ارزیابی اثرات ایمنی شاخص وضعیت رویه راه بر فراوانی تصادفات خروج از جاده

مهدی اکبری، دانشجوی دکتری راه و ترابری، دانشکده مهندسی عمران، دانشگاه سمنان
غلامعلی شفابخش*، دانشیار، دانشکده مهندسی عمران، دانشگاه سمنان
محمدرضا احدی، استادیار، پژوهشکده حمل‌ونقل، وزارت راه و شهرسازی، تهران

E-mail: shafabakhsh@semnan.ac.ir

دریافت: ۹٤/۰۱/۲۷ – پذیرش: ۹٤/۰٦/۱۵

چکیده

ساخت، نگهداری و بازسازی زیرساخت‌های حمل‌ونقل از جمله ضرورت‌های اصلی ایجاد و توسعه صنعت حمل‌ونقل پایدار به حساب می‌آیند. هزینه‌های گزاف و عملکرد نامناسب جنبه‌هایی مانند ایمنی و امنیت، مشکلاتی را پیش می‌آورند که اعتمادپذیری سیستم حمل‌ونقل را تهدید می‌کنند. مدل‌های پیش‌بینی تصادفات به عنوان ابزار تحلیل ایمنی، نقش به‌سزایی در تحلیل علل وقوع تصادفات ترافیکی دارند. در این مقاله، سعی شده است تا تأثیر خرابی رویه راه بر فراوانی تصادفات خروج از جاده، که جزو فراوان‌ترین و شدیدترین تصادفات در ایران و حتی جهان هستند، ارزیابی گردد. به همین منظور، شاخص وضعیت رویه راه برای ۱۵۰ کیلومتر از محور سمنان به تهران تعیین و با استفاده از فرایندهای GENMOD و NLMIXED در نرم‌افزار SAS، انواع مدل‌های خطی تعمیم‌یافته و غیرخطی رگرسیونی ارائه شدند. پس از ارزیابی نیکویی برازش مدل‌ها و تخمین خطای آنها، مناسب‌ترین مدل انتخاب می‌شود. برای انجام نیکویی برازش، از ابزارهایی مانند انحراف معیار، آماره کای‌دو پیرسون، $-2LL$، AIC، AIC_C و BIC استفاده شده است. همچنین، دو ابزار دیگر شامل میانگین خطای مطلق و جذر میانگین مربع خطاها، اعتبار خروجی مدل را تخمین زدند. در تجزیه باقیمانده‌ها به روش باقیمانده‌های استاندارد شده، میانگین خطاها حدود ۰/۱۰۴۸- و واریانس آنها برابر ۰/۹۲۵۹ تخمین زده شدند که به فرضیه‌های میانگین صفر و واریانس ۱ بسیار نزدیک هستند. نتایج مدل‌سازی نشان می‌دهد که مدل رگرسیون غیرخطی دوجمله‌ای منفی برازش بهتری دارد و ارتباط معنی‌دار بین شاخص وضعیت رویه راه و فراوانی تصادفات خروج از جاده را آشکار می‌کند. تحلیل حساسیت مدل نیز بیانگر آن است که به ازای افزایش هر واحد شاخص وضعیت رویه راه، نرخ فراوانی تصادفات خروج از جاده ۱/۹۳ درصد کاهش می‌یابد.

واژه‌های کلیدی: مدل پیش‌بینی تصادفات، شاخص وضعیت رویه راه، مدل خطی تعمیم‌یافته، مدل رگرسیون غیرخطی

۱. مقدمه

زیرساخت‌های حمل‌ونقل از جمله سرمایه‌های ملی هر کشور محسوب می‌شوند. هرچند ساخت، توسعه، تعمیر و نگهداری این سرمایه هزینه هنگفتی در بر دارد، اما انجام آن برای ایجاد یک صنعت حمل‌ونقلی پایدار یک

ضرورت است. با این وجود، مسائلی مانند وقوع تصادفات، بروز تأخیرات و عمر کوتاه زیرساخت‌ها نیز مطرح هستند که این صنعت را همواره تهدید می‌کنند. علاوه بر هزینه بسیار زیاد، پیامدهای ناشی از معضلات جسمی، روحی و عاطفی تصادفات نیز بسیار مهم هستند و فشار بالایی را به خانواده‌ها تحمیل می‌کنند.

رویه راه از نظر عملکرد تأثیر به‌سزایی بر کیفیت و راحتی سفر کاربران دارد. از این‌رو، اقدامات تعمیر و نگهداری و نیز بازسازی آن (در صورت لزوم) برای حفظ سطح خدمت‌دهی آن در حد مطلوب یک چالش مهم برای مهندسین راه‌وترابری محسوب می‌شود. اما نکته دیگری که در اینجا مطرح است، تأثیر وضعیت ظاهری و فیزیکی رویه بر فراوانی وقوع تصادفات است.

در بین تمام انواع تصادفات، تصادفات خروج از جاده همواره جزو شدیدترین و پرمخاطره‌ترین وقایع ترافیکی بوده‌اند؛ خصوصاً وقتی منجر به واژگونی می‌شوند بروز یک فاجعه دور از انتظار نیست (مورگان و مانرینگ، ۲۰۱۱؛ سازمان راهداری و حمل‌ونقل جاده‌ای، ۱۳۸۸). تصادفات خروج از جاده وقایعی را در بر می‌گیرند که در آن یک خودرو از امتداد مسیر عبوری خود خارج شده و به سمت حاشیه راه انحراف پیدا می‌کند و در نهایت به موانعی مانند حفاظ‌ها (گاردریل، بلوک‌های بتنی نیوجرسی یا حفاظ کابلی)، جدول کنار راه، پایه پل‌ها، تیرهای برق، درختان، ساختمان‌ها، خودروهای پارک شده، عابران و هر آن چه در مسیر آن قرار گیرد اصابت می‌کند.

در راستای آشکارسازی همبستگی بین شاخص وضعیت رویه راه به عنوان یک پارامتر مهم مقوله مدیریت و نگهداری زیرساخت راه از یک سو و از سویی فراوانی تصادفات خروج از جاده، مقاله حاضر سعی دارد تا با ارائه انواع توابع خطی تعمیم‌یافته و غیرخطی مدل‌های رگرسیون پواسونی و دوجمله‌ای منفی، میزان اثرگذاری این پارامتر را بر میزان فراوانی نوع خاصی از شدیدترین تصادفات، که همان خروج از جاده است، تخمین بزنند.

۲. مطالعات پیشین

تحقیقات اولیه در باره تأثیر رویه راه بر نرخ تصادفات به اوایل دهه ۱۹۷۰ میلادی برمی‌گردد. در آن زمان، مقاومت لغزندگی و عوامل دخیل مرتبط با آن بسیار مورد توجه بودند. از این‌رو، اولین مدل‌های تصادفات– وضعیت رویه بر مبنای توابع رگرسیون چندگانه خطی استاندارد (ساده) توسعه داده شدند (مور و هامفریز، ۱۹۷۲؛ رایزنبرگز و همکاران، ۱۹۷٦؛ آیوی و همکاران، ۱۹۷۷). علاوه بر متغیر اصطکاک، کارلافتیس و گولیاس (۲۰۰۲) در مدل خود پارامتر شاخص خدمت‌دهی موجود[1] (PSI) را نیز افزودند تا تأثیر آن بر نرخ کل تصادفات ارزیابی گردد.

بعدها، چن و همکاران (۲۰۱۰) بر تأثیرات وضعیت رویه آسفالتی بر تصادفات ترافیکی تمرکز کردند. آنها با ارائه مدل‌های دوجمله‌ای منفی، متغیرهای عمق شیارشدگی[2] (RD)، شاخص بین‌المللی ناهمواری[3] (IRI) و شاخص خدمت‌دهی موجود (PSI) را مورد ارزیابی قرار دادند. نتیجه این مطالعه بیان می‌کند که مدل‌های عمق شیارشدگی معنی‌داری کافی را برای تمام انواع تصادفات نشان ندادند؛ در صورتی که برعکس، پارامترهای IRI و PSI برازش خوبی را ارائه دادند.

در مطالعات جدیدتر، جیانگ و همکاران (۲۰۱۳ الف) تصادفات دو خودرویی را بر حسب فاکتورهای مهندسی ترافیک و مدیریت روسازی مورد آزمون قرار دادند. آنها دریافتند که در میان تمام پارامترهای رویه، پارامتر PSI تأثیر معکوسی بر شدت تصادفات دارد. به عبارتی، رویه‌های ناهموارتر منجر به تصادفات شدیدتر کمتری می‌شوند. از طرفی، سایر متغیرهای رویه از جمله شاخص خرابی رویه[4] (PDI)، عمق شیارشدگی (RD) و اختلاف عمق شیارشدگی بین چرخ‌های چپ و راست (RD_df)، اثرگذاری قابل توجهی بر تصادفات شدید ندارند.

