

کاربرد مدل‌های رگرسیون پرفر در مدل‌سازی تصادفات بزرگراه‌های درون‌شهری

اسماعیل آیتی^۱، احسان عباسی^{۲*}

۱- استاد دانشکده مهندسی دانشگاه فردوسی مشهد

۲- کارشناس ارشد راه و ترابری دانشگاه فردوسی مشهد

ehsan.abbasi395@gmail.com

چکیده - در این نوشتار به پژوهشی برای شناسایی عوامل مؤثر بر فراوانی و شدت تصادفات بزرگراه‌های درون‌شهری انجام شده پرداخته می‌شود و به عنوان مطالعه موردی از داده‌های تصادفات بزرگراه‌های شهری مشهد استفاده می‌شود. مدل‌های آماری که در این پژوهش استفاده شده، مدل‌های رگرسیون پواسن، دوجمله‌ای منفی، پواسن پرفر و دوجمله‌ای منفی پرفر است، از متغیرهای مربوط به جریان ترافیک و طرح هندسی راه به عنوان متغیرهای ناوابسته مدل‌ها استفاده شده است. کار ویژه‌ای که در این پژوهش انجام شده، جداسازی حجم کل ترافیک به حجم خودروهای سواری، خودروهای سنگین و خودروهای غیرسواری سبک است تا نقش حجم ترافیک در رخداد تصادفات به‌گونه‌ای ویژه و موشکافانه بررسی و روشن شود دقیقاً کدام بخش از ترافیک نقش مؤثر یا مؤثرتری در رخداد تصادفات دارد.

در این پژوهش کارایی مدل‌های رگرسیون پرفر در برابر مدل‌های رگرسیون ساده پواسن و دوجمله‌ای منفی در مدل‌سازی تصادفات بزرگراه‌های درون‌شهری بررسی می‌شود. نتایج پژوهش نشان می‌دهد شمار راه‌های دسترسی و قوس‌های افقی در بخش‌های بزرگراه و حجم خودروهای غیرسواری سبک، نقش افزایش‌دهنده‌ای در احتمال رخداد تصادفات دارد، همچنین مدل رگرسیون دوجمله‌ای منفی پرفر بهترین و برازنده‌ترین مدل، هم برای مدل‌سازی تصادفات با زیان مالی و هم برای تصادفات با زیان جانی است و بر این اساس می‌توان کفایت مدل‌های رگرسیون پرفر را در مدل‌سازی تصادفات بزرگراه‌های درون‌شهری مشهد تأیید کرد.

کلیدواژگان: تصادفات با زیان مالی، تصادفات با زیان جانی، بزرگراه‌های درون‌شهری، مدل‌سازی تصادفات، مدل‌های

رگرسیون پرفر

۱- مقدمه

مهندسی افزایش دهد [۱]. متغیرهای ناوابسته زیادی در رخداد تصادفات اثرگذار است که این عوامل معمولاً با حجم ترافیک، ویژگی‌های هندسی راه، وضعیت رویه راه، روشنایی، آب و هوا و رفتارهای رانندگان مرتبط است. پژوهشگران عموماً از مجموعه گسترده‌تری از متغیرهای

با وجود پیشرفت‌های مهم در زمینه افزایش ایمنی در بزرگراه‌ها، تصادف‌های چشم‌گیر با شدت‌های زیاد همچنان در بزرگراه‌ها رخ می‌دهند که مدل‌سازی می‌تواند توانایی شناسایی میزان تأثیر عوامل مؤثر در تصادف‌ها را برای

سن راننده [۵ و ۷] یا آب و هوای منطقه مانند بارش برف و باران [۳، ۵ و ۹] باشد.

در پژوهش‌های پیشین بیشتر از متغیر حجم متوسط ترافیک روزانه (ADT) یا حجم متوسط ترافیک روزانه در سال (AADT) به عنوان متغیر ناوابسته مربوط به جریان ترافیک در مدل‌سازی استفاده شده است [۶ و ۸] و یا برای در نظر گرفتن شمار خطوط در هر طرف، متغیر حجم متوسط ترافیک روزانه روی هر خط به کار رفته است [۵، ۹ و ۱۰]. پژوهشگران برای در نظر گرفتن نقش خودروهای سنگین در فراوانی و شدت تصادفات بزرگراه‌ها از متغیرهایی مانند درصد متوسط روزانه کامیون در ترکیب ترافیک یا حجم متوسط روزانه کامیون [۵، ۷ و ۹] و برای تفکیک نقش خودروهای سواری و خودروهای سنگین در رخداد و شدت تصادفات از متغیرهای حجم متوسط روزانه خودروهای سواری و حجم متوسط روزانه کامیون یا درصد کامیون در ترکیب ترافیک، به‌طور جداگانه در مدل‌سازی استفاده کرده‌اند [۷]. در این پژوهش برای بررسی دقیق‌تر نقش حجم ترافیک در رخداد تصادفات بزرگراه‌ها، حجم کل ترافیک به حجم خودروهای سواری، خودروهای سنگین و غیرسواری سبک تفکیک شده است تا روشن شود دقیقاً کدام بخش از ترافیک نقش مؤثر یا مؤثرتری در رخداد تصادفات دارد.

در این پژوهش و برخی پژوهش‌ها از مقدار سرعت جریان ترافیک در هر بخش از بزرگراه برای مدل‌سازی تصادفات استفاده شده است [۶]؛ در حالی که در بسیاری دیگر، مقدار سرعت مجاز روی تابلوهای نصب‌شده در بخش‌های مختلف بزرگراه‌ها به عنوان متغیر پیوسته یا نشانگر (ساختگی)، معیار قرار گرفته است [۴، ۵ و ۱۰]. در برخی پژوهش‌ها برای در نظر گرفتن متغیر شمار خطوط، عامل ADT یا AADT روی هر خط به عنوان متغیر جریان

ناوابسته، شمار کاهش یافته‌ای از متغیرها را برای ورود به مدل استخراج می‌کنند [۲]. در این نوشتار عوامل مؤثر بر فراوانی و شدت تصادفات بزرگراه‌های درون‌شهری پرداخته می‌شود و به عنوان مطالعه موردی از داده‌های تصادفات بزرگراه‌های درون‌شهری مشهد استفاده می‌شود. در این پژوهش از متغیرهای مربوط به جریان ترافیک، شامل حجم و سرعت و متغیرهای مربوط به طرح هندسی راه، شامل شمار خطوط، شمار قوس‌های افقی و شمار راه‌های دسترسی، به عنوان متغیرهای ناوابسته مدل‌ها استفاده شده است. مدل‌های آماری که در این پژوهش برای شناسایی عوامل مؤثر بر فراوانی و شدت تصادفات بزرگراه‌های درون‌شهری استفاده شده، مدل‌های رگرسیونی چهارگانه مطرح در مدل‌سازی تصادفات است؛ مانند مدل‌های رگرسیونی پواسن، دوجمله‌ای منفی، پواسن پرفر و دوجمله‌ای منفی پرفر. در این پژوهش کارایی مدل‌های رگرسیون پرفر در برابر مدل‌های رگرسیون ساده پواسن و دوجمله‌ای منفی در مدل‌سازی تصادفات بزرگراه‌های درون‌شهری بررسی می‌شود.

۲- مروری بر تحقیقات گذشته

پژوهش‌های فراوانی در سال‌های گذشته برای مدل‌سازی تصادفات به‌ویژه در آزادراه‌ها و بزرگراه‌های جاده‌ای و درون‌شهری انجام شده است تا عوامل مؤثر بر رخداد تصادفات و شدت آن‌ها شناسایی و برای بهبود شرایط ایمنی، مهار یا اصلاح شوند [۳، ۴، ۵، ۶، ۷، ۸ و ۹]. عوامل مؤثر در رخداد تصادفات می‌تواند مربوط به جریان ترافیک مانند حجم ترافیک و سرعت، طرح هندسی راه مانند پهنای خطوط، طرح قوس‌ها و دسترسی‌ها [۴، ۵، ۶، ۷، ۹ و ۱۰]، کیفیت و شرایط روسازی مانند ضریب اصطکاک و شیارشدگی [۳، ۵ و ۷]، ویژگی‌های رانندگان مانند جنس و

با جریان ترافیک بزرگراه تداخل می‌کند. کار ویژه‌ای که در این پژوهش انجام شده است، جداسازی حجم کل (معادل سواری) ترافیک به حجم خودروهای سواری، حجم خودروهای سنگین (معادل سواری) شامل تریلی، کامیون، اتوبوس و مینی‌بوس و حجم خودروهای غیرسواری سبک (معادل سواری) شامل تاکسی، وانت و موتورسیکلت است. با این کار، پژوهشگر می‌خواهد نگاهی ویژه و موشکافانه به نقش حجم ترافیک در رخداد تصادفات با زیان مالی و جانی دارد و درمی‌یابد دقیقاً کدام بخش از ترافیک نقش مؤثر یا مؤثرتری در رخداد تصادفات دارد.

