

مدل پیش‌بینی تصادفات موتورسیکلت در تقاطع‌های شهر تهران*

جلیل شاهی، دانشیار، دانشکده عمران، دانشگاه علم و صنعت ایران، تهران، ایران
محمود احمدی نژاد، استادیار، دانشکده عمران، دانشگاه علم و صنعت ایران، تهران، ایران
عبدالرضا شیخ‌الاسلامی، دانشجوی دکترای مهندسی و برنامه‌ریزی حمل و نقل، دانشکده عمران،
دانشگاه علم و صنعت ایران، تهران، ایران
E-mail: jalil@iust.ac.ir.

چکیده:

به دلیل نتایج و فواید قابل توجه مدل‌سازی تصادفات، تلاش‌های بسیاری در زمینه ارائه مدل‌های پیش‌بینی تصادفات در کشورهای مختلف صورت گرفته‌اند. چنین تحقیقاتی در کشور ما از سابقه قابل توجهی برخوردار نیستند و به ویژه در مورد موتورسیکلت‌ها تاکنون تحقیقی انجام نشده است. در این تحقیق با بهره‌گیری از ایجاد ارتباط بین بانک اطلاعات تصادفات تهران و مدل جامع حمل و نقل و ترافیک این شهر، یک پایگاه اطلاعاتی جامع شامل کلیه اطلاعات تردد، طرح هندسی، مشخصه‌های افراد و مشخصه‌های وسایل نقلیه درگیر در تصادفات در مورد ۲۴۱ تقاطع شهر تهران به دست آمده و با استفاده از آن مدل‌هایی در زمینه پیش‌بینی تعداد تصادفات موتورسیکلت‌ها در تقاطع و همچنین برآورد احتمال درگیر بودن موتورسیکلت در تصادفات تقاطع‌های شهر تهران ارائه شده است. از دو روش مدل‌سازی رگرسیون لاجستیک و رگرسیون خطی عمومی در این تحقیق بهره گرفته شده است. اصول مربوط به هریک و کنترل‌های مربوط به صحت کاربرد آنها نیز در مقاله ارائه شده‌اند. در این مقاله همچنین به بررسی اثر پارامترهای مختلف بر احتمال درگیر بودن موتورسیکلت در تصادف، تعداد تصادفات و مقایسه اثرات بین عوامل مختلف پرداخته شده است.

واژه‌های کلیدی: رگرسیون خطی عمومی، رگرسیون لاجستیک، تصادف، موتورسیکلت، تقاطع، مدل‌سازی

۱. مقدمه

حمل و نقل فراهم آورده است. با این حال، در بسیاری از موارد، پژوهشگران به مسائل آماری مربوط به تکنیک مدل‌سازی از قبیل کنترل فرضیات دربرگیرنده هر مدل و حوزه کاربرد صحیح نتایج آن توجه کافی نمی‌کنند و این امر منجر به ارائه نتایج نادرست و شتابزده می‌شود. در این مقاله، علاوه بر ارائه مدل مربوط به پیش‌بینی تصادفات موتورسیکلت، تمامی کنترل‌های مربوط به بررسی صحت مدل‌های رگرسیون نیز مورد اشاره قرار گرفته‌اند. علاوه بر این به بررسی اثر پارامترهای مختلف در افزایش احتمال وقوع تصادفات موتورسیکلت نیز پرداخته شده است.

تصادفات وقایعی کاملاً اتفاقی و غیرقابل پیش‌بینی هستند، و بنابراین پیش‌بینی رخداد یک تصادف منفرد عملی غیرممکن است. در عین حال اگر مجموعه‌ای از تصادفات در طول زمان قابل توجه (حداقل دوره زمانی مدل‌سازی تصادفات عموماً ۳ سال است) در نظر گرفته شوند، با توجه به حاکم بودن قانون اعداد بزرگ [۱] بر نمونه‌های آماری، می‌توان با کاربرد تکنیک‌های آماری، قابلیت پیش‌بینی تعداد تصادفات را به دست آورد. امروزه گسترش نرم‌افزارهای آماری باعث شده تا دستیابی به مدل‌ها سهل شده و این امر پیشرفت بسیاری را در مدل‌سازی در بخش ایمنی

۲. پیشینه پژوهش

به دلیل نتایج و فواید قابل توجه مدل‌سازی تصادفات، تاکنون تلاش‌های بسیاری در زمینه ارائه مدل‌های پیش‌بینی تصادفات صورت گرفته‌اند. ارزیابی و مدل‌سازی تصادفات همواره یکی از پر جاذبه‌ترین مسائل در مهندسی ترافیک بوده است که هم از نظر علمی و تحقیقاتی و هم از نظر فواید عملی و بازدهی مناسب، مورد توجه ویژه قرار داشته است. بازدهی نتایج حاصل از مطالعات مدل‌سازی تصادفات، که نشانگر زیانهای اقتصادی و اجتماعی هنگامت است، جلوگیری از رخداد حتی یک تصادف را به یک امر مهم و مؤثر اقتصادی و اجتماعی تبدیل می‌کند و بنابراین نقش مطالعات مدل‌سازی تصادفات و استفاده از نتایج آنها بسیار بارز است.

از آنجا که تمامی عوامل شامل انسان، راه و وسیله نقلیه به نوعی در تصادفات نقش دارند، پیش‌بینی تعداد تصادفات و مشخصات آنها به دلیل تعدد عوامل مؤثر و تعامل پیچیده آنها کار ساده‌ای نیست. این پیش‌بینی به ویژه در مورد معابر شهری که در آنها عوامل تأثیرگذار بیشتر از راه‌های برون‌شهری است از پیچیدگی بیشتری برخوردار است. به همین دلیل مدل‌های ارائه شده در بخش ایمنی، بیشتر در مورد راه‌های برون‌شهری ارائه شده‌اند و در کمتر تحقیق بین‌المللی به تصادفات موتورسیکلت توجه شده است [۲]. قابل توجه بودن تعداد تصادفات، توجه جهانی را به حساسیت مساله معطوف داشته و سالهای متمادی است که کشورهای توسعه‌یافته در پی تدوین برنامه‌های جامع ایمنی و تدوین استراتژی‌های مؤثر کاهش تصادفات و تلفات ناشی از آن هستند [۳]. تدوین چنین برنامه‌هایی در گام اول مستلزم شناخت و کسب قدرت پیش‌بینی است که هر دو مورد در صورت مدل‌سازی تصادفات قابل تحقق است. به همین دلیل، فعالیت گسترده‌ای در زمینه مدل‌سازی تصادفات طی سالهای اخیر در کشورهای توسعه یافته صورت گرفته و دستاوردهای مهمی به دنبال داشته است [۴]، اما در کمتر تحقیقی به مدل‌سازی تصادفات موتورسیکلت پرداخته شده است [۵].

۳. اطلاعات مورد استفاده در تحقیق

آشکار است که ایجاد تغییر در شرایط شبکه معابر شهر تهران به منظور انجام آزمایش‌های واقعی که منجر به شناخت عوامل مؤثر بر شدت تصادفات موتورسیکلت شود، امری محال است. دلایل

اصلی غیرممکن بودن انجام چنین آزمایش‌هایی، علاوه بر هزینه بسیار زیاد آن، نظم ناپایدار تردد در شبکه معابر شهری است که تغییرات ناچیز در وضعیت آن، موجب تراکم و بی‌نظمی می‌شود و در بسیاری موارد وضعیت تردد در محل مورد نظر را تا ساعتها دچار اختلال می‌سازد. از سوی دیگر، بنابر شرایط حاکم بر فرهنگ تردد، رفتار شهروندان شدیداً متأثر از این امر است که نظارتی بر اعمال خویش احساس کنند یا آن که فارغ از نظارت نهادهایی همچون پلیس راهنمایی و رانندگی، رفتار آزادانه داشته باشند. از آنجاکه ترتیب آزمایش‌های واقعی بدون همکاری پلیس امکان‌پذیر نیست، در نتیجه، صرف برقراری شرایط آزمایشی، منجر به ایجاد وضعیتی می‌شود که با وضعیت عادی تردد تفاوت دارد.