1- Present Serviceability Index (PSI)
2- Rut Depth (RD)
3- International Roughness Index (IRI)
4- Pavement Distress Index (PDI)

جیانگ و همکاران (۲۰۱۳ ب) به نقش عوامل ترافیکی و مدیریت روسازی بر فراوانی تصادفات نیز پرداختند. از میان تمام متغیرهای این مطالعه، نتایج تأیید کردند که متغیرهای PSI و PDI تأثیر معنی‌داری بر فراوانی تصادفات دارند. برخلاف مدل‌های شدت تصادفات در تحقیق قبلی، مدل‌های فراوانی جدید ارتباط مستقیمی را با ناهمواری راه نشان می‌دهند؛ یعنی ناهمواری کمتر روسازی منجر به کاهش فراوانی تصادفات خواهد شد.

کاوش در مطالعات پیشین آشکار می‌سازد که تحقیقات اندکی درباره تأثیر مشخصات فیزیکی/دینامیک رویه بر موضوع ایمنی راه‌های برون‌شهری، خصوصاً در مورد تصادفات از نوع خروج از جاده، صورت گرفته است. این موضوع مؤید آن است که همچنان یک شکاف علمی در مورد ارزیابی اثرات ایمنی شاخص وضعیت رویه راه بر فراوانی تصادفات خروج از جاده وجود دارد. از این جهت نیز ضرورت انجام تحقیق حاضر تأیید می‌شود.

۳. گردآوری داده‌ها

موقعیت جغرافیایی استان سمنان از دو جهت بسیار حائز اهمیت است. اول این که به عنوان بخشی از کریدور شرق به غرب، نقش به‌سزایی در جابجایی کالا و مسافر دارد و از طرفی، با قرارگیری در امتداد بزرگراه شماره ۱ آسیا (AH1) که نقش احیاگر جاده باستانی ابریشم را دارد، توانسته جایگاه این استان را در صنعت ترانزیت و جابجایی بین‌المللی کالا ارتقا دهد (ویکیپدیا، ۲۰۱۵). وجود خرابی‌های شدید روسازی و وقوع فراوان تصادفات خروج از جاده، علت اصلی انتخاب این محور برای انجام مدل‌سازی و ارزیابی وجود ارتباط احتمالی بین این پارامترها بوده است. داده‌های به‌کار رفته در این تحقیق شامل داده‌های فراوانی تصادفات خروج از جاده به عنوان متغیر وابسته و متغیرهای مستقل در معرض بودن (شامل متوسط سالانه ترافیک روزانه و طول قطعات همگن) و شاخص وضعیت رویه راه می‌شود. در جدول ۱، خلاصه‌ای از توصیف آماری این داده‌ها ارائه شده است. لازم به ذکر است که این توصیفات برای ۵۵ قطعه همگن به طول کل ۱۵۰/۵ کیلومتر تخمین زده شده است.

جدول ۱. توصیف آماری متغیرهای موجود در مدل پیش‌بینی تصادفات خروج از جاده

متغیرها و نمادها	واحد	حداقل	حداکثر	میانه	انحراف معیار استاندارد
متوسط سالانه ترافیک روزانه (AADT)×	خودرو در روز	۹/۵۷	۱۰/۳۸	۹/۸۳	۰/۳۸
طول قطعه همگن (Length)	کیلومتر	۰/۷	۶/۷	۲/۷۴	۱/۵۳
شاخص وضعیت رویه راه (PCI)	درصد	۱۲/۵	۹۹/۳۸	۶۹/۳۵	۳۰/۴
فراوانی تصادفات خروج از جاده (ROR-Accident)	عدد	۰	۶	۱/۱۶	۱/۲۶

× برای سهولت در محاسبات، بر حسب مضربی از ۱۰۰۰ تعیین شده است.

۳-۱. داده‌های تصادفات خروج از جاده

تصادفات خروج از جاده به وقایعی اطلاق می‌شود که در آنها خودروی سرگردان از مسیر خود منحرف شده و به خطوط مجاور و حاشیه راه تخطی می‌کند. این تخطی ممکن است سه نوع پیامد داشته باشد: سوانح غیربرخوردی (مانند واژگونی یا عبور از شیروانی کناره راه)، تصادف با اجسام متحرک در حاشیه راه (مانند عابران یا حیوانات) و برخورد با موانع یا اجسام ثابت در حاشیه راه (مانند حفاظ‌های گاردریلی یا بتنی،درختان، تیر برق، تابلوها و علائم ترافیک و پایه پل‌ها). در تحقیق حاضر، تمام موارد فوق به عنوان تصادفات خروج از جاده شناسایی و برآورد شده‌اند. داده‌های تصادفاتی که از طریق دفتر فاوای پلیس راهور ناجا به دست آمدند مربوط به سال ۱۳۹۰ است، یعنی همان سالی که داده‌های وضعیت

خرابی روسازی گردآوری شدند. در طول این دوره، حدود ۳۷۳ تصادف اتفاق افتاده بود که ۱٦٦ مورد آن شامل تصادفات خروج از جاده در امتداد محور سمنان به تهران است. البته این تصادفات صرفاً محدود به قطعات واقع در حوزه استحفاظی استان سمنان می‌شود.

۳-۲. در معرض بودن

شکی وجود ندارد که بدون پارامتر در معرض بودن[1]، نتایج مدل‌های پیش‌بینی تصادفات مورد تردید خواهند بود زیرا چطور ممکن است بدون وجود آن، تصادف یا سانحه‌ای رخ دهد. به عبارتی، اگر این متغیر برابر صفر فرض شود یا فراخوانده نشود، آنگاه فراوانی تصادفات خروج از جاده باید صفر شود. اینینک و همکاران (۲۰۰۸) اشاره می‌کنند که طول قطعه و میزان تردد در اکثر مدل‌های پیش‌بینی تصادفات به عنوان متغیرهای توصیفی در معرض بودن شناخته می‌شوند. از این‌رو، در تحقیق حاضر نیز این متغیرها به عنوان مولفه‌های در معرض بودن لحاظ شدند. یکی از واحدهای متداول متغیر در معرض بودن، واحد میلیون خودرو- کیلومتر پیموده شده (VKT) در سال[2] است که فرمول آن برای قطعه iام با طول و تردد معین در رابطه (۱) آمده است (هاروود و همکاران، ۲۰۰۰):

$$Exposure_i = \frac{AADT_i \times 365 \times Length_i}{10^6} \qquad (۱)$$

داده‌های تردد از سازمان راهداری و حمل‌ونقل جاده‌ای اخذ شد و پس از پردازش داده‌های برداشت مکانیزه، مقدار تردد روزانه مسیر تعیین گردید. داده‌های تردد شامل تمام انواع خودروهای شخصی، مسافری و باری می‌شود. بزرگی مقیاس تردد بر حسب ۱:۱۰۰۰ لحاظ شده است چون افزایش یا کاهش یک یا چند خودرو در مدل‌های پیش‌بینی تصادفات عملاً منجر به تغییرات بسیار ناچیز در مقادیر پیش‌بینی‌شده فراوانی یا شدت تصادفات خواهد شد. به همین دلیل، از بزرگی واحد تردد کاسته می‌شود تا خروجی مدل‌های پیش‌بینی تصادفات ملموس‌تر باشد.

قطعات کوتاه راه تأثیرات نامطلوبی بر مدل‌های رگرسیون می‌گذارند؛ از این‌رو، قطعات طویل ارجحیت داشته و قابل‌توصیه هستند (میائو و لوم، ۱۹۹۳). محققین پیشنهاد داده‌اند که نرخ تصادفات بایستی برای قطعاتی با طول حداقل ۸۰۰ متر یا بیشتر محاسبه شود تا به مدل‌های قابل‌اعتمادتری دست یافت (رسند و بینکوهال، ۱۹۹۷). طبق این ملاحظات، مسیر ۱۵۰ کیلومتری مورد مطالعه به ۵۵ قطعه همگن تقسیم شدند که مبنای این تقسیم‌بندی همانا ارزیابی یکنواختی وضعیت روسازی برای هر قطعه، یا به عبارتی نحوه تغییرات شاخص وضعیت روسازی در طول مسیر بوده است.

۳-۳. داده‌های وضعیت روسازی

داده‌های مربوط به وضعیت روسازی در این مطالعه به بررسی مشخصات فیزیکی رویه در قالب یکی از شاخص‌های مدیریت و نگهداری روسازی می‌پردازد. شاخص وضعیت رویه راه (PCI)، شاخص خدمت‌دهی موجود (PSI) و رتبه‌بندی خدمت‌دهی موجود[3] (PSR) از جمله این شاخص‌ها هستند که از آن میان شاخص PCI به عنوان معیار متداول در بررسی کیفیت روسازی در ایران به کار گرفته می‌شود. داده‌های خرابی را می‌توان به دو روش دستی یا خودکار به‌دست آورد. با توجه به شرایط غیرممکن برداشت دستی این داده‌ها از جاده‌های برون‌شهری و همچنین دقت بالایی که برداشت خودکار دارد، برداشت داده‌های خرابی به صورت خودکار توجیه‌پذیر می‌شود. مهمترین مزیت و شاید نقطه قوت برداشت خودکار این است که داده‌های خرابی کل قطعات به صورت کاملاً پیوسته و یکپارچه جمع‌آوری می‌شود.