برای گردآوری آمار تصادفات، بزرگراه‌ها به بخش‌هایی با حجم ترافیک ثابت تقسیم می‌شود. در هر بخش شمار تصادفات برای ساعت‌های مختلف شبانه‌روز گردآوری می‌شود. در واقع داده‌های تصادفات، داده‌هایی با دو سطح در نظر گرفته می‌شوند، سطح اول و اصلی آن‌ها بخش‌های راه است؛ یعنی یک بزرگراه به چندین قطعه یا بخش تقسیم می‌شود. مبنای قطعه‌بندی راه‌ها متفاوت است، قطعه‌بندی می‌تواند بر مبنای طول باشد؛ یعنی بزرگراه به قطعه‌های مساوی تقسیم می‌شود. اشکال این قطعه‌بندی آن است که نمی‌توان مقدار ثابتی برای حجم ترافیک در هر بخش [که متغیر مهم و اثرگذاری در این مطالعه است] داشت؛ از این رو ترجیح داده شده است که این قطعه‌بندی بر مبنای حجم کل ترافیک انجام شود؛ یعنی حجم ترافیک در هر بخش، ثابت است. سطح دوم، ساعت‌های شبانه‌روز است؛ به طوری که ساعت‌های اوج ترافیک به عنوان زیرسطح اول، ساعت‌های غیراوج روز، زیرسطح دوم و ساعت‌های غیر اوج شب به عنوان زیرسطح سوم در نظر گرفته شده است. برای انجام محاسبات آماری مربوط به مدل‌ها از نرم افزار SAS 9.1 استفاده شده است. پس از تحلیل‌های آماری، معلوم می‌شود در هر یک از مدل‌ها کدام متغیرها در رخداد

ترافیک، به جای ADT یا AADT برای مدل‌سازی به کار رفته است [۹ و ۱۰]؛ اما چنین متغیری در مطالعات بسیاری از محققان دیده می‌شود [۶، ۷ و ۸]. استفاده از متغیرهای شمار قوس‌های افقی و قائم در فاصله‌های معینی از بزرگراه‌ها، به عنوان متغیرهای مربوط به طرح هندسی راه در بسیاری از پژوهش‌ها دیده می‌شود [۵، ۷ و ۱۰]؛ هر چند در برخی تحقیقات از درجه انحنای متوسط قوس‌های افقی [۷ و ۹]، شیب یا طول قوس‌های قائم [۹ و ۱۰] در هر بخش از راه برای مدل‌سازی استفاده شده است. برخی پژوهشگران از متغیرهای شمار تقاطع‌های غیرهم‌سطح [۵]، تقاطع‌های هم‌سطح [۱۰] یا رمپ‌ها [۷] در فاصله‌های معین آزادراه‌ها یا بزرگراه‌ها، برای مدل‌سازی تصادفات استفاده کرده‌اند. در پژوهش پیش رو برای بررسی نقش ویژگی‌های هندسی بزرگراه در رخداد تصادفات از متغیرهای شمار خطوط، قوس‌های افقی و راه‌های دسترسی برای مدل‌سازی تصادفات استفاده شده است.

۳- روش تحقیق

در پژوهش پیش رو، تصادفات با زیان مالی (خسارتی) و تصادفات با زیان جانی (جرحی و فوتی) در بزرگراه‌های درون‌شهری مشهد به وسیله چهار مدل رگرسیون پواسن، دو جمله‌ای منفی، پواسن پرصفر (ZIP) و دو جمله‌ای منفی پرصفر (ZINB) مدل‌سازی شده است. متغیرهای توضیحی (ناوابسته) که در مدل‌ها به کار رفته است شامل متغیرهای مربوط به جریان ترافیک و متغیرهای مربوط به طرح هندسی راه می‌شود. متغیرهای مربوط به جریان ترافیک شامل حجم ترافیک و سرعت و متغیرهای مربوط به طرح هندسی راه شامل شمار خطوط، شمار قوس‌های افقی و شمار راه‌های دسترسی می‌شود. منظور از راه‌های دسترسی، راه‌هایی است که به بزرگراه، وارد یا از آن خارج می‌شود و

به دست می‌آید:

$$\text{BIC} = -2LL + k \log n \quad \text{AIC} = -2LL + 2k \quad (2)$$

LL ، لگاریتم درست‌نمایی، k شمار پارامترها و n ، شمار مشاهدات است. هرچه مقدار معیار آگاهی آکایک کمتر باشد مدل برازنده‌تر خواهد بود [۱۱]. نتایج به دست آمده از ارزیابی مدل‌ها با معیارهای AIC و BIC مشابه است و مقادیر آن‌ها به هم نزدیک است. نتایج مرحله دوم مقایسه، بیشتر مؤید نتایج مرحله اول است. پس از مقایسه مدل‌ها به معتبرسازی مدل بهتر برای پیش‌بینی تصادفات بزرگراه‌های درون‌شهری پرداخته می‌شود تا کارایی و کفایت آن در مدل‌سازی تصادفات بزرگراه‌های درون‌شهری ثابت شود.

۴- معرفی مدل‌های به کاررفته برای مدل‌سازی تصادفات

۴-۱ مدل رگرسیون پواسن

در مدل رگرسیون پواسن، مشاهده I_i - ام متغیر وابسته، y_i به عنوان یک متغیر تصادفی پواسن با میانگین λ_i مدل می‌شود:

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

ویژگی مهم مدل رگرسیون پواسن، برابری میانگین با واریانس توزیع است و از آن به عنوان کاستی عمده این مدل یاد می‌شود و جلوی پراکندگی زیاد داده‌های تصادفات را می‌گیرد. در مدل رگرسیون پواسن، واریانس شرطی برابر میانگین شرطی است:

$$V(y_i | x_i) = E(y_i | x_i) = \lambda_i = e^{x_i \beta} \quad (4)$$

تصادفات مؤثر و کدام نقش چندانی در رخداد تصادف ندارند. مهم‌ترین هدف پژوهشگر در این پژوهش، ارزیابی کارایی مدل‌های رگرسیون پرصفر در برابر مدل‌های رگرسیون پواسن و دوجمله‌ای منفی در مدل‌سازی تصادفات مالی و جانی در بزرگراه‌های درون‌شهری است. این کار با ارزیابی مدل‌ها و مقایسه آن‌ها با یکدیگر انجام می‌شود، نخست برای مقایسه مدل‌های پواسن و دوجمله‌ای منفی از معناداری پارامتر پراکندگی و نیز آزمون نسبت درست‌نمایی (LR) استفاده می‌شود آزمون نسبت درست‌نمایی با رابطه‌ی زیر بدست می‌آید:

$$LR = -2(LL_p - LL_{NB}) \quad (1)$$

که در آن LL_p و LL_{NB} به ترتیب لگاریتم درست‌نمایی مدل پواسن و دوجمله‌ای منفی است. این آماره یک توزیع کی دو با $k-1$ درجه آزادی دارد که k شمار ضرایب رگرسیون است. اگر مقدار آماره بزرگ‌تر از مقدار $\chi^2_{0.05}(k-1)$ باشد مدل دوجمله‌ای منفی بهتر از مدل پواسن خواهد بود. معمولاً نتیجه این آزمون برای مقایسه مدل‌های یاد شده، معادل ارزیابی معناداری پارامتر پراکندگی است. برای مقایسه مدل پواسن با پواسن پرصفر که در واقع ارزیابی کارایی مدل‌های پرصفر در برابر مدل‌های ساده است از آماره وانگ استفاده می‌شود. مدل‌های پواسن پرصفر و دوجمله‌ای منفی پرصفر را نیز می‌توان مانند مدل‌های پواسن و دوجمله‌ای منفی با ارزیابی معناداری پارامتر پراکندگی و آزمون نسبت درست‌نمایی (LR) با هم مقایسه کرد. اما این یک مرحله از مقایسه مدل‌هاست؛ مرحله دیگر این مقایسه، ارزیابی خوبی برازش مدل‌ها و مقایسه برازندگی آن‌ها است. برای این کار می‌توان از معیارهای آگاهی آکایک (AIC) یا بیز (BIC) استفاده کرد. مقدار AIC و BIC از روابط زیر

رگرسیون دوجمله‌ای منفی علاوه بر ضرایب رگرسیون، باید پارامتر پرکندگی ($\alpha = 1/r$) را نیز برآورد کرد. برای برآورد ضرایب رگرسیون در مدل رگرسیون دوجمله‌ای منفی نیز از روش درست‌نمایی بیشینه استفاده می‌شود. در مدل‌های دوجمله‌ای منفی، شمار پیش‌بینی شده تصادفات (\hat{y}_i)، میانگین شرطی یا شمار متوسط رخدادها به شرط X_i است. این مقدار همان λ_i است که میانگین متغیر تصادفی y_i با توزیع پواسن است. این مقدار معمولاً یک عدد درست نیست. در حالی که مقدار مشاهده شده y_i ، یک شمار است [۱۲].

