



ارزیابی اثرات دوربرگردان بر زمان وقوع تصادفات در بزرگراه چمران

عباس بابازاده¹، حامد صابونچی²

1- استادیار دانشکده مهندسی عمران، دانشگاه تهران تلفن: 021-61112176، فاکس: 021-66403808

2- دانشجوی کارشناسی ارشد راه و ترابری، دانشکده مهندسی عمران، دانشگاه تهران تلفن: 021-22279249

ababazadeh@ut.ac.ir

خلاصه

این مقاله به مطالعه اثر دوربرگردان در احتمال وقوع تصادفات شبانه در بزرگراهها می‌پردازد. اطلاعات مورد استفاده در این مطالعه نمونه‌ای شامل 3647 تصادف رخ داده در بزرگراه چمران در سال 1385 است. متغیر وابسته زمان وقوع تصادف است که بر اساس آن هر تصادف در یکی از دو دسته روزانه و شبانه قرار می‌گیرد. با توجه به طبیعت دوتایی متغیر وابسته، رویکرد رگرسیون لجستیک برای فرمولبندی احتمال وقوع تصادفات شبانه مناسب تشخیص داده شد. نتایج بدست آمده نشانگر افزایش احتمال وقوع تصادفات شبانه در محدوده دوربرگردان است.

کلید واژه‌ها: دوربرگردان، مدل‌سازی تصادف، رگرسیون لجستیک.

مقدمه

استفاده از دوربرگردان معمولاً به نواحی کم ترافیک و با سرعت پایین محدود می‌شود. دوربرگردان معمولاً قبل از تقاطعات همسطح، به منظور جلوگیری از گردش کامل وسایل نقلیه در تقاطع و در نتیجه افزایش کارایی آن، احداث می‌شوند. احداث دوربرگردان در بزرگراهها رایج نیست، زیرا موجب کاهش ناگهانی سرعت وسایل نقلیه در باند سرعت و نیز تغییر مسیر ناگهانی وسایل گردشی از باند کم سرعت به باند پر سرعت خواهد شد. در نتیجه، علاوه بر کاهش سرعت متوسط جابجایی‌های شهری، موجب تصادفات شدید خواهد شد.

هدف این مقاله ارزیابی اثرات دوربرگردان بر احتمال وقوع تصادفات شدید در بزرگراههای شهری تهران است. شواهد موجود در ادبیات ایمنی وجود رابطه مستقیم بین شدت تصادف و سرعت برخورد را به خوبی نشان میدهند [2و1]. افزایش سرعت وسایل نقلیه موجب افزایش زمان تصمیم‌گیری و واکنش راننده، و در نتیجه باعث شدت بیشتر برخورد در محل تصادف خواهد شد [3]. به علت عدم وجود اطلاعات سرعت وسایل نقلیه در کروکی تصادفات شهری تهران، در این مطالعه از زمان وقوع تصادف در شب یا روز به عنوان جایگزینی برای شدت برخورد استفاده می‌شود. به عبارت دیگر، هر تصادف در بزرگراه در یکی از دو دسته شبانه (شدید) یا روزانه (غیرشدید) قرار می‌گیرد.

در مجموع، در این مطالعه اثر دوربرگردان بر روی احتمال وقوع تصادفات شبانه ارزیابی می‌شود. این کار با پرداخت یک مدل رگرسیون لجستیک برای نمونه‌ای از تصادفات رخ داده در یک بزرگراه دارای دوربرگردان انجام می‌شود. مدل رگرسیون لجستیک وقتی استفاده می‌شود که متغیر پاسخ تنها دو مقدار داشته باشد. کاربردهای این مدل شامل مطالعاتی با متغیرهای پاسخ مرگ یا زندگی، داشتن یا نداشتن یک ویژگی خاص، و مانند آنهاست. در رگرسیون لجستیک متغیر خروجی دوتایی است، و هدف از تحلیل ارزیابی متغیرهای توضیح دهنده بر متغیر خروجی است. متغیرهای توضیح دهنده می‌توانند پیوسته و یا چندتایی (دسته بندی شده) باشند.

مجموعه اطلاعات استفاده شده در این مطالعه شامل یک نمونه 3647 عددی از تصادفات رخ داده در بزرگراه چمران شهر تهران در سال 1385 است. اطلاعات مورد استفاده برای هر تصادف شامل مشخصات سرنشینان و وسایل نقلیه مورد برخورد، نحوه برخورد، مکان و زمان برخورد، و آب و هوا از فرمهای کروکی ثبت شده در مرکز آمار تصادفات راهنمایی و رانندگی تهران بزرگ استخراج شده است.

