

مدلسازی فراوانی تصادفات در تقاطع های سه شاخه شهری

مقاله پژوهشی

محمد کوهی^{*}، دانش آموخته کارشناسی ارشد، دانشکده مهندسی عمران، دانشگاه پیام نور، تهران، ایران

شاهین شعبانی، دانشیار، گروه مهندسی عمران- راه و ترابری، دانشگاه پیام نور، تهران، ایران

^{*}پست الکترونیکی نویسنده مسئول: m.koohhi@gmail.com

دریافت: ۹۸/۱۱/۲۰ - پذیرش: ۹۸/۰۴/۰۵

صفحه ۵۰-۳۷

چکیده

ایمینی در تقاطع به دلیل نقاط برخورد احتمالی فراوان و مسیرهای مختلف موجود، شکل ویژه‌ای به خود میگیرد که نیاز به تحلیل‌های ویژه‌ای خواهد داشت. یکی از روشهای تحلیل ایمینی در تقاطع ها، ایجاد مدل‌های پیشبینی تصادفات با استفاده از مدل‌های رگرسیون پواسون و دوجمله‌ای منفی است. هر چند که استفاده از رگرسیون دوجمله‌ای منفی بدلیل رسیدگی به بیش پراکندگی داده‌ها ارجحیت دارد. اما تحقیق حاضر در ابتدای راه به جهت مدلسازی تصادفات تقاطع های شهری، هر دو مدل را مورد ارزیابی قرار می دهد. پس از آزمون مدل‌ها با معیارهای آکایک، انحراف و لکارتیم درستمایی، مدل رگرسیون دوجمله‌ای منفی برآزش بهتری نشان داد و به عنوان مدل بهتر انتخاب شد. برای مدلسازی، از داده های یک دوره ۵ ساله از تصادفات رخ داده در تقاطع های سه شاخه شهر بروجرد استفاده شده است. متغیرهای ورودی در مدل پس از انجام آزمون ارزیابی معناداری انتخاب شد. این متغیرها شامل تعداد خطوط عبوری، تعداد خطوط گردش به راست با جزیره ترافیکی و زاویه تقاطع بود. پس از انتخاب متغیرها، دقت مدل نیز مورد بررسی قرار گرفت. برای معتبرسازی مدل از معیارهای R^2 ، خطای جذر میانگین مربع، میانگین خطای مطلق و میانگین قدر مطلق درصد خطا استفاده شد. پس از ارزیابی دقت مدل در پیشبینی تصادفات، مشخص شد که مدل برای ارزیابی تصادفات نتایج قابل قبولی ارائه می‌دهد اما برای پیشبینی تصادفات، نتایج از دقت بالایی برخوردار نیست ($R^2 = 0.52$). مقدار R^2 برای مدل نشان می‌دهد که پیشبینی ها دقت بالایی نخواهند داشت اما هنوز پیشبینی‌های قابل قبولی می‌تواند ارائه دهد.

واژه‌های کلیدی: مدلسازی تصادفات، تقاطع‌های شهری، رگرسیون دوجمله‌ای منفی

۱-مقدمه

نقلیه جهت‌های مختلف، باید بصورت ویژه مد نظر باشد. یک راه برای تجزیه و تحلیل ایمینی تقاطع، مدلسازی عوامل مختلف انسانی، محیطی و هندسی، و ویژگی‌های وسیله نقلیه با تصادفات مشاهده شده با استفاده از مدل‌های آماری پیشبینی تصادف است. استفاده از روش‌های آماری در مدلسازی و پیشبینی تصادفات بطور چشمگیری رو به افزایش است. مدل‌های پیشبینی تصادف، با قابلیت‌هایشان برای تعیین رابطه بین فراوانی و یا شدت تصادفات، ابزاری بسیار مفید در ایمینی ترافیک هستند. در زمینه تحلیل ایمینی تقاطع با استفاده

تصادفات در تقاطع در مقایسه با تصادفات در قطعات راه، پیچیده‌تر هستند و به همین دلیل، راننده یک وسیله نقلیه باید هوشیارتر باشد تا قادر به بررسی چند رویداد متفاوت باشد، مانند توجه به علائم و چراغ راهنمایی، تنظیم سرعت، وسیله‌های نقلیه‌ای که از جهت مخالف می‌آیند یا در حال گردش هستند، وسیله‌های نقلیه‌ای که از چپ و راست می‌آیند و غیره. این عمل در تقاطع‌ها دارای دسترسی‌های شیب‌دار، حتی مشکل‌تر نیز می‌شود. ایمینی در تقاطع ها به دلیل ماهیت ریسک بالای برخوردها در تقاطع بین وسایل

از مدل‌های پیشبینی تصادف، یافته‌های متعددی بدست آمده است. مطالعات موجود، یک رابطه قوی بین هندسه راه، محیط و چراغ راهنمایی و فراوانی تصادفات در تقاطع‌ها را نشان می‌دهد. برخی از ویژگی‌های هندسی مهم که می‌تواند در برابر فراوانی تصادفات در تقاطع‌ها مدل کرد شامل تعداد خطوط دسترسی‌ها، عرض دسترسی، متوسط عرض خطوط، مسافت دید از خط توقف، زاویه‌های عمودی و افقی، وجود خطوط گردش به چپ و راست، عرض شانه، عرض میانه هستند. برای چراغ‌های راهنمایی عموماً تعداد فازهای هر سیکل، زمان سیکل، گردش به راست محافظت شده یا نشده استفاده می‌شوند. متغیرهای محیطی که عموماً استفاده می‌شوند، روز/شب، خشک/خیس، و نوع استفاده از زمین هستند. حجم ترافیک یکی از مهمترین عوامل در تجزیه و تحلیل داده‌های تصادف است. مطالعات نشان داده‌اند که AADT در شاخه‌های اصلی و فرعی تأثیری قابل توجه در فراوانی تصادفات دارد.

