تعیین عوامل مرتبط با تعداد تصادفات بر حسب شدت با استفاده از رگرسیون پواسون- لگ نرمال چند متغیره بیزی

ایرج کاظمی[۱]، مهدی تذهیبی[۲]، سمیه مومنیان[۳]، حسین حق‌شناس[۴]

چکیده

مقدمه: برازش مدل‌های تک متغیره در بسیاری از تحقیق‌های گذشته برای تحلیل داده‌های تصادف به کار رفته است. با توجه به آن که متغیر شدت در این مطالعه‌ها می‌تواند بیش از یک سطح باشد، از این رو در این مقاله مدل پواسون- لگ نرمال چند متغیره برای مدل‌سازی تعداد تصادفات بر حسب شدت استفاده شد. اگرچه برای مقایسه از مدل‌های تک متغیره نیز استفاده شد.

روش‌ها: استنباط آماری پارامترهای مدل توسط رهیافت بیز و استفاده از روش نمونه‌گیری گیبز و الگوریتم متروپلیس- هستینگز انجام شد. داده‌های تصادف مربوط به تقاطع‌های شهر اصفهان بود.

یافته‌ها: اثر حجم کل ترافیک بر تصادفات خسارتی در تمامی مدل‌ها معنی‌دار به دست آمد، اما اثر حجم کل ترافیک چپگردها بر تصادفات جراحتی و فوتی تنها در مدل پواسون تک متغیره معنی‌دار شد. بنابراین با فرض ثابت نگه‌داشتن بقیه متغیرهای توضیحی، انتظار می‌رود که افزایش حجم کل ترافیک باعث افزایش تصادفات خسارتی شود. همچنین افزایش حجم کل ترافیک چپگردها نیز باعث افزایش تصادفات جراحتی و فوتی در تقاطع‌ها می‌شود. به طور دقیق‌تر بر اساس مدل پواسون- لگ نرمال چند متغیره، افزایش ۱۰۰۰ وسیله نقلیه در متوسط حجم کل ترافیک باعث افزایش ۳۱ درصدی تصادفات خسارتی می‌شود. بنابراین با کاهش حجم کل ترافیک، کاهش هزینه تصادفات در دراز مدت پیش‌بینی می‌شود که بسیار مقرون به صرفه می‌باشد.

نتیجه‌گیری: با وجود بیش‌پراکنش در دو سطح شدت و وجود همبستگی بین این دو سطح، مدل پواسون- لگ نرمال چند متغیره برازش بهتری را نسبت به بقیه مدل‌ها ارایه می‌دهد.

واژه‌های کلیدی: استنباط بیزی، روش مونت کارلوی زنجیر مارکوفی، توزیع‌های پسین شرطی کامل، شدت تصادفات، ایمنی تقاطع‌ها

نوع مقاله: *پژوهشی*

دریافت مقاله: ۹۰/۱۰/۲۷ *پذیرش مقاله: ۹۱/۲/۲۶*

مقدمه

مرگ و جراحت‌های ناشی از تصادفات یکی از چالش‌های سلامت عمومی در سرتاسر جهان است (۱). به طور کلی در تصادفات وسایل نقلیه در سرتاسر جهان سالانه ۱/۲ میلیون نفر کشته و ۵۰ میلیون نفر مجروح می‌شوند (۲). قابل توجه بودن ارقام بالای تصادفات، توجه جهانی را به حساسیت مسأله معطوف داشته است و سال‌های متمادی است که فعالیت گسترده‌ای در زمینه مدل‌سازی تصادفات در کشورهای توسعه‌یافته صورت می‌گیرد و دستاوردهای مهمی به دنبال داشته است (۵-۳). داده‌های تصادف معمولاً بر حسب فراوانی‌های تصادف همراه با سطوح شدت آن جمع‌آوری می‌شوند. سطوح شدت مانند خسارت مالی، جراحت و آسیب

۱- استادیار، گروه آمار، دانشکده آمار، دانشگاه اصفهان، اصفهان، ایران

۲- استادیار، گروه آمار زیستی و اپیدمیولوژی، دانشکده بهداشت، دانشگاه علوم پزشکی اصفهان، اصفهان، ایران (نویسنده مسؤول)

Email: tazhibi@hlth.mui.ac.ir

۳- دانشجوی کارشناسی ارشد، کمیته تحقیقات دانشجویی، گروه آمار زیستی و اپیدمیولوژی، دانشکده بهداشت، دانشگاه علوم پزشکی اصفهان، اصفهان، ایران

۴- استادیار، گروه مهندسی عمران، دانشکده حمل و نقل، دانشگاه اصفهان، اصفهان، ایران

کم، جراحت و آسیب زیاد و مرگ است.

داده‌های تصادف بر حسب شدت اساساً در ماهیت، چند متغیره هستند، چون عامل‌های پنهان یا متغیرهای از قلم افتاده برای یک بخش مشخص از جاده یا تقاطع در تعداد تصادفات برای تمام سطوح به طور یکسان تأثیر دارند (۷، ۶). مدل داده‌های تعداد تصادفات می‌توانند به دو دسته کلی تقسیم شوند. این دو دسته شامل ۱- مدل‌های مرسوم مانند پواسون تک متغیره (Univariate Poisson و یا UVP) و مدل‌های مرتبط با آن مانند مدل دو جمله‌ای منفی تک متغیره (Univariate negative binomial یا UVNB) ۲- مدل‌های واقع بینانه‌تر مانند پواسون چند متغیره (Multivariate Poisson یا MVP)، پواسون- لگ نرمال چند متغیره (Multivariate Poisson-lognormal و یا MVPLN) و دو جمله‌ای منفی چند متغیره (MVNB یا Multivariate negative binomial) می‌باشند (۸-۶).