با توجه به امکان نداشتن انجام آزمایش‌های واقعی، در تحقیق حاضر، روش بهره‌گیری از سوابق اطلاعات گردآوری شده برای شناخت عوامل مؤثر در تصادفات موتورسیکلت، به‌عنوان شیوه مناسب برای تحلیل تصادفات انتخاب شده است. در این راستا تمامی امکانات موجود کشوری برای دستیابی به یک پایگاه یکپارچه و هماهنگ از عواملی که احتمال تأثیر آن بر تصادفات موتورسیکلت وجود دارد مورد بررسی قرار گرفته‌اند. باتوجه به بررسیهای به عمل آمده، پایگاه اطلاعاتی تصادفات راهنمایی و رانندگی نیروی انتظامی و اطلاعات مندرج در مدل جامع حمل و نقل و ترافیک شهر تهران به عنوان منابع مناسب اطلاعاتی مرتبط شناسایی، و اطلاعات آن گردآوری و پردازش شده است.

در شهر تهران، اطلاعات تصادفات بر مبنای فرم کام-۱۱۳ و فرم کروکی تصادفات گردآوری می‌شود. اطلاعات تصادفات در یک دوره سه ساله بین سالهای ۱۳۸۱ تا ۱۳۸۳ از این پایگاه اطلاعاتی استخراج شده که در مجموع ۴۸۰۷۵۱ رکورد اطلاعاتی را شامل می‌شده است. در بانک اطلاعاتی مذکور، اطلاعات متعددی در مورد مشخصات تصادف از قبیل شرایط آب و هوایی، افراد و وسایل نقلیه درگیر در تصادف وجود دارد که می‌تواند به عنوان متغیر مستقل در تحلیل وارد شود. اصلی‌ترین متغیر موجود در این بانک اطلاعاتی محل تصادف است که صرفاً در صورت وقوع تصادف در تقاطع‌های شهر، موقعیت دقیق آن قابل شناسایی است. به همین دلیل، اطلاعات مورد استفاده در این تحقیق منحصر به مواردی بوده که تصادف در تقاطع اتفاق افتاده است.

از سوی دیگر، با توجه به این که عوامل طرح هندسی و مشخصات تردد در بانک اطلاعات تصادفات وجود ندارد و این عوامل تأثیر بسیاری در وقوع تصادفات موتورسیکلت دارند،

$$\lambda_i = \sum_j \beta_j x_{ji} \quad (1)$$

در رابطه فوق، x نشان‌دهنده متغیرهای مستقل است و ذاتاً می‌تواند دارای فرم دلخواه و غیرخطی نیز باشد. در واقع خطی بودن مدل صرفاً به پارامترهای β (ضرایب معادله) مربوط می‌شود و x_{ji} ها عملاً می‌توانند هر فرمی از متغیرهای مستقل موردنظر (تابع لگاریتمی، مربع، مکعب یا مثلثاتی) را دارا باشند. اگر با جایگزینی $Y_i = \lambda_i + u_i$ رابطه (۱) بازنویسی شود، رابطه (۲) به دست می‌آید:

$$Y_i = \sum_{j=1}^J \beta_j x_{ji} + u_i \quad (2)$$

که در آن Y_i متغیر وابسته، x_{ji} متغیر مستقل، و u_i تغییرات تصادفی مدل است که باید دارای رابطه خطی با ضرایب مدل باشد. [۶] در حالت کلی، شرط صحت کاربرد رگرسیون خطی، وجود شرایط زیر در مورد u_i (مقادیر باقیمانده‌ها یا خطای معادله) است: [۸]

- مستقل بودن مقادیر باقیمانده‌ها از یکدیگر،
- حاکم بودن تابع توزیع نرمال بر مقادیر باقیمانده،
- صفر بودن مقدار میانگین مقادیر باقیمانده‌ها ($E u_i = 0$) و
- یکسان بودن مقدار انحراف معیار باقیمانده‌ها در مورد هر یک از مشاهدات $\forall_i \leftarrow (Var(u_i) = \sigma^2)$.

وجود خودوابستگی، واریانس خطای نابرابر، و خودهمبستگی از مشکلات متداول رگرسیون خطی است. در صورتی که بین عبارات خطای مربوط به نقاط متفاوت، کوواریانس غیرصفر وجود داشته باشد ($Cov(u_s, u_t) \neq 0$) مدل را دارای خودوابستگی می‌نامند. واریانس خطاها نیز وقتی ثابت نیست که در بین نمونه‌ها ($Var(u_s) \neq Var(u_t)$) باشد. خودهمبستگی به معنی آن است که متغیر وابسته تابعی از مقدار قبلی خود باشد (رابطه ۳):

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \alpha_p y_{t-p} + x_t + u_t \quad (3)$$

خودهمبستگی درجه اول در صورتی وجود دارد که:

$$\alpha_j = 0 \leftarrow \forall j > 1 \text{ و } \alpha_1 \neq 0 \quad (4)$$

کلاً y_t دارای خودهمبستگی درجه p است در صورتی که:

$$\alpha_j = 0 \leftarrow \forall j > p \text{ و } \alpha_p \neq 0 \quad (5)$$

فرم متداول‌تر مدل‌های بررسی تصادفات، مدل‌های خطی عمومی است. شکل عمومی تابع این مدل به صورت رابطه (۶) است:

$$h(\lambda_i) = \sum_j \beta_j x_{ji} \quad (6)$$

دستیابی به مدل مناسب، نیازمند افزودن چنین اطلاعاتی به بانک اطلاعات تصادفات بوده است. با توجه به این که در حال حاضر مدل جامع حمل و نقل و ترافیک شهر تهران شامل اطلاعات شبکه معابر و تردد است، مدل مزبور در محیط $EMME2^2$ که توسط شرکت مطالعات جامع حمل و نقل و ترافیک شهر تهران تهیه شده، مبنای دستیابی به اطلاعات تکمیلی قرار گرفته است.

اطلاعات مزبور صرفاً در مورد تقاطع‌ها از مدل استخراج و در قالب یک بانک اطلاعاتی مستقل تدوین شده است. درکل، بانک اطلاعاتی مزبور شامل ۵۵۱۹ رکورد اطلاعاتی (تعداد تقاطع‌های مدل شده در مورد شهر تهران) و ۵۴ ستون شده است که هر ستون نشان‌دهنده یکی از مشخصه‌های تردد یا طرح هندسی تقاطع بوده است [۶].

به منظور ایجاد ارتباط بین دو بانک اطلاعاتی تصادفات و مدل جامع شهر، از سیستم نمایش اطلاعات جغرافیایی استفاده شده است. به این منظور یک سامانه اطلاعات جغرافیایی در مورد شهر تهران طراحی شده که حاوی سه لایه نقشه، شامل نقشه گیتاشناسی شهر به همراه اسامی معابر، نقشه حاوی کد تقاطع‌ها در محیط $EMME/2$ و لایه تعداد تصادفات تقاطع‌هاست. از سیستم مزبور در جهت ایجاد ارتباط بین بانک‌های اطلاعاتی تصادفات و $EMME/2$ بهره‌گیری شده است. هر نشانی موجود در بانک اطلاعاتی تصادفات بر روی نقشه شبکه معابر شهر تهران جستجو و با استفاده از سیستم نمایش اطلاعات جغرافیایی، کد متناظر آن در محیط $EMME/2$ استخراج شده است. این عمل در مورد ۲۴۱ تقاطع نمونه در شهر تهران صورت گرفته و تقاطع‌های مزبور مبنای مدل‌سازی در تحقیق حاضر قرار گرفته‌اند. این تقاطع‌ها طی سه سال (سالهای ۱۳۸۱ تا ۱۳۸۳) دارای ۱۸۶۷۹ مورد تصادف بوده‌اند که از این بین ۱۷۰۲۴ مورد تصادف دارای تمامی اطلاعات مورد نیاز برای مدل‌سازی بوده و در تحلیل وارد شده‌اند. از بین تصادفات مزبور، در ۱۸۹۶ مورد (۱۱/۱ درصد)، موتورسیکلت در تصادف درگیر بوده و در ۱۵۱۲۸ مورد، موتورسیکلت دخالتی در تصادف به عنوان مقصر یا غیر مقصر نداشته است.

۴. روش مدل‌سازی

مدل‌سازی تصادفات با بهره‌گیری از اطلاعات گردآوری شده، از روشهای مختلفی انجام‌پذیر است. عموماً مدل‌های مورد استفاده در این بخش، مدل‌های رگرسیون خطی یا مشتقات آن هستند. رابطه عمومی مدل رگرسیون خطی به صورت رابطه (۱) است:

مدل‌های لگاریتمی - خطی مدل‌هایی هستند که در آنها لگاریتم متغیر وابسته برحسب تابع خطی از ضرایب ثابت سمت راست معادله و عبارت خطا بیان می‌شود:

$$\ln(y_i) = \sum_{j=1}^J \beta_j x_{ij} + u_i \quad (9)$$

این فرم تابع، شکل خاصی از مدل رگرسیون *Box-Cox* است که در آن $\omega = 0$ فرض شده و کلیه ψ_j ها از قبل به شکل مفروض معلوم فرض شده‌اند [10].