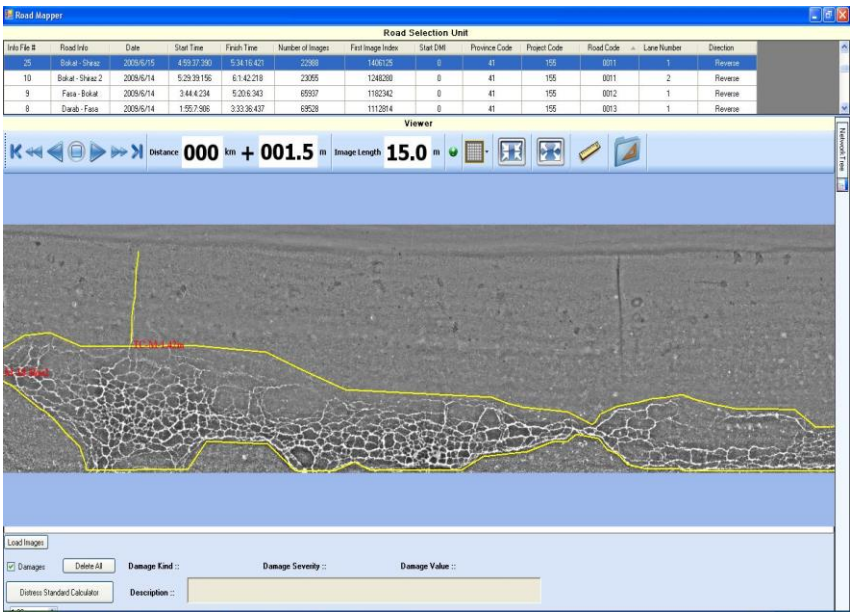
شاخص PCI یک معیار عددی است که شرایط موجود رویه را با لحاظ خرابی‌های مشاهده‌شده در سطح رویه تخمین می‌زند. این شاخص هیچ‌گونه شناختی در مورد

1- Exposure
2- Million Vehicle-Kilometer of Travel (VKT) per year
3- Present Serviceability Rating (PSR)

ظرفیت سازه‌ای یا ابزاری جهت سنجش مقاومت لغزندگی یا ناهمواری در اختیار نمی‌گذارد و صرفاً بیانگر وضعیت خرابی‌های عینی موجود است. مبنای محاسبه مقدار PCI همان استاندارد ASTM D6433-03 است که در آن انواع و شدت خرابی‌ها لحاظ می‌شوند (ASTM D6433، ۲۰۰۳؛ شاهین، ۲۰۰۵).

در این مطالعه، داده‌های خام خرابی رویه با انجام عکاسی پیوسته از سطح راه (در دو نوبت روز و شب) گردآوری شدند. به عبارتی، با استفاده از خودروی مجهز به دوربین‌های عکاسی در شب و روز، عکس‌هایی پیوسته با سطح پوششی معادل ۲ متر (در راستای طولی) در ۳/۸ متر (در راستای عرضی) تهیه شدند. البته هر عکس حدود ۰/۵ متر با عکس قبلی همپوشانی دارد تا بدین‌وسیله بتوان پیوستگی و یکپارچگی عکس‌برداری طولی مسیر را تأمین کرد (شکل ۱). پس از همپوشانی و به‌هم پیوستن تصاویر توسط نرم‌افزار Road Mapper، فرایند اندازه‌گیری خرابی‌ها برای قطعات ۱۰۰ متری صورت می‌پذیرد (شکل ۲). توصیف آماری داده‌های PCI مطالعه حاضر در جدول ۱ ارائه شده است.

شکل ۱. خودروی تجهیزشده جهت برداشت داده‌های خرابی رویه راه



شکل ۲. نمایی از محیط نرم‌افزار Road Mapper و نمونه اندازه‌گیری خرابی رویه راه

۴. روش‌شناسی

عبارت مدل پیش‌بینی تصادفات معمولاً توصیف‌کننده یک مدل چندمتغیره است که در آن داده‌های مربوط به متغیرهای توصیفی به داده‌های تصادفات برازش داده می‌شوند. فرایند کلی ارائه مدل پیش‌بینی تصادفات به شرح زیر است:

- تعیین و گردآوری داده‌های مربوط به متغیرهای مستقل و وابسته
- پیشنهاد فرم‌های ساختاری یا توابع عملکردی مناسب
- اجرای فرایند مدلسازی‌های پیشنهادی
- ارزیابی نیکویی برازش
- تخمین خطای مدل‌ها

در مطالعه حاضر، سه مدل پیش‌بینی تصادفات رگرسیونی شامل دو مدل خطی تعمیم‌یافته (پواسونی و دوجمله‌ای منفی) و یک مدل غیرخطی دوجمله‌ای منفی ارائه شده‌اند.

٤-١. مدل‌های خطی تعمیم‌یافته رگرسیونی

ضرایب مدل‌های پیش‌بینی تصادفات با استفاده از روش‌های حداکثر درستنمایی تعیین می‌شوند که در مدل‌های خطی تعمیم‌یافته[1] (GLMs) یا دیگر مدل‌های غیرخطی می‌توان آنها را یافت (لرد و پرساد، ۲۰۰۰؛ مایرز، ۲۰۰۰؛ مک‌کالاگ و نلدر، ۱۹۸۹). فرم ساختاری GLM مربوط به مدل پیش‌بینی تصادفات در قطعات بزرگراهی شامل حاصل‌ضرب پارامتر در معرض بودن در تابع نمایی متغیرهای مستقل می‌شود. تابع تبدیل خطی‌کننده برای این منظور نیز تابع تبدیل لگاریتمی است (صوالحا و سید، ۲۰۰٦). فرم ساختاری اولیه برای مدل خطی تعمیم‌یافته مطالعه حاضر در رابطه (۲) آمده است:

$$E(Y) = \beta_0 \times Exposure \times \exp \sum_i \beta_i . x_i \qquad (۲)$$

که $E(Y)$ فروانی پیش‌بینی شده تصادفات خروج از جاده، $Exposure$ طبق تعریف رابطه (۱)، x_i دیگر متغیرهای مستقل، اضافه بر متغیرهای در معرض بودن و β_i ضرایب پارامترهای مدل است که بایستی محاسبه شوند.

۴-۱-۱. مدل رگرسیون پواسونی

مدل رگرسیون پواسونی، مدلی از متغیر Y را توسعه می‌دهد که داده‌های آن از توزیع پواسون با میانگین λ پیروی می‌کند. فرض اصلی در مورد این مدل آن است که واریانس داده‌های توزیع پواسونی برابر است با میانگین آن. رابطه (۳) احتمال این که تعداد تصادفات برای i امین مورد برابر y_i باشد و رابطه (٤) اصل برابری واریانس با میانه را برای توزیع پواسونی بیان می‌کند (صوالحا و سید، ۲۰۰٦):

$$P(Y_i = y_i) = f_{Y_i}(y_i;\lambda_i) = \frac{e^{-\lambda_i}.\lambda_i^{y_i}}{y_i!} \qquad (۳)$$

$$Var(y_i) = \lambda_i \qquad (٤)$$

که y_i تعداد تصادفات خروج از جاده برای i امین قطعه مسیر، $P(y_i)$ احتمال این که تعداد y_i مورد تصادفات خروج از جاده در قطعه i رخ دهد و λ_i مقدار مورد انتظار (امید) y_i.

تابع تبدیل خطی‌کننده نیز همان لگاریتم متغیرهای توصیفی است که در رابطه (۵) ارائه شده است (مایرز، ۲۰۰۰):

$$\eta_i = g(\lambda_i) = \log(\lambda_i) = \beta_0 + \sum_{i=1}^{k} \beta_i . x_i \qquad (۵)$$

۴-۱-۲. مدل رگرسیون دوجمله‌ای منفی

مدل‌های پواسونی در اغلب موارد با بیش/کم پراکندگی داده‌ها مواجه هستند یعنی واریانس داده‌ها با میانگین آن مساوی نیست. در این شرایط، مدل‌های رگرسیون دوجمله‌ای منفی توصیه می‌شوند (لرد و مانرینگ، ۲۰۱۰؛ هیلب، ۲۰۱۱؛ وود و همکاران، ۲۰۱۳). ساختار خطی پیش‌بینی‌کننده و تابع تبدیل لگاریتمی برای هر دو مدل پواسونی و دوجمله‌ای منفی یکسان است. اما از لحاظ نظری دو تفاوت عمده وجود دارد:

[1] Generalized Linear Models (GLMs)

- مقدار مورد انتظار (امید) y_i از توزیع دوجمله‌ای منفی پیروی می‌کند
- در مدل دوجمله‌ای منفی، عبارت خطای $\exp(\varepsilon_i)$ افزوده شده است (رابطه ۶).