مرور منابع

در تحلیل رگرسیون معمولی، متغیر پاسخ عمدتاً پیوسته است. لذا این نوع رگرسیون برای متغیرهای پاسخ گسسته قابل استفاده نیست. از اینرو، محققان ایمنی در ترافیک از انواع دیگر رگرسیون برای پرداخت مدلهای شدت تصادف استفاده کرده‌اند.

هاتچینگز [4] اثرات بستن کمربند ایمنی در کاهش جراحات ناشی تصادفات وسایل نقلیه را با استفاده از مدلهای رگرسیون معادلات تخمین زنده تعمیم یافته (GEES) تحلیل می‌کند. گریب [5] از رویکرد مدلهای خطی تعمیم یافته (GLMs) برای تعیین ارتباط بین تعداد تصادفات و متغیرهای توضیح دهنده آن استفاده می‌کند. این پژوهشگر دو مدل پیش بینی تعداد تصادفات به تفکیک راه‌های شهری و تقاطعهای شهری ارائه داده است. ساوالها و ساید [6] مدل‌سازی تعداد تصادفات را با استفاده از مدلهای رگرسیون بواسون و رگرسیون دو جمله‌ای منفی انجام دادند. آنها اعتبار این دو



رویکرد مدلسازی را با ارایه شواهد کافی نشان دادند.

کوچلمان و کنون [7] نحوه استفاده از مدل‌های پربینت رتبه‌ای در مدلسازی احتمال وقوع سطوح مختلف جراحات در انواع تصادفات را در مطالعه خود تشریح کردند. وود [8] یک روش کاربردی برای رفع مشکل مقدار میانگین کم، که مانعی برای آزمون خوبی برآزش در مدل‌های خطی تعمیم یافته محسوب می‌شود، ارایه داد.

رگرسیون لوجستیک

رگرسیون لوجستیک رابطه بین یک متغیر پاسخ (وابسته) چندتایی را با مجموعه‌ای از متغیرهای توضیح دهنده (مستقل) بیان می‌کند. متغیرهای توضیح دهنده می‌توانند پیوسته یا گسسته باشند. رگرسیون لوجستیک امکان برآورد یک خروجی گسسته، مثل عضویت گروهی، را بر اساس مجموعه‌ای از متغیرهای دیگر، که ممکن است پیوسته، گسسته، دوتایی، یا مخلوطی از اینها باشند، فراهم می‌سازد [9]. متغیر وابسته در رگرسیون لوجستیک معمولاً دوتایی است، به این معنا که یا برابر 1 با احتمال موفقیت q ، یا برابر 0 با احتمال شکست $1-q$ است [10]. این نوع متغیر یک متغیر برنولی یا دو جمله‌ای نیز خوانده می‌شود.

متغیرهای مستقل، در رگرسیون لوجستیک، می‌توانند هر شکلی داشته باشند. به عبارت دیگر، این نوع رگرسیون هیچ پیش فرضی برای توزیع متغیرهای مستقل ندارد. در رگرسیون لوجستیک رابطه بین متغیر پاسخ و متغیرهای توضیح دهنده تابعی خطی نیست، بلکه به جای آن تابع رگرسیون لوجستیک زیر استفاده می‌شود [9]:

$$\pi(x) = \frac{e^{(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)}}{1 + e^{(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)}} \quad (1)$$

که در آن $\beta = (\beta_0, \beta_1, \beta_2, \dots, \beta_p)$ بردار ضرایب مدل و $x = (x_1, x_2, \dots, x_p)$ بردار متغیرهای پیش بینی کننده است. یک فرم جایگزین برای تابع رگرسیون لوجستیک بالا تبدیل لوجیت آن است. این تبدیل، بر حسب $\pi(x)$ ، به صورت زیر تعریف می‌شود:

$$g(x) = \text{logit}[\pi(x)] = \ln\left[\frac{\pi(x)}{1 - \pi(x)}\right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p \quad (2)$$

اهمیت این تبدیل آن است که تابع لوجیت $g(x)$ دارای بسیاری از خصوصیات مطلوب مدل‌های رگرسیون خطی است. به علاوه، $g(x)$ ، که بر حسب پارامترهای خطی است، می‌تواند در دامنه $-\infty$ تا $+\infty$ پیوسته باشد.