نتیجه استفاده از مدل‌های پواسون و به خصوص مدل رگرسیون دوجمله‌ای منفی در مدلسازی تصادفات جایگزین مدل‌های رگرسیون خطی سنتی شد. برای مثال سالیفو مدل‌های رگرسیون دوجمله‌ای منفی را برای مطالعه روابط بین تعداد تصادفات و جریان ترافیک، کنترل ترافیک و ویژگی‌های هندسی تقاطع‌های شهری در غنا ایجاد کرد. مجموع ۹۱ تقاطع شامل ۵۳ تقاطع ۳ شاخه و ۴۷ تقاطع چهارشاخه مقایسه شدند. مدل ایجاد شده برای تقاطع‌های سه شاخه نشان داد که متغیرهای حجم ترافیک شاخه‌های اصلی و فرعی و انحراف استاندارد سرعت نقطه‌ای متوسط در دسترسی‌های اصلی یک تأثیر مثبت بر تعداد تصادفات دارد (Nambuusi et al., 2015). هارنن و همکاران مدل‌های رگرسیون دوجمله‌ای منفی و پواسون را برای بررسی رابطه بین تصادفات موتورسیکلت در تقاطع‌های بدون چراغ شهری و متغیرهای توصیفی مربوط به جریان ترافیک، کنترل ترافیک، هندسه تقاطع و استفاده از زمین در مالزی به کار گرفتند. نتایج این مطالعه تنها برای مدل دوجمله‌ای منفی ارائه شد. نتایج نشان داد که متغیرهای مربوط به جریان ترافیک و کنترل ترافیک سبب افزایش تصادفات و هندسه ترافیک سبب کاهش تصادفات شده بود (Nambuusi et al., 2015). گریبی چند یافته مهم از مطالعه بر روی مدل‌های پیشبینی تصادف برای تقاطع‌های شهری تعریف کرد. گریبی مدل‌های تصادف ساده و عملی که می‌تواند تعداد مورد انتظار تصادفات در تقاطع‌های شهری را پیشبینی کند ایجاد کرد. این مدل‌ها می‌توانند برای شناسایی عوامل علی‌ایمنی راه برای شناسایی نقاط پرتصادف برای تحلیل ایمنی شبکه شهری به کار رود. مدل‌های پیشبینی تصادف بر اساس داده‌های متشکل از ۱۰۳۶ تقاطع شکل گرفتند. در ساخت این مدل‌ها، مدلسازی خطی تعمیم یافته با مربوط کردن فراوانی تصادف به

۲- پیشینه تحقیق

یک تقاطع که در واقع یک نقطه تلاقی جریان‌های ترافیکی از جهت‌های مخالف است، دارای چندین متغیر است که می‌تواند بر تصادفات اثر بگذارد. با این حال، به منظور مدل کردن انواع مختلف تصادفات بصورت جداگانه، وجود مقادیر زیادی از داده‌ها ضروری است، که ممکن است برای تمام مکان‌ها آسان نباشد. گاهی اوقات داده‌های ترافیک ممکن است به آسانی برای همه شاخه‌های تقاطع موجود نباشد، در نتیجه داده‌های حجم ترافیک مسیرهای موازی یا شاخه اصلی ممکن است برای تخمین حجم ترافیک شاخه فرعی یک تقاطع استفاده شود (Dixon et al., 2015; Shankar and Madanat, 2015). مدل‌های پیشبینی تصادف از سال‌های دور تا به حال، بطور مداوم مورد نقد و بررسی قرار گرفته‌اند. مدل‌های پیشبینی اولیه عموماً مبتنی بر یک رگرسیون خطی با فرض نرمال بودن عبارت خطا، واریانس ثابت برای باقیمانده‌ها، و رابطه خطی بین متغیرهای مستقل و وابسته بودند (Obaidat & Ramadan, 2012). با این حال محققان چندین مسئله با مدل‌های رگرسیون خطی را نشان داده‌اند (Lord & Mannering,

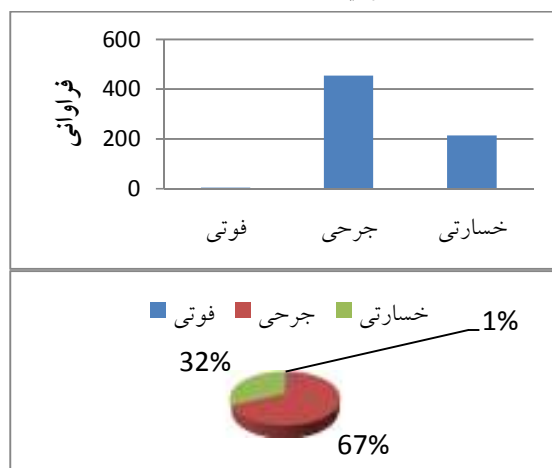
نقلیه در شاخه اصلی قابل بکارگیری است. همچنین این مدلها برای تقاطع های T شکل با حجم ترافیک روزانه ۱۸۶۰ تا ۱۳۶۰۰۰ برای شاخه فرعی، و ۶۷۳۰ تا ۱۷۳۸۶۰ برای شاخه اصلی نیز قابل کاربرد است (Karuppanen, 2015). با توجه به کاربرد گسترده مدل رگرسیون دوجمله‌ای منفی برای مدلسازی تصادفات، تحقیق حاضر نیز برای مدلسازی تصادفات در تقاطع‌های سه شاخه شهری، از مدلسازی رگرسیون دوجمله‌ای استفاده می‌کند. با توجه به ادبیات موضوع، برجسته‌ترین مزیت رگرسیون دوجمله‌ای منفی در برابر پواسون، رسیدگی به بیش پراکندگی داده‌ها می‌باشد.

۳- داده‌ها

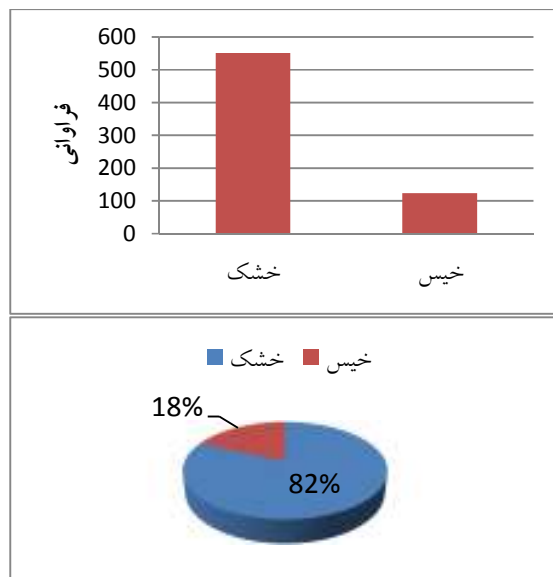
برای تحقیق حاضر یک دوره ۵ ساله برای جمع آوری داده‌ها و تحلیل آنها در نظر گرفته شده است. تعداد کلی ۶۷۲ تصادف در ۴۱ تقاطع در نظر گرفته شده در این تحقیق شناسایی شد. یک تحلیل توصیفی از داده‌های تصادف انجام شد. داده‌های تصادف در نظر گرفته شده در این تحقیق مربوط به دوره ۵ ساله از ابتدای ۲۰۰۵ تا انتهای ۲۰۰۶ است. تقاطع های T شکل در نظر گرفته شده این تحقیق در منطقه شهری بروجرد است. تحلیل اولیه داده‌های تصادفات در شکل‌های (۱) تا (۵) نشان داده شده است. منظور از کدهای ۱، ۲، ۳ در جدول (۵) به ترتیب، تصادفات جلو به عقب با وسایل نقلیه در جهت موافق، تصادفات وسایل نقلیه عبوری و در حال گردش به راست از جهت‌های مخالف و تصادفات جلو به عقب با وسایل نقلیه در حال گردش به چپ در خط موافق است.