با توجه به وجود پراکندگی، واریانس مدل دوجمله‌ای منفی به صورت رابطه (۷) در خواهد آمد و تابع توزیع دوجمله‌ای منفی به صورت رابطه (۸) است و تابع جرم احتمال آن نیز با رابطه (۹) تعریف می‌شود (چن و همکاران، ۲۰۱۰).

$$E(y_i) = \lambda_i = \exp(\beta_i.X_i + \varepsilon_i) \quad (٦)$$

$$Var(y_i) = \lambda_i + \lambda_i^2/\phi \quad (٧)$$

$$P(Y_i = y_i) = f_{Y_i}(y_i;\lambda_i) = \frac{e^{-\lambda_i.\exp(\varepsilon_i)}.\lambda_i^{y_i}}{y_i!} \quad (٨)$$

$$P(Y_i = y_i) = f_{Y_i}(y_i;\phi;\lambda_i) = \frac{\Gamma(\phi + y_i)}{\Gamma(\phi)\times\Gamma(y_i+1)}.\left(\frac{\phi}{\phi+\lambda_i}\right)^{\phi}.\left(\frac{\lambda_i}{\phi+\lambda_i}\right)^{y_i} \quad (٩)$$

که $\Gamma(.)$ تابع گاما و ϕ معکوس پارامتر پراکندگی دوجمله‌ای منفی (κ) است.

در مدل دوجمله‌ای منفی، مقادیر ضرایب و پارامتر پراکندگی κ با استفاده از تخمین حداکثر درستنمایی به‌دست می‌آیند که به طور مستقیم نشان‌دهنده چگونگی برازش مدل به داده‌ها خواهند بود.

با وجود این که در مدل رگرسیون پواسونی، پارامتر پراکندگی برآورد نمی‌شود اما برای تخمین نحوه پراکندگی داده‌ها و تعیین ضرورت اجرای مدل‌سازی دوجمله‌ای منفی، می‌توان از پارامتر پراکندگی σ_d (رابطه ۱۰) استفاده کرد که تابعی از مقدار کای‌دو پیرسون و درجه آزادی مدل است (رابطه ۱۱). با بررسی نحوه نوسان این پارامتر در اطراف مقدار واحد می‌توان مشخص کرد که داده‌ها دچار بیش‌پراکندگی یا کم‌پراکندگی هستند. مقادیر بسیار بیشتر از واحد می‌تواند توجیهی برای اجرای مدل‌سازی دوجمله‌ای منفی باشد (مک‌کالاگ و نلدر، ۱۹۸۹).

$$\sigma_d = \frac{Pearson\chi^2}{\text{degree of freedom (DOF)}} \quad (١٠)$$

$$Pearson\chi^2 = \sum_{i=1}^{n}\frac{(y_i - \hat{E}(Y_i))^2}{Var(Y_i)} \quad (١١)$$

که DOF تفاضل تعداد مشاهدات از تعداد متغیرهای مدل، y_i تعداد تصادفات مشاهده شده، $\hat{E}(y_i)$ تعداد تصادفات مورد انتظار (امید) تصادفات در قطعه i ام و $Var(Y_i)$ واریانس متغیر وابسته است.

٤-۲. مدل غیرخطی رگرسیون دوجمله‌ای منفی

روک و کاردوسو (۲۰۱٤) بر این اعتقادند که دقت و نیکویی برازش مدل با داده‌های تصادفات بسیار مهمتر از سادگی فرایند مدل‌سازی در مدل‌های خطی تعمیم‌یافته است؛ حتی اگر فرایند توسعه مدل غیرخطی بسیار پیچیده و طولانی باشد. ساختار مدل غیرخطی پیش‌بینی تصادفات بر آن تأکید دارد که تأثیر متغیرهای مستقل در هم ضرب می‌شوند و دیگر ضرورت ندارد فرم ساختاری مدل به یک تابع تبدیل مشخص پایبند باشد (رابطه ۱۲). یعنی می‌توان انواع توابع را برای کسب بالاترین نیکویی برازش در مدل‌سازی لحاظ کرد.

$$E(y) = f(Exposure)\times g(PCI) \quad (١٢)$$

که $E(y)$ فراوانی مورد انتظار (امید) تصادفات خروج از جاده و $g(PCI)$ فرم‌های ساختاری پیشنهادی برای متغیر توصیفی مستقل پیشنهادی است.

رویکرد مدل‌سازی رگرسیونی غیرخطی به حداکثرسازی تقریبی لگاریتم درستنمایی روی توابع رگرسیونی با تأثیرات تصادفی[1] می‌پردازد و از برخی تکنیک‌های بهینه‌یابی بهره می‌برد. تابع لگاریتم درستنمایی مورد نظر که باید روی داده‌ها با توزیع دوجمله‌ای منفی حداکثرسازی شود را می‌توان به صورت روابط (۱۳) و (۱٤) بیان کرد (موسسه SAS، ۲۰۰۹):

$$L(y,\mu,\kappa) = \sum_i l_i \quad (١٣)$$

$$l_i = \log(f(y_i,\mu_i,\kappa)) = y_i.\log(\kappa.\mu_i) - (y_i+1/\kappa).\log(1+\kappa.\mu_i) + \log\left(\frac{\Gamma(y_i+1/\kappa)}{\Gamma(y_i+1).\Gamma(1/\kappa)}\right) \quad (١٤)$$

1- Random effects

که L تابع لگاریتم درستنمایی، l_i سهم پارامتر در لگاریتم درستنمایی، y_i مقدار پاسخ مدل μ_i میانه تخمین زده شده و κ پارامتر پراکندگی دوجمله‌ای منفی می‌باشد.

۴-۳. تخمین نیکویی برازش، انتخاب فرم ساختاری مناسب و تخمین خطای مدل‌ها

سنجش نیکویی برازش جهت مقایسه مدل‌های (و یا فرم‌های ساختاری) پیش‌بینی تصادفات صورت می‌پذیرد.

۴-۳-۱. نیکویی برازش مدل‌های خطی تعمیم‌یافته رگرسیونی

از میان انواع ابزارهای نیکویی برازش، در مطالعه حاضر از آماره کای‌دو پیرسون و انحراف معیار مقیاس‌بندی شده برای برازش مدل‌های خطی تعمیم‌یافته استفاده شده است. آماره کای‌دو پیرسون در رابطه (۱۱) ارائه شد که در آن، هرچه مقدار این معیار به درجه آزادی مدل نزدیک‌تر باشد (یا به عبارتی، نسبت این دو به ۱ نزدیک‌تر باشد)، برازش بهتر و دقیق‌تری از مدل مورد انتظار خواهد بود. اما در مورد انحراف معیار مقیاس‌بندی شده بر حسب این فرض که ساختار خطاها از توزیع پواسون یا دوجمله‌ای منفی تبعیت می‌کند، دو حالت پیش می‌آید که به ترتیب با استفاده از روابط (۱۵) و (۱۶) مقدار آنها تخمین زده می‌شود. در مقایسه نیکویی برازش دو مدل مختلف، یا بررسی میزان بهبود مدل قبلی با مدلی که پارامتر جدید به آن افزوده شده، هرچه مقدار انحراف معیار مقیاس‌بندی شده مدل کمتر باشد، آن مدل دقیق‌تر و بهتر است (مک‌کالاگ و نلدر، ۱۹۸۹؛ ماهر و سامرزگیل، ۱۹۹۶؛ گریب، ۲۰۰۳).

$$SD = 2\sum_{i=1}^{n} y_i.\ln\left(\frac{y_i}{E(Y_i)}\right) \equiv 2\{L(y_i) - L(E(Y_i))\} \quad (۱۵)$$

$$SD = 2\sum_{i=1}^{n}\left[y_i.\ln\left(\frac{y_i}{E(Y_i)}\right) - (y_i+\kappa).\ln\left(\frac{y_i+\kappa}{E(Y_i)+\kappa}\right)\right] \quad (۱۶)$$

۴-۳-۲. نیکویی برازش مدل‌های غیرخطی رگرسیونی

برای ارزیابی نیکویی برازش مدل‌های غیرخطی رگرسیونی نیز روش‌های متنوع‌تری وجود دارد که دلیل عمده آن دخیل دادن عبارات و عوامل تاوان[1] در ارزیابی برازش‌ها است. در این روش‌ها، عبارات و ضرایبی به عنوان تاوان تعریف می‌شوند تا تأثیر مواردی از قبیل اختلاف تعداد متغیرها (یا در واقع تأثیر تعداد ضرایب) یا تعداد مشاهدات بر تخمین نیکویی برازش مدل‌های غیرخطی اعمال شوند. معیارهای به‌کار رفته در تحقیق حاضر شامل انحراف معیار مقیاس‌بندی شده ($-2LL$)، AIC، AIC_C و BIC می‌باشند. تمام این ابزارها از یک قانون تبعیت می‌کنند؛ هرچه مقدار شاخص نیکویی برازش کمتر باشد، مدل بهتر است.