مدل‌های تک متغیره تعداد تصادفات وسایل نقلیه را در سطوح مختلف شدت به طور جداگانه برآورد می‌کند که نتیجه آن نادیده گرفتن ساختار همبستگی بین سطوح شدت است (۶، ۴). دخالت ندادن ساختار همبستگی این پیشامدهای چند متغیره، برآورد واریانس ضرایب رگرسیونی را تحت تأثیر قرار می‌دهد؛ در صورتی که در مدل‌های چند متغیره به حساب آوردن این همبستگی باعث افزایش دقت برآوردها می‌شود (۹). معمولاً در داده‌های تصادف بیش پراکنش وجود دارد، چون مقادیر صفر مشاهده شده در داده‌ها از مقادیر صفر که می‌تواند توسط مدل پواسون تولید شود، بیشتر است (۱۱، ۱۰).

مدل رگرسیون پواسون چند متغیره توسط Tsionas گسترش پیدا کرد تا اثر کواریانس‌های مختلف روی داده‌های تصادف بر حسب شدت ارزیابی شود (۱۲)، اما این مدل مشکل بیش پراکنش را که اغلب در داده‌های تصادف وجود دارد، در تحلیل‌ها به حساب نمی‌آورد. همچنین تنها ساختار همبستگی مثبت و یکسان بین داده‌ها را می‌پذیرد. مدل رگرسیون دو جمله‌ای منفی چند متغیره می‌تواند مشکل بیش پراکنش در داده‌ها را حل کند، اما این مدل نیز نمی‌تواند ساختار همبستگی منفی بین داده‌ها را بپذیرد (۴). مدل رگرسیون پواسون- لگ نرمال چند متغیره توسط Chib و Winkelmann گسترش پیدا کرد (۱۳). این مدل می‌تواند به عنوان جایگزین مناسبی برای مدل MVP و MVNB در تحلیل داده‌های تصادف چند متغیره به کار رود. دلیل آن به حساب آوردن بیش پراکنش داده‌ها و به کارگیری ساختار همبستگی کامل مشاهده‌ها است (۶).

برازش مدل رگرسیونی MVPLN توسط روش‌های متداول برآوردیابی امکان‌پذیر نیست و نیاز به نرم افزارهای آماری پیشرفته دارد (۱۴). در عمل استفاده از روش‌های بیزی می‌تواند کمک شایانی به استنباط آماری پارامترهای مدل کند. بنابراین در این مطالعه از روش شبیه‌سازی مونت کارلوی زنجیر مارکوفی (Markov chain monte carlo و یا MCMC) تحت قالب بیزی استفاده شد تا مشکلات محاسباتی حل انتگرال‌های چندگانه در تابع درست‌نمایی حاشیه‌ای از بین برود (۱۷-۱۵).

در این تحقیق متغیرهای مربوط به حجم ترافیک تقاطع‌ها را با فراوانی تصادفات در دو سطوح شدت در نظر گرفته شد و با استفاده از مدل پواسون- لگ نرمال چند متغیره تحلیل شد. داده‌های مورد نظر مربوط به تقاطع‌های شهر اصفهان به مدت ۶ ماه می‌باشد. برای برآورد پارامترهای مورد نظر نیز از روش نمونه‌گیری گیبز و الگوریتم متروپلیس- هستینگز استفاده شد که توسط نرم‌افزار Open BUGS نسخه ۳.۲.۱ قابل اجرا است.

روش‌ها

برای یک مجموعه داده‌ها در n مکان که در هر مکان شدت تصادفات به k طبقه تقسیم می‌شود، بردار مشاهدات به صورت ($y_i = y_{i1}, y_{i2}, \ldots, y_{iK}$) تعریف می‌شود. در آن y_{ik} تعداد تصادفات در iامین مکان و در k سطح شدت است. بنابراین y یک ماتریس n در k بعدی است. تعداد تصادفات در هر یک از n مکان از یکدیگر مستقل و در داخل طبقات به هم وابسته هستند. توزیع y_{ik}ها هر کدام به صورت

$$f(y_{ik} \mid \lambda_{ik}) = \frac{\lambda_{ik}^{y_{ik}} e^{-y_{ik}}}{y_{ik}!} \quad \begin{array}{l} i = 1,\ldots, n \\ k = 1,\ldots, K \end{array}$$

فرض می‌شود. میانگین توزیع پواسون را به صورت تابعی از متغیرهای توضیحی در نظر گرفته می‌شود.

www.mui.ac.ir

$$\lambda_{ik} = \exp(x_i \beta_k + \varepsilon_{ik})$$

J تعداد متغیرهای توضیحی را نشان می‌دهد و بردار $x_i = (x_{i1},\ldots, x_{iJ})$ یک بردار سطری J بعدی است و در حقیقت iامین سطر یک ماتریس n در J بعدی X است.

$$\beta_k = \begin{bmatrix} \beta_{1k} \\ . \\ . \\ . \\ \beta_{Jk} \end{bmatrix}$$ یک بردار ستونی J بعدی است $$\beta_k X = \begin{bmatrix} x_1 \\ x_2 \\ . \\ . \\ . \\ x_n \end{bmatrix}$$

و kامین ستون ماتریس J در K بعدی ضرایب رگرسیون می‌باشد $B = (\beta_1,\ldots, \beta_K)$.