متغیرهای توصیفی به کار گرفته شده است (GRIEBE, 2003). قبادی، حسن زاده و زراعت پیما به بررسی عوامل موثر بر ایمنی تقاطع‌های هم سطح شهری و ارزیابی آنها پرداخته و سعی در ارائه مدلی آماری برای پیشبینی تعداد تصادفات و ارائه راهکارهایی برای بهبود ایمنی تقاطع‌ها داشتند. آنها پس از مقایسه تقاطع های منتخب شهر تهران از نظر پرتصادف بودن و انتخاب متغیرهای موثر در تصادفات و انتخاب مدل مناسب دریافتند که دلیل عدم برابری میانگین داده‌ها با واریانس داده‌ها در دوجمله‌ای منفی و ضریب همبستگی بهتر و جواب منطقی‌تر نسبت به مدل خطی و پواسون، بهتر است از مدل دوجمله‌ای منفی استفاده شود (Ghobadi, Hasanzadeh, Zeraatpeima, 2016).

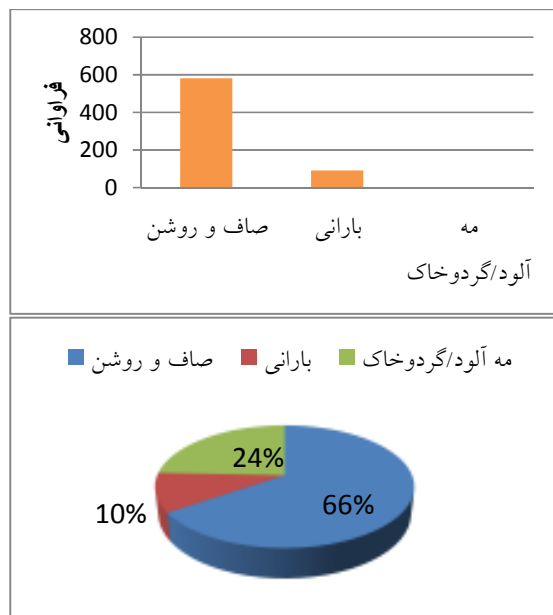
کلاروس و همکاران اقدام به ایجاد مدل پیشبینی تصادف مخصوص ایالت میسوری برای تقاطع های چراغ دار کرده‌اند. آنها مدل‌های پیشبینی تصادف مخصوص میسوری را برای شدت‌های فوتی و جرحی و خسارتی ایجاد کرده‌اند. متغیرهای پیشبینی کننده در نظر گرفته شده شامل AADT تقاطع، سرعت مجاز عبوری، نوع کنترل، خطوط گردش به راست و گردش به چپ انحصاری، گردش به راست در زمان چراغ قرمز و تسهیلات موجود در فاصله ۱۰۰۰ فوتی از تقاطع (ایستگاه اتوبوس، مدارس، و مراکز فروش الکل) بوده است. آنها نتیجه گرفته اند که هر ایالت باید مدل‌های پیشبینی خاص خود را ایجاد کنند (Claros et al., 2018). کاروپانن برای ایجاد مدل‌های پیشبینی تصادف برای ارزیابی ایمنی تقاطع های شهری، ۱۰۶ تقاطع در چنانی را انتخاب کرد و داده‌های حجم ترافیک را از مطالعات قبلی بدست آورد. او اشاره می‌کند که مدل‌های پیشبینی شده در تحقیقش برای تقاطع های چهارشاخه با حجم ترافیک ۲۵۰۴ تا ۸۳۵۰۳ وسیله نقلیه برای شاخه فرعی و ۷۵۱۳ تا ۱۶۴۸۸۰ وسیله



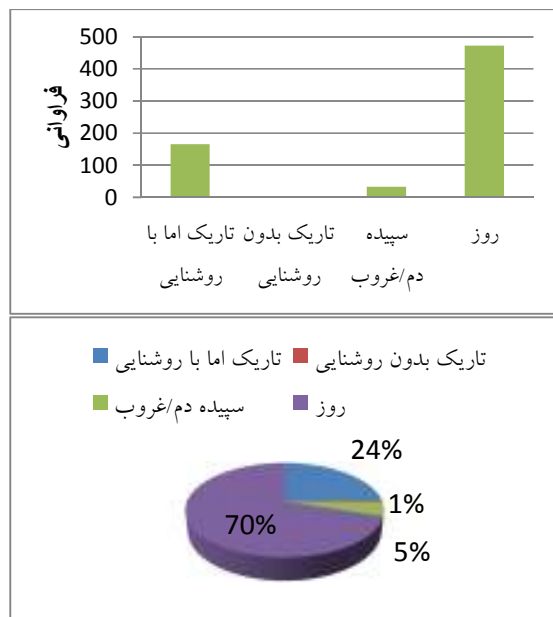
شکل ۱. نمودارهای فراوانی و درصد فراوانی شدت تصادفات



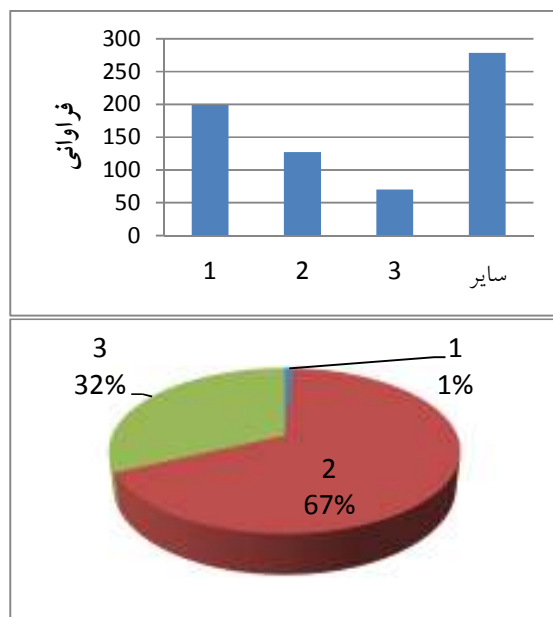
شکل ۲. نمودارهای فراوانی و درصد فراوانی تصادفات بر اساس شرایط سطح راه



شکل ۳. نمودارهای فراوانی و درصد فراوانی شرایط جوی تصادفات



شکل ۴. نمودارهای فراوانی و درصد فراوانی تصادفات بر اساس شرایط روشنایی



شکل ۵. نمودارهای فراوانی و درصد فراوانی تصادفات بر اساس نوع تصادف

تصادفات مربوط به وسایل نقلیه عبوری در جهت موافق است.

از تحلیل توصیفی فوق، نتیجه می شود که برای تقاطع های سه شاخه در نظر گرفته شده در این تحقیق، تصادفات جلو به عقب دارای بیشترین احتمال رخ دادن هستند و بیشتر

۴- تحلیل داده‌ها و مدل‌سازی

میانگین توزیع پواسون فوق λ است. این معادله را می‌توان برای تجزیه و تحلیل رگرسیون داده‌های تصادف جاده‌ای استفاده کرد.

