as a ”time series”. Many variables, such as the number of accidents, crime victims, and other similar events, exhibit a discrete nature. Therefore, the analysis and forecasting of these types of data are more appropriately conducted using count time series models. In this article, count time series models are applied to analyze daily intercity accident data from the counties of South Khorasan province. These data span a three-year period, from 2021 to 2024. The objective of this research is to identify the most suitable model that provides the best fit to the data. After comparing various models, the one that demonstrates the best match with the existing data is selected, and based on it, the necessary forecasts are made.

Keywords: Counting timeseries , Forecasting, Thinning operator.

Mathematics Subject Classification (2020): 62M10.