فرض می‌شود که اثرات تصادفی ناهمگون غیر قابل مشاهده $\varepsilon_{i1},\ldots,\varepsilon_{iK}$ بین سطوح مختلف شدت برای هر مکان وجود دارد تا ناهمگونی خارج از توزیع پواسون برای هر مکان را بپوشاند. فرض می‌شود که بردار K بعدی ε_i دارای توزیع نرمال K متغیره با بردار میانگین صفر و ماتریس کواریانس Σ باشد؛ به طوری که Σ یک ماتریس با ساختار کلی است تا همبستگی بین ε_{ik}ها را بپوشاند. توجه کنید که اگر Σ یک ماتریس قطری باشد، آنگاه این مدل به حالت تک متغیره تبدیل می‌شود.

با توجه به این فرض‌ها بردارهای $\lambda_i = (\lambda_{i1},\ldots,\lambda_{iK})'$ دارای توزیع لگ نرمال K متغیره است که این منتبع به آن است که y_iها دارای توزیع پواسون- لگ نرمال K متغیره باشد. تابع چگالی احتمالی حاشیه‌ای بردار مشاهدات y_i به صورت

$$f(y_i \mid \lambda_i, \Sigma) = \int \prod_{k=1}^{K} f_{poisson}(y_{ik} \mid \lambda_{ik}) \phi_k(\varepsilon_i \mid 0, \Sigma) d\varepsilon_i$$

به دست می‌آید که در آن ϕ_k نشان‌دهنده تابع چگالی نرمال K متغیره با میانگین صفر و واریانس Σ است.

چگالی حاشیه‌ای تعداد تصادفات y_i به صورت صریح نمی‌تواند به دست آید، چون به محاسبه انتگرال K متغیره توزیع پواسون نیاز دارد که با توزیع ε_i آمیخته شده است. با این وجود امید ریاضی، واریانس y_iها و کواریانس‌های مربوط می‌تواند بدون محاسبه انتگرال‌ها به دست آید.

Chib و Winkelmann روابط

$$E(y_{ik}) = \mu_{ik} \exp(0.5\sigma_{kk})$$
$$\mathrm{var}(y_{ik}) = E(y_{ik}) + (E(y_{ik}))^2 (\exp(\sigma_{kk}) - 1)$$
$$\mathrm{cov}(y_{ik}, y_{ij}) = \lambda_{ik} \exp(0.5\sigma_{kk})(\exp(\sigma_{kj}) - 1)\lambda_{ij} \exp(0.5\sigma_{jj})$$

را نشان دادند. مشاهده می‌شود که در مدل فوق به شرط $\sigma_{kk} > 0$، واریانس بیشتر از میانگین داده‌ها است (۱۳). همچنین با توجه به علامت σ_{kj}، کواریانس می‌تواند مثبت یا منفی شود. بنابراین مدل پیشنهادی می‌تواند ساختار کلی همبستگی در داده‌ها را در برگیرد.

برآورد پارامترها در مدل رگرسیونی MVPLN به صورت صریح امکان‌پذیر نیست و هیچ کدام از نرم‌افزارهای آماری موجود توانایی برآورد پارامترها به عنوان یک تابع موجود را ندارند. بنابراین از روش شبیه‌سازی زنجیر مارکوفی در قالب بیزی استفاده شد تا برآوردی از پارامترها به دست آید. در مرحله اول برای پارامترهای (β، $\sum$) توزیع پیشین در نظر گرفته شد. متداول‌ترین توزیعی که برای پارامترهای رگرسیون فرض می‌شود بصورت نرمال، با میانگین صفر و واریانس بزرگ است، که در آن Σ^{-1} دارای توزیع ویشارت است. به عبارت دیگر

$$\beta_{jk} \sim N(0,1000)$$
$$\Sigma^{-1} \sim wishart(R, n)$$

می‌باشد که در آن R ماتریس اطلاع و n معمولاً برابر با رتبه ماتریس R فرض می‌شود. بر اساس نظریه بیز توزیع پسین پارامترها متناسب با حاصل‌ضرب توزیع‌های پیشین و تابع درستنمایی است. در مرحله دوم با استفاده از این نظریه و جهت به کارگیری الگوریتم‌های مبتنی بر MCMC، لازم است که توزیع‌های پسین شرطی کامل پارامترها به دست آید.