در این تحقیق، مدل‌های خطی عمومی در زمینه پیش‌بینی تصادفات موتورسیکلت به کار گرفته شده‌اند. درمورد متغیرهای وابسته و مستقل نیز توابع انتقال مختلف مورد بررسی قرار گرفته و بهترین فرم متغیرها در مدل وارد شده است.

در تحقیق حاضر، ابتدا به ارائه مدلی در خصوص پیش‌بینی احتمال درگیربودن موتورسیکلت در تصادفات پرداخته شده و در این مورد از مدل رگرسیون لاجستیک بهره گرفته شده است.

رگرسیون لاجستیک را می‌توان یک مدل رگرسیون خطی تعمیم‌یافته در نظر گرفت که توزیع بینم یا دوجمله‌ای بر متغیرهای پاسخ آن حاکم بوده و دارای تابع انتقال لاجیت است. لاجیت *Logit* یا لگاریتم شانس موفقیت^۴ تابع انتقالی است که به صورت زیر تعریف می‌شود [11]:

$$\eta_i = \text{Logit}(\pi_i) = \text{Log} \frac{\pi_i}{1 - \pi_i} \quad (10)$$

اثر کاربرد این تابع انتقال آن است که سمت چپ معادله آن همواره بین صفر و یک است و بنابراین درمورد پیش‌بینی احتمال وقوع یک رخداد (π_i) کاربرد مناسبی دارد. لاجیت، تابع انتقالی است که احتمالات واقع در بازه (۰، ۱) را به کلیه مقادیر اعداد حقیقی می‌نگارد.

مدل رگرسیون لاجستیک با فرض آن که لاجیت احتمال درعوض خود احتمال، از مدل خطی تبعیت نماید و توزیع دوجمله‌ای بر داده‌ها حاکم باشد تعریف می‌شود. در این حالت نیز ضرایب ثابت (β) همانند ضرایب متغیرها در مدل خطی قابل تعبیر هستند، با این تفاوت که سمت چپ معادله درعوض مقادیر پاسخ، لاجیت مقادیر است. بنابراین β_j بیانگر تغییر در مقدار لاجیت احتمال ناشی از ایجاد یک واحد تغییر در متغیر J ام با فرض ثابت بودن سایر متغیرهاست. براین اساس، در صورتی که متغیر J ام با فرض ثابت بودن سایر متغیرها یک واحد تغییر کند، شانس موفقیت‌ها

h در رابطه فوق، یک تابع ارتباط تلقی می‌شود. در واقع در این حالت با استفاده از برخی توابع یکنواخت عمومی (h)، مقدار مورد انتظار متغیر وابسته، به کمیتی که قابل مدل‌سازی توسط رگرسیون خطی باشد تبدیل می‌شود. به عنوان مثال از آنجاکه تعداد تصادفات، کمیتی است که الزاماً نمی‌تواند منفی باشد، در بسیاری از موارد (خصوصاً درمورد تعداد تصادفات کم‌شمار)، به منظور اعمال وضوح در مدل، شکل تابع f به نحوی که نتواند دارای مقدار منفی شود، تغییر داده می‌شود و به عنوان مثال لگاریتم تعداد تصادفات، تابعی خطی از متغیرهای مستقل در نظر گرفته می‌شود. در این مدل رگرسیون لگاریتمی - خطی^۳ ماهیت روابط بین متغیر وابسته و متغیرهای مستقل، مضرب است. امری که در اکثر مدل‌های تصادفات به چشم می‌خورد [7].

توزیع تغییرات تصادفی $u_i = (y_i - \lambda_i)$ در این حالت باید با یکی از توابع متداول توزیع آماری (به عنوان مثال فرم نمایی) منطبق باشد. مدل‌های اطلاعات شمارش شده (مثل پواسون و بینم منفی) در این گروه می‌گنجد. [۲] در کل، مدل‌های مبتنی بر اطلاعات شمارش شده از احتمال دارابودن خودهمبستگی برخوردار نیستند. در مورد توابع انتقال h ، یکی از متداول‌ترین توابع در مدل‌سازی تصادفات، تابع انتقال *Box-Cox* است [8].

این تابع در زمینه ساخت مدل *DRAG* برای شهر کبک ارائه شده است. تابع تغییر شکل *Box-Cox* به شکل رابطه (۷) ارایه می‌شود [۹]:

$$x^{(\psi)} = \begin{cases} \frac{x^\psi - 1}{\psi} & \psi \neq 0 \\ \ln(x) & \psi = 0 \end{cases} \quad \text{در صورتی که} \quad (7)$$

تابع احتمال مزبور، یک سری از توابع انتقال متفاوت را دربرمی‌گیرد و در حالت خاص، شامل تابع مکعب ($\psi = 3$)، مربع ($\psi = 2$)، خطی ($\psi = 1$)، جذری ($\psi = 0/5$)، لگاریتمی ($\psi = 0$)، و معکوس ($\psi = -1$) می‌شود.

فرم مدل رگرسیون *Box-Cox* به صورت رابطه (۸) است:

$$y_i^{(w)} = \sum_{j=1}^J \beta_j x_{ij}^{(\psi_j)} + u_i \quad (8)$$

در حالت ایده‌آل، تمامی پارامترهای w, ψ_j, β_j و $j = 1, 2, \dots, J$ به صورت همزمان برآورد می‌شوند. بنابراین با استفاده از اطلاعات گردآوری شده، نه تنها ضرایب متغیرها، بلکه شکل بهینه تابع نیز مشخص می‌شود.

$$\frac{P}{1-p} = e^{-5.47+1.79SHEDAT+0.13SHAKHE_E+\dots+0.17DAY5} \quad (12)$$

که در آن p احتمال درگیر بودن موتورسیکلت در تصادف است. با توجه به استقلال متغیرهای مستقل مورد استفاده در این مدل‌ها، بررسی اثر هر متغیر مستقل بر شانس احتمال درگیر بودن موتورسیکلت در تصادف تقاطع‌ها امکان پذیر می‌شود. این اثر برابر با $\exp\{\beta_j\}$ است که مقدار آن در جدول ۱ درج شده است. بر این اساس مهم‌ترین پارامتری که بر احتمال درگیر بودن موتورسیکلت در تصادفات اثر می‌گذارد شدت تصادف است که مقدار اثر آن بر شانس احتمال درگیر بودن موتورسیکلت در تصادف، حدود ۶ است. شدت تصادف در تحقیق حاضر به چهار دسته‌ی ترتیبی (یا منظم) به صورت زیر تقسیم شده است.

$$= Y$$

- در تصادف خسارتی غیر چندوسیله‌ای
- در تصادف خسارتی چندوسیله‌ای
- در تصادف جرحی
- در تصادف فوتی

متغیر $SPCL$ بیانگر وجود خط ویژه اتوبوس در معابر منتهی به تقاطع است. مقدار این متغیر در سیستم بانک اطلاعاتی به این صورت تعریف شده که اگر در معابر منتهی به تقاطع، خط ویژه وجود داشته باشد مقدار آن "یک" و در غیر این صورت مقدار آن "صفر" خواهد بود. بر اساس یافته‌های حاصل از مدل‌سازی، وجود خط ویژه در معابر منتهی به تقاطع، باعث می‌شود تا شانس احتمال درگیر بودن موتورسیکلت در تصادفات حدود $1/65$ برابر شود. علت این امر را می‌توان در گسترش خطوط ویژه اتوبوس در سطح شهر و عدم برخورد با موتورسیکلت‌های متخلف که به استفاده غیرمجاز از این خطوط می‌پردازند جستجو کرد.

متغیر $Shakhe - E$ به این صورت تعریف شده که اگر تعداد شاخه‌های ورودی با تعداد شاخه‌های خروجی تقاطع برابر باشد مقدار آن "صفر" و در غیر این صورت "یک" است. در واقع یک بودن متغیر مزبور نشان‌دهنده آن است که بین تعداد شاخه‌های ورودی تقاطع و تعداد شاخه‌های خروجی آن تساوی برقرار نیست. بر اساس یافته‌های حاصل از مدل‌سازی، در اثر یکسان نبودن شاخه‌های ورودی و خروجی تقاطع، شانس احتمال درگیر بودن موتورسیکلت در تصادفات حدوداً $1/2$ برابر می‌شود.