۴-۳-۳. تخمین خطای خروجی مدل‌ها (تجزیه باقیمانده‌ها)

تخمین خطای خروجی به درک بهتر نحوه انطباق و اعتبار خروجی[2] داده‌ها به مدل کمک می‌کند. این‌گونه ابزارها غالباً برای آزمودن اعتبار خروجی مدل‌ها به کار می‌روند. در این مطالعه، از ابزارهای میانگین خطای مطلق[3] (MAE) و جذر میانگین مربع خطاها (RMSE)[4] جهت تخمین خطای مدل‌های خطی تعمیم‌یافته و غیرخطی پیش‌بینی تصادفات بهره گرفته شده است. هرچه مقدار معیار MAE (رابطه ۱۷) به صفر نزدیک‌تر باشد، نشانگر آن است که مدل پیشنهادی، پیش‌بینی‌های خوبی روی داده‌های مشاهده شده داشته است (اوه و همکاران، ۲۰۰۳). معیار RMSE نیز در واقع ریشه دوم ابزار میانگین مربع خطای پیش‌بینی (MSPE)[5] است که به عنوان یک

1- Penalty Terms
2- External Validation
3- Mean Absolute Error (MAE)
4- Root Mean Squared Error (RMSE)
5- Mean Squared Prediction Error (MSPE)

شاخص متداول ارزیابی خطا محسوب می‌شود. مشابه معیار MAE، هرچه مقدار RMSE (رابطه ۱۸) به صفر نزدیک‌تر باشد، نتایج انطباق خوبی با داده‌ها داشته‌اند.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|O_i - E_i| \qquad (۱۷)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(O_i - E_i)^2} \qquad (۱۸)$$

که n تعداد مشاهدات، O_i مقدار مشاهدات میدانی یا همان تعداد تصادفات مشاهده شده و E_i مقدار محاسبه شده یا همان خروجی مدل پیش‌بینی تصادفات است.

برای تشخیص این که در فرضیات مدل‌های رگرسیونی نقضی وجود دارد یا خیر، لازم است تا تجزیه باقی‌مانده‌ها نیز صورت پذیرد. روش‌های مختلفی برای نشان دادن باقیمانده‌ها وجود دارد. اما در مطالعه حاضر از باقیمانده‌های استانداردشده[1] (رابطه ۱۹) بهره گرفته شده است که در آن، باقیمانده‌ها به وسیله مقادیر میانگین مربعات خطاها (RMS)[2] (رابطه ۲۰) برآورد می‌شود. در تحلیل نتایج آن باید گفت که اگر باقیمانده‌های استانداردشده را d_i بنامیم، برای مدل با برازش قابل قبول باید میانگین باقیمانده‌های استاندارد به سمت صفر و واریانس آنها به سمت ۱ همگرا شوند.

$$RMS = \frac{\sum_{i=1}^{n}(e_i - \bar{e})^2}{n - p - 1} \qquad (۱۹)$$

$$d_i = \frac{e_i}{\sqrt{RMS}} \qquad (۲۰)$$

که e_i باقیمانده (خطا) مشاهده i ام، $\bar{e}$ میانگین باقیمانده‌ها، n تعداد مشاهده‌ها، p تعداد پارامترهای مدل، RMS میانگین مربعات خطاها و d_i باقیمانده استانداردشده مشاهده i ام است.

۵. نتایج مدل‌سازی

۵-۱. مدل‌های خطی تعمیم‌یافته رگرسیونی

برای اجرای مدل‌های خطی تعمیم‌یافته رگرسیونی، شامل پواسونی و دوجمله‌ای منفی، می‌توان از نرم‌افزارهای مختلفی استفاده نمود. یکی از مناسب‌ترین و البته معتبرترین نرم‌افزارهای آماری، نرم‌افزار SAS می‌باشد. فرایند GENMOD در نرم‌افزار SAS، مقادیر حداکثر درستنمایی مربوط به مدل‌های خطی تعمیم‌یافته رگرسیونی را تخمین می‌زند. نکته حائز اهمیت در مورد فرم مدل رگرسیونی آن است که لگاریتم متغیر در معرض بودن به عنوان یک پارامتر آفست در نظر گرفته شده است. از این‌رو، تابع λ_i برای هر دو مدل پواسونی و دوجمله‌ای منفی با افزودن متغیر در معرض بودن بازنویسی گردیده و بنابراین، تابع η_i بایستی به صورت رابطه (۲۱) ویرایش گردد (اشپیگلمن و همکاران، ۲۰۱۱):

$$\eta_i = \log(Exposure) + \beta_0 + \sum_{i=1}^{k}\beta_i.x_i \qquad (۲۱)$$

بر اساس رابطه (۲۱)، داده‌های تصادفات خروج از جاده و متغیرهای مستقل مدل در فرایند GENMOD، مدل‌های رگرسیون پواسونی (PR) و رگرسیون دوجمله‌ای منفی (NBR) ساخته شدند و ضرایب مربوط به متغیرهای در معرض بودن و شاخص وضعیت رویه راه به صورت جدول ۲ ارائه شدند. ضرایب متغیرهای نمایی در فرایند GENMOD بیانگر آن هستند که میزان اثرگذاری تغییرات متغیر شاخص وضعیت رویه راه بر فراوانی تصادفات خروج از جاده چقدر است. این درصد تغییر در تصادفات را می‌توان با محاسبه $(\exp(\beta_i)-1)\times 100\%$ به دست آورد (اشپیگلمن و همکاران، ۲۰۱۱). در جدول ۳ نیز وضعیت مدل‌ها از نظر نیکویی برازش مورد ارزیابی قرار گرفته‌اند که در آنها انحراف معیار و آماره کای‌دو پیرسون و همچنین نسبت آنها به تعداد درجه آزادی به عنوان شاخص‌های نیکویی برازش با هم مقایسه شده‌اند.

1- Standardized Residuals
2- Residual Mean Squares (RMS)

جدول ۲. تحلیل پارامتر و ضرایب مدل‌های خطی تعمیم‌یافته رگرسیونی تصادفات خروج از جاده

نوع GLM	مقدار محاسبه‌شده		خطای استاندارد		حدود اطمینان ۹۵٪ Wald		آماره کای‌دو		مقدار آماره P	
	PR[1]	NBR[2]	PR	NBR	PR	NBR	PR	NBR	PR	NBR
مقدار ثابت	۴/۹۱۲۹	۴/۹۹۸۱	۰/۳۰۸	۰/۴۱۳۹	۴/۳۰۹۲	۴/۱۸۶۹	۵/۵۱۶۵	۵/۸۰۹۲	۲۵۴/۴۵	۱۴۵/۸۴	<۰/۰۰۰۱	<۰/۰۰۰۱
PCI	−۰/۰۰۲	−۰/۰۰۱۳	۰/۰۰۴۱	۰/۰۰۵۵	−۰/۰۱۰۱	−۰/۰۱۲۲	۰/۰۰۶	−۰/۰۰۹۵	۰/۲۴	۰/۰۶	۰/۶۲۰۹	۰/۸۱۱۴
پراکندگی	–	۰/۵۵۲۲	–	۰/۲۶	–	۰/۰۴۲۶	–	۱/۰۶۱۹	–	–	–	–

۱- رگرسیون پواسونی؛ ۲- رگرسیون دوجمله‌ای منفی

جدول ۳. تحلیل نیکویی برازش مدل‌های خطی تعمیم‌یافته

	اندازه نیکویی برازش		نسبت اندازه نیکویی برازش به درجه آزادی	
	رگرسیون پواسونی	رگرسیون دوجمله‌ای منفی	رگرسیون پواسونی	رگرسیون دوجمله‌ای منفی
انحراف معیار	۸۷/۷۱۲۶	۵۳/۵۹۴۴	۱/۶۵۵۰	۱/۰۱۱۲
کای‌دو پیرسون	۱۵۳/۲۲۰۴	۹۱/۵۸۲۵	۲/۸۹۱۰	۱/۷۲۸۰
لگاریتم درستنمایی	−۶۶/۲۸۵۲	−۶۱/۷۱۰۸	–	–

ضریب منفی PCI در جدول ۲ نشان می‌دهد که قطعاتی که دارای مقادیر شاخص وضعیت رویه بیشتری هستند با نرخ تصادفات خروج از جاده کمتری مواجه‌اند. ضریب تابع نمایی PCI در مدل پواسونی برابر ۰/۹۹۸ (یا همان معادل $\exp(-0.002)$) است. این ضریب دلالت بر این دارد که با افزایش ۱ واحد از شاخص وضعیت رویه راه، تعداد تصادفات خروج از جاده به میزان ۰/۲- درصد (معادل $(\exp(-0.002)-1)\times 100\%$) کاهش می‌یابد. در مدل دوجمله‌ای منفی نیز ضریب تابع نمایی PCI برابر ۰/۹۹۸۷۰ به دست می‌آید و نرخ تغییرات به ازای ۱ واحد تغییر در PCI معادل ۰/۱۳- درصد خواهد بود.