El-Basyouny و Sayed نشان دادند که توزیع‌های پسین هر یک از پارامترها به شرط بقیه پارامترهای مدل و متغیرهای پنهان به صورت زیر محاسبه می‌شود (۷).

www.mui.ac.ir

$$f(\lambda_i \mid Y, X, B, \Sigma) \propto \exp\{(-0.5(\lambda_i^* - \mu_i^*)'\Sigma^{-1}(\lambda_i^* - \mu_i^*) + \sum_{k=1}^{K}[(y_{ik} - 1)\ln(\lambda_{ik}) - \lambda_{ik}]\} \quad (۱)$$

$$f(B \mid Y, X, \lambda, \Sigma) \propto \exp\{(-0.5\sum_{i=1}^{n}(\lambda_i^* - \mu_i^*)'\Sigma^{-1}(\lambda_i^* - \mu_i^*) + \sum_{k=1}^{K}[(y_{ik} - 1)\ln(\lambda_{ik}) - \lambda_{ik}]\} f(B) \quad (۲)$$

$$f(\Sigma^{-1} \mid Y, X, \lambda, B) = wishart\ (p + \sum_{i=1}^{n}(\lambda_i^* - \mu_i^*)(\lambda_i^* - \mu_i^*)', k + n) \quad (۳)$$

در آن‌ها

$$\lambda_i = (\lambda_{i1}, \ldots, \lambda_{ik})' \quad \lambda_i^* = (\ln \lambda_{i1}, \ldots, \ln \lambda_{ik})'$$
$$\mu_i = (\mu_{i1}, \ldots, \mu_{ik})' \quad \mu_i^* = (\ln \mu_{i1}, \ldots, \ln \mu_{ik})'$$

می‌باشد. حال این توزیع‌های پسین شرطی کامل با استفاده از روش MCMC نمونه‌گیری می‌شوند. توزیع ویشارت در (۳) با استفاده از نمونه‌گیری گیبز شبیه‌سازی می‌شود. توزیع‌های پسین (۱) و (۲) که صریح نبوده و شناخته شده نیستند، با استفاده از الگوریتم متروپلیس- هستینگز شبیه‌سازی می‌شوند.

روش نمونه‌گیری گیبز که مبتنی بر MCMC است، از توزیع‌های پسین شرطی کامل پارامترها نمونه‌های تصادفی می‌گیرد؛ به طوری که این نمونه‌برداری‌ها به صورت متوالی انجام می‌گیرد و توزیع نمونه گرفته شده فقط به مقدار نمونه قبلی بستگی دارد تا سرانجام به توزیع‌های پسین پارامترها همگرا شود. جهت استقلال نمونه تولید شده از مقادیر اولیه زنجیر ۱۰۰۰۰ نمونه تولید شده دور ریخته می‌شود. سپس ۱۰۰۰۰۰ نمونه برای به دست آوردن برآورد پارامترها تولید می‌شود. همگرایی زنجیر با استفاده از نمودار اثر پارامترهای مدل و خطای مونت کارلوی آن‌ها بررسی می‌شود. با توجه به یک قانون کلی، نسبت خطای مونت کارلو پارامتر به برآورد انحراف معیار آن باید کم‌تر از ۰/۰۵ باشد. بعد از همگرایی بر اساس محاسبه میانگین، انحراف معیار و صدک‌های ۲/۵ و ۹۷/۵ نمونه‌ها، استنباط بیزی پارامترها انجام می‌شود.

مطالعه حاضر از نوع مطالعه مقطعی است که با توجه به اطلاعات شش ماه اول سال ۱۳۹۰ انجام شد. جامعه مورد مطالعه ۲۸۰ تقاطع شهر اصفهان بود که با مراجعه به سازمان‌های مربوطه تنها تقاطع‌هایی که اطلاعات مورد نیاز آن‌ها موجود بود، وارد مطالعه شد. بنابراین روش نمونه‌گیری، آسان و حجم نمونه مورد مطالعه ۶۵ تقاطع بود. اطلاعات مورد نیاز برای این مطالعه مربوط به تعداد تصادفات و حجم ترافیکی هر تقاطع استخراج شد. متغیر پاسخ، تعداد تصادفات رخ داده شده در هر تقاطع بود. تصادفات بر حسب شدت به دو دسته خسارتی و جراحتی/فوتی تقسیم شد. متغیرهای مستقل شامل حجم کل ترافیک تقاطع و حجم کل ترافیک چپ‌گردهای تقاطع بود. حجم کل ترافیک تقاطع به صورت حجم ورودی ترافیک به تقاطع از تمام پایه‌های آن تعریف شد. حجم کل ترافیک چپ‌گردهای تقاطع نیز به صورت حجم ورودی ترافیک به تقاطع از تمام چپ‌گردهای آن تعریف شد.

یافته‌ها

تعداد ۱۸۲ تصادف خسارتی و ۶۶ تصادف جراحتی و فوتی در ۶۵ تقاطع شهر اصفهان در طول مدت ۶ ماه وجود داشت. جدول ۱ آماره‌های توصیفی متغیرهای مستقل و وابسته را نشان می‌دهد. وقتی سطح شدت در داده‌های تصادف وجود دارد، مدل MVPLN برازش بهتری را ارایه می‌دهد. در این جا از دو سطح برای شدت تصادفات استفاده شد. پارامترهای