X_i ، بردار متغیرهای ناوابسته مشاهده شده و β ، بردار ضرایب رگرسیون است. برای برآورد ضرایب رگرسیون در مدل رگرسیون پواسن از روش درست‌نمایی بیشینه استفاده می‌شود. مقدار پیش‌بینی شده \hat{y}_i ، میانگین شرطی یا شمار متوسط رخدادها به شرط X_i است. این مقدار همان λ_i است که میانگین متغیر تصادفی y_i با توزیع پواسن است. این مقدار معمولاً یک عدد درست نیست. در حالی که مقدار مشاهده شده y_i ، یک شمار است [۱۲].

۴-۳- مدل‌های رگرسیون پرفر

مشکل دیگری که داده‌های تصادفات معمولاً با آن روبرو می‌شود، فزونی داده‌های صفر است. به سخن دیگر، شمار داده‌های صفر، بیش از شمار مورد انتظار در مدل‌های پواسن و دوجمله‌ای منفی است. برای برداشتن این کاستی، پژوهشگران به مدل‌های رگرسیونی تازه‌ای برای در نظر گرفتن این داده‌های پرشمار صفر روی آوردند. فرض اساسی در توزیع پرفر آن است که اجزای راه مانند تقاطع‌ها و بخش‌های راه، دو حالت داشته باشد [۱۴]:

- ۱- حالت داده تصادف واقعاً صفر یا حالت ذاتاً ایمن. هرچند برخی آن‌را حالت ایمن مجازی تعریف کرده‌اند تا مجبور به دفاع از ادعای ایمنی کامل آن بخش از راه نشوند.
- ۲- حالت داده غیرصفر. در این حالت ممکن است در برخی دوره‌های زمانی، داده‌های صفر هم گزارش شوند. این حالت دارای یک توزیع آماری پواسن (ZIP) یا دوجمله‌ای منفی (ZINB) است.

در حالت نخست تنها داده‌های صفر وجود دارد و در حالت دوم، شمار تصادفات مانند داده‌هایی با توزیع پواسن یا دوجمله‌ای منفی است. حالت اول با احتمال ω و حالت دوم با احتمال $1-\omega$ روی می‌دهد. داده‌های تصادفات معمولاً داده‌هایی با دو سطح است که یک سطح آن بیشتر

۴-۲- مدل رگرسیون دوجمله‌ای منفی

در مدل رگرسیون دوجمله‌ای منفی y_i ، مشاهده i -م متغیر وابسته، دارای تابع توزیع احتمال زیر است:

$$p(y_i) = \frac{\Gamma(y_i+r)}{y_i! \Gamma(r)} \left[\frac{\mu_i}{\mu_i+r} \right]^{y_i} \left[\frac{r}{\mu_i+r} \right]^r \quad (5)$$

میانگین شرطی y_i ، برای بردار متغیرهای ناوابسته مشاهده شده، X_i از بستگی زیر به دست می‌آید:

$$E(y_i | x_i) = \mu_i e^{x_i \beta} \quad (6)$$

علت روی آوردن پژوهشگران عرصه از توزیع پواسن به توزیع دوجمله‌ای منفی، کاستی توزیع پواسن در برابری میانگین و واریانس توزیع است. رابطه میان میانگین و واریانس در توزیع دوجمله‌ای منفی به شکل زیر است:

$$V(Y) = \mu + \frac{1}{r} \mu^2 \quad (7)$$

واریانس توزیع دوجمله‌ای منفی همواره بزرگ‌تر از میانگین آن است. از این رو برای داده‌های با واریانس بزرگ‌تر از میانگین، این توزیع مناسب است. در مدل

برای آزمودن مناسب بودن استفاده از یک مدل پرفر به جای مدل‌های رگرسیون پواسن یا دوجمله‌ای منفی از آماره وانگ استفاده می‌شود.

۵- گردآوری داده‌ها

برای مدل‌سازی تصادفات بزرگراه‌های درون‌شهری از داده‌های تصادفات بزرگراه‌های شهر مشهد در سال‌های ۸۵ و ۸۶ استفاده شده است. این اطلاعات به وسیله سازمان حمل و نقل و ترافیک مشهد به کمک GIS و با استفاده از گزارش‌های پلیس گردآوری شده است. آمار تصادفات در دو گروه با زیان مالی (خسارتی) و زیان جانی (جرحی و فوتی) گردآوری شده است. متغیرهای ناوابسته مدل‌ها شامل متغیرهای مربوط به جریان ترافیک و متغیرهای مربوط به طرح هندسی راه می‌شود. متغیرهای مربوط به جریان ترافیک شامل حجم و سرعت و متغیرهای مربوط به طرح هندسی راه شامل شمار خطوط، شمار قوس‌های افقی و شمار راه‌های دسترسی می‌شود. حجم ترافیک به حجم خودروهای سواری، غیرسواری سبک و غیرسواری سنگین تقسیم شده است. آمار تصادفات و متغیرهای ناوابسته مدل‌ها برای ۱۵۶ بخش بزرگراه‌های درون‌شهری مشهد در ۹۳۶ زیربخش گردآوری شده است. شمار کل تصادفات مالی، جرحی و فوتی رخ داده در بزرگراه‌های درون‌شهری مشهد در سال‌های ۸۵ و ۸۶ به ترتیب ۶۶۱۴، ۷۶۲ و ۱۴۷ است.

سازمان هر چند سال یک‌بار یک طرح جامع آمارگیری حمل و نقل و ترافیک در آبان‌ماه انجام می‌دهد و حجم عبوری انواع خودروها از بخش عمده‌ای از راه‌های شهری در ساعت‌های شبانه‌روز را به دست می‌آورد. در آبان‌ماه وضعیت ترافیک مشهد عادی است و مدل‌هایی که در این پژوهش ساخته شده تغییرات فصلی ترافیک را در نظر نمی‌گیرند؛ زیرا آمار مربوط به حجم ترافیک در فصل‌های

بخش معینی از راه و سطح دیگر می‌تواند دوره زمانی معینی از سال باشد. اگر y_{ij} ، شمار تصادفات رخ داده در بخش i ام راه در دوره زمانی j ام، دارای توزیع پرفر پواسن (ZIP) باشد، تابع جرم احتمال آن به شکل زیر است:

$$P(y_{ij}) = \begin{cases} \omega_{ij} + (1 - \omega_{ij}) \exp(-\lambda_{ij}), & y_{ij} = 0 \\ (1 - \omega_{ij}) \frac{\exp(-\lambda_{ij}) \lambda_{ij}^{y_{ij}}}{y_{ij}!}, & y_{ij} > 0 \end{cases} \quad (8)$$

در مدل رگرسیون ZIP، میانگین توزیع پواسن λ_{ij} با تابع پیوندی لگاریتم طبیعی به رگرسیونی از بردار متغیرهای ناوابسته X_{ij} و احتمال فرآیند صفر (ω_{ij}) با تابع پیوندی لاجیت به رگرسیونی از بردار متغیرهای ناوابسته Z_{ij} پیوند داده می‌شود:

$$\ln\left(\frac{\omega_{ij}}{1 - \omega_{ij}}\right) = Z_{ij}'\gamma \quad \text{و} \quad \ln \lambda_{ij} = X_{ij}'\beta \quad (9)$$

β و γ ، برداری از ضرایب رگرسیون است. معمولاً برای سادگی برآورد، برای بردار Z_{ij} همان متغیرهای ناوابسته بردار X_{ij} در نظر گرفته می‌شود. اگر داده‌های تصادفات علاوه بر مشکل صفرهای زیاد، دچار پراکندگی زیاد نیز باشد، مانند آنچه در مدل‌های رگرسیون دوجمله‌ای منفی بود از مدل رگرسیون دوجمله‌ای منفی پرفر (ZINB) برای مدل‌سازی استفاده می‌شود. متغیر شمارشی y_{ij} با توزیع دوجمله‌ای منفی پرفر تابع جرم احتمال زیر را دارد:

[۱۰ و ۱۵]:

(۱۰)

$$P(y_{ij}) = \begin{cases} \omega_{ij} + (1 - \omega_{ij})(1 + \alpha \lambda_{ij})^{-1/\alpha}, & y_{ij} = 0 \\ (1 - \omega_{ij}) \frac{\Gamma(y_{ij} + 1/\alpha)}{y_{ij}! \Gamma(1/\alpha)} (1 + \alpha \lambda_{ij})^{-1/\alpha} \left(1 + \frac{1}{\alpha \lambda_{ij}}\right)^{-y_{ij}}, & y_{ij} > 0 \end{cases}$$

P_i ، درصد خودروهای نوع i است. چون شمار انواع خودروها به طور ساعتی آمارگیری می شود؛ این شمار، حجم آن‌ها برحسب خودرو در ساعت (Veh/hr) است. از رابطه (۱۲) در برنامه Excel برای به دست آوردن شمار هر نوع خودرو از روی حجم معادل سواری و با داشتن درصد هر نوع خودرو در بخش مورد نظر راه، و در ساعت مورد نظر استفاده شده است. داده‌های سرعت در بخش‌های مختلف بزرگراه‌ها برای ساعات اوج همراه با داده‌های حجم به وسیله‌ی سازمان ترافیک و برای ساعات غیراوج روز و شب با حرکت در جریان ترافیک در ساعات مورد نظر گردآوری شده است. اطلاعات مربوط به شمار خطوط و راه‌های دسترسی در بخش‌های مختلف بزرگراه‌های درون شهری به طور عینی در بزرگراه‌ها و اطلاعات مربوط به شمار قوس‌های افقی از روی نقشه مسیر بزرگراه‌ها گرد آمده است.

