

영과잉을 고려한 중심상업지구 교통사고모형 개발에 관한 연구

Safety Performance Functions for Central Business Districts Using a Zero-Inflated Model

이 상 혁 Lee, Sang Hyuk | 정회원 · 한국건설기술연구원 도로연구소 수석연구원 (E-mail : sanghyuklee8612@gmail.com)
우 용 한 Woo, Yong Han | 정회원 · 경일대학교 건설공학부 부교수 · 교신저자 (E-mail : yhwoo333@naver.com)

ABSTRACT

PURPOSES : The purpose of this study was to develop safety performance functions (SPFs) that use zero-inflated negative binomial regression models for urban intersections in central business districts (CBDs), and to compare the statistical significance of developed models against that of regular negative binomial regression models.

METHODS : To develop and analyze the SPFs of intersections in CBDs, data acquisition was conducted for dependent and independent variables in areas of study. We analyzed the SPFs using zero-inflated negative binomial regression model as well as regular negative binomial regression model. We then compared the results by analyzing the statistical significance of the models.

RESULTS : SPFs were estimated for all accidents and injury accidents at intersections in CBDs in terms of variables such as AADT, Number of Lanes at Major Roads, Median Barriers, Right Turn with an Exclusive Turn Lane, Turning Guideline, and Front Signal. We also estimated the log-likelihood at convergence and the likelihood ratio of SPFs for comparing the zero-inflated model with the regular model. In the SPFs, estimated log-likelihood at convergence and the likelihood ratio of the zero-inflated model were at -836.736, 0.193 and -836.415, 0.195. Also estimated the log-likelihood at convergence and likelihood ratio of the regular model were at -843.547, 0.187 and -842.631, 0.189, respectively. These figures demonstrate that zero-inflated negative binomial regression models can better explain traffic accidents at intersections in CBDs.

CONCLUSIONS : SPFs that use a zero-inflated negative binomial regression model demonstrate better statistical significance compared with those that use a regular negative binomial regression model.

Keywords

Safety performance functions, urban intersections, central business districts, negative binomial regression model, zero inflation

Corresponding Author : Woo, Yong Han, Associate Professor
Kyungil University, 50, Gamasil-gil, Hayang-eup, Gyeongsan-si,
Gyeongbuk, 38428, Korea
Tel : +82.53.600.5427 Fax : +82.53.600.5439
E-mail : yhwoo333@naver.com

International Journal of Highway Engineering

<http://www.ksre.or.kr/>

ISSN 1738-7159 (print)

ISSN 2287-3678 (Online)

Received May, 24, 2016 Revised Jul, 12, 2016 Accepted Jul, 15, 2016

1. 서론

1.1. 연구의 배경 및 목적

1990년대 들어서면서 우리나라의 자동차등록대수와 차량이용이 급격히 증가함에 따라 교통사고 발생건수도 꾸준히 증가하여 2000년도에 교통사고 발생건수 290,481

건, 사망자수 10,236명으로 정점을 기록하였다. 이후 정부의 교통사고 다발지역 개선사업 등 교통사고 발생건수와 교통사고 사망자수를 줄이기 위한 노력으로 교통사고 발생건수가 2014년 223,552건, 사망자수 4,762명으로 감소하였다(Korean National Police Agency, 2015).

하지만 교통사고 사망자수는 OECD 국가 평균보다 약 2배 정도 높은 수준으로 교통사고 발생건수와 교통사고 사망자수를 감소시키기 위한 대책이 필요한 실정이다.

특히, 2014년 발생한 전체 교통사고 중 99,068건은 교차로 내에서 발생한 교통사고로서 전체 교통사고 발생의 약 44%를 차지하고 있으며, 이는 가로구간이나 횡단보도 교통사고에 비해 발생건수가 높게 나타나고 있다. 또한, 교차로 내의 교통사고는 신호위반이나 차량간의 상충 등으로 인해 다른 교통사고유형과 달리 중상과 사망사고와 관련된 교통사고가 많이 발생하고 있다(Retting et al., 1999).

특히, 도시지역 중심상업지구(central business districts, CBD)는 도시 내에서 상업, 금융, 서비스 시설이 집중되어 있는 핵심지역으로 CBD 내 교차로는 교통량과 보행자수 증가에 대비한 기하구조 및 교통운영체계 개선이 이루어지지 못하고 있어 교통사고 발생위험이 다른 지역에 비해 높게 나타나고 있다. 그럼에도 불구하고 도시지역 CBD 내 교차로에 대한 교통사고발생 원인을 분석하고자 하는 연구가 활발하게 이루어지지 못하고 있다.