این بخش به روش ایجاد یک مدل پیشبینی تصادف در تقاطع‌های سه شاخه شهر بروجرد می‌پردازد.

۴-۲- استراتژی مدل‌سازی وقوع تصادف

مطالعات اولیه‌ای که بر روی داده‌های تصادفات جاده‌ای انجام شده، از رگرسیون خطی برای تجزیه و تحلیل اثرات متغیرهای مستقل مختلف بر فراوانی تصادف استفاده کرده‌اند. رگرسیون خطی در نهایت با مدل‌های داده شمارشی جایگزین شده است. ماهیت تصادفات به شکلی است که همیشه گسسته، غیرمنفی و پراکنده هستند. برای مثال، یک تقاطع ممکن است هیچ تصادفی در یک سال یا چندین تصادف در یک سال نداشته باشد. این تصادفات ممکن است در یک روز، یک هفته یا در تمام طول سال اتفاق افتاده باشد. مدل رگرسیون پواسون یک مدل داده شمارشی پایه‌ای است. در این مدل، داده‌های تصادف باید توزیع پواسون را دنبال کنند. رگرسیون پواسون توسط محققان مختلفی برای مدل‌سازی داده‌های فراوانی تصادف استفاده شده است. فرضیه اساسی توزیع پواسون این است که فرض می‌کند میانگین متغیر تصادفی (تصادف) برابر با واریانس است. هنگامی که واریانس داده‌های تصادف بیشتر از میانگین آنها باشد، بیش پراکندگی اتفاق می‌افتد و هنگامی که کمتر از میانگین شود پراکندگی کمتر از حد رخ می‌دهد. داده‌های تصادف معمولاً بیش پراکنده هستند. استفاده از رگرسیون پواسون برای مدل داده‌های تصادف بیش پراکنده می‌تواند تخمین‌های غلطی از خطای استاندارد ارایه دهد، که می‌تواند منجر به معنادار شدن متغیرهای کم اهمیت شود. نتیجه احتمالی، یک استنتاج غلط درباره متغیرهای مستقل است. مدل‌های رگرسیون دوجمله‌ای منفی^۱ (NB) یا پواسون-گاما^۲ برای مدل‌کردن داده‌های تصادف بیش پراکنده استفاده می‌شوند. مدل‌های NB یک مولفه خطا اضافه شده در مدل دارند که مشکل بیش پراکندگی را مرتفع می‌سازد.

۴-۱- فرآیند وقوع تصادف

در یک دوره زمانی هزاران وسیله نقلیه از یک قطعه راه یا تقاطع عبور می‌کنند. برای هر وسیله نقلیه یک ریسک تصادف وجود دارد، با این حال تصادفات به ندرت روی می‌دهند. احتمال وقوع یک تصادف به تعداد وسایل نقلیه در راه بستگی دارد. این احتمال، برای تقاطع‌ها بدلیل شانس برخورد‌های ترافیکی بیشتر بین وسایل نقلیه، کمی بیشتر است. مجلی را در نظر بگیرید، هر زمان که یک وسیله نقلیه به این محل وارد می‌شود، شانس برابری برای رخ دادن یا رخ ندادن تصادف وجود دارد. وقتی وسیله نقلیه دیگری به محل وارد می‌شود، دوباره همین شانس برابر برای رخ دادن یا ندادن تصادف وجود دارد. دو رویداد به هم مرتبط نیستند و کاملاً مستقل از یکدیگرند و می‌تواند با آزمون‌هایی بصورت یک فرآیند برنولی مقایسه شود. اگر فرض کنیم که در یک تعداد از فرآیندهای برنولی n (وسایل نقلیه ورودی در محل)، احتمال p داشتن یک تصادف با متغیر Y نشان داده شود، پس آزمون‌های برنولی را می‌توان با توزیع دوجمله‌ای نشان داد. احتمال P داشتن یک تصادف بصورت زیر بیان می‌شود:

$$P(Y = y) = \binom{n}{y} p^y (1 - p)^{(n-y)} \quad (1)$$

که $y = 0, 1, 2, \dots, n$ میانگین توزیع دوجمله‌ای فوق برابر $E[Y] = np$ و واریانس آن بصورت $Var(Y) = np(1 - p)$ است. توزیع دوجمله‌ای را می‌توان بصورت یک توزیع پواسون به صورتی که احتمال داشتن یک تصادف در یک موقعیت (p) نادر و تعداد وسایل نقلیه عبوری از موقعیت (n) خیلی زیاد است، تقریب زد. هنگامی که $n \rightarrow \infty$ ، معادله (۱) را می‌توان بصورت معادله (۲) نوشت:

$$P(Y = y) = \lim_{n \rightarrow \infty} \binom{n}{y} \left(\frac{\lambda}{n}\right)^y \left(1 - \frac{\lambda}{n}\right)^{(n-y)} = \frac{\lambda^y e^{-\lambda}}{y!} \quad (2)$$

۴-۳- ایجاد مدل

مدل رگرسیون پواسون مدل پایه ای خانواده مدل‌های خطی تعمیم یافته است که برای مدل کردن داده شمارشی مانند داده‌های فراوانی تصادف استفاده می‌شود. مدل

بیشینه برای انتخاب بهترین مدل انتخاب شدند. برای یک نوع توزیع مفروض، مدل و داده، تخمین‌های MLE مجموعه ای از مقادیر پارامترهای مدل است که تابع درستنمایی را بیشینه می‌کنند. درستنمایی، تابعی از پارامترهای یک مدل آماری است. آماره های مختلف استفاده شده بصورت زیر است:

- انحراف
- کای مجذور پیرسون
- لگاریتم درستنمایی

انحراف، یک آزمون آماری است که برای چک کردن تناسب مدل بوسیله روش درستنمایی بیشینه استفاده می‌شود و مشابه با استفاده از مجموع باقیمانده‌های مربع در حداقل مربعات معمولی است (McCullagh, 1989). برای مدل رگرسیون پواسون، انحراف (Dp) بصورت معادله (۸) و برای مدل رگرسیون دوجمله ای منفی بصورت معادله (۹) محاسبه می‌شود:

$$D_p = 2 \sum_{i=1}^n \left(y_i \ln \left(\frac{y_i}{\hat{\lambda}_i} \right) - (y_i - \hat{\lambda}_i) \right) \quad (8)$$

(۹)

$$D_p = 2 \sum_{i=1}^n \left(y_i \ln \left(\frac{y_i}{\hat{\lambda}_i} \right) - \left(y_i - 1/k \right) \ln \left(\frac{y_i + 1/k}{\hat{\lambda}_i + 1/k} \right) \right)$$