۶- ساخت مدل‌ها

در این بخش از روی آمار گردآمده تصادفات و اطلاعات مربوط به متغیرهای ناوابسته، مدل‌های پواسن، دوجمله‌ای منفی و مدل‌های پرفر پواسن و دوجمله‌ای منفی برای تصادفات با زیان مالی و تصادفات با زیان جانی ساخته می شود تا مشخص شود برای هر گروه از تصادفات کدام متغیرها نقش مؤثری در رخداد تصادفات ایفا می کند و کدام اثر چندانی روی احتمال رخداد تصادفات ندارد. سپس با ارزیابی خوبی برازش و مقایسه مدل‌ها، کارایی و کفایت مدل‌های رگرسیون پرفر را در برابر مدل‌های پواسن و دوجمله‌ای منفی برای مدل‌سازی تصادفات با زیان مالی و تصادفات با زیان جانی در بزرگراه‌های درون شهری بررسی می شود. متغیرهای ناوابسته مدل‌ها شامل حجم خودروهای سواری، خودروهای سنگین، خودروهای غیرسواری سبک،

مختلف سال در دسترس نبود و داده‌های حجم ترافیک به ساعت‌های مختلف شبانه‌روز مربوط می شود. از روی این آمارها می توان به ترکیب ترافیک و درصد انواع خودروها پی برد. سازمان همچنین هر سال، حجم کل ترافیک در بخش‌های مختلف راه‌ها به ویژه بزرگراه‌ها را در ساعات اوج ترافیک برآورد می کند. از روی این حجم کل معادل و با داشتن ضریب همسنگ انواع خودروها و درصد هر یک در ترکیب ترافیک در ساعت‌های مختلف شبانه‌روز و نسبت حجم کل ترافیک در ساعت‌های غیراوج به ساعت اوج می توان حجم هر یک را در ساعت اوج و ساعت نماینده غیراوج در روز و در شب به دست آورد و از روی آن حجم خودروهای سواری و حجم معادل خودروهای غیرسواری سبک و سنگین در ساعت اوج و ساعت‌های غیراوج در روز و در شب به دست خواهد آمد. این محاسبات به وسیله‌ی برنامه Excel انجام شده است. ضریب همسنگ سواری انواع خودروها در جدول ۱ آمده است [۱۶]:

جدول (۱) ضریب همسنگ سواری انواع خودروها

سواری	وسایل سنگین	مینی بوس	اتوبوس غیرواحد
۱	۲.۵	۲	۲.۵
تاکسی	اتوبوس واحد	وانت	موتور و دوچرخه
۲	۵	۱	۰.۵

حجم کل (معادل سواری) از بستگی زیر به دست می آید:

$$V_{te} = e_1 N_1 + \dots + e_n N_n \quad (11)$$

که در آن V_{te} ، حجم معادل سواری و e_i و N_i به ترتیب ضریب همسنگ و شمار خودروهای نوع i است. شمار هر نوع خودرو از بستگی زیر به دست می آید:

$$N_i = \frac{P_i V_{te}}{e_1 P_1 + \dots + e_n P_n} \quad (12)$$

$$\lambda_{ij} = \exp(-0.6119 + 0.00082x_{ij,3} - 0.1492x_{ij,5} + 0.5369x_{ij,6} + 0.07797x_{ij,7}) \quad (14)$$

با توجه به مقدار برآورد شده λ_{ij} ، احتمال مشاهده y_{ij} تصادفات با زیان مالی و جانی به شکل زیر خواهد بود:

$$P(y_{ij}) = \frac{e^{-\lambda_{ij}} \lambda_{ij}^{y_{ij}}}{y_{ij}!} \quad (15)$$

۶-۲- مدل رگرسیون دوجمله‌ای منفی

برای برآورد بردار ضرایب رگرسیون و پارامتر پراکنندگی، مانند مدل رگرسیون پواسن، روش تکرار نیوتن-رافسون به کار می‌رود. ارزیابی معناداری متغیرهای ناوابسته مدل رگرسیون دوجمله‌ای منفی نیز مانند مدل پواسن است. پارامترهای برآورد شده مدل و ارزیابی معناداری آن‌ها برای تصادفات با زیان مالی و تصادفات با زیان جانی در جدول ۱ آمده است. مقدار μ_{ij} برای تصادفات با زیان مالی از بستگی زیر به دست خواهد آمد:

$$\mu_{ij} = \exp(-0.8108 + 0.00039x_{ij,1} + 0.00065x_{ij,3} + 0.01051x_{ij,4} + 0.2364x_{ij,5} + 0.3753x_{ij,6} + 0.1810x_{ij,7}) \quad (16)$$

همچنین مقدار μ_{ij} برای تصادفات با زیان جانی دارای بستگی زیر است:

$$\mu_{ij} = \exp(-0.9212 + 0.00101x_{ij,3} + 0.5018x_{ij,6} + 0.09643x_{ij,7}) \quad (17)$$

با توجه به مقدار برآورد شده μ_{ij} ، $r=1/\alpha=0.923$ برای تصادفات با زیان مالی و $r=1.272$ برای تصادفات با زیان جانی، به دست آمد احتمال مشاهده y_{ij} تصادفات با زیان مالی و جانی مانند زیر خواهد بود:

سرعت، شمار خطوط، قوس‌های افقی و راه‌های دسترسی در بخش آم بزرگراه در دوره زمانی زام شبانه‌روز، به ترتیب از $x_{ij,1}$ ، $x_{ij,2}$ ، $x_{ij,3}$ ، $x_{ij,4}$ ، $x_{ij,5}$ ، $x_{ij,6}$ و $x_{ij,7}$ است.

۶-۱- مدل رگرسیون پواسن

برای برآورد پارامترهای مجهول مدل از روش تکرار نیوتن-رافسون استفاده می‌شود. برای ارزیابی معناداری متغیرهای ناوابسته، معکوس ماتریس هسین در آخرین تکرار نیوتن-رافسون، ماتریس واریانس (- کوواریانس) بی‌پایان‌گرا خواهد بود. واریانس ضرایب برآورد شده، عناصر قطری ماتریس و ریشه دوم آن‌ها، انحراف معیار ضرایب است. آماره t برای هر ضریب هم برابر نسبت ضریب برآورد شده به انحراف معیار به دست آمده است. برای به دست آوردن مقادیر p -value هم از توزیع t با درجه آزادی برابر شمار مشاهدات استفاده می‌شود. اگر مقدار p -value حساب شده کمتر از سطح معناداری خواسته شده (۰.۰۵ و کمتر) باشد، متغیر مربوط به ضریب برآورد شده، معنادار است و در مدل خواهد ماند؛ در دیگر صورت از آن چشم‌پوشی شده و در مدل نخواهد ماند [۱۲]. همه مراحل مربوط به برآورد ضرایب رگرسیون و ارزیابی معناداری متغیرهای ناوابسته، به وسیله نرم‌افزار SAS انجام شده است. پارامترهای برآورد شده مدل و ارزیابی معناداری آن‌ها برای تصادفات با زیان مالی و تصادفات با زیان جانی در جدول ۲ آمده است. بدین ترتیب مقدار λ_{ij} برای تصادفات با زیان مالی از بستگی زیر به دست خواهد آمد:

$$\lambda_{ij} = \exp(0.0003x_{ij,1} + 0.0005x_{ij,3} + 0.00665x_{ij,4} + 0.1945x_{ij,5} + 0.3196x_{ij,6} + 0.1365x_{ij,7}) \quad (18)$$

همچنین مقدار λ_{ij} برای تصادفات با زیان جانی دارای بستگی زیر خواهد بود:

جدول (۲) برآورد پارامترهای مدل‌های رگرسیون پواسن و دوجمله‌ای منفی

تصادفات با زیان جانی				تصادفات با زیان مالی				متغیرها
دوجمله‌ای منفی		پواسن		دوجمله‌ای منفی		پواسن		
P-value	برآورد	P-value	برآورد	P-value	برآورد	P-value	برآورد	
۰/۰۰۵۷	-۰/۹۲۱۲	۰/۰۱۰۷	-۰/۶۱۱۹	۰/۰۰۳۷	-۰/۸۱۰۸	۰/۰۷۷۲	-۰/۱۶۰۷	ثابت عددی
۰/۱۸۴۵	-۰/۰۰۰۱	۰/۳۴۱۰	-۰/۰۰۰۱	<۰/۰۰۰۱	۰/۰۰۰۳۹	<۰/۰۰۰۱	۰/۰۰۰۳	حجم خودروهای سواری
۰/۲۴۱۷	-۰/۰۰۰۲	۰/۰۵۴۸	-۰/۰۰۰۲	۰/۹۱۹۵	-۰/۰۰۰۰۱	۰/۵۳۰۲	۰/۰۰۰۰۲	حجم خودروهای سنگین
<۰/۰۰۰۱	۰/۰۰۱۰	<۰/۰۰۰۱	۰/۰۰۰۸	<۰/۰۰۰۱	۰/۰۰۰۶۵	<۰/۰۰۰۱	۰/۰۰۰۵	حجم خودروهای سبک
۰/۱۶۰۴	۰/۰۰۴۴	۰/۱۲۱۴	۰/۰۰۳۶	<۰/۰۰۰۱	۰/۰۱۰۵	<۰/۰۰۰۱	۰/۰۰۶۶۵	سرعت
۰/۲۲۳۱	-۰/۰۹۶۶	۰/۰۱۰۰	-۰/۱۴۹۲	۰/۰۰۰۶	۰/۲۳۶۴	<۰/۰۰۰۱	۰/۱۹۴۵	شمار خطوط
<۰/۰۰۰۱	۰/۰۵۱۸	<۰/۰۰۰۱	۰/۵۳۶۹	<۰/۰۰۰۱	۰/۳۷۵۳	<۰/۰۰۰۱	۰/۳۱۹۶	شمار قوس های افقی
۰/۰۰۲۶	۰/۰۹۶۴۳	۰/۰۰۰۹	۰/۰۷۷۹۷	<۰/۰۰۰۱	۰/۱۸۱۰	<۰/۰۰۰۱	۰/۱۳۶۵	شمار راههای دسترسی
<۰/۰۰۰۱	۰/۷۸۶۳			<۰/۰۰۰۱	۱/۰۸۳۷			پارامتر پراکنندگی

۶-۳-۱- مدل رگرسیون پواسن پرفسر

با توجه به جدول ۲، مقدار λ_{ij} برای تصادفات با زیان مالی از بستگی زیر به دست خواهد آمد:

$$\lambda_{ij} = \exp(0.7884 + 0.00021 x_{ij,1} + 0.00013 x_{ij,2} + 0.00038 x_{ij,3} + 0.0039 x_{ij,4} + 0.0925 x_{ij,5} + 0.2267 x_{ij,6} + 0.0943 x_{ij,7}) \quad (20)$$

همچنین ω_{ij} ، احتمال مشاهده صفر تصادف در بخش نام بزرگراه در دوره زمانی t ، از بستگی زیر به دست می‌آید:

$$\omega_{ij} = \frac{\exp(\tau x'_{ij}\beta)}{1 + \exp(\tau x'_{ij}\beta)} \quad (21)$$

به طوری که:

$$\exp(\tau x'_{ij}\beta) = \exp(4.1307 - 0.00081 x_{ij,1} + 0.00067 x_{ij,2} - 0.00074 x_{ij,3} - 0.01832 x_{ij,4} - 0.8403 x_{ij,5} - 1.0028 x_{ij,6} - 0.5835 x_{ij,7}) \quad (22)$$

$$p(y_{ij}) = \frac{\Gamma(y_{ij} + r)}{y_{ij}! \Gamma(r)} \left[\frac{\mu_{ij}}{\mu_{ij} + r} \right]^{y_{ij}} \left[\frac{r}{\mu_{ij} + r} \right]^r \quad (18)$$

۶-۳-۲- مدل‌های رگرسیون پرفسر

روابط پیوندی میانگین توزیع، λ_{ij} و ω_{ij} ، احتمال فرایندی که تنها داده‌های صفر تولید می‌کند با بردار متغیرهای ناوابسته، X_{ij} به شکل زیر است:

$$\ln\left(\frac{\omega_{ij}}{1 - \omega_{ij}}\right) = \tau x'_{ij}\beta \quad \text{و} \quad \ln \lambda_{ij} = x'_{ij}\beta \quad (19)$$

برای برآورد ضرایب رگرسیون و ارزیابی معناداری متغیرهای رگرسیون مانند گذشته عمل می‌شود، ولی این بار متغیرهای معنادار رابطه پیوندی λ_{ij} با متغیرهای معنادار

رابطه پیوندی ω_{ij} متفاوت است. پارامترهای برآورد شده مدل‌ها و ارزیابی معناداری آن‌ها برای تصادفات با زیان مالی و تصادفات با زیان جانی در جدول ۳ آمده است.

$$\exp(\tau x'_{ij}\beta) = \exp(5.0675 - 0.00186 x_{ij,2} - 0.00672 x_{ij,4} + 2.9237 x_{ij,5} - 1.3949 x_{ij,6} + 2.1289 x_{ij,7}) \quad (28)$$

با توجه به مقدار برآورد شده λ_{ij} ، ω_{ij} و $\alpha = 0.748$ برای تصادفات مالی و $\alpha = 0.6365$ برای تصادفات جانی احتمال مشاهده y_{ij} تصادف با زیان مالی و جانی از بستگی (۱۰) به دست می‌آید.

۶-۴- ارزیابی خوبی برازش و مقایسه مدل‌ها

پس از ساخت مدل‌ها و شناسایی عوامل مؤثر به رخداد تصادفات مالی و جانی، ارزیابی و مقایسه مدل‌ها انجام شود تا کارایی مدل‌های رگرسیون پرفر در برابر مدل‌های ساده پواسن و دوجمله‌ای منفی برای مدل‌سازی تصادفات در بزرگراه‌های درون‌شهری روشن شود. نتایج ارزیابی خوبی برازش مدل‌ها و مقایسه آن‌ها با یکدیگر برای تصادفات مالی و جانی در جدول ۳ آمده است. برای ارزیابی خوبی برازش مدل‌ها از معیار آگاهی آیک (AIC) استفاده می‌شود. هرچه مقدار AIC کمتر باشد، مدل برازنده‌تر خواهد بود؛ اما برازندگی مدل‌های رگرسیونی مورد بحث را نمی‌توان به‌طور مطلق ارزیابی کرد بلکه این برازندگی با مقایسه مقدار AIC مدل‌ها ارزیابی می‌شود. نتایج ارزیابی خوبی برازش مدل‌ها و مقایسه آن‌ها با یکدیگر در جدول ۴ داده شده. برای تصادفات با زیان مالی، پارامتر پراکندگی در مدل دوجمله‌ای منفی معنادار و مقدار منفی دوبرابر لگاریتم درست‌نمایی برای مدل پواسن خیلی بزرگ‌تر از مدل دو جمله‌ای منفی است. مقدار آماره نسبت درست‌نمایی (LR) برابر $LR = -2(LL_p - LL_{NB}) = 3377.3$ است که بسیار بزرگ‌تر از مقدار $\chi^2_{0.05}(7) = 14.07$ است.

مقدار λ_{ij} برای تصادفات با زیان جانی نیز از بستگی زیر به دست خواهد آمد:

$$\lambda_{ij} = \exp(0.00074 x_{ij,3} + 0.4913 x_{ij,6} + 0.1071 x_{ij,7}) \quad (23)$$

ω_{ij} برای تصادفات جانی هم از بستگی (۲۱) به دست می‌آید به‌طوری که:

$$\exp(\tau x'_{ij}\beta) = \exp(-1.3377 - 0.00069 x_{ij,3} + 0.4276 x_{ij,5}) \quad (24)$$

از روی مقدار برآورد شده λ_{ij} و ω_{ij} ، احتمال مشاهده y_{ij} تصادف با زیان مالی و جانی از بستگی (۸) بدست می‌آید.