باید در ضریب $\exp\{\beta_j\}$ ضرب شود. بنابراین تابع نمائی ضرایب ثابت $\exp\{\beta_j\}$ ، بیانگر اثر متغیر J ام بر شانس احتمال است [12].

۵. مدل احتمال درگیر بودن موتورسیکلت در

تصادف‌های تقاطع‌های شهر تهران

درگیر بودن موتورسیکلت در یک تصادف و یا عدم دخالت موتورسیکلت در تصادف مزبور، می‌تواند به عنوان یک متغیر تصادفی دوگانه در نظر گرفته شود که در حالت اول مقدار آن برابر یک و در حالت دوم صفر است. چنین رخدادی یک آزمون برنولی است که کسب موفقیت در آن به منزله مقدار "یک" متغیر مربوط و دخیل بودن حداقل یک موتورسیکلت در تصادف مورد بررسی است. مقدار "صفر" این متغیر نیز نشان دهنده آن است که هیچ موتورسیکلتی در تصادف درگیر نبوده است.

جدول (۱)، نشان دهنده مشخصات دو مدل استخراج شده در مورد پیش‌بینی متغیر فوق تحت عنوان احتمال درگیر بودن موتورسیکلت در تصادفات تقاطع‌های شهر تهران و به روش رگرسیون لاجستیک است. علت ارائه دو مدل مختلف در این زمینه آن است که شرط محاسبه اثر متغیرهای مستقل بر متغیر پاسخ، عدم وابستگی متغیرهای مستقل موجود در مدل با یکدیگر است. بررسی صحت این دو مدل نشانگر آن است که در ۹۲ درصد از موارد، کاربرد مدل‌های مزبور منجر به پیش‌بینی صحیح احتمال درگیر بودن موتورسیکلت در تصادفات می‌شود. آزمون معنی دار بودن آماری کاربرد متغیرها در مدل‌نشان دهنده آن است که مقدار آماره X^2 دو در این مورد با ۱۲ درجه آزادی ۳۹۴۳ بوده که بیانگر اعتبار آماری مدل در سطح اطمینان ۹۵ درصد است. با استفاده از ضرایب ارائه شده در این جدول امکان بازنویسی مدل لاجستیک فراهم می‌شود. به عنوان مثال در مورد اول، مدل پیش‌بینی احتمال درگیر بودن موتورسیکلت در تصادفات در تقاطع‌های شهر تهران به صورت زیر است:

$$(11)$$

$$\begin{aligned} \logit(Y) &= \log \frac{p}{1-p} = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n \\ &= -5.47 + 1.79SHEDAT + 0.13SHAKHE_E + \dots + 0.17DAY5 \end{aligned}$$

که در آن Y متغیر وابسته تحلیل است. فرم مدل مزبور را می‌توان تبدیل و به صورت معادله زیر نوشت:

جدول ۱. نتایج مدل سازی احتمال درگیر بودن موتورسیکلت در تصادف تقاطع‌ها

متغیرهای مستقل تحلیل													درصد صحت پیش‌بینی مدل	شماره مدل					
پارامترهای مدل											Log likelihood	Square							
β_0	ضریب ثابت	SHEDAT	SPCL	SHAKHE-E	VOL-CA	SPEED-CA	ncbd	Light	MDN-DOU	Typ 3	VALJ3-1	VAL29-1	VALI6-1	V23-1-7	DAY 3	DAY 4	DAY 5	۱	۲
																		β	Exp (β)
-۵/۴۵		۱/۸۱		۱/۱۰			۷/۴۰	۰/۶۲	-۰/۳۲	۰/۱۸	۰/۴۳		۰/۱۶	-۰/۴۷	۰/۱۸	۰/۱۹	۰/۱۷	۸۹۳۳۸	۸۹۳۳۸
		۵/۹۵		۴/۱۴			۶/۰۱	۰/۸۰	۰/۸۲	۰/۴۸	۱/۳۵		۱/۱۸	۰/۶۳	۱/۱۹	۱/۲۱	۱/۱۸	۰/۳۱	۰/۳۱
-۶/۱۵		۱/۸۱		۰/۲۷		-۰/۱۵			۰/۳۵			۰/۸۷	۰/۱۸	-۰/۴۷	۰/۱۸	۰/۲۱	۰/۱۷	۸۹۳۳۸	۸۹۳۳۸
		۵/۹۵		۱/۳۸		۰/۶۷			۰/۸۱			۲/۱۵	۱/۲۰	۰/۶۲	۱/۲۰	۱/۲۳	۱/۱۹	۰/۳۱	۰/۳۱

ساعت اوج آنها بیشتر از مقدار متوسط است، عموماً در تقاطع‌های خارج از مرکز شهر رخ می‌دهند که احتمال وجود موتورسیکلت در آنها کمتر است.

متغیر *ncbd* متغیری دوگانه است که نشان دهنده واقع شدن تقاطع مورد نظر در محدوده طرح ترافیک و یا خارج از محدوده مزبور است. مقدار "یک" این متغیر نشان‌دهنده آن است که تقاطع مورد نظر در محدوده طرح ترافیک واقع شده است. براساس نتیجه حاصل از مدل‌سازی، چنین حالتی (واقع شدن تقاطع در محدوده طرح ترافیک) منجر به ۱/۶ برابر شدن احتمال درگیر بودن موتورسیکلت در تصادفات می‌شود. این ضریب قابل توجه در افزایش احتمال درگیر بودن موتورسیکلت در تقاطع‌ها در محدوده طرح ترافیک، اثر عدم اعمال ممنوعیت تردد موتور سواران در محدوده طرح مزبور را به وضوح نشان می‌دهد.

متغیر *Light* نشان دهنده چراغ‌دار بودن تقاطع یا غیرچراغ‌دار بودن آن است. از ۲۴۱ تقاطع مورد بررسی، ۷۴ تقاطع مجهز به چراغ راهنما و ۱۶۷ تقاطع بدون چراغ راهنما بوده‌اند. نتایج مدل‌سازی صورت گرفته در این مورد که در جدول ۱ نشان داده شده‌اند مؤید این نتیجه‌اند که کنترل تقاطع با چراغ راهنمایی، باعث کاهش اثر احتمال درگیر بودن موتورسیکلت در تصادفات به مقدار ۰/۸ برابر شود.

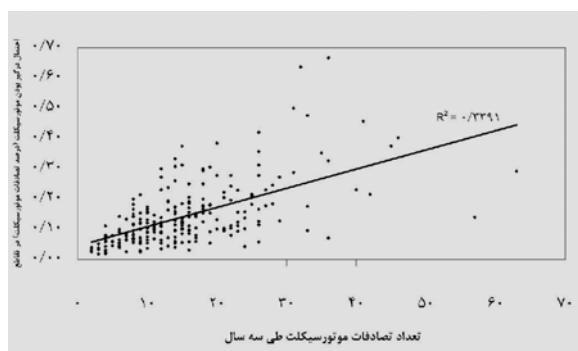
متغیر *MDN_Doub* متغیری است که نشان دهنده وجود داشتن یا نداشتن جان‌پناه وسط در معابر منتهی به تقاطع است. این متغیر یک متغیر دوگانه است که "یک" بودن مقدار آن نشان دهنده این امر است که کلیه معابر منتهی به تقاطع، جداکننده میانی دارند.

متغیر *Vol-Ca* بیانگر دسته‌بندی حجم ترافیک ورودی به تقاطع است. در مورد هر یک از تقاطع‌های تحت بررسی، کل حجم ترافیک ورودی به تقاطع در طی ساعات اوج صبح از کلیه شاخه‌های منتهی به تقاطع محاسبه شده است، سپس برحسب آن که حجم ترافیک کل ورودی به تقاطع از حدود مقدار میانگین تقاطع‌های شهر (۶۰۰۰) وسیله نقلیه معادل سواری در ساعات اوج صبح) بیشتر یا کمتر باشد، به ترتیب کدهای "یک" و "یا" صفر" به متغیر *Vol-Ca* درمورد تقاطع مزبور اختصاص داده شده است. بر اساس یافته‌های حاصل از مدل‌سازی، در صورت بیشتر بودن حجم ترافیک ورودی به تقاطع نسبت به مقدار میانگین، شانس احتمال دخالت موتورسیکلت در تصادفات تقاطع‌ها ۱/۳۸ برابر می‌شود.