فرض کنید که نمونه‌ای شامل n مشاهده مستقل (x_i, y_i) ، $i = 1, 2, \dots, n$ ، وجود دارد که y_i مقدار یک متغیر خروجی دو تایی، و x_i مقدار بردار متغیر توضیح دهنده i ام است. برای پرداخت مدل رگرسیون لوجستیک (1) برای اطلاعات موجود، لازم است که بردار ضرایب β برآورد شود. این کار در رگرسیون لوجستیک با استفاده از روش درستنمایی بیشینه صورت می‌گیرد. در این روش بردار ضرایب β بنحوی تعیین می‌شود که احتمال پیش‌بینی مشاهدات بیشینه شود. برای این کار ابتدا تابعی به نام تابع درستنمایی تعریف شود. این تابع احتمال اطلاعات مشاهده شده را به صورت تابعی از بردار ضرایب β بیان می‌کند. این ضرایب طوری تعیین می‌شوند که تابع اخیر بیشینه شود. لگاریتم تابع درستنمایی بر حسب β به صورت زیر تعریف می‌شود [9]:

$$L(\beta) = \sum_{i=1}^n \{y_i \ln[\pi(x_i)] + (1 - y_i) \ln[1 - \pi(x_i)]\} \quad (3)$$

برای یافتن β که $L(\beta)$ را بیشینه کند کافی است گرادیان $L(\beta)$ برابر صفر قرار داده شود. β بدست آمده از حل معادله اخیر تخمین درستنمایی بیشینه بردار ضرایب نامیده می‌شود و با $\hat{\beta}$ نشان داده می‌شود. تخمین درستنمایی بیشینه تابع رگرسیون لوجستیک $\pi(x)$ با $\hat{\pi}(x)$ نشان داده می‌شود، و برابر مقدار حاصل از رابطه (1) به ازای $\hat{\beta}$ است. به همین صورت، $\hat{g}(x)$ تخمین درستنمایی بیشینه تابع لوجیت $g(x)$ است که با استفاده از $\hat{\beta}$ در رابطه (2) تعیین می‌شود.

پس از تخمین ضرایب، نوبت به ارزیابی اهمیت متغیرها در مدل می‌رسد. این کار معمولاً با آزمون یک یا چند فرض آماری صورت می‌پذیرد. در این مقاله از آزمون والد و آزمون نسبت درستنمایی برای این منظور استفاده می‌شود

آزمون والد



آزمون والد (Wald) برای ارزیابی اهمیت آماری هر ضریب β_j در مدل استفاده می‌شود. آماره آزمون والد به صورت زیر محاسبه می‌شود [9]:

$$W_j = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \quad (4)$$

که در آن $\hat{\beta}_j$ تخمین درست‌نمایی بیشینه ضریب β_j و $SE(\hat{\beta}_j)$ برآوردی از انحراف معیار آن است. تحت فرض $\beta_j = 0$ (فرض صفر)، نسبت بالا دارای توزیع نرمال استاندارد خواهد بود. با تعریف z به عنوان یک متغیر تصادفی دارای توزیع نرمال استاندارد، اگر $P(|z| > W_j) < \alpha$ آنگاه فرض صفر را می‌توان در سطح اهمیت α رد کرد.

آزمون نسبت درست‌نمایی

آزمون نسبت درست‌نمایی برای ارزیابی اهمیت کلی ضرایب (p ضریب) مربوط به متغیرهای توضیح دهنده انجام می‌شود. این آزمون بر اساس آماره G زیر است [9 و 10]:

$$G = -2[L(C) - L(\hat{\beta})]. \quad (5)$$

در رابطه بالا، لگاریتم درست‌نمایی $L(\hat{\beta})$ با استفاده از $\hat{\pi}(x)$ در رابطه (3) حساب می‌شود. همچنین، $L(C)$ لگاریتم درست‌نمایی برای مدل بدون متغیرهای توضیح دهنده، یعنی $\hat{g}(x) = \beta_0$ است. آماره G ، تحت فرض صفر $\beta = 0$ ، دارای توزیع کای-مجدور با p درجه آزادی است. اگر $\chi^2(p)$ نشانگر یک متغیر تصادفی کای-مجدور با p درجه آزادی باشد، آنگاه در صورت $P(\chi^2(p) > G) < \alpha$ فرض صفر در سطح اهمیت α رد می‌شود. در حالت عمومی، آزمون درست‌نمایی برای ارزیابی اهمیت آماری ضرایب مربوط به تعدادی از متغیرهای توضیح دهنده مدل نیز استفاده می‌شود. در این حالت، برای تعیین اهمیت هر زیرمجموعه $K = \{x_{j_1}, x_{j_2}, \dots, x_{j_{|K|}}\}$ از مجموعه متغیرهای توضیح دهنده، $K \subseteq \{x_1, x_2, \dots, x_p\}$ ، آماره G به صورت زیر تعریف می‌شود:

$$G = -2[L(\hat{\beta}_{\bar{K}}) - L(\hat{\beta})], \quad (6)$$

که در آن، $L(\hat{\beta}_{\bar{K}})$ لگاریتم درست‌نمایی برای مدل بدون متغیرهای توضیح دهنده K است. تحت فرض صفر $(\beta_{j_1}, \beta_{j_2}, \dots, \beta_{j_{|K|}}) = 0$ ، آماره G دارای توزیع کای-مجدور با $|K|$ درجه آزادی خواهد بود.

توضیح اطلاعات

اطلاعات مورد استفاده در این مقاله شامل یک نمونه 3647 عددی از تصادفات رخ داده در بزرگراه چمران در سال 1385 است. برای هر تصادف یک برگه کروکی شامل اطلاعات مربوط به آن تصادف موجود است. با توجه به اینکه هدف مطالعه تعیین اثرات دوربرگردان و نیز سایر عوامل مؤثر در زمان وقوع تصادف (بر حسب شبانه یا روزانه) می‌باشد، 4 متغیر از اطلاعات اخیر گردآوری شدند. توصیف و دسته‌بندی این متغیرها در جدول 1 داده شده‌اند. متغیر پاسخ (وابسته) متغیر شماره 1 به نام ACCIDENT است، که طبیعتی دوتایی دارد. دو دسته برای این متغیر موجود است: 1 اگر تصادف شبانه باشد، و 0 اگر تصادف روزانه باشد. متغیرهای توضیح دهنده 2، 3 و 4 نیز دسته بندی شده‌اند. چون دو متغیر توضیح دهنده 3 و 4 دارای بیش از دو دسته بنامهای 1، 2، 3 و 4 هستند، مجموعه‌ای از متغیرهای طراحی (متغیرهای 0 یا 1) برای نمایش متغیرها در تابع رگرسیون لجستیک و نیز تنظیم فرمت فایل ورودی به نرم افزار SPSS [11] (نرم افزار مورد استفاده در این مطالعه) لازم است. متغیر شماره 2 محل تصادف (LOCATION) است که با توجه به موضوع این مقاله شامل دو دسته داخل دوربرگردان و خارج دوربرگردان می‌باشد. هر تصادف با توجه به نمایش یا عدم نمایش دوربرگردان در برگه کروکی آن تصادف در یکی از دو دسته اخیر قرار می‌گیرد. متغیر شماره 3 سن راننده مقصر (AGE) است که برای آن چهار دسته کمتر از 30 سال، 30 تا 40 سال، 40 تا 50 سال، و بالای 50 سال در نظر گرفته شده است. بالاخره، متغیر شماره 4 یعنی نوع برخورد (MANNER) دارای چهار دسته طبق جدول 1 است.

هر متغیر دارای k دسته را می‌توان توسط $K-1$ متغیر طراحی در تابع $g(x)$ نمایش داد. برای نمونه، متغیر سن (AGE) که دارای 4 دسته است را

جدول 1- توضیح متغیرهای مطالعه

شماره	توصیف	شماره	ارزش	نام متغیر	تعداد
-------	-------	-------	------	-----------	-------



متغیر	دسته			
1	تصادف	○	روزانه (7 تا 22) شبانه (22 تا 7)	3122 526
2	محل وقوع	1	داخل دوربرگردان	726
		2	خارج دوربرگردان	2921
3	سن راننده مقصر	1	کمتر از 30 سال	1816
		2	30 تا 40 سال	838
		3	40 تا 50 سال	651
		4	بالای 50 سال	342
4	نحوه برخورد	1	با جسم ثابت	253
		2	با موتور	123
		3	جلو به پهلو	847
		4	جلو به عقب	2424

می‌توان با 3 متغیر طراحی AGE_1، AGE_2 و AGE_3 تعریف کرد. وقتی که سن راننده بالای 50 سال است، هر سه متغیر طراحی اخیر برابر صفر قرار داده می‌شوند. وقتی سن راننده زیر 30 سال باشد، AGE_1 برابر 1 قرار داده می‌شود ولی دو متغیر طراحی دیگر هنوز برابر 0 هستند. متغیرهای طراحی AGE_2 و AGE_3 نیز به همین ترتیب مقدار دهی می‌شوند.