۶-۳-۲- مدل رگرسیون دوجمله‌ای منفی پرفر

با توجه به جدول ۲ مقدار λ_{ij} برای تصادفات با زیان مالی دارای رابطه زیر خواهد بود:

$$\lambda_{ij} = \exp(0.00027 x_{ij,1} + 0.00064 x_{ij,3} + 0.0081 x_{ij,4} + 0.0569 x_{ij,5} + 0.3096 x_{ij,6} + 0.1179 x_{ij,7}) \quad (25)$$

همچنین ω_{ij} از بستگی (۲۱) بدست می‌آید به‌طوری که:

$$\exp(\tau x'_{ij}\beta) = \exp(7.358 - 1.9668 x_{ij,5} - 1.5138 x_{ij,6} - 1.8 x_{ij,7}) \quad (26)$$

مقدار λ_{ij} برای تصادفات با زیان جانی نیز از بستگی زیر بدست خواهد آمد:

$$\lambda_{ij} = \exp(-1.2635 + 0.00094 x_{ij,3} + 0.4456 x_{ij,6} + 0.0695 x_{ij,7}) \quad (27)$$

ω_{ij} برای تصادفات جانی هم از بستگی (۲۱) بدست می‌آید به‌طوری که:

جدول (۳) برآورد پارامترهای مدل‌های پرفر

تصادفات با زیان مالی				تصادفات با زیان جانی				ضرایب
پواسن پرفر		دوجمله‌ای منفی پرفر		پواسن پرفر		دوجمله‌ای منفی پرفر		
بخش پواسن		بخش دوجمله‌ای منفی		بخش پواسن		بخش دوجمله‌ای منفی		
P-value	برآورد	P-value	برآورد	P-value	برآورد	P-value	برآورد	
<۰/۰۰۰۱	-۱/۲۶۳۵	۰/۱۲۹۷	-۰/۵۵۸۸	۰/۳۳۷۹	۰/۲۶۴۸	<۰/۰۰۰۱	۰/۷۸۸۴	ثابت عددی
۰/۱۹۴۱	-۰/۰۰۰۱	۰/۱۵۸۲	-۰/۰۰۰۱	۰/۰۰۰۱	۰/۰۰۰۳	<۰/۰۰۰۱	۰/۰۰۰۲	حجم خودروهای سواری
۰/۲۶۴۳	-۰/۰۰۰۲	۰/۴۰۹۶	-۰/۰۰۰۱	۰/۵۲۳۶	۰/۰۰۰۱	۰/۰۰۰۵	۰/۰۰۰۱	حجم خودروهای سنگین
<۰/۰۰۰۱	۰/۰۰۰۹	<۰/۰۰۰۱	۰/۰۰۰۷	<۰/۰۰۰۱	۰/۰۰۰۶	<۰/۰۰۰۱	۰/۰۰۰۴	حجم خودروهای سبک
۰/۶۱۲۱	۰/۰۱۲۹	۰/۶۱۲۴	۰/۰۰۱۹	۰/۰۰۰۳	۰/۰۰۸۱	<۰/۰۰۰۱	۰/۰۰۳۹	سرعت
۰/۱۹۱۳	-۰/۰۹۷۰	۰/۹۹۲۹	-۰/۰۰۰۶	۰/۰۳۷۳	۰/۰۵۶۹	<۰/۰۰۰۱	۰/۰۹۲۵	شمار خطوط
<۰/۰۰۰۱	۰/۴۴۵۶	<۰/۰۰۰۱	۰/۴۹۱۳	<۰/۰۰۰۱	۰/۳۰۹۶	<۰/۰۰۰۱	۰/۲۲۶۷	شمار قوس‌های افقی
۰/۰۲۱۰	۰/۰۶۹۵	۰/۰۰۰۵	۰/۱۰۷۱	<۰/۰۰۰۱	۰/۱۱۷۹	<۰/۰۰۰۱	۰/۰۹۴۳	شمار راه‌های دسترسی
<۰/۰۰۰۱	۰/۶۳۵۶			<۰/۰۰۰۱	۰/۷۴۸۰			پارامتر پراکندگی

جدول (۴) ارزیابی و مقایسه مدل‌ها

دوجمله‌ای منفی پرفر		پواسن پرفر		دوجمله‌ای منفی		پواسن	
P-value	مقدار			P-value	مقدار		
تصادفات با زیان مالی							
<۰/۰۰۰۱	۰/۷۴۸۰			<۰/۰۰۰۱	۱/۰۸۳۷		alpha
	۵۲۵۸	۷۶۷۳		۵۳۶۹		۹۰۴۶	-2LL
		۱۰/۵۶					آماره وانگ
	۵۲۹۲	۷۷۰۵		۵۳۸۷		۹۰۶۲	AIC
	۵۳۷۴	۷۷۸۲		۵۴۳۰		۹۱۰۱	BIC
تصادفات با زیان جانی							
<۰/۰۰۰۱	۰/۶۳۶۵			<۰/۰۰۰۱	۰/۷۸۶۳		alpha
	۲۳۰۲	۲۳۶۱		۲۳۲۸		۲۴۷۷	-2LL
		۳/۸۰					آماره وانگ
	۲۳۱۸	۲۳۹۳		۲۳۴۶		۲۴۹۳	AIC
	۲۴۱۸	۲۴۷۱		۲۳۹۰		۲۵۳۲	BIC

است، پس مدل پواسن پرفر بهتر از مدل پواسن است. معناداری پارامتر پراکندگی در مدل دوجمله‌ای منفی پرفر و نیز بزرگ‌تر بودن مقدار منفی دو برابر لگاریتم درست‌نمایی مدل پواسن پرفر از مدل دوجمله‌ای منفی پرفر، به میزان

پس می‌توان نتیجه گرفت که مدل رگرسیون دوجمله‌ای منفی برای تصادفات با زیان مالی، بهتر از مدل پواسن است. اما برای مقایسه مدل پواسن با مدل پواسن پرفر از آماره وانگ استفاده می‌شود. مقدار آماره برابر ۱۰.۵۶، بزرگ‌تر از ۱.۹۶

بزرگراه در سال‌های ۸۵ و ۸۶ در ۲۰ بخش و ۶۰ زیربخش، مجموعاً ۱۲۰ بار مشاهده تصادف همراه آمار مربوط به متغیرهای ناوابسته [مدل شامل متغیرهای مربوط به جریان ترافیک و طرح هندسی بزرگراه] گردآوری شده است. از ساخت مدل، شمار پیش‌بینی شده تصادفات با زیان مالی و جانی به ترتیب از روابط (۲۵) و (۲۷) به دست می‌آید. برای مقایسه شمار مشاهده‌شده و پیش‌بینی شده تصادفات، آزمون مانده‌های پیرسون به کار می‌رود [۱۳]. داریم:

$$S = \sum_{i=1}^n \sum_{j=1}^3 s_{ij} = \sum_{i=1}^n \sum_{j=1}^3 \left(\frac{y_{ij} - E(y_{ij})}{\sqrt{V(y_{ij})}} \right)^2 \quad (29)$$

$$= \sum_{i=1}^n \sum_{j=1}^3 \frac{(y_{ij} - \lambda_{ij})^2}{\lambda_{ij} + \alpha \lambda_{ij}^2}$$

y_{ij} ، شمار مشاهده‌شده و λ_{ij} ، شمار پیش‌بینی شده تصادفات در بخش i ام بزرگراه و در دوره زمانی j ام شبانه‌روز، α پارامتر پراکندگی و n ، شمار بخش‌های بزرگراه است. S دارای یک توزیع کی‌دو با $N-k$ درجه آزادی است که در آن N ، شمار مشاهدات و k ، شمار ضرایب رگرسیون است. هرچه تعداد صفحات مشاهده‌شده و پیش‌بینی شده به هم نزدیک‌تر باشد مدل معتبرتر است و در حالت بحرانی، مقدار S ، برابر مقدار $\chi^2_{0.05}(N-k)$ است و مقدار S کمتر از آن، نشان از معتبر بودن مدل خواهد داشت. برای تصادفات با زیان مالی، مقدار S برابر 36.64 ، کمتر از مقدار $\chi^2_{0.05}(120-6) = 139.41$ است که نشان از معتبر بودن مدل دارد. اما برای تصادفات با زیان جانی، مقدار S برابر 33.83 ، کمتر از مقدار $\chi^2_{0.05}(120-4) = 142$ است که نشان از معتبر بودن مدل رگرسیون دوجمله‌ای منفی پرفسفر در مدل‌سازی تصادفات با زیان جانی دارد.