جدول ۱: آماره‌های توصیفی متغیرها

	مینیمم	ماکسیمم	میانگین	انحراف معیار
تعداد تصادفات خسارتی	۰	۱۷	۲/۸	۳/۹۸۱
تعداد تصادفات جراحتی و فوتی	۰	۸	۱/۰۱۵۴	۱/۵۱۵
حجم کل ترافیک تقاطع	۱۱۸۴	۶۶۸۹	۴۰۵۳/۰۳	۱۳۶۸/۲۶۵
حجم کل ترافیک چپ‌گردهای تقاطع	۱۷۹	۲۵۵۸	۱۲۰۵/۰۴	۵۳۲/۵۲۳

www.mui.ac.ir

مدل با استفاده از نرم‌افزار OpenBUGS نسخه ۳.۲.۱ برآورد شد. نتایج ارایه شده در جداول ۲ تا ۴ نشان می‌دهند که اختلاف کمی بین برآورد پارامترها تحت مدل MVPLN، پواسون تک متغیره و پواسون- لگ نرمال تک متغیره وجود دارد. همچنین نشان می‌دهند که برآورد بعضی از پارامترها در سطح ۰/۹۵ معنی‌دار است. برای نمونه اثر حجم کل ترافیک بر تصادفات خسارتی در هر سه مدل معنی‌دار است، اما بر تصادفات جراحتی و فوتی در هیچ کدام از آن‌ها معنی‌دار نیست.

جدول ۲: برآورد پارامترها تحت مدل پواسون تک متغیره

	مدل پواسون تک متغیره	برآوردها	انحراف معیار	فاصله اطمینان ۹۵ درصد	
				حد بالا	حد پایین
سطح شدت ۱ (تصادفات خسارتی)	مقدار ثابت	-۰/۵۳۲۱	۰/۴۰۱۹	۰/۲۲۰۱	-۱/۳۸۹
	حجم کل ترافیک	۰/۰۰۰۱۳۴	۰/۰۰۰۰۵۲	۰/۰۰۰۲۴	۰/۰۰۰۰۱۶
	حجم کل ترافیک چپ‌گردها	۰/۰۰۰۰۶۶	۰/۰۰۰۱۶۹	۰/۰۰۰۴۱	-۰/۰۰۰۲۵
	DIC		۴۰۸/۱		
	Pearson' Chi-Square/DF		۵/۶۰		
سطح شدت ۲ (تصادفات جراحتی و فوتی)	مقدار ثابت	-۰/۳۲۹	۰/۶۰۳۳	۰/۷۳۲۱	-۱/۵۹۱
	حجم کل ترافیک	۰/۰۰۰۱۲	۰/۰۰۰۱۳	۰/۰۰۰۲۱۶	-۰/۰۰۰۳۸
	حجم کل ترافیک چپ‌گردها	۰/۰۰۰۹۸	۰/۰۰۰۳۲۲	۰/۰۰۱۷۳۸	۰/۰۰۰۴۲
	DIC		۱۹۹/۴		
	Pearson' Chi-Square/DF		۱/۹۴		

جدول ۳: برآورد پارامترها تحت مدل پواسون- لگ نرمال تک متغیره

	مدل پواسون- لگ نرمال تک متغیره	برآوردها	انحراف معیار	فاصله اطمینان ۹۵ درصد	
				حد بالا	حد پایین
سطح شدت ۱ (تصادفات خسارتی)	مقدار ثابت	-۱/۱۸۹۰۰	۱/۰	۰/۸۱۶۷۰	-۳/۳۱۹۰۰
	حجم کل ترافیک	۰/۰۰۰۲۲	۰/۰۰۰۰۸۳	۰/۰۰۰۳۸	۰/۰۰۰۰۶۷
	حجم کل ترافیک چپ‌گردها	-۰/۰۰۰۰۳۳	۰/۰۰۰۲۷۰	۰/۰۰۰۴۵	-۰/۰۰۰۵۶۰
	۱۱σ	۱/۶۰۷۰۰۰	۰/۵۰۴۲۰۰	۲/۸۰۵۰	۰/۸۵۵۵۰
	DIC		۲۳۸/۶		
سطح شدت ۲ (تصادفات جراحتی و فوتی)	مقدار ثابت	-۱/۴۱۰۰۰	۱/۰۲۵۰۰	۰/۴۴۶۶	-۳/۷۴۵
	حجم کل ترافیک	۰/۰۰۰۱۱	۰/۰۰۰۱۳۴	۰/۰۰۰۴۱	-۰/۰۰۰۱۴۱
	حجم کل ترافیک چپ‌گردها	۰/۰۰۰۵۸	۰/۰۰۰۳۸	۰/۰۰۱۳۳۶	-۰/۰۰۰۱۳
	۲۲σ	۱/۱۰۳۰۰	۰/۵۳۶۰۰	۲/۴۰۰۰۰	۰/۳۴۲۲۰
	DIC		۱۷۰/۲		

www.mui.ac.ir

اثر حجم کل ترافیک چپ‌گردها بر تصادفات جراحتی و فوتی تنها در مدل پواسون تک متغیره معنی‌دار است. بنابراین با فرض ثابت نگه داشتن بقیه متغیرهای توضیحی، انتظار می‌رود که افزایش حجم کل ترافیک باعث افزایش تصادفات خسارتی شود. همچنین افزایش حجم کل ترافیک چپ‌گردها نیز باعث افزایش تصادفات جراحتی و فوتی در تقاطع‌ها می‌شود. به طور دقیق‌تر افزایش ۱۰۰۰ وسیله نقلیه در متوسط حجم کل ترافیک باعث افزایش ۳۱ درصدی تصادفات خسارتی می‌شود. افزایش ۱۰۰ وسیله نقیله در متوسط حجم کل ترافیک چپ‌گردها باعث افزایش ۱۰ درصدی تصادفات جراحتی و فوتی می‌شود.