توضیح مدل

متغیر ACCIDENT به عنوان متغیر وابسته در این مطالعه متغیری دوتایی و بیانگر شدت تصادف است. هر تصادف در اطلاعات نمونه این مطالعه در یکی از دو دسته روزانه (Daily) یا شبانه (Nightly) قرار می‌گیرد. مدل لوجستیک استفاده شده به شرح زیر است:

$$P(\text{Nightly accident}) = \pi(x) = \frac{e^{(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)}}{1 + e^{(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)}} \quad (7)$$

در نتیجه

$$P(\text{Daily accident}) = 1 - \pi(x) = \frac{1}{1 + e^{(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)}} \quad (8)$$

در این مدل، با توجه به جدول 1، بردار متغیرهای توضیح دهنده X می‌تواند شامل حداکثر 7 متغیر طراحی باشد ($p = 7$). در رگرسیون لوجستیک، بردار ضرایب β با استفاده از روش درستیابی بیشینه بنحوی تعیین می‌شود که احتمال پیش بینی مشاهدات بیشینه شود. سپس، آزمونهای نظیر آزمونهای والد و نسبت درستیابی به عنوان معیارهایی برای حذف متغیرهای توضیح دهنده کم اهمیت از مدل مورد استفاده قرار می‌گیرند.

ساخت مدل

در این مطالعه از روش انتخاب رو به عقب برای تعیین متغیرهای مستقل مهم در مدل رگرسیون لوجستیک استفاده می‌شود. به منظور حذف متغیرهای بی‌اهمیت، در گام 1 کلیه متغیرهای توضیح دهنده بر اساس آزمون والد آزمایش می‌شوند. نتایج پرداخت یک مدل رگرسیون لوجستیک برای متغیر طراحی مربوط به متغیرهای توضیح دهنده AGE، LOCATION و MANNER در جدول 2 نشان داده شده است. در این جدول، متغیر طراحی مربوط به LOCATION با LOCATION_1، سه متغیر طراحی مربوط به AGE با AGE_1، AGE_2 و AGE_3، و سه متغیر طراحی مربوط



جدول 2- نتایج پرداخت یک مدل رگرسیون لوجستیک با استفاده از متغیرهای AGE، LOCATION و MANNER در مطالعه اثر دوربردگان بر زمان وقوع تصادفات

متغیر	$\hat{\beta}_j$	$SE(\hat{\beta}_j)$	W_j	$P(z > W_j)$	$\exp(\hat{\beta}_j)$
LOCATION_1	0/528	0/133	15/68	0/000	1/695
AGE_1	-0/004	0/169	0/00	0/980	0/996
AGE_2	-0/290	0/189	2/36	0/125	0/748
AGE_3	-0/062	0/192	0/10	0/747	0/940
MANNER_1	0/799	0/175	20/86	0/000	2/222
MANNER_2	-1/982	0/716	7/66	0/006	0/138
MANNER_3	1/069	0/104	106/02	0/000	2/912
Constant	-2/493	0/196	161/05	0/000	0/083

$$L(\hat{\beta}) = 1430/8295$$

به MANNER با MANNER_1، MANNER_2، MANNER_3 و MANNER نشان داده شده‌اند. ضرایب برآورد شده توسط روش درستنمایی بیشینه برای 7 متغیر طراحی اخیر در ستون دوم جدول 2 نشان داده شده است. تخمین درستنمایی بیشینه تابع لوجیت مربوط به این مدل به شرح زیر است:

$$\hat{g}_1(x) = -2/493 + 0/528 \text{LOCATION}_1 - 0/004 \text{AGE}_1 - 0/290 \text{AGE}_2 - 0/062 \text{AGE}_3 + 0/799 \text{MANNER}_1 - 1/982 \text{MANNER}_2 + 1/069 \text{MANNER}_3, \quad (9)$$

لگاریتم درستنمایی داده شده در پایین جدول 2 برابر مقدار حاصل از رابطه (3) به ازای ضرایب برآورد شده $\hat{\beta}$ است. چهار ستون دیگر نیز در جدول 2 ارائه شده‌اند. ستونهای سوم و چهارم به ترتیب خطای معیار و آماره آزمون والد هر یک از ضرایب مدل پرداخت شده را نشان می‌دهند. مقادیر احتمال $P(|z| > W_j)$ نیز در ستون پنجم این جدول آورده شده‌اند. اگر سطح اهمیت 0/05 مورد استفاده قرار گیرد، می‌توان نتیجه گرفت که متغیرهای طراحی LOCATION و MANNER در این سطح مهم هستند، در حالی که متغیر AGE مهم نیست. بالاخره، ستون آخر نشانگر عدد e به توان ضریب پرداخت شده است، که در بخش تفسیر مدل مورد استفاده قرار می‌گیرد. به منظور دستیابی به بهترین مدلی که در عین حال دارای کمترین تعداد پارامتر نیز باشد، قدم منطقی بعدی پرداخت کردن مدلی خلاصه شده شامل فقط متغیرهای توضیح دهنده مهم، و سپس مقایسه آن با مدل کامل شامل تمام متغیرها است. نتایج پرداخت این مدل خلاصه شده در جدول 3 نشان داده شده است.