فاکتور انحراف و درجه آزادی، پراکندگی در یک مدل را نشان می‌دهد. پراکندگی در داده‌ها می‌تواند منجر به بیشتر شدن فاصله اطمینان شود، که می‌تواند بر معنادار بودن یک متغیر بی اهمیت تاثیر داشته باشد. اگر این مقدار نزدیک به ۱ باشد، هیچ پراکندگی در داده وجود ندارد. یک مقدار بیش از ۱ بیش پراکندگی را نشان می‌دهد و یک مقدار کمتر از ۱ پراکندگی کمتر از حد را نشان می‌دهد. آزمون آماری کای مجذور پیرسون نیز می‌تواند برای ارزیابی تفاوت بین یک توزیع فراوانی مشاهده شده و یک توزیع نظری استفاده شود. این آزمون آماری بصورت معادله (۱۰) محاسبه می‌شود:

$$\chi^2 = \sum_{i=0}^n \frac{(y_i - \hat{\lambda}_i)^2}{\hat{\lambda}_i} \quad (10)$$

یکی از آزمون‌های دیگر مورد استفاده لگاریتم درستنمایی است. لگاریتم درستنمایی، لگاریتم طبیعی تابع درستنمایی است. درستنمایی تابعی از پارامترهای یک مدل آماری

رگرسیون پواسون-گاما (دوجمله ای منفی) اگر داده‌ها بیش پراکندگی نشان دهند استفاده می‌شود (Miau, 1994). مطابق توزیع پواسون، احتمال $P(y_i)$ یک تقاطع i دارای y_i تصادف با معادله (۳) ارایه می‌شود:

$$P(y_i) = e^{-\lambda_i} \lambda_i^{y_i} / y_i! \quad (3)$$

$$\lambda_i = E[y_i] = \mu_i \quad (4)$$

که λ_i پارامتر پواسون، $E[y_i]$ فراوانی تصادفات مورد انتظار در تقاطع i ، μ_i میانگین تعداد تصادفات در تقاطع i هستند. متداولترین رابطه بین پارامتر پواسون و متغیرهای مستقل، مدل خطی لگاریتمی زیر است:

$$\lambda_i = EXP(\beta X_i) \quad (5)$$

$$\lambda_i = E[y_i] = \mu_i = VAR[y_i] \quad (6)$$

که EXP نمایی، β بردار پارامترهای قابل تخمین، X_i بردار متغیرهای مستقل قابل تخمین است. مشکل مدل رگرسیون پواسون ساده این است که فرض می‌کند میانگین داده‌های تصادف برابر با واریانس آنها است. مدل NB یا پواسون-گاما برای محاسبه یافتن بیش پراکندگی در داده‌های تصادف طراحی شده است. مدل NB می‌تواند به بیش پراکندگی رسیدگی کند چون یک عبارت خطای وصل شده به پارامتر پواسون دارد که مدل را از این فرض که میانگین داده‌های تصادف باید برابر با واریانس باشد، بصورت زیر آزاد می‌کند:

$$\lambda_i = EXP(\beta X_i + \varepsilon_i) \quad (7)$$

عبارت $EXP(\varepsilon_i)$ توزیع گاما با میانگین ۱ و واریانس α است. واریانس α پارامتر بیش پراکنده نیز نامیده می‌شود. مدل NB وقتی مقدار α برای داده‌های غیرپراکنده نزدیک به صفر می‌شود، به یک مدل پواسون ساده کاهش می‌یابد.

۴-۴- خویی برازش

نرم افزار آماری استفاده شده در این تحقیق، نرم افزار SPSS است. در این تحقیق آماره های مبتنی بر درستنمایی

شرکت های بیمه اخذ و با هم ادغام شد. دلیل این کار این است که برخی از تصادفات خسارتی خفیف با توافق رانندگان و حضور نمایندگان بیمه در صحنه تصادف حل و فصل می شود و گزارشی به پلیس داده نمی شود.

۴-۷- معیار جمع آوری متغیرهای مستقل

به منظور یافتن رابطه بین فراوانی تصادفات و متغیرهای مستقل مربوط به تقاطع، فهرستی از متغیرهای مستقل قابل اندازه گیری و قابل سنجش آماده شد. از آنجایی که تنها تقاطع های T شکل برای تجزیه و تحلیل در نظر گرفته شدند، بنابراین، داده های هندسه تقاطع و داده های مربوطه تنها برای تقاطع های T شکل جمع آوری شد. معیار این تحقیق برای جمع آوری متغیرها بر اساس موارد زیر بود:

- متغیرهای استفاده شده مطالعات پیشین و یافتن معنی داری آنها
 - آسان بودن درک و تفسیر متغیر
 - عملی بودن جمع آوری داده های متغیر
- با توجه به نکات فوق، اطلاعات تعدادی از متغیرها جمع آوری شد. متغیرهای مستقل در نظر گرفته شده برای تجزیه و تحلیل در ادامه ارائه می شود.

۴-۸- متغیرهای مستقل تقاطع

اگر چه اطلاعات در مورد بسیاری از متغیرها جمع آوری شد، اما تنها متغیرهایی ارائه شده در جدول (۱) با توجه به معیارهای انتخاب و جمع آوری داده ها مهم تلقی می شوند. تمام متغیرهای مستقل بصورت متغیرهای دسته بندی شده تحلیل شدند، بجز برای متوسط عرض خط (که پیوسته است) و حجم ترافیک، که لگاریتم طبیعی آنها بصورت همسنگ در نظر گرفته شده است.

است (McCullagh, 1989). یک مدل با مقدار لگاریتم درستنمایی بالاتر (که به عنوان درستنمایی بیشینه نیز شناخته می شود)، داده ها را بهتر برازش می کند. برای رگرسیون بواسون، تابع لگاریتم درستنمایی بردار پارامتر β بصورت زیر محاسبه می شود (Washington, Karlaftis and Mannering, 2010):

(۱۱)

$$LL(\beta) = \sum_{i=1}^n [-e^{\beta X_i} + y_i \beta X_i - \ln(y_i!)]$$

که $LL(\beta)$ لگاریتم درستنمایی بردار پارامترها و β بردار پارامترها است. برای رگرسیون دوجمله ای منفی، تابع درستنمایی بصورت معادله (۱۲) محاسبه می شود (Washington, Karlaftis and Mannering, 2010):

(۱۲)

$$L(\lambda_i) = \prod_i \frac{\Gamma((1/\alpha) + y_i)}{\Gamma(1/\alpha)^{y_i}} \alpha \left(\frac{1/\alpha}{(1/\alpha) + \lambda_i} \right)^{1/\alpha} \left(\frac{\lambda_i}{(1/\alpha) + \lambda_i} \right)^{y_i}$$

که $\Gamma(\cdot)$ یک تابع گاما است.