کافی نشان از بهتر بودن مدل دوجمله‌ای منفی پرفسفر نسبت به مدل پواسن پرفسفر دارد. مقایسه مقدار معیار آگاهی آکایک (AIC) برای مدل‌های چهارگانه نشان می‌دهد مدل رگرسیون دوجمله‌ای منفی پرفسفر با کمترین مقدار AIC، بهترین و برانده‌ترین مدل برای تصادفات با زیان مالی است. در تصادفات با زیان جانی نیز به همین ترتیب، پارامتر پراکندگی معنادار و مقدار آماره نسبت درست‌نمایی $LR = -2(LL_p - LL_{NB}) = 148.4$ است که نشان می‌دهد مدل رگرسیون دوجمله‌ای منفی بهتر از مدل پواسن است. همچنین مقدار آماره وانگ برابر $3/8$ ، بزرگ‌تر از $1/96$ نشان می‌دهد که مدل پواسن پرفسفر بهتر از مدل رگرسیون پواسن است. اما از آن‌جا که پارامتر پراکندگی مدل رگرسیون دوجمله‌ای منفی پرفسفر معنادار و مقدار آماره نسبت درست‌نمایی برابر $LR = -2(LL_{ZIP} - LL_{ZINB}) = 59.2$ ، بزرگ‌تر از مقدار $\chi^2_{0.05}(7) = 14.07$ برابر است، مدل دوجمله‌ای منفی پرفسفر بهتر از مدل پواسن پرفسفر است. تأییدکننده این نتایج، ارزیابی و مقایسه مدل‌ها با استفاده از معیار آگاهی آکایک (AIC) است. مقدار AIC برای مدل رگرسیون دوجمله‌ای منفی پرفسفر، کمتر از دیگر مدل‌هاست و این نشان می‌دهد که این مدل بهترین و برانده‌ترین مدل برای مدل‌سازی تصادفات با زیان جانی در بزرگراه‌های درون‌شهری است.

۶-۵- معتبرسازی مدل دوجمله‌ای منفی پرفسفر

برای معتبرسازی مدل رگرسیون دوجمله‌ای منفی پرفسفر (ZINB) در مدل‌سازی تصادفات بزرگراه‌های درون‌شهری، از این مدل برای پیش‌بینی شمار و احتمال رخداد تصادفات با زیان مالی و زیان جانی در بزرگراه درون‌شهری دیگری استفاده می‌شود تا با مقایسه شمار مشاهده‌شده و پیش‌بینی شده تصادفات در هر بخش از بزرگراه و در دوره زمانی معین، اعتبار مدل آزموده شود. آمار تصادفات این

۷- نتایج و ارائه راه کار

نتایج به دست آمده از ساخت مدل های رگرسیون چهارگانه شامل ضرایب رگرسیون و معناداری متغیرهای ناوابسته برای تصادفات با زیان مالی و تصادفات با زیان جانی، در جدول ۵ آمده است.

چنانکه از جدول ۵ برمی آید، با وجود و افزایش شمار راه های دسترسی، شمار تصادفات افزایش می یابد؛ زیرا وجود راه های دسترسی در بخش های بزرگراه های درون شهری سبب تداخل با جریان ترافیک بزرگراه می شود و احتمال رخداد تصادف را افزایش می دهد و هرچه شمار راه های دسترسی به بخش های بزرگراه ها بیشتر باشد، تداخل با جریان ترافیک بزرگراه بیشتر و احتمال رخداد تصادف بالاتر می رود؛ این تصادفات می تواند هم تصادفات با زیان مالی باشد و هم تصادفات با زیان جانی.

نتایج مدل سازی نشان می دهد با وجود و افزایش شمار قوس های افقی در بخش های راه، احتمال رخداد تصادفات با زیان مالی و تصادفات با زیان جانی افزایش می یابد. نتیجه ای که به دست آمده با یافته های برخی پژوهشگران همخوانی ندارد [۵ و ۹]. برای نمونه Chang (2005) در مطالعات خود روی فراوانی تصادفات آزادراه ها به این نتیجه رسید که وجود قوس های افقی در بخش های آزادراه، احتمال رخداد تصادف را کاهش می دهد. توجیه او هم این بود که رانندگان در قوس های افقی یا همان پیچ ها با احتیاط بیشتری می رانند. پژوهشگر بر این باور است که دست کم در بزرگراه های درون شهری کشور نمی توان چندان روی کاهش احتمال رخداد تصادفات در بخش های دارای قوس افقی به سبب احتیاط بیشتر رانندگان تکیه کرد.

جدول (۵) نتایج مدل سازی تصادفات

دوجمله ای منفی پرفسر		پواسن پرفسر		دوجمله ای منفی	پواسن	
دوجمله ای منفی	پرفسر	پواسن	پرفسر			
تصادفات با زیان مالی						
۰/۰۰۰۳	-۰/۰۰۰۹	۰/۰۰۰۲	-۰/۰۰۰۸	۰/۰۰۰۳۹	**۰/۰۰۰۳	حجم خودروهای سواری
۰/۰۰۰۱	۰/۰۰۰۵	۰/۰۰۰۱	۰/۰۰۰۷	-۰/۰۰۰۰۱	۰/۰۰۰۰۲	حجم خودروهای سنگین
۰/۰۰۰۶	-۰/۰۰۰۳	۰/۰۰۰۴	-۰/۰۰۰۷	۰/۰۰۰۶۵	۰/۰۰۰۰۵	حجم خودروهای سبک
۰/۰۰۸۱	-۰/۰۱۷۸	۰/۰۰۳۹	-۰/۰۱۸۳	۰/۰۱۰۵	۰/۰۰۶۶	سرعت
۰/۰۵۶۹	-۱/۹۶۶۸	۰/۰۹۲۵	-۰/۸۴۰۳	۰/۲۳۶۴	۰/۱۹۴۵	شمار خطوط
۰/۳۰۹۶	-۱/۵۱۳۸	۰/۲۲۶۷	-۱/۰۰۲۸	۰/۳۷۵۳	۰/۳۱۹۶	شمار قوس های افقی
۰/۱۱۷۹	-۱/۸۰۰۰	۰/۰۹۴۳	-۰/۵۸۳۵	۰/۱۸۱۰	۰/۱۳۶۵	شمار راه های دسترسی
تصادفات با زیان جانی						
-۰/۰۰۰۱	۰/۰۰۱۱	-۰/۰۰۰۱	-۰/۰۰۰۱	-۰/۰۰۰۱	-۰/۰۰۰۱	حجم خودروهای سواری
-۰/۰۰۰۲	-۰/۰۰۱۹	-۰/۰۰۰۱	۰/۰۰۰۲	-۰/۰۰۰۲	-۰/۰۰۰۲	حجم خودروهای سنگین
۰/۰۰۰۹	۰/۰۰۰۸	۰/۰۰۰۷	-۰/۰۰۰۷	۰/۰۰۱۰	۰/۰۰۰۸	حجم خودروهای سبک
۰/۰۱۲۹	-۰/۰۰۶۷	۰/۰۰۱۹	-۰/۰۰۶۸	۰/۰۰۴۴	۰/۰۰۳۶	سرعت
-۰/۰۹۷۰	۲/۹۲۳۷	-۰/۰۰۰۶	۰/۴۲۷۶	-۰/۰۹۶۶	-۰/۱۴۹۲	شمار خطوط
۰/۴۴۵۶	-۱/۳۹۴۹	۰/۴۹۱۳	-۰/۱۳۴۳	۰/۵۰۱۸	۰/۵۳۶۹	شمار قوس های افقی
۰/۰۶۹۵	۲/۱۲۸۹	۰/۱۰۷۱	۰/۰۱۰۹	۰/۰۹۶۴	۰/۰۷۸۰	شمار راه های دسترسی

*ضرایب برجسته، معنادار در سطح ۰/۰۵ می باشند.

و خودروهای سنگین نقش چندانی در رخداد تصادفات با زیان جانی نداشت.

پس از تحلیل نتایج مدل‌سازی، ارزیابی خوبی برآزش مدل‌ها و مقایسه آن‌ها با یک‌دیگر انجام شد و نتیجه نشان داد که مدل رگرسیون دوجمله‌ای منفی پرفر بهترین و برزنده‌ترین مدل، هم برای مدل‌سازی تصادفات با زیان مالی و هم برای تصادفات با زیان جانی است. سپس به معتبرسازی مدل برای پیش‌بینی تصادفات بزرگراه‌های درون‌شهری پرداخته شد تا بر این اساس بتوان کفایت مدل‌های رگرسیون پرفر را در مدل‌سازی تصادفات بزرگراه‌های درون‌شهری مشاهده تأیید کرد.

از نتایج پژوهش می‌توان دریافت که این خودروهای غیر سواری سبک یعنی تاکسی‌ها و موتورسیکلت‌ها است که نقشی عمده در بروز تصادفات بزرگراه‌های درون‌شهری ایفا می‌کند و نه خودروهای سنگین. نتایج به‌دست‌آمده، قابل استناد در طرح‌های آینده ساماندهی ترافیک بزرگراه‌های درون‌شهری کشور است در این طرح‌ها باید تمهیداتی برای مهار نقش تاکسی‌ها و موتورسیکلت‌ها در رخداد تصادفات اندیشید. این تمهیدات می‌تواند ساماندهی و یا تفکیک این نوع خودروها از جریان عادی ترافیک بزرگراه‌های درون‌شهری باشد. همچنین با کاهش شمار راه‌های دسترسی و قوس‌های افقی می‌توان از احتمال بروز تصادفات در بزرگراه‌ها کاست. از سوی دیگر با کنترل مستمر رعایت سرعت مجاز در بزرگراه‌های درون‌شهری از طریق حضور پلیس بزرگراه‌ها، سرعت سنج‌ها یا دوربین‌های کنترل سرعت می‌توان به میزان چشم‌گیری از شمار تصادفات بزرگراه‌های درون‌شهری کاست.