جدول 3- نتایج پرداخت یک مدل رگرسیون لوجستیک با استفاده از متغیرهای LOCATION و MANNER در مطالعه اثر دوربردگان بر زمان وقوع تصادفات

متغیر	$\hat{\beta}_j$	$SE(\hat{\beta}_j)$	W_j	$P(z > W_j)$	$\exp(\hat{\beta}_j)$
LOCATION_1	0/537	0/133	16/27	0/000	1/710
MANNER_1	0/793	0/175	20/61	0/000	2/209
MANNER_2	-1/965	0/716	7/53	0/006	0/140
MANNER_3	1/067	0/104	105/91	0/000	2/907
Constant	-2/574	0/135	362/42	0/000	0/076

$$L(\hat{\beta}) = 1434/1000$$

ضرایب برآورد شده برای 4 متغیر طراحی مدل خلاصه شده طبق ستون دوم جدول 3 است. بر این اساس، تخمین تابع لوجیت مربوط به این مدل به شرح زیر خواهد بود:

$$\hat{g}_2(x) = -2/574 + 0/537 \text{LOCATION}_1 + 0/793 \text{MANNER}_1 - 1/965 \text{MANNER}_2 + 1/067 \text{MANNER}_3, \quad (10)$$

اختلاف مدل بالا با مدل داده شده در جدول 2 همان حذف 3 متغیر طراحی مربوط به AGE است. استفاده از آزمون درستنمایی برای مقایسه این دو مدل با تعریف آماره G طبق رابطه (6) امکانپذیر است. تحت این فرض که ضرایب متغیرهای طراحی حذف شده برابر صفر است، آماره G این آزمون دارای توزیع کای-مجدور با 3 درجه آزادی است. مقدار این آماره با استفاده از مقادیر لگاریتم درستنمایی داده شده در جداول 2 و 3 به صورت زیر



محاسبه می‌شود:

$$G = -2[1433/600 - 1430/8295] = 5/541$$

که با 3 درجه آزادی دارای احتمال $P(\chi^2(3) > 5/541) > 0/05$ است. چون این احتمال از 0/05 بیشتر است، حذف متغیر AGE از مدل کامل در سطح اهمیت 0/05 تایید می‌شود. به عبارت دیگر، مدل خلاصه شده به خوبی مدل کامل است. آمارهای آزمون والد مربوط به مدل خلاصه شده نیز در ستون چهارم جدول 3 داده شده‌اند. مقادیر احتمالهای مربوط به این آمارها در ستون پنجم نشان می‌دهند که هر دو متغیر LOCATION و MANNER این مدل در سطح اهمیت 0/05 مهم هستند.

قدم آخر این روش مدل‌سازی همان مقایسه مدل اخیر با مدلی فقط شامل متغیر LOCATION است. نتایج پرداخت این مدل در جدول 4 ارائه شده است. همانطور که در این جدول ملاحظه می‌شود، متغیر طراحی LOCATION_1 در سطح اطمینان 0/05 مهم است. بر این اساس این جدول، تخمین تابع لوجیت مربوط به این مدل به شرح زیر خواهد بود:

$$\hat{g}_2(x) = -2/089 + 0/375 \text{LOCATION}_1 \quad (11)$$

جدول 4- نتایج پرداخت یک مدل رگرسیون لوجستیک با استفاده از متغیر LOCATION در مطالعه اثر دوربرگردان در زمان وقوع تصادفات					
متغیر	$\hat{\beta}_j$	SE($\hat{\beta}_j$)	W_j	p_j	$\exp(\hat{\beta}_j)$
LOCATION_1	0/375	0/129	8/43	0/004	1/455
Constant	-2/089	0/119	310/58	0/000	0/124