متغیر *Speed-Ca* نشان‌دهنده دسته‌بندی سرعت سفر در معابر متصل به تقاطع است. برای محاسبه پارامتر مزبور، ابتدا شاخه‌ای از تقاطع که دارای بیشترین سرعت سفر است شناسایی و سرعت سفر آن در صورتی که کمتر یا برابر با ۴۰ کیلومتر در ساعت باشد، با کد صفر مشخص شده است. به صورت متقابل، وجود کد یک در مورد متغیر *Speed-Ca* نشان‌دهنده آن است که حداقل یک معبر منتهی به تقاطع وجود دارد که سرعت میانگین سفر در ساعات اوج صبح آن بیشتر از ۴۰ کیلومتر در ساعت باشد. چنین حالتی بر اساس یافته‌های حاصل از مدل‌سازی، منجر به افزایش احتمال درگیر بودن موتورسیکلت در تصادفات نمی‌شود، بلکه شانس احتمال درگیر بودن موتورسیکلت را ۰/۸۶ برابر می‌کند. این امر به دلیل آن است که چنین موقعیت‌هایی که سرعت حرکت در

مدل پیش‌بینی تصادفات موتورسیکلت در تقاطع‌های شهر تهران

در مدل‌سازی فراهم نمی‌شود. به همین دلیل، در این بند ابتدا به بررسی اثر متقابل این متغیر و متغیر وابسته تحلیل می‌پردازیم. شکل ۱ ارتباط بین متغیر تعداد تصادفات موتورسیکلت طی سه سال و احتمال درگیر بودن موتورسیکلت در تقاطع را نشان می‌دهد. همان‌طور که در نمودار مزبور مشاهده می‌شود، فرض خطی بودن ارتباط بین این پارامترها تقریباً محقق می‌شود. سعی و خطای بسیاری برای یافتن بهترین تابع انتقال برای متغیر وابسته و ترکیب متغیرهای مستقل به نحوی که بتواند بیشترین ارتباط خطی بین متغیر وابسته و مجموعه متغیرهای مستقل را ایجاد کند به کار رفته و از جمله تابع انتقال Box-cox با کاربرد دامنه وسیعی از ثابت معادله و کاربرد برنامه رایانه‌ای تدوین شده توسط نرم افزار Visual Basic مورد استفاده قرار گرفته است.



شکل ۱. ارتباط بین تعداد تصادفات موتورسیکلت و احتمال درگیر بودن موتورسیکلت در تقاطع

بررسی‌های انجام شده ثابت می‌کنند که در صورت کاربرد تابع انتقال لگاریتمی در مورد متغیر وابسته تحلیل، لازم است که متغیر مستقل نشان‌دهنده احتمال درگیر بودن موتورسیکلت در تصادفات نیز متقابلاً با تبدیل مقادیر توسط تابع مزبور در مدل وارد شود. شکل ۲ نمودار ارتباط بین متغیرهای مزبور در صورت کاربرد تابع انتقال لگاریتمی در مورد مقادیر این متغیرها را نشان می‌دهد. همان‌گونه که در این نمودار مشاهده می‌شود، کاربرد تابع انتقال لگاریتمی در مورد متغیرهای مزبور به میزان زیادی ارتباط خطی بین متغیرها را افزایش می‌دهد.

لازم به ذکر است که متغیر وابسته، تحلیل بین ۲ تا ۶۳ مورد تصادفات موتورسیکلت را در مورد تقاطع‌های بررسی شده نشان داده و نتیجه‌گیری می‌کند که حداقل نیمی از داده‌ها، دارای تعداد تصادفات کمتر از ۱۳ مورد طی سه سال بوده‌اند. مقدار متغیر احتمال درگیر بودن موتورسیکلت در تصادفات تقاطع نیز بین ۰/۰۲ تا

"صفر" بودن متغیر مزبور به این معنی است که حداقل یکی از معابر منتهی به تقاطع جداکننده ندارد. بر اساس نتایج حاصل از تحلیل، اگر تمامی معابر منتهی به تقاطع دارای جداکننده میانی باشند، اثر احتمال درگیری موتورسیکلت در تصادفات حدود ۰/۷ برابر می‌شود.

متغیر $TYP3$ نشان دهنده وجود کد ۳ در مورد نوع درجه‌بندی معابر منتهی به تقاطع در سیستم $EMME2$ است. در واقع، نظر به این‌که کد ۳ در مدل جامع به معابر شریانی درجه دو اطلاق می‌شود، اگر مقدار متغیر $TYP3$ برابر یک باشد، مؤید آن است که یکی از معابر منتهی به تقاطع، از نوع شریانی درجه دو است. عدد صفر در مورد متغیر $TYP3$ نشان دهنده وجود نداشتن معبر از نوع شریانی درجه دو در بین معابر متصل به تقاطع است. بر اساس نتایج مدل‌سازی، "یک" بودن متغیر $TYP3$ باعث می‌شود که احتمال دخیل بودن موتورسیکلت در تصادفات حدود ۰/۸ برابر شود.

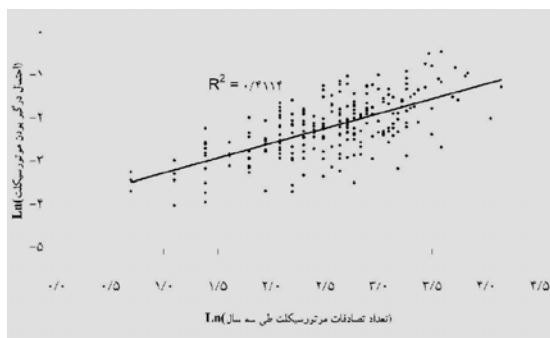
$Val13_1$ نشان دهنده صاف بودن وضع هوا در هنگام تصادفات است. بر اساس نتایج حاصل از مدل‌سازی، صاف بودن هوا، احتمال درگیر بودن موتورسیکلت در تصادفات در تقاطع‌ها را حدود ۱/۵ برابر می‌کند. علاوه بر آن، خشک و بدون مشکل بودن روسازی که اثر آن در متغیر $Val29_1$ وارد شده است، احتمال درگیر بودن موتورسیکلت در تصادفات تقاطع‌ها را حدود ۱/۲ برابر می‌کند. متغیر $V23_1_7$ نشان می‌دهد که وجود نقص در معبر، بیش از آن‌که موجب تشدید تصادفات موتورسیکلت شود بر بروز تصادفات سایر وسایل نقلیه اثر می‌گذارد.

بر اساس نتایج حاصل از مدل‌سازی، وقوع تصادفات در روزهای میانی هفته (در فاصله دوشنبه تا چهارشنبه) احتمال درگیر بودن موتورسیکلت در تصادفات تقاطع‌ها را حدود ۱/۲ برابر می‌کند. متقابلاً نتیجه‌گیری می‌شود که در روزهای آغازین و پایانی هفته احتمال وجود موتورسیکلت در تصادفات $1/(1/2)$ یعنی حدود ۰/۸۳ برابر می‌شود.

۶. مدل پیش‌بینی تعداد تصادفات موتورسیکلت در تقاطع‌ها

بررسی‌های انجام شده درباره‌ی ارتباط بین متغیرهای مختلف قابل وارد کردن در مدل پیش‌بینی تعداد تصادفات موتورسیکلت در تقاطع‌های شهر تهران نشان می‌دهند که بدون کاربرد متغیر احتمال درگیر بودن موتورسیکلت در تصادفات، امکان کسب نتایج صحیح

۰/۶۷ متغیر بوده و حداقل در نیمی از موارد بررسی شده مقدار آن کمتر از ۰/۱۳ بوده است.



شکل ۲. ارتباط بین مقادیر انتقال یافته تعداد و احتمال تصادفات موتور سیکلت

براساس نتایج حاصل از سعی و خطای صورت گرفته در مورد اثر متقابل پارامترها، یکی دیگر از متغیرهایی که باید در تحلیل وارد شود، کل حجم ترافیک ورودی به تقاطع در ساعات اوج است. بررسی ارتباط بین این متغیر و متغیر وابسته تحلیل (تعداد تصادفات طی سه سال در تقاطع) نشان دهنده وجود ارتباط خطی ضعیفی بین این دو متغیر است. بررسی انجام شده نشان می دهد که خطی بودن این ارتباط در صورت کاربرد تابع لگاریتمی در مورد تبدیل مقادیر متغیر وابسته و استفاده از فرم X^α در مورد متغیر حجم ترافیک ورودی به تقاطع با فرض α برابر با ۱/۲، به میزان قابل توجهی افزایش می یابد.