교통사고 발생원인은 교통사고모형(safety performance functions, SPFs)을 통하여 분석할 수 있다. 교통사고모형은 특정지점 또는 지역의 교통사고 발생빈도 또는 교통사고 심각도와 이에 영향을 미치는 요인들과의 관계를 분석하기 위해 모형화한 것이다. 교통사고모형은 교통사고 발생건수 또는 교통사고 심각도를 종속변수로 하고 교통량(annual average daily traffic, AADT), 도로 기하구조(geometric information), 교통운영정보(traffic operation) 등 교통사고 발생요인을 독립변수로 하여 개발할 수 있다.

교통사고모형개발 시 교통사고데이터는 항상 양의 수(positive) 또는 영(0)이고, 교통사고는 이산적으로 발생하므로 교통사고모형으로 선형모형은 적합하지 않다. 또한, 교통사고데이터의 오차항(error term)이 선형모형처럼 정규분포(normal distribution) 형태를 띠지 않기 때문에 교통사고모형은 포아송 회귀모형(Poisson regression model)이나 음이항 회귀모형(negative binomial regression model)을 이용하여 개발하는 것이 보통이다(Miaou and Lum, 1993; Jovanie and Chang, 2002; Kweon 2012; Lee et al., 2015).

포아송 회귀모형은 주어진 장소에서 발생하는 사건(events)이 각각 독립적이고 주어진 시간 또는 기간을

대상으로 사건이 발생할 확률을 모형화한 것이다. 하지만, 포아송 회귀모형의 경우 종속변수의 분산이 분산의 평균과 같아야 하는 조건을 만족해야 한다. 만약 종속변수의 분산이 분산의 평균보다 크다면 이를 과분산(over-dispersion)이라고 하며, 종속변수가 과분산 시 계수 벡터 β 는 편향(biased)되는 특징이 있다(Carlin and Louis, 2009; Lee et al., 2016). 이러한 경우 교통사고모형은 포아송 회귀모형 대신 음이항 회귀모형을 이용하여 개발할 수 있다(Chin and Quddus, 2003, Lee et al., 2015).

또한, 교통사고데이터가 조사대상연도나 조사대상구간 또는 교차로에서 교통사고가 전혀 발생하지 않는 경우가 있다. 이러한 경우 영과잉(Zero Inflation)을 고려하여 교통사고모형을 개발하는 것이 적절하다(Lord et al., 2005; Na and Park, 2012; Lee et al., 2012).

본 연구에서는 국내 교통사고발생데이터(traffic accident frequency data)와 교차로 교통량, 교차로 지형정보, 교통운영정보 등을 이용하여 도시지역 CBD의 교통사고모형을 개발하여 CBD의 교통사고 발생 원인을 분석하고자 한다. 특히, 교통사고모형의 적합성과 교통사고 원인분석의 정확성을 높이고자 영과잉(Zero Inflation)모형을 적용하여 모형을 개발하고 적용하지 않은 모형과 비교·분석하였다.

1.2. 연구의 범위 및 내용

본 연구에서는 대전광역시 CBD 내의 31개의 교차로를 대상으로 자료를 수집·가공하여 교통사고모형을 개발하였다. 교통사고발생데이터는 경찰청과 도로교통공단에서 개발하여 운영 중인 교통사고분석시스템(Traffic Accident Analysis System, TAAS)에서 제공하고 있는 2007년~2014년까지 8년간의 교통사고발생 중 사망사고(Fatality), 중상사고(Injury A), 경상사고(Injury B) 데이터를 이용하였다. 교통량(AADT)은 현장조사와 '대전광역시 교통조사·분석 보고서'를 통해 수집하였으며, 연구대상 교차로의 지형정보 및 기하구조정보는 현장조사와 Google Map을 통하여 수집하였다.

교통사고모형은 전세계에서 범용적으로 이용되고 있는 IBM SPSS Statistics 20과 NLOGIT 4.0 econometrics을 이용하여 개발하였다. 영과잉을 고려한 교통사고모형은 교통사고발생에 영향을 미친다고 판단되는 변수의 종류와 개수를 여러 가지로 조합·대입하여 최적의 모형 적합도를 가진 모형을 선택할 수 있는 기법인 단계적 방법(stepwise manner)을 이용하여 개발하였다. 개발된

영과잉을 고려한 모형과 그렇지 않은 모형에 대한 효용성을 분석하여 모형의 적용성을 비교·분석하였다.