تحلیل نتایج نشان می‌دهد که هرچند مقادیر ضرایب، نوسان بالایی ندارند اما مدل دوجمله‌ای منفی با لحاظ پارامتر پراکندگی داده‌ها در محاسبات خود توانسته برازش بسیار بهتری نسبت به مدل رگرسیون پواسونی ارائه کند؛ به طوری که نسبت اندازه نیکویی برازش به درجه آزادی مربوط به ابزار انحراف معیار مدل دوجمله‌ای منفی به مقدار واحد بسیار نزدیک شده است (۱/۰۱۱۲ در مقابل ۱/۶۵۵). در مورد دیگر ابزارهای سنجش نیکویی برازش نیز شرایط مشابهی برقرار است و همگی مؤید برازش بهتر مدل دوجمله‌ای منفی هستند. از این‌رو، مدل رگرسیون دوجمله‌ای منفی به عنوان مدل خطی تعمیم‌یافته برتر انتخاب و مدل آن به صورت رابطه (۲۲) ارائه می‌شود:

$$ROR-Accident_{GLM\ prediction}=e^{4.9981}\times Exposure\times e^{-0.0013\times PCI} \quad (۲۲)$$

۵-۲. مدل غیرخطی رگرسیون دوجمله‌ای منفی

برای تخمین ضرایب پارامترها و همچنین مقدار پارامتر پراکندگی، دو گام به صورت فرایند پیشرونده اعمال می‌شود. ابتدا مدل پایه یا همان فرم ساختاری اولیه تشکیل می‌گردد. در گام دوم، برای تعیین مدل نهایی، لازم است تا چندین فرم ساختاری (یا در واقع همان روابط جبری و ریاضی) پیشنهاد گردد و پس از ارزیابی شاخص‌های نیکویی برازش برای هر مدل، بهترین مدل برگزیده شود. از آنجا که فرض بر آن بوده که مدل پیش‌بینی تصادفات مورد نظر دارای ساختاری غیرخطی است، از فرایند NLMIXED در نرم‌افزار SAS استفاده گردید. این فرایند در مدل‌سازی داده‌های دوجمله‌ای بیش‌پراکنده بسیار کارآمد بوده و عملکرد خوبی دارد. جهت حداکثرسازی تخمین لگاریتم درستنمایی، فرایند NLMIXED می‌تواند از ۷ تکنیک بهینه‌سازی بهره بگیرد که البته از الگوریتم

دوگانه شبه‌نیوتنی[1] به صورت پیش‌فرض استفاده می‌کند. این روش جایگزینی است برای روش نیوتن در مواقعی که روش ژاکوبین[2] (هنگامی که فرایند در جستجوی مقادیر صفر است) یا هزین[3] (هنگامی که فرایند در جستجوی مقادیر اکسترمم است) در دسترس نباشند یا نیازمند محاسبات بسیار پیچیده‌ای در هر تلاش[4] باشند. این روش برای بهینه‌سازی از بردار گرادیان بهره می‌گیرد و برای مسائل بهینه‌سازی با مقیاس متوسط تا نسبتاً بزرگ مناسب است (مؤسسه SAS، ۲۰۰۹).

۵-۲-۱. مدل (فرم ساختاری) اولیه

فرم ساختاری اولیه مدل پیش‌بینی تصادفات در واقع تنها متغیرهای دخیل در پارامتر در معرض بودن را در بر می‌گیرد. این مدل به عنوان یک مدل مرجع برای ارزیابی و مقایسه نیکویی برازش مدل‌های پیشنهادی بعدی است که در آنها پارامترهای جدید به مدل اضافه خواهند شد. با فرض آفست بودن پارامتر در معرض بودن برای مدل پیش‌بینی تصادفات خروج از جاده، فرایند NLMIXED به اجرا در آمد که پس از ۱۵ بار تکرار، ضریب متغیر در معرض بودن برابر ۱۳۵/۳۹ تخمین زده شد. دیگر مقادیر مربوط به نیکویی برازش که مبنای سنجش مدل‌های آتی خواهد بود نیز در جدول ٤ آمده‌اند.

۵-۲-۲. مدل (توسعه‌یافته) نهایی

در این مرحله، متغیر شاخص وضعیت رویه راه به مدل افزوده می‌شود. اما در اینجا علاوه بر فرم نمایی موجود در مدل‌های خطی تعمیم‌یافته، از فرم‌های ساختاری یا توابع ریاضی دیگری نیز بهره گرفته شده است تا با بررسی نتایج مربوط به نیکویی برازش و پارامتر پراکندگی بتوان بهینه‌ترین مدل از نظر برازش با داده‌های فراوانی تصادفات خروج از جاده را انتخاب نمود. علاوه بر توابع پیشنهادی، مقادیر مربوط به ضرایب پارامترها، ابزارهای سنجش نیکویی برازش و پارامتر پراکندگی در جدول ۵ ارائه شده‌اند.

همان‌طور که در جدول ۵ مشاهده می‌شود، با مقایسه مقادیر ابزارهای نیکویی برازش می‌توان دریافت که فرم ساختاری شماره ٤ دارای بهترین برازش روی داده‌های تصادفات خروج از جاده و شاخص وضعیت رویه راه می‌باشد. در این جدول، مقادیر مربوط به ضریب β_{1-1}، در واقع همان ضریب پارامتر در معرض بودن است که برای شرایط فرم ساختاری جدید تخمین زده شده است.

در مدل‌های غیرخطی فوق‌الذکر، شاخص LRT (تفاضل مقادیر $-2LL$ مربوط به مدل پایه و بهبودیافته که برابر ۰/۵ است) در مقایسه با شاخص کای‌دو با یک درجه آزادی (که برای سطح معنی‌داری ۹۰٪ برابر ۲/۷۱ است) دارای مقدار کمتری است که نشان از عدم معنی‌داری در سطح مورد نظر را دارد. اما با این حال، رابطه جدید توانسته یک بهبودی نسبی را در برازش کسب کند. با در نظر گرفتن این موضوع، می‌توان فرضیه صفر را رد کرد و مدل جدید را به عنوان جایگزین قبلی در نظر گرفت. ذکر این نکته حائز اهمیت است که در تفسیر مقادیر LRT، هنگامی که مقدار آن کمتر از حد مورد نظر فرضیه صفر باشد، فرد باید در مورد رد یا پذیرش آن دقت بیشتری داشته باشد (مؤسسه SAS، ۲۰۰۹).

ضریب منفی متغیر شاخص وضعیت رویه راه در اینجا نیز نشان از وجود رابطه معکوس بین آن و فراوانی تصادفات خروج از جاده دارد. ضریب نمایی متغیر برابر ۰/۹۸۰۹ است و درصد تغییرات در نرخ تصادفات خروج از جاده با یک درصد افزایش شاخص وضعیت رویه راه معادل ۱/۹۳- درصد (حاصل عبارت $1.01\times(\exp(-0.0193)-1)\times100\%$) است. با توجه به فرم ساختاری شماره ٤ در جدول ۵ و مقادیر ضریب متغیرها، فرم مدل نهایی مدل رگرسیون غیرخطی دوجمله‌ای منفی به صورت رابطه (۲۳) حاصل می‌شود:

1- Dual quasi-Newton Algorithm
2- Jacobian
3- Hessian
4- Iteration

$$ROR - Accident_{NLMIXED\ prediction} = 8.4118 \times Exposure \times PCI \times e^{-0.0193 \times PCI} \quad (۲۳)$$

جدول ۴. مقادیر معیارهای سنجش نیکویی برازش برای فرم ساختاری اولیه

متغیر	فرم ساختاری	ضریب متغیر	خطای استاندارد	مقدار آماره P	نیکویی برازش				پراکندگی (κ)
					-2LL	AIC	AIC_C	BIC	
در معرض بودن	$E(Y) = \beta_{1-1} \times Exposure$	۱۳۵/۳۹	۲۳/۱۲۱۱	<۰/۰۰۰۱	۱۷۲/۵	۱۷۶/۵	۱۷۶/۷	۱۸۰/۵	۰/۵۵۷۹

جدول ۵. مقادیر خروجی فرایند NLMIXED مدل پیش‌بینی تصادفات خروج از جاده مبتنی بر پارامتر PCI