۸- مراجع

[۱] ابی ترابی، م، رضایی مقدم، ف، "مدل‌سازی شدت

نتایج مدل‌سازی تصادفات در جدول ۵، نشان می‌دهد با افزایش شمار خطوط، احتمال رخداد تصادفات با زیان مالی افزایش می‌یابد؛ در حالی که همین شمار خطوط، اثر چندانی روی احتمال رخداد تصادفات با زیان جانی ندارد. این نتایج با یافته‌های دیگر پژوهشگران سازگار است [۶ و ۹]، هرچند با افزایش شمار خطوط، جریان ترافیک روان‌تر می‌شود اما از سوی دیگر باعث افزایش تداخل ترافیکی و مانورهای رانندگی می‌شود و همین احتمال، رخداد تصادف را بالا می‌برد.

نتایج مدل‌سازی نشان می‌دهد که عامل سرعت، نقشی مؤثر در رخداد تصادفات با زیان مالی ایفا می‌کند؛ به‌طوری که با افزایش سرعت، احتمال رخداد تصادفات با زیان مالی بالا می‌رود اما نقش چندانی در رخداد تصادفات با زیان جانی ندارد. نتایج برخی پژوهشگران نشان می‌دهد که تجاوز از سرعت مجاز و تجاوز از سرعت مطمئنه در شرایط رانندگی باعث افزایش احتمال رخداد تصادفات با زیان جانی می‌شود اما اثر چندانی روی احتمال رخداد تصادفات با شدت کم ندارد [۱۰ و ۱۷]. این یافته‌ها با نتایج به‌دست‌آمده در این پژوهش سازگاری ندارد. شاید علت این باشد که مطالعات دیگر پژوهشگران بیشتر روی بزرگراه‌های جاده‌ای یا درون‌شهری با سرعت‌های مجاز بالا متمرکز است؛ در حالی که این پژوهش روی بزرگراه‌های درون‌شهری با سرعت‌های مجاز کمتر انجام شده است.

نتایج پژوهش نشان می‌دهد حجم خودروهای سواری و خودروهای غیرسواری سبک، نقش افزایش‌دهنده‌ای در احتمال رخداد تصادفات با زیان مالی دارد؛ اما حجم خودروهای سنگین اثر چندانی روی احتمال رخداد تصادفات با زیان مالی در بزرگراه‌های درون‌شهری ندارد. همچنین حجم خودروهای سبک باعث افزایش احتمال رخداد تصادفات با زیان جانی می‌شود، در حالی که حجم خودروهای سواری

- Data Analysis", Paper 371-2008.
- [12] Anselin, L. (2002) "Appendix C: Ordinary Least Square and Poisson Regression Models", University of Illinois Champaign-Urbana, IL.
- [13] Agresti, A. (2002) "Categorical Data Analysis: Case Study - Negative Binomial Regression", Wiley, Nascar Lead Changes 1975-1979. www.nascar.com.
- [14] Lord, D., Washington, S.P. and Ivan, J.N. (2007) "Further notes on the application of Zero-inflated models in highway safety", Accident Analysis and Prevention, 39, pp. 53-57.
- [15] Xiang, L., Lee, A.H., Yau, K.K.W. and McLachlan, G.J. (2007) "A score test for overdispersion in zero-inflated poisson mixed regression model", Statistics in Medicine, 26, pp. 1608-1622.
- [۱۶] دفتر مطالعات جامع حمل و نقل "چهارمین آمارنامه حمل و نقل شهر مشهد"، سازمان حمل و نقل و ترافیک مشهد. (۱۳۸۷)
- [17] Chang, L.-Y. and Wang, H.-W. (2006) "Analysis of traffic injury severity: An application of non-parametric classification tree techniques", Accident Analysis and Prevention, 38, pp. 1019-1027.
- تصادف‌ها در بزرگراه‌های درون‌شهری"، پژوهشنامه حمل و نقل، شماره ۱۸، ص. ۱-۱۲. (۱۳۸۸)
- [2] Caliendo, C., Guida, M. and Parisi, A.. (2007) "A crash-prediction model for multilane roads", Accident Analysis and prevention 39, 657-6.
- [3] Lee, J.Y., Chung, J.-H. and Son, B. (2008) "Analysis of traffic accident size for Korean highway using structural equation models", Accident Analysis and Prevention, 40, pp. 1955-1963.
- [4] Malyshkina, N.V. and Mannering, F.L. (2009) "Empirical assessment of the impact of highway design exceptions on the frequency and severity of vehicle accidents", Accident Analysis and Prevention, 43, pp. 1-9.
- [5] Milton, J.C., Shankar, V.N. and Mannering, F.L. (2008) "Highway accident severities and the mixed logit model: An exploratory empirical analysis", Accident Analysis and Prevention, 40, pp. 260-266.
- [6] Wang, C., Quddus, M.A. and Ison, S.G. (2009) "Impact of traffic congestion on road accidents: A spatial analysis of the M25 motorway in England", Accident Analysis and Prevention, 41, pp. 798- 808.
- [7] Anastasopoulos, C. and Mannering, F.L. (2009) "A note on modeling vehicle accident frequencies with random-parameters count models", Accident Analysis and Prevention, 41, pp. 153-159.
- [8] Ma, J., Kockelman, K.M. and Damien, P. (2008) "A multivariate Poisson-lognormal regression model for prediction of crash counts by severity, using Bayesian methods", Accident Analysis and Prevention, 40, pp. 964-975.
- [9] Chang, L.-Y. (2005) "Analysis of freeway accident frequencies: Negative binomial regression versus artificial neural network", Safety Science, 43, pp. 541-557.
- [10] Lee, J. and Mannering, F. (2002) "Impact of roadside features on the frequency and severity of run-off-roadway accidents: an empirical analysis", Accident Analysis and Prevention, 41, pp. 798-808.
- [11] Liu, W. and Cela, J. (2008) "Count Data Models in SAS. SAS Global Forum 2008: Statistics and

Application of Zero-Inflated Regression Models in Modeling Accidents on Urban Highways

E. Ayati¹, E. Abbasi^{2*}

1- Full Professor, Department of Engineering, Ferdousi University of Mashhad

2- M.Sc., Department of Engineering, Ferdousi University of Mashhad

ehsan.abbasi395@gmail.com

Abstract:

We handle in this paper, the research that have been performed to recognize the factors that affect crash frequency and severity in urban highways and use crash data of Mashhad urban highways as a case study. Statistical models that have been used in this research include Poisson, Negative binomial, Zero-inflated poisson and Zero-inflated negative binomial regression models. Traffic flow related variables and road geometric related variables have been used as independent variables of models. We are interested in this study, to inspect the efficiency of Zero-inflated models against simple Poisson and Negative binomial regression models in modeling accidents on urban highways. Special task that have been done in this research, is separation of total traffic volume into passenger cars, heavy vehicles and light non-passenger vehicles volume. Through this special, Researcher intend to have an especial look at the role of traffic volume in accident occurrence to see precisely, which part of traffic have effective role or more effective role in crash occurrence.

Accident data are two-level data, the first level is road segments i.e. highway is divided into several segments. The segmentation is based on total traffic volume i.e. each segment has a constant volume. The second level is daily hours; peak hour traffic considered as the first sublevel, day non-peak hour traffic the second and night non-peak hour traffic as the third sublevel. SAS 9.1 software has been used to fulfill statistical computations. It turns up, after statistical analyses, which factors affect crash occurrence and which do not have much effect. Comparisons between obtained results and other researchers' results have been made then. The main object of researcher is to assess the efficiency of Zero-inflated models against Poisson and Negative binomial regression models in modeling urban highways crashes. This aim is followed by, with evaluating goodness of fit and making comparison between models.

The Results of study show that the presence and number of access roads and horizontal curves on highway segments increase the likelihood of accidents, both no injury and more severe. Also increment of speed and number of lanes increase the likelihood of no injury accidents, but not more severity ones. The conclusions also demonstrate that the volume of passenger cars and light non-passenger car vehicles increase the likelihood of no injury accidents, but heavy vehicles volume does not have much effect on occurrence of no injury accidents, also light vehicles increase the likelihood of more severe accidents, but passenger cars and heavy vehicles volume does not have much effect on occurrence of severe accidents. Finally, the results of research indicate that Zero-inflated negative binomial regression model is best fitting the modeling of accidents, whether no injury or more severe and consequently, the efficiency of zero-inflated models in modeling accidents on urban highways is approved.

Keywords: No injury accidents, More severe accidents, Urban highways, Modeling accidents, Zero-inflated regression models