به همراه متغیرهای حجم ترافیک و احتمال درگیر بودن موتور سیکلت در تصادفات، تعدادی متغیرهای گسسته که به صورت مجازی در سیستم تعریف شده اند و نشان دهنده وجود داشتن یا نداشتن یک حالت ویژه هستند نیز قابل منظور کردن در مدل هستند که به کاربرد تابع انتقال برای این متغیرها نیازی نیست. (لازم به ذکر است که چنین متغیرهایی عموماً دارای کدهای صفر یا یک به عنوان وجود داشتن یا وجود نداشتن یک حالت هستند که کاربرد تابع انتقال در مورد آنها مقدور نیست). با کاربرد متغیرهای یاد شده، مدل رگرسیون پیشنهادی به صورت معادله زیر نوشته می شود:

$$(۱۳)$$

$$Ln(y) = B_0 + B_1 Ln(x_1) + B_2 X_2^\alpha + B_3 x_3 + \dots + B_n x_n$$

که در آن،

y = تعداد تصادفات موتورسیکلت طی سه سال در یک تقاطع،

x_1 = احتمال درگیر بودن موتورسیکلت در تصادفات تقاطع،

x_2 = کل حجم ترافیک ورودی به تقاطع برحسب معادل سواری

وسایل نقلیه،

x_3 تا x_n = متغیرهای مستقل دسته بندی شده و

α و B_0 تا B_n = ضرایب ثابت.

در صورتی که دو طرف معادله مزبور به توان e رسانده شود، تعداد تصادفات موتورسیکلت طی سه سال از رابطه زیر قابل محاسبه است:

$$y = e^{B_0 + B_1 m(x_1) + B_2 x_2^\alpha + B_3 x_3 + \dots + B_n x_n}$$

$$\Rightarrow y = e^{B_0} \times e^{B_1 Ln(x_1)} \times e^{B_2 x_2^\alpha} \times e^{B_3 x_3} \times \dots \times e^{B_n x_n}$$

$$y = e^{B_0} \times x_1^{B_1} \times e^{B_2 x_2^\alpha} \times e^{B_3 x_3} \times \dots \times e^{B_n x_n}$$

که به صورت یک تابع مضربی که قادر به پیش بینی تعداد تصادفات است در می آید و با استفاده از نرم افزار آماری SPSS ضرایب B_0 تا B_n آن محاسبه می شوند.

با بررسی ترکیب مختلف متغیرهای قابل در نظر گرفتن در مدل، متغیرهای مذکور مطابق با جدول ۲ به دست می آیند. با کاربرد متغیرهای تشریح شده در جدول ۲ و با فرض $\alpha = 1/2$ ، ضرایب متغیرها برابر با مقادیر جدول ۳ به دست آمده اند.

جدول ۲. شرح متغیرهای به کار برده شده در مدل پیش بینی

تعداد تصادفات موتورسیکلت در تقاطعها

ردیف	شرح متغیر	متغیر
۰	ثابت معادله رگرسیون	ضریب ثابت
۱	احتمال درگیر بودن موتورسیکلت در تصادف	$MOT\ 45 - MG$
۲	حجم معادل سواری ورودی به تقاطع در ساعت اوج	$PCU - In$
۳	بیشترین عرض سواره رو ورودی در بین شاخه های متصل به تقاطع	$Width - A$
۴	در صورتی که در سیستم درجه بندی معابر، معبر دارای درجه بندی بالاتر از شریانی درجه دو (کد ۳) به تقاطع وارد نشود=یک؛ در غیراین صورت= صفر	$Typ\ A - 23$
۵	در صورتی که تمامی شاخه های ورودی به تقاطع یک طرفه باشند=صفر؛ در غیراین صورت=یک	$Twoder - A$
۶	در صورتی که تعداد شاخه های ورودی تقاطع با تعداد شاخه های خروجی آن برابر باشد= صفر؛ در غیراین صورت=یک	$Shakhe - e$

جدول ۳. ضرایب متغیرها در مدل پیش‌بینی تعداد تصادفات

موتورسیکلت در تقاطع‌ها

ردیف	کد	متغیر	ضریب B	انحراف معیار B		فاصله اطمینان ۹۵٪ ضرایب
				حد پائین	حد حداکثر	
۰	-	constant	۳/۶۴۱	۰/۱۵۳	۳/۳۳۹	۳/۹۴۴
۱	Ln(X ₁)	Mot 45-Ln	۰/۶۴۹	۰/۰۴۵	۰/۵۶۱	۰/۷۳۷
۲	(x ₂) ^{1/2}	Volume	۵/۹ × ۱۰ ^{-۶}	۰/۰۰۰	۰/۰۰۰	۰/۰۰۰
۳	(x ₃)	Width_A	۰/۰۱۹	۰/۰۱۱	-۰/۰۰۳	۰/۰۴۰
۴	(x ₄)	TypA-23	۰/۲۰۵	۰/۰۷۱	-۰/۳۴۴	-۰/۰۶۵
۵	(x ₅)	Twodr_A	۰/۱۵۲	۰/۰۶۸	۰/۰۱۷	۰/۲۸۶
۶	(x ₆)	Shakhe_e	-۰/۱۵۰	۰/۰۶۱	-۰/۲۷۱	۰/۰۲۹

۷. کنترل صحت مدل رگرسیون پیش‌بینی تعداد

ماده خام موردنیاز برای کنترل مدل رگرسیون، باقیمانده‌های r_i است که به‌عنوان اختلاف مقادیر مشاهده شده و پیش‌بینی شده تعریف می‌شود:

$$r_i = y_i - \hat{y}_i$$

در معادله فوق y_i پاسخ مشاهده شده و $\hat{y}_i = X_i' \hat{\beta}$ مقادیر پیش‌بینی شده برای مورد i ام است. در شکل ماتریسی، مقادیر پیش‌بینی شده می‌تواند به صورت $\hat{y} = X \hat{\beta}$ نوشته شود. مقادیر پیش‌بینی شده را می‌توان به صورت $\hat{y} = H y$ نوشت که در آن:

$$H = X(X'X)^{-1}X'$$

ماتریس H به‌عنوان مبدل y به \hat{y} عمل می‌کند و به همین دلیل ماتریس *hat* نامیده می‌شود. بر این اساس می‌توان نشان داد که مقادیر پیش‌بینی شده دارای میانگین $E(\hat{y}) = \mu$ و ماتریس واریانس-کوواریانس $Var(\hat{y}) = H\sigma^2$ هستند.

یکی از مواردی که ممکن است منجر به صحت نداشتن استنتاج‌های حاصل از رگرسیون شود، وجود تمایل در بین باقیمانده‌های مدل است. این مورد عموماً توسط آزمون Durbin-Watson کنترل می‌شود. کنترل این آزمون به این صورت است که مقدار آماره مربوط، محاسبه، و با شاخص مندرج در جدول مقایسه می‌شود. در صورتی که آماره محاسبه شده، بزرگ‌تر از شاخص جدول باشد، فرض وجود تمایل در بین باقیمانده‌ها رد می‌شود. عموماً اگر مقدار آماره محاسبه شده در جریان مدل‌سازی، بیش از عدد ۲ باشد، بدون مراجعه به جدول آزمون مزبور می‌توان نتیجه گرفت که تمایل بین باقیمانده‌ها وجود نداشته است [۱۳]. در مورد مدل معرفی شده در این بند، مقدار آماره Durbin-Watson برابر با ۲/۱۹۰ به دست آمده است. با توجه به بیشتر بودن مقدار این آماره از مقدار شاخص، احتمال وجود تمایل باقیمانده‌ها در مدل مزبور منتفی و مدل فاقد اشکال است.

یکی دیگر از فرضیات مدل‌های رگرسیون خطی که نیاز به کنترل آن وجود دارد، فرض صفر بودن میانگین باقیمانده‌ها در مدل است. براساس نتایج حاصل از مدل‌سازی، مقدار باقیمانده‌های مدل بین ۱/۰۳۵ - و ۱/۲۳ قرار داشته و میانگین مقادیر آن دقیقاً صفر بوده است. انحراف معیار مقادیر میانگین نیز ۰/۴۲۵ محاسبه شده است.

مدل ارائه شده در این بخش، دارای R برابر ۰/۷۳، R^2 برابر با ۰/۵۳۴ و R^2 اصلاح شده برابر با ۰/۵۲۲ است که در بین مدل‌های ارائه شده در مطالعات ایمنی، میزان انطباق آن با واقعیت مناسب است.