مقدار آماره بیشتر از ۱ ۲χ Pearson' تقسیم بر درجه آزادی $\chi^2 = \sum_{i=1}^{n} \frac{(y_i - \mu_i)^2}{\mathrm{var}(y_i)} \Big/ n - p_x$ در مدل پواسون تک متغیره نشان‌دهنده مشکل بیش پراکنش در داده‌ها می‌باشد. جدول ۲ نشان می‌دهد که این مقدار برای دو سطح شدت بیشتر از ۱ است. بنابراین مشکل بیش پراکنش در داده‌ها وجود دارد که مدل پواسون تک متغیره قادر به حل این مشکل نمی‌باشد. جدول ۴ نیز نشان می‌دهد که برآورد پارامترهای بیش پراکنش پواسون در مدل MVPLN بیشتر از صفر است. همان طور که Chib و Winkelmann نشان دادند تا هرگاه پارامترهای بیش پراکنش پواسون در مدل MVPLN بیشتر از صفر باشد، آنگاه واریانس بیشتر از میانگین است. بنابراین مدل MVPLN می‌تواند بیش پراکنش در داده‌ها را پوشش دهد (۱۳).

مشکلی که بیش پراکنش در داده‌ها ایجاد می‌کند، کم برآورد شدن انحراف معیار پارامترهای مدل می‌باشد. در این گونه موارد انحراف معیار برآورد درستی از عدم قطعیت واقعی نیست و فاصله اطمینان‌های مربوطه مقادیر واقعی پارامترها را نمی‌پوشاند. البته باید تأکید شود که تنها برای برآوردهای نااریب مقادیر کوچک‌تر انحراف معیار به معنای دقت بیشتر برآوردها می‌باشد (۴). با مقایسه برآورد انحراف معیار ضرایب در سه جدول ۲، ۳ و ۴ می‌توان دریافت که تحت مدل پواسون تک متغیره مقادیر انحراف معیار پارامترها کم برآورد شدند. به همین علت متغیرهای بیشتری تحت این مدل نسبت به مدل MVPLN و پواسون- لگ نرمال تک متغیره معنی‌دار شدند. همچنین نتایج نشان دادند که برآورد پارامترهای بیش پراکنش تحت مدل MVPLN نسبت به مدل پواسون- لگ نرمال تک متغیره کوچک‌تر است. با توجه به روابط زیر و این که دقت متناسب با معکوس واریانس فراوانی مشاهدات است، مدل MVPLN نسبت به مدل پواسون- لگ نرمال تک متغیره دقیق‌تر است (۷).

$$\mathrm{var}(y_{ik}) = E(y_{ik}) + (E(y_{ik}))^2(\exp(\sigma_{kk}) - 1)$$
$$\mathrm{var}(y_{ik}) \cong (E(y_{ik}))^2(\exp(\sigma_{kk}) - 1)$$
$$\frac{\mathrm{var}(y_{ik} \mid MVPLN)}{\mathrm{var}(y_{ik} \mid PLN)} \cong \frac{(\exp(\hat{\sigma}_{kk}) - 1)}{(\exp(\hat{\sigma}_{kk}) - 1)}$$

جدول ۴: برآورد پارامترها تحت مدل پواسون- لگ نرمال چند متغیره

	مدل پواسون-لگ نرمال چند متغیره	برآوردها	انحراف معیار	فاصله اطمینان ۹۵ درصد	
				حد بالا	حد پایین
	مقدار ثابت	-۱/۱۶۴	۰/۷۹۳	۰/۴۷۷۳	-۲/۵۱۲
سطح شدت ۱	حجم کل ترافیک	۰۰۰۳۱/۰	۰۰۰۱/۰	۰۰۰۵۱/۰	۰۰۰۱۳/۰
(تصادفات خسارتی)	حجم کل ترافیک چپ‌گردها	۰۰۰۱۴/۰-	۰۰۰۳۲/۰	۰۰۰۴۳/۰	۰۰۰۷۸/۰-
	σ11	۱/۵۷۳	۰/۴۶۸۵	۲/۶۸۸	۰/۸۶۲۷
	مقدار ثابت	-۱/۳۲۳	۰/۸۳۵۳	۲۹۶۶/۰	-۲/۹۴۸
سطح شدت ۲	حجم کل ترافیک	۰۰۰۰۶۸/۰	۰۰۰۲۲/۰	۰۰۰۴/۰	۰۰۰۳۱/۰-
(تصادفات جراحتی و فوتی)	حجم کل ترافیک چپ‌گردها	۰۰۰۷۳/۰	۰۰۰۵۱/۰	۰/۰۰۱۷۰۲	۰۰۰۲۷/۰-
	۲۲σ	۱/۰۹۱	۰/۴۵۷۸	۲/۲۰۳	۰/۴۴۹۹
	Covariance(σ۱۲)	۰/۹۳۸	۰/۳۵۲	۱/۷۴۴	۰/۳۷۵۱
	Correlation (ρ)		۰/۷۱۲		
	DIC		۴۰۲/۷		

www.mui.ac.ir

همبستگی میان تعداد تصادفات در دو سطح شدت تحت مدل MVPLN مقدار ۰/۷۱ برآورد شد که به طور قابل توجهی معنی‌دار است. به این دلیل که حتی بعد از کنترل کردن توسط متغیرهای توضیحی مدل، انتظار می‌رود احتمال رخداد تصادف برای هر دو سطح شدت در یک تقاطع به علت متغیرهای غیر قابل مشاهده یا وجود نقوص یکسان در طراحی تقاطع افزایش پیدا کند. استفاده از مدل MVPLN این همبستگی را در برآورد پارامترها لحاظ می‌کند. به همین علت نسبت به مدل‌های تک متغیره، برآوردهای دقیق‌تری ارایه می‌دهد.