۴-۵- انتخاب مدل

معیار اطلاعات آکایک^۳ (AIC) یک معیار انتخاب مدل مفید است. AIC مدل های مختلف مقایسه می شود و مدل با کمترین AIC بعنوان مدل با کمترین مقدار اطلاعات از دست رفته در نظر گرفته می شود (Liu and Cela, 2008). AIC بصورت زیر محاسبه می شود:

$$AIC = (-2 LL(\beta) + 2k) \quad (13)$$

که $LL(\beta)$ لگاریتم درستنمایی بردار پارامترها است، و k تعداد پارامترها در مدل است.

۴-۶- داده ها

تعداد کلی ۴۱ تقاطع سه شاخه شهر بروجرد انتخاب شد. تصادفات برای ۵ سال از ۱۳۹۲ تا ۱۳۹۶ برای تجزیه و تحلیل در نظر گرفته شد. اطلاعات تصادف از مراجع پلیس و

جدول ۱. متغیرهای مستقل تقاطع

نماد	شرح
ATL	تعداد کلی خطوط عبوری در تقاطع
ART	تعداد کلی خطوط گردش به راست در تقاطع
ASL	تعداد کلی خطوط گردش به راست با جزیره ترافیکی در تقاطع
AM	تعداد کلی میانه ها در تقاطع
SP	تعداد کلی فازهای چراغ راهنمایی در تقاطع
ANG	زاویه ایجاد شده بوسیله جاده فرعی بر جاده اصلی
AAVLW	متوسط میانگین عرض خطوط در هر دسترسی
AV	حجم کلی ترافیک ورودی به تقاطع

اطلاعات متغیرهای دسته بندی شده در جدول (۲) ارایه شده است. آماره های توصیفی متغیرهای استفاده شده برای مدل سازی نیز، در جدول (۳) ارایه شده است. در جدول (۳)، شده است. آماره های توصیفی متغیرهای استفاده شده برای تصادف به کل تصادفات دوره ۵ ساله اشاره دارد.

جدول ۲. متغیرهای دسته بندی شده تقاطع

متغیر مستقل	دسته ها	فراوانی	درصد
تعداد کلی خطوط عبوری	خط عبوری ۶	۷	٪۱۷/۱
	خط عبوری ۵	۷	٪۱۷/۱
	خط عبوری ۴	۲۰	٪۴۸/۷
	خط عبوری ۳	۴	٪۹/۸
	خط عبوری ۲	۳	٪۷/۳
تعداد کلی خطوط گردش به راست	خط گردش به راست ۴	۷	٪۱۷/۱
	خط گردش به راست ۳	۱۰	٪۲۴/۴
	خط گردش به راست ۲	۲۴	٪۵۸/۵
تعداد کلی خطوط گردش به راست با جزیره ترافیکی	خط ۲	۱۳	٪۳۱/۷
	خط ۱	۱۸	٪۴۳/۹
	بدون خط	۱۰	٪۲۴/۴
تعداد کلی میانه ها	۳ میانه	۳۲	٪۷۸
	۲ میانه	۶	٪۱۴/۷
	بدون میانه	۳	٪۷/۳
تعداد فازهای هر سیکل	۴ فاز در هر سیکل	۹	٪۲۲
	۳ فاز در هر سیکل	۳۰	٪۷۳/۱
	۲ فاز در هر سیکل	۲	٪۴/۹
زاویه تقاطع (بین راه فرعی و راه اصلی)	۶۰ درجه	۸	٪۱۹/۵
	۳۰ درجه	۲۰	٪۴۸/۸
	۹۰ درجه	۱۳	٪۳۱/۷

جدول ۳. توصیف آماری متغیرهای تقاطع

متغیر	حداقل	حداکثر	میانگین	انحراف معیار
ATL	۲	۶	۴/۲۷	۱/۰۹۶
ART	۲	۴	۲/۵۹	۰/۷۷۴
ASL	۰	۲	۱/۰۷	۰/۷۷۵
AM	۰	۳	۲/۶۳	۰/۸۲۹
SP	۲	۴	۳/۱۷	۰/۴۹۵
ANG	۰	۶۰	۲۶/۳۴	۲۱/۴۱۹
AANW	۸	۱۴	۱۰/۳۹	۱/۵۶۳
AV	۱۲۳۰۲	۶۷۷۷۹	۳۲۷۳۶	۱۳۲۹۳
تصادف	۳	۴۸	۱۶/۴۴	۹/۹۲

۴-۹- تخمین و انتخاب مدل

با کل تصادفات بعنوان متغیر وابسته، و هشت متغیر، نقاط داده ۴۱ تقاطع با استفاده از مدل خطی تعمیم یافته از SPSS تحلیل شد. مدلها شامل تمام هشت متغیر ذکر شده هستند. حجم ترافیک جریان ورودی (AV) بصورت همسنگ در مدل استفاده شد. متغیرهای بی اهمیت یک به یک حذف

جدول ۴. انتخاب مدل و آماره های خوبی برازش

مدل	انحراف	انحراف/df	لگاریتم درستنمایی	AIC
پواسون	۹۷/۳	۳/۰۴	- ۱۴۰	۲۹۹
دوجمله ای منفی	۵۲/۹	۱/۶۵	- ۱۳۰	۲۷۹

(دوجمله ای منفی) برای مدلسازی تصادفات در تقاطع را نشان می‌دهد.

جدول ۵. متغیرهای معنادار با سطح معناداری

معناداری	متغیرهای مستقل در مدل
۰/۰۰۰	تعداد کلی خطوط عبوری
۰/۰۱۰	خطوط گردش به راست با جزیره ترافیکی
۰/۰۵۰	زاویه تقاطع

هر دو مدل نهایی پواسون و دوجمله ای منفی، بصورت مشابه متغیرهای مستقل را معنادار نشان می‌دهد. طبق جدول (۵) میانگین تخمین زده شده برای تعداد خطوط عبوری، تعداد خطوط گردش به راست با جزیره ترافیکی و زاویه تقاطع به مقدار ناچیزی متفاوت است.