$L(\hat{\beta}) = 1500/1300$

آزمون درست‌نمایی بیشینه برای مقایسه این مدل با مدل شامل متغیرهای طراحی LOCATION و MANNER نتیجه می‌دهد

$$G = -2[1500/1300 - 1433/600] = 133/06,$$

و احتمال مربوط به این آزمون با 3 درجه آزادی برابر $P(\chi^2(3) > 133/06) < 0/05$ خواهد بود، که در سطح $\alpha = 0/05$ مهم است. در این حالت فرض صفر رد می‌شود. در نتیجه، سه متغیر طراحی حذف مربوط به MANNER مهم خواهند بود. حذف آنها تایید نمی‌شود. شایان ذکر است که حذف یا افزودن متغیرها به مدل تنها بر پایه اهمیت آماری استوار نیستند. بلکه، برخی برخی متغیرها ممکن است به دلیل اهمیت کاربردی، حتی در صورت عدم اهمیت آماری، به مدل اضافه شوند. با توجه به موضوع این مطالعه، یعنی بررسی اثر محل تصادف (داخل یا خارج دوربرگردان) بر زمان وقوع آن، متغیر LOCATION دارای اهمیتی کاربردی است و تحت هر شرایطی در مدل حفظ می‌شود. بر اساس روش مدل‌سازی ارائه شده در بالا، مدل رگرسیون لوجستیک زیر برای تخمین احتمال وقوع یک تصادف شبانه پیشنهاد می‌شود:

$$P(\text{Nightly accident}) \approx \hat{\pi}(x) = \frac{e^{\hat{g}_2(x)}}{1 + e^{\hat{g}_2(x)}}$$

که در آن $\hat{g}_2(x)$ تخمین تابع لوجیت طبق رابطه (7) است.

تفسیر مدل

پس از پرداخت یک مدل و ارزیابی اهمیت آماری ضرایب آن، نوبت به تفسیر مقادیر ضرایب می‌رسد. تفسیر ضرایب هر مدل همان استنتاج‌های عملی است که بر پایه آن ضرایب و به منظور پاسخ دادن به پرسش‌های مطالعه صورت می‌گیرد. در مدل رگرسیون لوجستیک، مقدار هر ضریب $\hat{\beta}_j$ نشانگر میزان تغییر در تابع لوجیت به ازای یک واحد افزایش متغیر مستقل x_j است (یا $\hat{\beta}_j(x_j + 1) - \hat{\beta}_j x_j = \hat{\beta}_j$). در واقع، تفسیر صحیح این ضریب در مدل نیازمند یافتن معنایی برای اختلاف بین دو تابع لوجیت است. به منظور تفسیر این اختلاف از معیاری به نام نسبت شانس استفاده می‌شود. در



صورتیکه $\hat{\beta}_j$ یک متغیر طراحی \circ یا 1 باشد، این معیار طبق تعریف برابر با نسبت شانس وقوع پاسخ در صورت $x_j = 1$ به شانس وقوع آن در صورت $x_j = \circ$ است. شانس متغیر پاسخ برای $x_j = 1$ به صورت $\hat{\pi}(x_1, \dots, x_j = 1, \dots, x_p) / [1 - \hat{\pi}(x_1, \dots, x_j = 1, \dots, x_p)]$ و شانس متغیر پاسخ در صورت $x_j = \circ$ به صورت $\hat{\pi}(x_1, \dots, x_j = \circ, \dots, x_p) / [1 - \hat{\pi}(x_1, \dots, x_j = \circ, \dots, x_p)]$ تعریف می‌شوند. تقسیم دو رابطه اخیر نتیجه می‌دهد:

$$\frac{e^{(\hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_j + \dots + \hat{\beta}_p x_p)} / [1 + e^{(\hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_j + \dots + \hat{\beta}_p x_p)}]}{e^{(\hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_j + \dots + \hat{\beta}_p x_p)} / [1 + e^{(\hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_j + \dots + \hat{\beta}_p x_p)}]} = \frac{e^{(\hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_j + \dots + \hat{\beta}_p x_p)} / [1 + e^{(\hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_j + \dots + \hat{\beta}_p x_p)}]}{e^{(\hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_j + \dots + \hat{\beta}_p x_p)} / [1 + e^{(\hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_j + \dots + \hat{\beta}_p x_p)}]} = e^{\hat{\beta}_j}$$

در نتیجه، برای رگرسیون لجستیک، نسبت شانس برای هر متغیر طراحی x_j برابر خواهد شده با

$$OR = e^{\hat{\beta}_j}$$

نسبت شانس تقریب می‌زند که احتمال پاسخ به ازای $x_j = 1$ چند برابر احتمال پاسخ برای $x_j = \circ$ خواهد بود. رابطه ساده فوق برای تعیین نسبت شانس از جمله نقاط قوت مدل رگرسیون لجستیک محسوب می‌شود.