2. 교통사고모형의 이론적 배경

교통사고모형은 도로의 특정지역, 구간, 지점 등에서 발생하는 교통사고를 예측하거나 발생한 교통사고를 설명하기 위하여 교통사고와 관련 있는 교통정보, 지형 및 기하구조 정보, 교통운영정보를 독립변수로 하는 통계학적 모형이다. 본 연구에서는 대전광역시 CBD 내의 교차로에서 발생한 전체교통사고(사망·중상·경상교통사고)와 부상교통사고(중상·경상교통사고)에 대한 교통사고모형을 개발하여 CBD 내 교차로에서 발생하는 교통사고 원인을 분석하였다. 교통사고모형을 개발하기 위하여 교통사고발생빈도를 종속변수로 하고, 교차로 전체 AADT, 교차로 주간선도로 AADT, 교차로 보조간선도로 AADT, 보행신호 카운트다운기 운영여부, 교차로 주변 토지이용, 제한속도(교차로에 접근하는 주간선도로), 교통섬 설치 유무, 차로수(교차로에 접근하는 주간선도로), 중앙분리대 설치 유무, 교차로 내 횡단보도 수, 좌회전 전용차로 설치 유무, 우회전 전용차로 설치 유무, 교차로 내 차량유도선 설치 유무, 전방신호등 설치 유무 등을 독립변수로 하여 교통사고모형을 개발하였다. 특히 AADT의 경우 연도별·교차로별 교통량 변화와 교통사고발생빈도의 상관관계가 높아 교통사고모형 개발 시 연도별·교차로별 AADT를 반영하여 모형을 개발하였다(Jonsson et al., 2009; Mitra and Washington, 2012).

교통사고모형은 포아송 회귀모형과 음이항 회귀모형을 이용하여 개발하였다. 포아송 모형은 각 CBD 내 교차로에서 발생하는 교통사고가 독립적이고 일정시간 또는 기간 내에 교통사고가 발생할 확률을 모형화한 통계학적 분포로 일반식은 다음과 같다(Lee et al., 2016).

$$P(Y = y_i) = \frac{\mu_i^{y_i} e^{-\mu_i}}{y_i!} \quad (1)$$

$$\mu_i = \exp(\beta_0 + \sum_{j=1}^n \beta_j x_{ij}) \quad (2)$$

여기서, $P(Y = y_i)$ 는 주어진 기간 동안 i 교차로에서 y_i 번의 교통사고가 발생할 확률, μ_i 는 주어진 기간동안 i 교차로에서 발생한 기대 교통사고빈도, x_{ij} 는 설명변수, β_i 는 계수의 벡터값이다.

포아송 모형을 교통사고모형 개발에 적용하기 위하여 우도함수(likelihood function)를 적용한 최우추정법(standard maximum likelihood method)을 이용하여 포아송 회귀모형을 개발할 수 있다.

포아송 회귀모형의 기본식은 다음과 같다.

$$L(\beta) = \sum_{i=1}^n (y_i \log \mu_i - \mu_i - \log y_i) \quad (3)$$

여기서, β 는 계수($\beta_0, \beta_1, \dots, \beta_n$)의 벡터값이고, y_i 는 i 교차로에서 관측된 교통사고발생빈도, μ_i 는 i 교차로에서 예측된 교통사고발생빈도이다.

포아송 회귀모형을 교통사고모형에 적용하기 위하여 종속변수의 분산과 분산의 평균이 같아야 한다는 제약 조건을 따라야 한다. 만약 종속변수의 분산이 분산의 평균보다 크다면 이를 과분산(over dispersion)이라고 하며, 과분산 시 계수의 벡터값 β 는 편향이 되어 모형의 적합도가 떨어지게 된다. 이러한 경우 교통사고모형을 포아송 회귀모형 대신 음이항 회귀모형을 이용하여 교통사고발생빈도를 예측할 수 있다(Jovanis et al., 1986; Lee et al., 2012; Lee, et al., 2016).

음이항 분포는 과분산을 모형에 반영하여 교통사고발생확률을 모형화한 통계학적 분포로 일반식은 다음과 같다.

$$P(Y = y_i) = \frac{\Gamma\left[\frac{1}{\alpha} + y_i\right]}{\Gamma\left[\frac{1}{\alpha}\right] * y_i!} \cdot \left[\frac{1}{(1 + \alpha\mu_i)}\right]^{\frac{1}{\alpha}} \cdot \left[1 - \frac{1}{1 + \alpha\mu_i}\right]^{y_i} \quad (4)$$

$$Var(Y) = \mu_i(1 + \alpha\mu_i) \quad (5)$$

여기서, $P(Y = y_i)$ 는 주어진 기간동안 i 교차로에서 y_i 번의 교통사고가 발생할 확률, μ_i 는 주어진 기간동안 i 교차로에서 발생한 기대 교통사고빈도, α 는 최우추정법에 의해 예측된 과포화계수, Γ 는 감마함수, $Var(Y)$ 는 분산이다.