انحراف یکی از معیارهای خوبی برازش است. انحراف مدل دوجمله ای منفی کمتر از مدل پواسون است. انحراف، باقیمانده در برازش داده ها برای خط رگرسیون را آرایه می کند. انحراف کمتر مدل دوجمله ای منفی نشان می دهد که داده های مدل دوجمله ای منفی بهتر از مدل پواسون برازش شده اند. با این حال، مقدار انحراف/درجه آزادی (df) برای مدل پواسون برابر ۳/۰۴ بود که بیشتر از مقدار استاندارد ۱ بوده و نشان می‌دهد که مدل پواسون بیش پراکنده بوده است. لگاریتم درستنمایی مدل دوجمله ای منفی بیشتر از مدل پواسون است و به این معنا است که داده ها در مدل دوجمله ای منفی دارای برازش بهتری هستند. مقدار AIC برای مدل دوجمله ای منفی کمتر از مدل پواسون است. وقتی دو مدل مقایسه می شوند، مدلی که مقدار AIC کمتری دارد ارجح تر است و بنابراین مدل دوجمله ای منفی داده‌ها را بهتر برازش می‌کند. جدول (۵) متغیرهای معنادار برای بهترین مدل

۴-۱۰- متغیرهای معنادار در تقاطع

مدلسازی فراوانی تصادف در تقاطع با استفاده از مدل دوجمله‌ای منفی و مدل پواسون نشان داد که تنها ۳ متغیر مستقل بصورت معنادار مربوط به تصادفات در تقاطع هستند. این متغیرهای مستقل عبارتند از:

- تعداد خطوط عبوری در یک تقاطع
- تعداد خطوط گردش به راست با جزیره ترافیکی در یک تقاطع
- زاویه تقاطع

۴-۱۱- معبر سازی مدل

مدلسازی داده‌های تصادف در تقاطع با رگرسیون دوجمله‌ای منفی، سه متغیر معنادار را نشان داد: تعداد خطوط عبوری، تعداد خطوط گردش به راست با جزیره ترافیکی و زاویه اریب. به منظور بدست آوردن یک بینش در دقت پیشبینی مدل، تعدادی از آماره‌های خوب تشخیص داده شده که شامل R^2 ، خطای جذر میانگین مربع، میانگین خطای مطلق و میانگین قدر مطلق درصد خطا هستند، استفاده می‌شوند این آماره‌ها بصورت گسترده برای مقایسه برآزش مقادیر پیشبینی شده از مدل‌های آماری با توجه به مقادیر مشاهده شده، استفاده می‌شوند (Qu et al., 2014).

R^2 مجذور همبستگی بین مقادیر مشاهده شده و مقادیر پیشبینی شده است و بصورت معادله (۱۴) محاسبه می‌شود.

$$R^2 = \frac{\sum(O-O_m)(P-P_m)}{\sqrt{\sum(O-O_m)^2 \sum(P-P_m)^2}} \quad (14)$$

که O تصادف مشاهده شده، P تصادف پیشبینی شده، O_m میانگین تصادفات مشاهده شده و P_m میانگین تصادفات پیشبینی شده است. مقدار R^2 به دست آمده برای تصادفات پیشبینی شده در برابر تصادفات مشاهده شده در تقاطع ۰/۵۶ است. اگر داده‌های تصادف پیش‌بینی شده دقیقاً برابر با

داده‌های تصادف مشاهده شده باشد، باید مقدار R^2 ۱ به دست آید. مقدار ۰/۵۶ پیشنهاد می‌کند که داده‌های تصادف پیشبینی شده خط رگرسیون را بخوبی برازش نمی‌کنند اما مدل هنوز هم یک تاثیر پیشبینی خوب برای شمارش تصادف دارد. خطای جذر میانگین مربع^۴ (RMSE) بصورت معادله (۱۵) محاسبه می‌شود، هر چه مقدار RMSE به صفر نزدیکتر باشد، نتایج انطباق خوبی با داده‌ها داشته‌اند (Akbari, Shafabakhsh and Ahadi, 2015):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (O_i - E_i)^2} \quad (15)$$

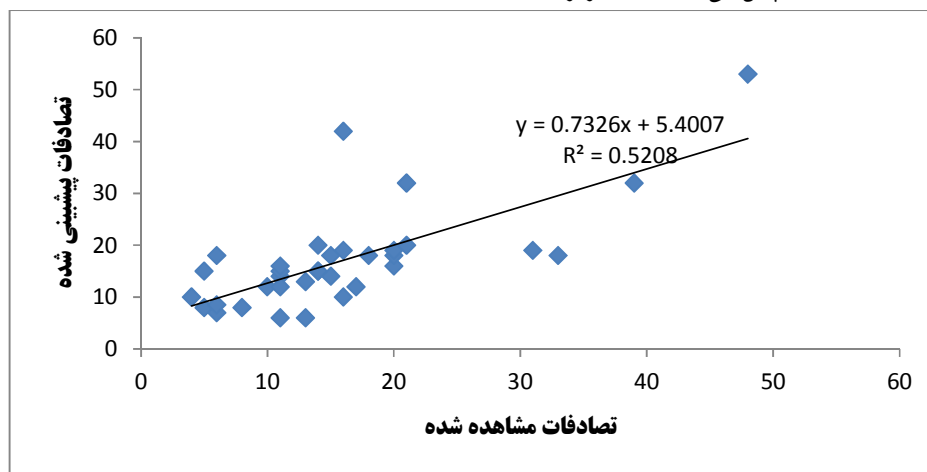
میانگین خطای مطلق^۵ (MAE) بصورت معادله (۱۶) محاسبه می‌شود، هر چه میزان MAE به صفر نزدیکتر باشد، نشان می‌دهد که مدل پیشنهادی، پیشبینی‌های بهتری روی داده‌های مشاهده شده داشته است (Akbari, Shafabakhsh and Ahadi, 2015):

$$MAE = \frac{1}{n} \sum_{i=1}^n |O_i - E_i| \quad (16)$$

میانگین قدر مطلق درصد خطا^۶ (MAPE) بصورت معادله (۱۷) محاسبه می‌شود، این معیار نیز همانند معیار MAE عمل می‌کند:

$$MAPE = \frac{1}{n} \sum_{i=1}^n |O_i - P_i| / O_i \quad (17)$$

نمودار R مجذور تصادفات پیشبینی شده در برابر تصادفات مشاهده شده (شکل ۶) یک مقدار ۰/۵۲ را نشان می‌دهد. اگر داده‌های تصادف پیشبینی شده دقیقاً برابر با داده‌های تصادف مشاهده شده بود، مقدار R مجذور ۱ باید بدست می‌آمد. مقدار ۰/۵۲ پیشنهاد می‌دهد که داده‌های تصادف پیشبینی شده به خوبی خط رگرسیون را برازش نمی‌کند اما هنوز یک تاثیر پیشبینی خوب برای تعداد تصادف دارد.



شکل ۶. نمودار R مجذور تقاطع

- 2-Poisson-Gamma
- 3-Akaike Information Criterion
- 4-Root Mean Square Error
- 5-Mean Absolute Error
- 6-Mean Absolute Percentage Error

RMSE برای داده های پیشبینی شده و مشاهده شده محاسبه شد و مقدار $6/69$ بدست آمد. اگر داده های پیشبینی شده نزدیک به داده های مشاهده شده باشد، RMSE باید نزدیک به صفر باشد. مقدار RMSE $6/69$ مشخص می کند که داده های پیشبینی شده از خط رگرسیون خیلی دور است و به این معنی است که یک تفاوت قابل توجه بین تصادفات مشاهده شده و پیشبینی شده است. بصورت مشابه، MAE و MAPE محاسبه شده و بترتیب مقادیر $5/3171$ و $0/4123$ بدست آمد.