شماره مدل	فرم ساختاری برای تابع PCI	ضرایب متغیرها				نیکویی برازش				تعداد تلاش	مقدار κ
		β_{1-1}	β_{2-1}	β_{2-2}	β_{2-3}	-2LL	AIC	AIC_C	BIC		
۱	$1+\beta_{2-1}.x$	۱۴۸/۵۱	−۰/۰۰۱۳			۱۷۲/۴	۱۷۸/۴	۱۷۸/۹	۱۸۴/۵	۴۲	۰/۵۵۱۹
۲	$x^{\beta_{2-1}}$	۱۵۶/۱۰	−۰/۰۳۴۸			۱۷۲/۵	۱۷۸/۵	۱۷۹	۱۸۴/۵	۳۴	۰/۵۵۵۴
۳	$\beta_{2-1}.x^{\beta_{2-2}}$	۱۲/۴۹۴۱	۱۲/۴۹۴۱	−۰/۰۳۴۸		۱۷۲/۵	۱۸۰/۵	۱۸۱/۳	۱۸۸/۵	۲۲	۰/۵۵۵۴
۴	$x.e^{\beta_{2-1}.x}$	۸/۴۱۱۸	−۰/۰۱۹۳			۱۷۲	۱۷۸	۱۷۸/۵	۱۸۴	۲۲	۰/۵۴۲۸
۵	$x^{\beta_{2-1}}.e^{\beta_{2-2}.x}$	۴/۴۵۶۰	۱/۲۲۰۹	−۰/۰۲۳۲		۱۷۲	۱۸۰	۱۸۰/۸	۱۸۸	۴۳	۰/۵۴۲۲
۶	$e^{\beta_{2-1}.x}$	۱۴۸/۱۳	−۰/۰۰۱۳			۱۷۲/۴	۱۷۸/۴	۱۷۸/۹	۱۸۴/۵	۳۳	۰/۵۵۲۲
۷	$\beta_{2-1}.e^{\beta_{2-2}.x}$	۱۲/۱۷۰۷	۱۲/۱۷۰۷	−۰/۰۰۱۳		۱۷۲/۴	۱۸۰/۴	۱۸۱/۲	۱۸۸/۵	۲۴	۰/۵۵۲۲
۸	$1/(1+e^{-\beta_{2-1}.x})$	۱۳۵/۳۹	۱/۰۰۰۷			۱۷۲/۵	۱۷۸/۵	۱۷۹	۱۸۴/۵	۱۵	۰/۵۵۷۹
۹	$1/(1+\beta_{2-1}.x)$	۱۴۷/۷۴	۰/۰۰۱۳۵			۱۷۲/۴	۱۷۸/۴	۱۷۸/۹	۱۸۴/۵	۲۵	۰/۵۵۲۵

۵-۳. ارزیابی ساختار خطای مدل‌های خطی تعمیم‌یافته و غیرخطی

مقایسه نتایج مدل غیرخطی با مدل‌های خطی تعمیم‌یافته این نکته را آشکار می‌کند که در شدت اثرگذاری تغییرات مربوط به متغیر وضعیت رویه راه بر فراوانی تصادفات خروج از جاده، اختلاف نسبتاً زیادی وجود دارد (مقدار ۱/۹۳- در مقابل ۰/۲- یا ۰/۱۳- درصد). این وضعیت باعث می‌شود تا جهت قضاوت در مورد اعتبارسنجی ظاهری و دقت کلی مدل‌ها، تحلیل‌هایی روی مقادیر باقیمانده‌ها و خطای مدل‌ها اقداماتی صورت پذیرد.

جدول ٦ اطلاعاتی در مورد تخمین خطای مدل‌ها ارائه کرده است. همان‌طور که قبلاً اشاره شد، ابزارهای MAE و RMSE به عنوان شاخص‌های مطرح انتخاب شدند. طبق فرض نزدیک بودن این مقادیر به صفر، همان‌طور که مشاهده می‌شود، مقدار این شاخص‌ها نسبتاً نزدیک به صفر هستند. با این حال، با انجام فرایند تجزیه باقیمانده‌ها بهتر می‌توان در مورد ساختار باقیمانده‌ها (خطاها) تصمیم‌گیری کرد که در اینجا از روش باقیمانده‌های استانداردشده بهره گرفته شد.

جدول ۶. ساختار خطای ظاهری مدل‌ها و رتبه‌بندی آنها

نوع مدل	ابزار تخمین خطا		رتبه مدل پیشنهادی	
	MAE	RMSE	MAE	RMSE
رگرسیون دوجمله‌ای منفی	۱/۰۹۱۴	۱/۴۹۹۸	۲	۲
غیرخطی	۱/۰۷۶۱	۱/۴۹۷۱	۱	۱

۵-۳-۱. تجزیه باقیمانده‌های مدل رگرسیون خطی تعمیم‌یافته دوجمله‌ای منفی

تحلیل باقیمانده‌ها نشان می‌دهد که میانگین آنها برابر ۰/۱۶۸۲- است که به فرضیه صفر بودن میانگین باقیمانده‌ها بسیار نزدیک است. اما برای تجزیه و تحلیل باقیمانده‌ها از روش باقیمانده‌های استانداردشده استفاده شد که در آن، مقدار RMS برای مدل با ۵۵ نمونه و ٤ متغیر برابر ۲/٤٤۳۳ به‌دست آمد که با جایگذاری در رابطه (۲۰)، مقدار باقیمانده استانداردشده (d_i) برای هر مشاهده حاصل می‌شود. تجزیه باقیمانده‌های استانداردشده نشان می‌دهد که میانگین آنها برابر ۰/۱۰۷٦- و واریانس آنها برابر ۰/۹۲۵۹ برآورد شده‌اند که بیانگر عدم نقض فرضیات مدل رگرسیونی است و در واقع انحراف بین مقادیر مشاهده شده و برازش شده ناچیز است.

۵-۳-۲. تجزیه باقیمانده‌های مدل رگرسیون غیرخطی

در مورد مدل غیرخطی نیز تحلیل داده باقیمانده‌ها نشان می‌دهد که میانگین آنها برابر ۰/۱٦۳٦- است که نشان می‌دهد این مدل دقت بالاتری نسبت به مدل خطی تعمیمی‌یافته دارد. تحلیل باقیمانده‌ها به روش باقیمانده‌های استانداردشده نیز بیانگر آن است که مقدار RMS مدل برابر ۲/٤۳٦۲ است. در تجزیه باقیمانده‌های استانداردشده نیز مشخص می‌شود که میانگین آنها برابر ۰/۱۰٤۸- و واریانس آنها برابر ۰/۹۲۵۹ تخمین زده شده‌اند که نزدیک‌تر بودن آنها به ترتیب به مقادیر صفر و ۱، برازش بهتر مدل غیرخطی را بار دیگر اثبات می‌کند.

۶. خلاصه و نتیجه‌گیری

در تحقیق حاضر، سعی شد تا با بررسی دو مجموعه از مدل‌های رگرسیونی شامل مدل‌های خطی تعمیم‌یافته و مدل غیرخطی، میزان تأثیر پارامتر شاخص وضعیت رویه راه بر فراوانی تصادفات خروج از جاده ارزیابی گردد. ضمن این که مقایسه‌ای نیز روی مدل مناسب پیش‌بینی تصادفات صورت گرفت. به همین منظور، از داده‌های حقیقی گردآوری‌شده حدود ۱۵۰ کیلومتر از مسیر سمنان به تهران، که یکی از پروقوع‌ترین محورها از نظر بروز تصادفات خروج از جاده محسوب می‌شود و در ضمن وضعیت خرابی‌های آن نیز شدید بود، بهره گرفته شود. این داده‌ها شامل داده‌های پارامتر در معرض بودن (شامل متوسط سالانه تردد روزانه و طول قطعات همگن از نظر شاخص وضعیت رویه راه)، مشخصات فیزیکی رویه راه (شامل شاخص وضعیت رویه راه) و همچنین داده‌های تصادفات خروج از جاده می‌شود.

مدل‌های رگرسیونی پواسونی و دوجمله‌ای منفی از جمله مدل‌های متداول و پرکاربرد از مجموعه مدل‌های خطی تعمیم‌یافته هستند که در این مقاله نیز نتایج آنها مورد ارزیابی قرار گرفت. با این وجود، این مدل‌ها همیشه برآورده‌کننده انتظارات پژوهشگران نبوده و برای کسب مدلی با نیکویی برازش بهتر لازم است تا مدل‌های غیرخطی نیز مورد ارزیابی قرار گیرند. از این‌رو برخی فرم‌های ساختاری (مدل ریاضی) پیشنهاد و نتایج آنها مورد سنجش و مقایسه قرار گرفت تا بهترین مدل غیرخطی از نظر نیکویی برازش انتخاب شود.