음이항 모형을 교통사고모형 개발에 적용하기 위하여 포아송 회귀분석과 동일하게 우도함수를 적용한 최우추정법을 이용하여 음이항 회귀모형을 개발할 수 있다.

음이항 회귀모형의 기본식은 다음과 같다.

$$LL(\beta) = \sum_{i=1}^n \left\{ \left[\sum_{j=0}^{y_i} \log(1 + \alpha \cdot j) \right] - \log(1 + \alpha \cdot y_i) \right. \\ \left. + y_i \cdot \log \mu_i - (y_i + \alpha^{-1}) \log(1 + \alpha \cdot \mu_i) \right. \\ \left. - \log(y_i!) \right\} \quad (6)$$

여기서, α 는 과포화계수, β 는 계수($\beta_0, \beta_1, \dots, \beta_n$)의

벡터값, y_i 는 i 교차로에서 관측된 교통사고발생빈도, μ_i 는 i 교차로에서 예측된 교통사고발생빈도이다.

교통사고발생데이터는 연도별·교차로별·교통사고 심각도별 발생빈도를 정리한 자료이다. 하지만 교통사고 발생빈도가 연도별·교차로별·교통사고 심각도별로 전혀 없는 경우(zero accident)가 발생할 수 있다. 이러한 경우 영과잉모형을 적용하면 모형의 적합도를 향상시킬 수 있다. 영과잉모형은 반응값이 0인 부분과 0이 아닌 부분으로 나누어 분석하는 방법으로 확률변수 Y 는

$$Y \sim \begin{cases} 0 & \text{with Probability } p_i \\ P(Y=y_i) & \text{with Probability } 1-p_i \end{cases}$$

이다. 여기서, $0 \leq p < 1$ 이다(Lord et al., 2005; Na and Park, 2012; Lee et al, 2012).

영과잉을 고려한 포아송 모형의 기본식은 다음과 같다.

$$P(Y=0) = p_i + (1-p_i)e^{-\mu_i} \quad (7)$$

$$P(Y=y_i \neq 0) = (1-p_i) \frac{\mu_i^{y_i} e^{-\mu_i}}{y_i!} \quad (8)$$

또한, 영과잉을 고려한 음이항 모형의 기본식은 다음과 같다.

$$P(Y=0) = p_i + (1-p_i) \left[\frac{1}{(1+\alpha\mu_i)} \right]^{\frac{1}{\alpha}} \quad (9)$$

$$P(Y=y_i \neq 0) = (1-p_i) \cdot \frac{\Gamma\left[\frac{1}{\alpha} + y_i\right]}{\Gamma\left[\frac{1}{\alpha}\right] * y_i!} \cdot \left[\frac{1}{(1+\alpha\mu_i)} \right]^{\frac{1}{\alpha}} \cdot \left[1 - \frac{1}{1+\alpha\mu_i} \right]^{y_i} \quad (10)$$

교통사고모형의 적합도는 Akaike information criterion(AIC), Bayesian information criterion(BIC), Vuong test를 시행하여 검증하였다. AIC는 모형개발 시 단계적 방법(stepwise manner)를 적용하였을 경우 모형의 적합도(goodness of fit)을 검증하는 방법이고 BIC는 우도함수를 기준으로 한 모형선정 기준이다. 교통사고모형의 선정은 단계적 방법으로 산정된 여러 모형들 중에서 AIC와 BIC값이 가장 작은 값을 가지는 모형을 선택하여 교통사고모형 분석에 이용하였다.

AIC와 BIC의 기본식은 다음과 같다.

$$AIC = -2\ln L + 2K \quad (11)$$

$$BIC = -2\ln L + k \ln n \quad (12)$$

여기서, L 은 산정된 모형의 우도함수의 최대값이고, K 는 모형의 계수의 수, n 은 표본의 크기이다.