ستون آخر آن جداول 2 تا 4 نشانگر نسبت‌های شانس مربوط ضرایب مدل‌های رگرسیون لجستیک داده شده در آن جداول است. همانطور که ملاحظه می‌شود، نسبت‌های شانس مربوط به متغیر طراحی LOCATION_1 در هر سه مدل اخیر بزرگتر از 1 است، که تخمین می‌زنند تصادفات داخل دوربرگردان با احتمال بیشتری نسبت به تصادفات خارج دوربرگردان در ساعات شب رخ می‌دهند. در مدل رگرسیون لجستیک برگزیده شده در این مقاله، یعنی مدل (10) که نتایج پرداخت آن در جدول 3 ارائه شده است، نسبت شانس اخیر برابر 1/7 تعیین شده است. به عبارت دیگر، احتمال وقوع تصادف شبانه در میان تصادفات داخل دوربرگردان 1/7 برابر احتمال آن در میان تصادفات خارج دوربرگردان است.

نتیجه گیری

استفاده از دوربرگردان به جای تقاطع چراغدار، وقتی حجم ترافیک حرکت‌های گردش سبک باشد، در کشورهای مختلف دنیا مرسوم است. در بزرگراه‌های شهر تهران اخیراً دوربرگردان‌های متعددی احداث شده است که علت احداث آنها معمولاً کم بودن میانگین سرعت وسایل نقلیه در ساعات اوج عنوان می‌شود. این در حالی است که استفاده از این دوربرگردانها در ساعات غیر اوج، با توجه به سرعت زیاد وسایل نقلیه در باند سمت چپ، باعث افزایش وقوع تصادفات شدید به ویژه در ساعات شب شده است.

در این مطالعه با استفاده از آمار غیرهمفزود تصادفات بزرگراه چمران در سال 1385 (در مجموع تعداد 3658 تصادف) یک مدل رگرسیون لجستیک پرداخت شد که متغیر وابسته آن احتمال شبانه بودن تصادف و متغیرهای مستقل آن شامل محل وقوع تصادف (داخل دوربرگردان یا خارج دوربرگردان) و نیز نحوه برخورد (با جسم ثابت، با موتور، جلو به پهلو، و جلو به عقب) بوده‌اند. این مدل نشان داد که احتمال وقوع تصادف شبانه در میان تصادفات داخل دوربرگردان 1/7 برابر احتمال آن در میان تصادفات خارج دوربرگردان است. به عبارت دیگر، احداث دوربرگردان در بزرگراه چمران موجب افزایش احتمال وقوع تصادفات شبانه، که اغلب شدید هستند، شده است.

مراجع

1. Comte, S.L. and Jamson, A.H. (2000) Traditional and innovative speed-reducing measures for curves: an investigation of driver behaviour using a driving simulator. *Safety Science*, **36**, 137–150.
2. Finch, D.J., Kompfner, P., Lockwood, C.R. and Maycock, G. (1994) Speed, speed limits and accidents. *Project Report 58 of Transportation Research Laboratory*, Crowthorne, UK.
3. Navon, D. (2003) The paradox of driving speed: two adverse effects on highway accident rate. *Accident Analysis and Prevention*, **35**, 361–367.
4. Hutchings, C.B., Knight, S. and Reading, J.C. (2003) The use of generalized estimating equations in the analysis of motor vehicle crash data, *Accident Analysis and Prevention*, **35**, 3–8.
5. Greibe, P. (2003) Accident prediction models for urban roads. *Accident Analysis and Prevention*, **35**, 273–285.



6. Sawalha, Z. and Sayed, T. (2003). Statistical issues in traffic accident modeling. *2003 Annual Meeting of the Transportation Research Board*, Washington, DC.
7. Kockelman, K.M. and Kweon, Y.J. (2001) Driver injury severity: An application of ordered probit models. Re-Submitted to *Accident Analysis and Prevention*.
8. Wood, G.R. (2002). Generalized linear accident models and goodness of fit testing. *Accident Analysis and Prevention*, **34**, 417–427.
9. Hosmer, D.W. and Lemeshow, S. (2000) *Applied Logistic Regression*. John Wiley & Sons Inc., New York.
10. Myers, R.H., Montgomery, D.C. and Vining, G.G. (2002) *Generalized Linear Models: with Applications in Engineering and Science*, John Wiley & Sons Inc., New York.
11. *SPSS Statistical Software* (2004) Version 13.0 for Windows. SPSS Inc., Chicago.