در نهایت، نتایج مدل‌سازی‌ها را می‌توان به شرح زیر خلاصه کرد:

- در بین مدل‌های خطی تعمیم‌یافته و مبتنی بر قیاس معیارهای ۱) نسبت اندازه انحراف معیار به درجه آزادی، ۲) نسبت اندازه کای‌دو پیرسون به درجه آزادی و ۳) حداکثر درستنمایی، مدل رگرسیون دوجمله‌ای منفی برازش بهتری روی داده‌ها داشته است. بررسی نسبت‌های نیکویی برازش نشان می‌دهد که مدل دوجمله‌ای منفی در مقایسه با مدل پواسونی حدود ۶۰٪ بهبود داشته و به میزان قابل توجه و قابل قبولی به مقدار واحد نزدیک‌تر است (۱/۰۱۱۲ و ۱/۷۲۸ در مقابل ۱/۶۵۵ و ۲/۸۹۱ به ترتیب برای ابزارهای انحراف معیار و کای‌دو پیرسون). مدل نهایی خطی تعمیم‌یافته به شکل رابطه (۲۲) ارائه شده است.

- در بین مدل غیرخطی دوجمله‌ای منفی پیشنهادی، فرم ساختاری شماره ۴ در جدول ۵ دارای بهترین نیکویی برازش از نظر تمام شاخص‌های 2LL-، AIC، AICC و BIC می‌باشد. از بین این معیارها، ابزار BIC از آن جهت که بیشترین تاوان را لحاظ می‌کند به عنوان معیار اصلی برای مقایسه در نظر گرفته می‌شود. مقایسه این مدل با مدل ساختاری اولیه (مدل توسعه‌یافته تنها بر اساس پارامتر در معرض بودن) نشان می‌دهد که فرم ساختاری ۴ بهترین برازش و بهبود را نسبت به مدل پایه داشته است. به طوری که در واقع میزان بیشترین کاهش در معیار - 2LL و همچنین کمترین افزایش را در معیار BIC داشته است. مدل نهایی غیرخطی در رابطه (۲۳) داده شده است.
- برای تشخیص این که کدامیک از مدل‌های خطی تعمیم‌یافته یا غیرخطی بهتر است، می‌توان از ابزارهای MAE و RMSE استفاده کرد. مقایسه مقادیر جدول ۶ نشان می‌دهد که مدل غیرخطی تطابق بهتری با داده‌های اولیه دارد. انجام تجزیه باقیمانده‌ها به روش باقیمانده‌های استانداردشده نشان داد که مدل غیرخطی کمترین خطا را داشته است.
- در نهایت، بر اساس انتخاب رابطه (۲۳) به عنوان مدل بهینه، به میزان اثرگذاری پارامتر شاخص وضعیت رویه راه پرداخته شده است. افزایش شاخص وضعیت رویه راه، یا به عبارتی میزان خرابی‌های رویه، تأثیر معکوس بر فراوانی تصادفات خروج از جاده دارد. این مدل نشان می‌دهد که افزایش کیفیت روسازی موجب کاهش فراوانی تصادفات خواهد شد؛ به طوری که با یک درصد افزایش شاخص وضعیت رویه راه، میزان فراوانی تصادفات خروج از جاده معادل ۱/۹۳ درصد کاهش می‌یابد و این نشانگر اهمیت حفظ سطح خدمت‌دهی روسازی در سطح مطلوب است.
- همان‌طور که مدل‌سازی مقاله تأثیر معنی‌دار وضعیت رویه بر فراوانی تصادفات خروج از جاده را نشان می‌دهد، به عنوان نکته آخر و جهت ارائه راهکار اجرایی توصیه می‌شود که سطح خدمت‌دهی روسازی در حد قابل قبول حفظ گردد و اقدامات تعمیر و نگهداری روسازی حتماً در اولویت اقدامات تصمیم‌گیرندگان باقی بماند.

۷. مراجع

سازمان راهداری و حمل‌ونقل جاده‌ای. ۱۳۸۸. "نمونه‌هایی از تحلیل تصادفات واژگونی در شبکه راه‌ها در دو بخش ملی و استانی". ادارات کل/سازمان‌های حمل‌ونقل و پایانه‌های سراسر کشور.

ASTM D6433, 2003. "Standard Practice for Roads and Parking Lots Pavement Condition Index (PCI) Surveys". American Society for Testing and Materials, USA.

Chan, C. Y., Huang, B., Yan, X. and Richards, S. 2010. "Investigating effects of asphalt pavement conditions on traffic accidents in Tennessee based on the pavement management system (PMS)". J. Adv. Transport. 44(3): 150-161.

Eenink, R., Reurings, M., Elvik, R., Cardoso, J., Wichert, S. and Stefan, C. 2008. "Accident Prediction Models and Road Safety Impact Assessment: Recommendations for using These Tools". Road Infrastructure Safety Protection- Core-Research and Development for Road Safety in Europe, The Netherlands.

Greibe, P. 2003. "Accident prediction models for urban roads". Accid. Anal. Prev. 35: 273-285.

Harwood, D. W., Council, F. M., Hauer, E., Hughes, W. E. and Vogt, A. 2000. "Prediction of the Expected Safety Performance of Rural Two-lane Highways". Federal Highway Administration, FHWA-RD-99-207, USA.

Hilbe, J. M., 2011. "Negative Binomial Regression". Cambridge University Press, UK.

Ivey, D. L., Griffin III, L. I., Newton, T. M., Hunkins, K. C., Blumentritt, C. W. and Lytton, R. L. 1977. "Development of a Wet Weather Safety Index". Federal Highway Administration, FHWATX77-221/F Final Report,USA.

Jiang, X., Huang, B., Yan, X., Zaretzki, R. L. and Richards, S. 2013a. "Two-vehicle injury severity models based on integration of pavement management and traffic engineering factors". Traffic Inj. Prev. 14(5): 544 - 553.

Jiang, X., Huang, B., Zaretzki, R. L., Richards, S. and Yan, X., 2013b. "Estimating safety effects of pavement management factors utilizing Bayesian random effect models". Traffic Inj. Prev. 14(7): 766-775.

Karlaftis, M. G. and Golias, I. 2002. “Effects of road geometry and traffic volumes on rural roadway accident rates”. Accid. Anal. Prev. 34: 357-365.

Lord, D. and Mannering, F. 2010. “The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives”. Transport. Res. Part A: Policy and Practice 44(5): 291-305.

Lord, D. and Persaud, B. N., 2000. “Accident prediction models with and without trend: Application of the generalized estimating equations (GEE) procedure”. Transport. Res. Board: Transport. Res. Rec. 1717: 102-108.

Maher, M. and Summersgill, I. 1996. “A comprehensive methodology for the fitting of predictive accident models”. Accid. Anal. Prev. 28(3): 281-296.

McCullagh, P. and Nelder, J. A. 1989. “Generalized Linear Models”. Chapman and Hall/CRC Press, London.

Miaou, S. P., and Lum, H. 1993. “Modeling vehicle accidents and highway geometric design relationships”. Accid. Anal. Prev. 25(6): 689-703.

Moore, A. B. and Humphreys, J. B. 1972. ”A Study of Pavement Skid Resistance at High Speed and at Locations Shown to be Focal Points of Accidents”. National Technical Information Service, RP-7-FR Final Report, USA.

Morgan, A. and Mannering, F. L. 2011. “The effects of road-surface conditions, age, and gender on driver-injury severities”. Accid. Anal. Prev. 43: 1852-1863.

Myers, R. 2000. “Classical and Modern Regression with Applications”. Duxbury Press, CA, USA.

Oh, J., Lyon, C., Washington, S., Persaud, B. and Bared, J. 2003. “Validation of FHWA crash models for rural intersections”. Transport. Res. Board: Transport. Res. Rec. 1840: 41-49.

Resende, P. T. V. and Benekohal, R. F. 1997. “Effects of Roadway Section Length on Accident Model”. ASCE, Proc. of Traffic Congestion and Traffic Safety in the 21st Century Conference, USA.

Rizenbergs, R. L., Burchett, J. L., Napier, C. T. and Deacon, J. A. 1976. “Accidents on rural interstate and parkway roads and their relation to pavement friction”. Transport. Res. Board: Transport. Res. Rec. 584: 22-36.

Roque, C. and Cardoso, J. L. 2014. “Investigating the relationship between run-off-the-road crash frequency and traffic flow through different functional forms”. Accid. Anal. Prev. 63: 121-132.

SAS Institute Inc. 2009. “SAS/STAT® 9.2 User’s Guide”. 2nd Edition, Cary, NC, USA.

Sawalha, Z. and Sayed, T. 2006. “Traffic accident modeling: Some statistical issues”. Can. J. Civil Eng. 33(9): 1115-1124.

Shahin, M. Y. 2005. “Pavement Management for Airports, Roads, and Parking Lots”. Springer, New York.

Spiegelman, C. H., Park, E. S. and Rilett, L. R. 2011. “Transportation Statistics and Microsimulation”. Chapman and Hall/CRC Press, USA.

Wikipedia, 2015. “Asian Highway 1 (AH1)”. Wikipedia, the free encyclopedia. https://en.wikipedia.org/wiki/AH1.

Wood, A. G., Mountain, L. J., Connors, R. D., Maher, M. J. and Ropkins, K. 2013. “Updating outdated predictive accident models”. Accid. Anal. Prev. 55: 54-66.