또한, 영과잉 모형 적합도는 Vuong test를 시행하여 검증할 수 있다. Vuong test는 영과잉 포아송 또는 음이항 회귀모형과 그렇지 않은 모형을 비교하기 위하여 산정하는 검증방식이며 기본식은 다음과 같다(Vuong, 1989; Mouatassim and Ezzahid, 2012).

$$V = \frac{\sqrt{n} \left[\left(\frac{1}{n} \right) \sum_{i=1}^n m_i \right]}{\sqrt{\left(\frac{1}{n} \right) \sum_{i=1}^n (m_i - \bar{m})^2}} = \frac{\sqrt{n} (\bar{m})}{S_m} \quad (13)$$

$$m_i = \text{Log} \left(\frac{P_1(y_i | x_i)}{P_2(y_i | x_i)} \right) \quad (14)$$

여기서, $P_N(y_i | x_i)$ 은 두 개의 비교모형의 예측확률이고, \bar{m} 과 S_m 는 m_i 의 평균과 표준편차, n 은 표본의 크기이다.

Vuong test에서 산정된 값 V 가 신뢰수준 0.05에서 1.96보다 크다면 첫 번째 모형이, V 가 -1.96보다 작다면 두 번째 모형이 적합한 것으로 판단할 수 있다. 만약 $|V|$ 이 1.96보다 작다면 비교하는 두 모형은 동일할 것으로 판단할 수 있다.

3. CBD 내 교차로 교통사고 모형개발

3.1. 독립변수의 정의 및 분석

본 연구는 대전광역시 중심상업지구(CBD) 내의 31개 교차로를 중심으로 교통사고발생데이터와 교통정보, 교차로 기하구조정보 등을 수집·분석하였다. 교통사고발생데이터는 횡단면 시계열데이터(cross-sectional time series data)로서 2007년부터 2014년까지 각 연구대상 교차로 8년간의 교통사고발생빈도를 연도별로 나누어 종속변수로 이용하였으며, 교통사고발생빈도는 전체교통사고발생빈도(사망교통사고발생빈도+중상교통사고발생빈도+경상교통사고발생빈도)와 전체부상사고발생빈도(중상교통사고발생빈도+경상교통사고발생빈도)로 나누어 교통사고모형을 개발하였다.

교통사고모형개발에 이용된 독립변수는 교통사고발생빈도에 영향을 미친다고 판단되는 교차로 년평균일교통량(AADT), 교차로 주간선도로 년평균일교통량(MAADT),

교차로 보조간선도로 년평균일교통량(SAADT), 보행신호 카운트다운기 운영여부(CD), 교차로에 접근하는 주간선 도로의 제한속도(SPLMT), 교통섬 설치 유무(ISLAND), 교차로 주간선도로의 차로수(NLANE), 중앙분리대 설치 유무(BAR), 교차로 내 횡단보도수(PED), 좌회전 전용차로 설치 유무(LP), 우회전 전용차로 설치 유무(RP), 교차로 내 차량유도선 설치 유무(GL), 전방신호등 설치 유무(FS) 등을 독립변수로 하여 모형을 개발하였다.

Table 1은 교통사고모형개발에 이용된 종속변수와 독립변수의 정의 및 기초통계분석 결과를 나타낸 것으로, 주요 변수의 기초통계분석 결과를 살펴보면, 연구대상 CBD 내 교차로에서 발생한 총 교통사고발생빈도(TOTAL)는 교차로당 연평균 15.863건으로 나타났으며, 이 중 사망교통사고발생빈도(FATAL)는 교차로당 연평균 0.234건으로 나타났다. 이는 같은 기간 대전광역시 88개 주요교차로에서 발생한 총 교통사고발생빈도(11.152건/교차로/년)와 사망교통사고발생빈도(0.168건/교차로/년)에 비해 약 30% 정도 높은 것으로 나타나 일반교차로와 비교하여 CBD 내의 교차로가 교통사고발생위험에 취약한 것으로 나타났다.

교통사고모형개발에 이용한 독립변수를 살펴보면 교차로 연평균일교통량은 평균 87,768대/일로 나타났으며, 최대 연평균일교통량은 148,443대/일, 최소 연평균일교통량은 25,341대/일로 CBD 특성상 교차로 통과교통량이 많은 것으로 나타났다. 본 연구에서 연평균일교통량은 기존 연구에서 이용된 교통량변수와 같이 로그값을 취하여 그 결과를 모형개발의 독립변수로 이용하였다(Mitra and Washington, 2012; Moutassim and Ezzahid, 2012; Lee et al., 2015, 2016).