در مقوله نیکویی انتخاب مدل از معیار اطلاع انحراف (DIC) استفاده شد. همان طور که نتایج ارایه شده در جدول ۲، ۳ و ۴ نشان می‌دهد، این معیار برای مدل MVPLN بسیار کم‌تر از مجموع DICها برای مدل پواسون تک متغیره و پواسون- لگ نرمال تک متغیره در دو سطح شدت می‌باشد. بنابراین مدل MVPLN برازش بهتری را نسبت به دو مدل دیگر فراهم می‌کند.

بحث

مدل‌های تک متغیره که معمولاً برای تحلیل داده‌های تصادف استفاده می‌شود، تعداد تصادفات را در سطوح مختلف شدت به طور جداگانه برآورد می‌کند یا سطوح شدت را در نظر نمی‌گیرد و تعداد تصادفات را به طور کلی برآورد می‌کند. در این مقاله از مدل MVPLN برای تحلیل داده‌ها در دو سطح شدت استفاده شد که داده‌های مربوط به ۶۵ تقاطع شهر اصفهان بود. علاوه بر آن برای مقایسه از مدل پواسون تک متغیره و پواسون لگ- نرمال تک متغیره نیز برای هر یک از سطوح به طور جداگانه برازش شد. نتایج نشان داد که مدل MVPLN به دو مدل دیگر ارجحیت دارد، چون مدل‌های تک متغیره همبستگی که بین دو سطح شدت وجود دارد را نادیده می‌گیرد. این در حالی است که مدل MVPLN به همان خوبی که بیش پراکنش را در داده‌ها می‌پذیرد، ساختار همبستگی کلی را نیز در بین سطوح مختلف شدت می‌پذیرد (۶).

برآورد مدل توسط رهیافت بیز با استفاده از نمونه‌گیری گیبز و الگوریتم متروپلیس- هستینگز به دست آمد. در مورد برآورد پارامترهای مدل و واریانس آن‌ها در هر سه مدل بحث شد. همان طور که انتظار می‌رفت، نتایج توصیه‌هایی را برای بهبود ایمنی تقاطع‌های شهری ارایه داد. برای نمونه کاهش حجم کل ترافیک در مقوله کاهش هزینه تصادفات در دراز مدت پیش‌بینی می‌شود که بسیار مقرون به صرفه است. همبستگی بین دو سطح شدت ۰/۷۱ برآورد شد. بنابراین استفاده از مدل MVPLN به دلیل لحاظ این همبستگی نسبت به دو مدل تک متغیره دیگر، برآوردهای دقیق‌تری ارایه داد. در مقوله انتخاب مدل نیز مدل MVPLN برازش بهتری را نسبت به مدل پواسون تک متغیره و پواسون- لگ نرمال تک متغیره فراهم کرد.

این مقاله بر اساس داده‌هایی با اندازه نمونه کم و برای مدت زمان کوتاه است. با وجود این محدودیت، نتایج این مقاله با آن چه که در مقاله‌های دیگر است، مطابقت دارد (۷، ۴). با این وجود برای تحقیقات بیشتر نیز چالش‌هایی وجود دارد. به عنوان نمونه داده‌های تصادف معمولاً میانگین نمونه و اندازه نمونه پایینی دارد و مدل‌های آماری وقتی برای این نوع داده‌ها مورد استفاده قرار می‌گیرد، بسیار ناپایدار هستند (۴). بنابراین پایداری مدل MVPLN باید مورد بررسی قرار گیرد. همچنین حساسیت مدل MVPLN برای مشخصات مختلف توزیع‌های پیشین نیز باید تحلیل شود.

مدل‌های مربوط به حجم ترافیک از اریبی مربوط به متغیرهای حذف شده از مدل نیز رنج می‌برد، اما با این حال این نوع مدل‌ها در ادبیات مربوط به ایمنی ترافیک به رسمیت شناخته شدند (۷). در این مقاله دو سطح برای شدت تصادفات در نظر گرفته شد. با این حال سطوح بیشتری برای شدت مانند مرگ و میر، جراحت شدید، جراحت کم و خسارت مالی می‌تواند استفاده شود. البته نوع برخوردهای مختلف مانند با زاویه، از سر، از پشت و از پهلو نیز می‌تواند در نظر گرفته شود. همچنین ویژه‌گی‌های طراحی هندسی تقاطع که در مجموعه داده‌ها در دسترس نبود، نیز می‌تواند در مدل گنجانده شود. بنابراین

www.mui.ac.ir

مهندسین حمل و نقل توسط روش‌های آماری که اجازه تحلیل دقیق‌تر را می‌دهند و همچنین با استفاده از مجموعه داده‌های مناسب، قادر خواهند بود تا ارتباط بین علت وقوع تصادفات و عوامل مرتبط با آن‌ها را بهتر درک کنند.