۷-مراجع

-اکبری، م. شفاعخش، غ.ع. احدی، محم.ر.، (۱۳۹۴)، "ارزیابی اثرات شاخص وضعیت رویه راه بر فراوانی تصادفات خروج از جاده"، مهندسی زیرساخت های حمل و نقل، سال اول، شماره سوم، پاییز.

-قبادی، م. حسن زاده، م.ر. و زراعت پیمان، ف.، (۱۳۹۵)، "پیشبینی تعداد تصادفات در تقاطع های هم سطح شهری"، فصلنامه مطالعات مدیریت ترافیک، شماره ۴۰، بهار.

-Claros, B. Sun, C. Edara, P., (2018), "Missouri-Specific Crash Prediction Model for Signalized Intersections, Journal of the transportation research board", Transportation Research Record.

-Dixon, K., C. Monsere, R. Avelar, J. S. Barnett, and P. Escobar, (2015), "Improved Safety Performance Functions for Signalized Intersections. Report No. FHWA-OR-RD-16-03. Oregon Department of Transportation.

-Eenink, R., Reuring, M., Elvik, R., Cardoso, J., Wichert, S., & Stefan, C., (2007), "Accident Prediction Models and Road Safety Impact Assessment: Recommendations for using these tools", Result of the pilot studies. RI-SWOV-WP2-R4-V2-Results. RIPCOD-ISEREST Deliverable D2. BAST. Retrieved from <http://ripcord.bast.de/pdf/RIPCORD-ISEREST-Deliverable-D2-Final.pdf>.

-Griebe, P., (2003), "Accident prediction models for urban roads", Accident Analysis and Prevention, Vol. 35, pp. 273-285.

-Harnen, S., Umar, R., R., Wong, S. & Hashim, W., (2004), Development of prediction models for motorcycle crashes at signalized intersections on urban roads in

۵- نتیجه گیری

تجزیه و تحلیل توصیفی داده های تصادف تقاطع نشان داد که تصادفات جلو به عقب، گروه غالب تصادفات است. هر چند تصادفات فوتی کم است، اما تعداد تصادفات جرحی بیشترین نوع تصادف است. برای مدلسازی تصادفات در تقاطع های سه شاخه شهر بروجرد، مدل های پواسون و دو جمله ای منفی در نظر گرفته شد. بر اساس معیارهای خوبی برازش مختلف، مدل دو جمله ای منفی به عنوان بهترین مدل برای مدلسازی تصادفات در تقاطع انتخاب شد. متغیرهای معنادار مشخص و در مدل وارد شد. نتایج نشان داد که مدلسازی آماری فراوانی تصادفات در تقاطع در عبارت ارزیابی قابل قبول است. اما پیشبینی در تقاطع همانگونه که بوسیله مقدار R^2 $0/52$ ، MAE $5/3171$ و $6/69$ RMSE مستند شد، خیلی دقیق نیست. نتایج آزمون دقت مدل نشان می دهد که هر چند مدل های رگرسیون دو جمله ای منفی در پیشبینی تصادفات تقاطع های سه شاخه شهری نتایج قابل قبولی ارائه می دهد، اما نتایج پیشبینی از دقت بالایی برخوردار نخواهد بود.

۶- پی نوشت ها

1-Negative Binomial (NB) Regression

- models for road intersections, Diepenbeek, 2008”, Steunpunt Mobiliteit & Openbare Werken Spoor Verkeersveiligheid.
- Obaidat, M., & Ramadan, T., (2012), “Traffic Accidents at Hazardous Locations of Urban Roads”, *Jordan Journal of Civil Engineering*, 6(4).
- QU, X., Kuang, Y., OH, E. & JIN, S., (2014), “Safety evaluation for expressways: A comparative study for macroscopic and microscopic indicators”, *Traffic Injury Prevention*, 15, pp.89-93.
- Shankar, V., and S. Madanat, (2015), “Methods for Identifying High Collision Concentrations for Identifying Potential Safety Improvements: Development of Safety Performance Functions for California”, Report No. CA 15-2317. California Department of Transportation,
- Washington, P. Karlaftis, M. G. & Mannering, F. L., (2010), “Statistical and econometric methods for transportation data analysis”, CRC press.
- Winkelmann, R., (2003), “Econometric Analysis of Count Data, 4th ed”, Springer Verlag, Heidelberg.
- Malaysia. *Journal of transportation and ststistics*, 7, pp.27.
- Karuppanen, Gunasekaran, (2015), “Development of Accident Prediction Models For Safety Evaluation of Urban Intersections”, *Indian Highways*, May.
- Liu, W. and Cela, J., (2008), “Count Data Models in SAS. SAS Global Forum 2008: Statistics and data Analysis”, paper pp.371.
- Lord, D., & Mannering, F., (2010), “The Statistical Analysis of Crash-Frequency Data: A Review and Assessment of Methodological Alternatives”, *Transportation Research Part A: Policy and Practice*, 44(5), pp.291-305.
- McCullagh, P. & J. A., N., (1989), “Generalized Linear Models, 2nd ed.”, Chapman and Hall, NewYork.
- Miaou, S.P., (1994), “The Relationship between truck accidents and geometric design of road sections: Poisson versus Negative Binomial Regressions”, *Accident Analysis & Prevention*, 26-4, pp.471-482.
- Nambuusi, B.B., Brijs, T., Hermans, E., (2008), “A review of accident prediction

Modeling Crashes Frequency at Urban Three Leg Intersections

Mohammad Koohi, M.Sc., Grad. Civil Engineering Faculty, P.N.U. University, Tehran, Iran.

Shahin Shaabani, Associate Professor, Civil Engineering Faculty, P.N.U. University, Tehran, Iran.

E-mail: m.koohhi@gmail.com

Received: February 2020 -Accepted: June 2020

ABSTRACT

One of the intersection safety analysis methods is creating crash prediction models using the negative binomial and Poisson regression models. However, the use of negative binomial regression is preferable due to its more data dispersion. But, the research evaluates both models at first for modeling urban intersections crashes. After testing the models by criteria of Akaike information, deviation and log-likelihood, the negative binomial regression model showed a better fit and was chosen as the better model. For modeling, the data of a 5-year period of crashes occurred at the three leg intersections of Boroujerd city has been used. The input variables of the model were selected after a significance evaluation test. These variables included the number of passing lanes, the number of right turn with traffic island lanes and the skew angle. After selecting the variables, the accuracy of the model was also studied. To validate the model, the criteria of R², root mean square error, mean absolute error, and mean absolute percentage error were used. After evaluating the model accuracy in the prediction of accidents, it was found that the model provides acceptable results for evaluating crashes, but the results are not very accurate for predicting them (R² = 0.52). The R² value for the model shows that the predictions will not be precise enough, but reasonable predictions can still be provided.

Keywords: Crashes Modeling, Urban Intersections, Negative Binomial Regression