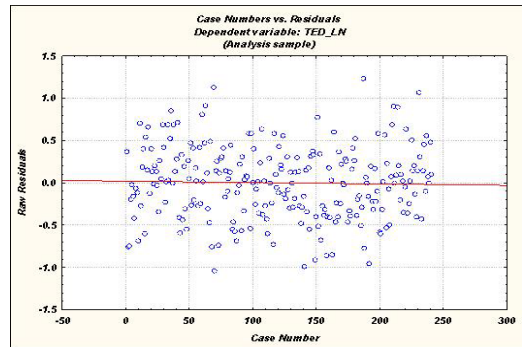
معمولاً به دلیل آن که میانگین مقدار متغیرها بر ضریب پارامتر مربوط تاثیر می‌گذارد و مقایسه اثر پارامترهای مختلف را برحسب ضریب ثابت آنها غیرممکن می‌سازد، از ضریب استاندارد شده که در آن، مقادیر متغیر مستقل مورد نظر از مقدار میانگین متغیر مربوط کم و بر انحراف معیار تقسیم شده و سپس ضریب ثابت معادله برآورد شده است استفاده می‌شود. جدول (۴) مقدار ضرایب استاندارد شده پارامترها را نشان می‌دهد.

بر اساس ارقام مندرج در جدول (۴)، پارامترهایی که دارای بیشترین میزان تاثیر نسبی بر تعداد تصادفات موتورسیکلت در تقاطع‌ها هستند، به‌ترتیب شامل احتمال درگیر بودن موتورسیکلت، درجه‌بندی معبر، حجم ترافیک، تفاوت تعداد شاخه‌های ورودی-خروجی، دوطرفه بودن، و عرض معبر هستند.

جدول ۴. مقادیر استاندارد شده ضرایب ثابت معادله رگرسیون

ردیف	کد متغیر	متغیر	مقدار ضریب استاندارد شده متغیر
۱	Ln (X ₁)	Mot 45-LN	۰/۶۹۹
۲	(x ₂) ^{1/2}	Volume	۰/۱۵۷
۳	(x ₃)	Width_A	۰/۰۸۸
۴	(x ₄)	TypeA_23	-۰/۱۶۱
۵	(x ₅)	Twodr_A	۰/۱۰۵
۶	(x ₆)	Shakhe_e	-۰/۱۱۴

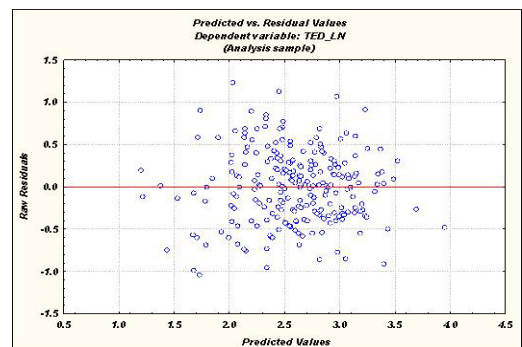
نمودار توزیع مقدار باقیمانده مدل رگرسیون درمورد نمونه‌های مختلف در شکل ۳ نشان داده شده است.



شکل ۳. نمودار مقدار باقیمانده‌های مدل رگرسیون در برابر شماره نمونه

یکی دیگر از روش‌های مفید کنترل صحت مدل، رسم نمودار باقیمانده‌ها در مقابل مقادیر پیش‌بینی شده \hat{Y}_i است. مواردی که در خصوص این نمودار باقیمانده‌ها باید مورد بررسی و توجه قرار گیرند عبارتند از:

- نبود روند خاص در نمودار از قبیل وجود روند در باقیمانده‌های منفی درمورد مقادیر \hat{Y}_i های کوچک و در باقیمانده‌های مثبت درمورد \hat{Y}_i های بزرگ.
 - نبود تمایل در نمودار که نشان‌دهنده غیرخطی بودن ارتباط است.
 - پراکندگی غیریکنواخت باقیمانده‌ها.
- شکل ۴ نمودار مقدار باقیمانده‌ها در برابر مقادیر پیش‌بینی شده توسط مدل را نشان می‌دهد که فاقد اشکال وجود تمایل و هم‌دستگی مذکور در فوق است.



شکل ۴. نمودار مقدار باقیمانده‌ها در برابر مقادیر پیش‌بینی شده توسط مدل

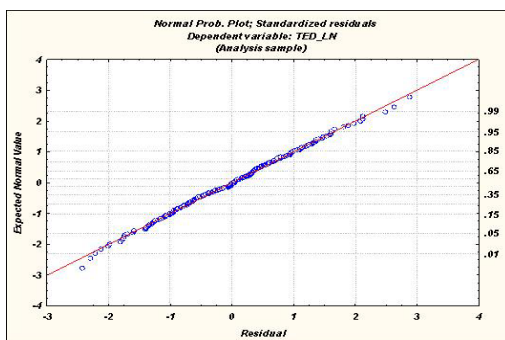
یکی دیگر از فرضیات مدل که نیاز به کنترل دارد، فرض حاکم بودن توزیع نرمال بر باقیمانده‌ها است. روش مفید کنترل معرفی شده در این مورد، رسم نمودار احتمالات است. نمودار احتمالات، نموداری است که مقدار باقیمانده‌ها را در برابر آماره‌های ترتیبی مورد انتظار از توزیع نرمال استاندارد نشان می‌دهد. این نمودار، در صورت نمایش چارک‌های^۶ توزیع، نمودار $Q-Q$ و در صورت رسم صدک‌ها، نمودار $P-P$ نامیده می‌شود.

اولین گام در تهیه چنین نموداری، مرتب نمودن باقیمانده‌ها از کوچک به بزرگ است. بر اساس چنین ترتیبی، i ، $N(i)$ ، i ، $N(i)$ کوچکترین باقیمانده‌ها است و آماره ترتیبی خوانده می‌شود.

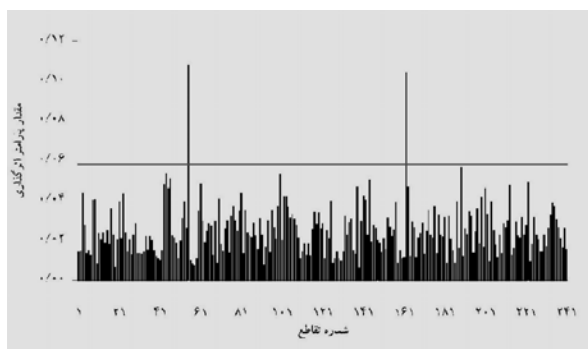
گام بعد، استخراج تصویری از توزیع نرمال درمورد نمونه‌ای با ابعاد n و محاسبه آماره‌ای ترتیبی است که به‌عنوان $Z(i)$ نشان داده می‌شود. مقادیر این آماره ترتیبی، اغلب به‌عنوان رانکیت^۷ نامیده می‌شود. تقریب مناسب i ، $N(i)$ رانکیت درمورد نمونه با ابعاد n به صورت زیر می‌باشد:

$$E(Z(i)) \approx \Phi^{-1} \left[\left(i - \frac{3}{8} \right) / \left(n + \frac{1}{4} \right) \right] \quad (۱۴)$$

که در آن Φ^{-1} نشانگر تابع معکوس توزیع نرمال است. در صورتی که مشاهدات از توزیع نرمال استخراج شده باشند، توزیع آماره ترتیبی مشاهدات، باید مشابه توزیع رانکیت یا آماره ترتیبی مورد انتظار باشد. مشخصاً در صورتی که نمودار آماره ترتیبی در برابر رانکیت‌ها ترسیم شود، باید تقریباً یک خط راست به دست آید. شکل ۵ نشان‌دهنده نمودار $P-P$ باقیمانده‌ها در مورد مدل رگرسیون حاضر است. نمودار مزبور، یک خط راست و بدون هرگونه تقعر یا تحدب است که بیانگر حاکم بودن توزیع نرمال بر مقادیر باقیمانده است.



شکل ۵. نمودار $P-P$ مقادیر باقیمانده‌ها



شکل ۶. نمودار مقدار leverage نمونه‌ها

این دو مورد مربوط به شماره‌های ۵۵ و ۱۶۲ در ترتیب نمونه‌ها بوده‌اند. به منظور بررسی اثر حذف نقاط مزبور از تحلیل، مدل رگرسیون مجدداً و پس از حذف دو نقطه مزبور مورد محاسبه قرار گرفته است. به دلیل آن که نقاط پرت مدل منحصر به دو نقطه از بین ۲۴۱ مورد مشاهده شده، بوده است، این نقاط تأثیر قابل توجهی بر تحلیل نداشته و با حذف آنها نیز مجدداً همان نتایج قبلی در مورد رگرسیون حاصل شود.