تشکر و قدردانی

نویسندگان این مقاله از سازمان کنترل ترافیک و پلیس راهنمایی و رانندگی به خاطر همکاریشان جهت فراهم کردن داده‌ها برای این تحقیق تشکر می‌کنند.

References

1. Abdalla IM. Fatality risk assessment and modeling of drivers responsibility for causing traffic accidents in Dubai. J Safety Res 2002; 33(4): 483-96.
2. Eksler V, Lassarre S, Thomas I. Regional analysis of road mortality in Europe. Public Health 2008; 122(9): 826-37.
3. Lord D, Ivan JN, Washington S. Statistical Challenges with Modeling Motor Vehicle Crashes: Understanding the Implications of Alternative Approaches. Texas: Texas Transportation Institute; 2004.
4. Park ES, Lord D. Multivariate Poisson-lognormal models for jointly modeling crash frequency by severity [Online]. 2007; Available from: URL: http://pubsindex.trb.org/view.aspx?id=801837. 2012.
5. Lord D, Mannering F. The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. Transportation Research Part A: Policy and Practice 2010; 44(5): 291-305.
6. Ma J, Kockelman KM, Damien P. A multivariate Poisson-lognormal regression model for prediction of crash counts by severity, using Bayesian methods. Accid Anal Prev 2008; 40(3): 964-75.
7. El-Basyouny K, Sayed T. Collision prediction models using multivariate Poisson-lognormal regression. Accid Anal Prev 2009; 41(4): 820-8.
8. Ma J, Kockelman KM. Bayesian Multivariate Poisson Regression for Models of Injury Count, by Severity [Online]. 2006; Available from: URL: http://www.ce.utexas.edu/prof/kockelman/public_html/TRB06MVPBayesian.pdf. 2012.
9. Aguero-Valverde J, Jovanis PP. Bayesian Multivariate Poisson Lognormal Models for Crash Severity Modeling and Site Ranking [Online]. 2009; Available from: URL: http://pubsindex.trb.org/view.aspx?id=880997. 2012.
10. Miranda-Moreno LF. Statistical Models and Methods for Identifying Hazardous Locations for Safety Improvements. Ontario: University of Waterloo; 2006.
11. Lord D, Washington SP, Ivan JN. Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory. Accid Anal Prev 2005; 37(1): 35-46.
12. Tsionas EG. Bayesian multivariate Poisson regression. Communications in Statistics - Theory and Methods 2001; 30(2): 243-55.
13. Chib S, Winkelmann R. Markov chain Monte Carlo analysis of correlated count data. Journal of Business and Economic Statistics 2001; 19(4): 428-35.
14. Tierney L. Markov Chains for Exploring Posterior Distributions. Annals of Statistics 1994; 22: 1701-28.
15. Tunaru R. Hierarchical Bayesian Models for Multiple Count Data. Austrian Journal of Statistics 2012; 31(2-3): 221-9.
16. Gilks WR, Richardson S, Spiegelhalter D. Markov Chain Monte Carlo in Practice: Interdisciplinary Statistics. New York: Taylor & Francis; 1995.
17. Liu JS. Monte Carlo Strategies in Scientific Computing. London: Springer; 2008.

Factors Determining the Number of Accidents According to Severity Using Bayesian Multivariate Log-Normal Poisson Regression

Iraj Kazemi[1], Mehdi Tazhibi[2], Somayeh Momenyan[3], Hossein Haghshenas[4]

Abstract

Introduction: Univariate models have been used to analyse crash data in previous studies. With regards to the severity variable in these researches that can be more than one level, in the present study multivariate Poisson-log normal model has been recommended to model crash count by severity, although it has been also used to compare the univariate models.

Methods: The statistical inference of model parameters was conducted by Bayesian method, via a Gibbs Sampler and the Metropolis-Hastings (M-H) algorithms. The crash data were related to the intersections of Isfahan, Iran.

Results: Despite over-dispersion at two levels of severity and correlation among these two levels, the multivariate Poisson-lognormal model offers a better fit than other models. Also the effect of the total AADT (average annual daily traffic) on the property was significant in all of the models. But the effect of total left turn AADT on injuries and fatalities was significant just in the univariate Poisson model. Under multivariate Poisson-lognormal model, an increase of 1000 vehicles in average total AADT was predicted to result in 31% more property damage.

Discussion: Reducing the total volume of traffic in the cost of accidents reduction was predicted to be very affordable in long term.

Keywords: Bayesian Method, Markov Chain Monte Carlo Methods, Full Conditional Posterior Distribution, Severity of Crashes, Intersection Safety

1- Assistant Professor, Department of Statistics, School of Statistics, University of Isfahan, Iran
2- Assistant Professor, Department of Biostatistics and Epidemiology, School of Public Health, Isfahan University of Medical Sciences, Isfahan, Iran (Corresponding Author) Email: tazhibi@hlth.mui.ac.ir
3- MSc Student, Student Research Committee, Department of Biostatistics and Epidemiology, School of Public Health, Isfahan University of Medical Sciences, Isfahan, Iran
4- Assistant Professor, Department of Civil Engineering, School of Transportation, University of Isfahan, Isfahan, Iran