۸. نتیجه‌گیری

در تحقیق حاضر، با بهره‌گیری از رگرسیون لاجستیک، مدلی برای پیش‌بینی احتمال درگیر بودن موتورسیکلت در تصادفات تقاطع‌ها ارائه شده است. نتایج مدل مزبور نشان می‌دهد که شدت تصادف، خشک بودن سطح روسازی، وجود خط ویژه اتوبوس، واقع شدن تقاطع در محدوده طرح ترافیک و ... عواملی هستند که بیشترین اثر را بر احتمال درگیر بودن موتورسیکلت در تقاطع‌های شهر تهران دارند. استفاده از پارامتر احتمال درگیر بودن موتورسیکلت در تصادفات و کاربرد مدل‌های خطی عمومی و توابع انتقال مناسب، منجر به ارائه مدلی شده است که توسط آن پیش‌بینی تعداد تصادفات موتورسیکلت در تقاطع‌های شهر امکان‌پذیر می‌شود. بررسی صحت مدل مزبور بیانگر موفقیت تلاش‌های صورت گرفته در زمینه پیش‌بینی تصادفات موتورسیکلت در تقاطع‌ها است. براساس نتایج حاصل، در مورد احتمال دخیل بودن موتورسیکلت در تصادف، درجه‌بندی معبر، حجم ترافیک، تفاوت تعداد شاخه‌های ورودی و خروجی، دوطرفه بودن، و عرض سواره‌رو پارامترهایی هستند که بیشترین تأثیر را بر تعداد تصادفات موتورسیکلت در تقاطع‌ها دارند.

جهت کمی نمودن تحلیل در این مورد می‌توان از آزمون فیلیبن استفاده کرد. بر اساس اطلاعات استخراج شده پایه برای رسم نمودار (۵)، آزمون فیلیبن به برآورد ضریب همبستگی پیرسون بین دو سری اطلاعات به‌کاررفته در مورد محورهای عمودی و افقی نمودار $(Z(i), I(i))$ می‌پردازد. محاسبات انجام شده نشان می‌دهند که ضریب همبستگی پیرسون بین این دوسری اطلاعات، ۰/۹۹۹ است. چون مقدار این ضریب بیشتر از مقدار شاخص آزمون آماری Filliben است که در جدول توزیع مزبور به ازاء ۲۴۰ درجه آزادی ارایه شده است، فرضیه صفر آزمون فیلیبن قبول و مشخص می‌شود که توزیع نرمال بر اطلاعات حاکم است. علاوه بر کنترل‌های ذکر شده تا این بخش، بررسی اثر گذاری نقاط مختلف و شناسایی نقاط پرت بر اساس معیار leverage نیز در این تحقیق مورد نظر قرار گرفته است.

در صورتی که اجزاء ماتریس H که در ابتدای این بند تعریف شده، در نظر گرفته شود، واریانس باقیمانده‌ها اصل ضرب $(1 - h_{ii})$ و σ^2 است. بنابراین تمایل h_{ii} ها به سمت یک، باعث نزدیک شدن باقیمانده‌ها به صفر می‌شود. این امر نشان دهنده آن است که مقادیر پیش‌بینی شده \hat{y}_i مجبور به نزدیک‌تر شدن به مقادیر مشاهده شده y_i می‌شوند. بر اساس این دیدگاه، مقادیر h_{ii} به‌عنوان معیار سنجش اثرگذاری^۱ یا اثر بالقوه مورد i ام شناخته می‌شود. موارد دارای معیار سنجش اثرگذاری زیاد، نیاز به توجه خاص دارند زیرا ممکن است وابستگی قابل توجهی به آنها به‌وجود آید و مدل را از حالت واقعی خارج کنند [13].

عموماً اگر یک مورد خاص دارای h_{ii} بیشتر از $\frac{2p}{n}$ باشد به‌عنوان موردی که دارای معیار اثربخشی زیاد است شناخته می‌شود. p تعداد متغیرهای مستقل به همراه ضریب ثابت و n تعداد کل موارد مشاهده شده است. در این مورد $P = 7$ و $n = 241$ می‌باشد و بنابراین $\frac{2P}{n} = 0/058$ است.

مقادیر leverage محاسبه شده توسط مدل، بین ۰/۰۰۶ تا ۰/۱۰۸ متغیر بوده و میانه مقادیر آن در حدود ۰/۰۲۳ بوده است. شکل ۶ نمودار مقادیر leverage را در برابر شماره تقاطع‌ها نشان می‌دهد. براساس نمودار مزبور، دو مورد از ۲۴۱ مورد تحت بررسی دارای شاخص اثرگذاری نامناسب به عنوان نقاط پرت تحلیل رگرسیون بوده‌اند.

۹. مراجع

10. Guadry, M. (1995a) "Road safety modeling: the DRAG approach and emerging research network", Newsletter of the World Conference on Transport Research Society, 8, 1,2-3.
11. Rodriguez. G. (2001) "wws 509- Generalized linear models", Princeton University.
12. SPSS Inc. (2003) "SPSS 12.0 Command Syntax Reference", Chicago, n.p.
13. Cox, D.R., and Wermuth, N. (1992) "A comment on the coefficient of determination for binary responses", the American Statistician, Vol-46, No. 1, pp. 1-4
14. Ridgeway, G.(2003) "Strategies and methods for prediction", Chapter 6 of the Handbook of data mining, NewJersey, Lawrence-Erlbaum Associates.
۱. بهبودیان، جواد (۱۳۸۳) "آمار و احتمال مقدماتی"، مشهد. دانشگاه امام رضا.
2. Pang, Toh Yen, Umar, R.S.R.; A.A.Azhar. "Relative risk of fatal injury in high-performance-small-motorcycle crashes in Malaysia", Journal of Crash Prevention and Injury Control, Vol.2, No.4, 2001.
۳. قاسم زاده، احسان (۱۳۸۳) "مدل‌های پیش‌بینی تصادفات راه‌های شهری"، مجله پیام پیارک شماره ۷، ص. ۳۸ - ۴۹.
4. Kim, Karl and Boski. J. "Finding fault in motorcycle crashes in Hawaii: environmental, temporal, spatial and human factors", TRB Paper 01-2295.
5. Umar, R.S.R., Mackay, M. and Hills, B. (1999) "Multivariate analysis of motorcycle accidents and the effects of exclusive motorcycle lanes in Malaysia", Journal of Crash Prevention and Injury Control, Vol. 2, No.1, pp. 11-17.
۶. شرکت مطالعات جامع حمل و نقل و ترافیک تهران (۱۳۸۱) "سیستم ایجاد سناریو در محیط EMME2 و تولید جداول گزارش"، گزارش شماره ۴۱۲.
7. OECD (1997) "Road safety principles and models: review of descriptive, predictive, risk and accident consequence models", OCDE / GD (97)/153 Organization for Economic Co-operation and Development-Paris.
۸. آریان‌زاد، میربهادرقلی، ذهبیون، محمد (۱۳۷۲) "احتمالات و آمار کاربردی"، تهران، دانشگاه علم و صنعت ایران.
9. Boosche Vanden F., Wets, G. (2003) "Macro models in traffic safety and the DRAG family: literature review", RA- 2003-08, onderzoekslijn Kennis Verkeersonveiligheid.

پانویس‌ها:

۱- Law of Large Numbers: براساس قضیه ای مشهور در احتمال، به نام قانون اعداد بزرگ یا قانون معدل‌ها، ثابت می‌شود که برای هر $\epsilon > 0$ ، با بزرگ شدن n (ابعاد نمونه تصادفی)، احتمال $p(|\bar{x} - \mu| > \epsilon)$ به سمت صفر میل می‌نماید. به عبارت دیگر هر چه ابعاد نمونه تصادفی بزرگتر شود، شانس نزدیک شدن معدل نمونه (\bar{x} ها) به μ میانگین واقعی جامعه بیشتر می‌گردد.

2- Multimodal equilibrium

3- Log – Linear regression

4 -Log – odds

5 -Odds ratio

۶-Quantiles – "چندک‌ها" مقادیری هستند که از تقسیم توزیع

به چند قسمت مساوی به دست می‌آیند. به عنوان مثال دهک‌های متغیر تصادفی X به وسیله تقسیم یک توزیع به ۱۰ قسمت مساوی بدست می‌آید. معمولاً چارک‌ها را با Q و صدک‌ها را با P نشان می‌دهند.

7-Rankits

8- Leverage