차로수(NLANE)는 교차로에 접근하는 차로 중 주간선도로의 차로수로 정의하였으며, 평균 4.839차로/교차로, 최대 차로수 7차로/교차로, 최소 차로수 3차로/교차로로 나타나 CBD의 특성상 많은 교통량으로 인한 넓은 교차로 면적을 가지는 것으로 분석되었다. 중앙분리대(BAR)와 전방신호등(FS)의 경우 CBD 내 교차로의 차로수가 다른 지구(districts) 교차로의 차로수에 비해 많이 중앙분리대 설치와 전방신호등 설치가 각각 평균 0.581개/교차로와 0.516개/교차로로 높게 나타났다.

더불어, 본 연구에서 이용된 독립변수의 정의 및 기초통계량은 Table 1과 같다.

Table 1. Descriptive Statistics of Variables

Variables	Description	Mean	Standard deviation	Minimum	Maximum
TOTAL	Number of total accidents (/Intersection/Year)	15.863	9.734	0	52
FATAL	Number of fatal accidents (/Intersection/Year)	0.234	0.478	0	2
IA	Number of serious injury accidents (/Intersection/Year)	5.238	3.528	0	19
IB	Number of slight injury accidents (/Intersection/Year)	10.391	7.284	0	40
IAB	Number of all injury accidents (/Intersection/Year)	15.629	9.690	0	52
AADT	Annual average daily traffic (/Intersection/Year)	87767.800	25503.500	39790	148443
LAADT	Ln(AADT)	11.339	0.301	10.591	11.908
MAADT	Major road AADT	57996.300	18393.500	25341	103061
SAADT	Minor road AADT	29771.500	14931.200	7958	61239
CD	Intersections with countdown timer	0.613	0.488	0	1
SPLMT	Speed limits at major road	64.516	7.567	50	80
ISLAND	Traffic island in intersection	0.484	0.501	0	1
NLANE	Number of lanes at major road	4.839	1.112	3	7
BAR	Median barrier	0.581	0.494	0	1
PED	Number of pedestrian signal	3.387	1.290	0	5
LP	Left turn with an exclusive turn lane	0.710	0.455	0	1
RP	Right turn with an exclusive turn lane	0.613	0.488	0	1
GL	Turning guideline	0.968	0.177	0	1
FS	Front signal	0.516	0.501	0	1

3.2. 독립변수의 상관성 분석

교통사고모형개발 시 각 독립변수가 종속변수를 정확하게 설명할 수 있는 모형을 개발하여야 CBD 교차로에서 발생한 교통사고에 대한 분석 및 예측이 가능해진다. 이러한 경우, 독립변수 간의 상호 관련성이 없거나 낮아야 한다. 하지만, 독립변수 간의 상호 관련성이 존재하거나 높으면, 각 독립변수의 특성이 중복적으로 종속변수에 영향을 미쳐 표준오차가 큰 불안정한 회귀계수가 산정될 가능성이 높아진다. 따라서 본 연구에서는 독립변수 간의 상호 연관성을 파악하기 위하여 다중공선성 분석(multicollinearity test)을 시행하였다. 다중공선성은 모형의 독립변수 간 높은 상관성을 가진 통계학적 현상으로 정의할 수 있다.

다중공선성 분석을 위하여 공차(tolerances)와 분산팽창인자(variance inflation factor, VIF)를 산정하여야 한다. 공차와 VIF의 기본식은 다음과 같다.

$$tolerance = 1 - R_j^2 \quad (15)$$

$$VIF = \frac{1}{tolerance} \quad (16)$$

여기서, R_j^2 은 j 번째 독립변수를 종속변수로 하여 나머지 독립변수에 대해 회귀분석을 시행하였을 때 산정되는 모형결정계수이다. Table 2는 모형개발에 이용된 독립변수 간의 다중공선성 분석 결과이다.

Table 2. Results of Multicollinearity Test

Variables	Total accidents	
	Collinearity statistics	
	Tolerance	VIF
LAADT	0.244	4.104
CD	0.723	1.384
SPLMT	0.249	4.010
ISLAND	0.644	1.554
NLANE	0.325	3.073
BAR	0.457	2.189
PED	0.696	1.436
LP	0.244	4.096
RP	0.443	2.257
GL	0.586	1.706
FS	0.677	1.478

각 독립변수의 공차의 값은 0.10과 같거나 커야 하고, VIF 값은 10과 같거나 작아야 각 독립변수 간의 상관관

계가 없거나 낮다고 판단할 수 있다. 본 연구에서의 다중공선성 분석결과를 살펴보면 모형개발에 이용된 모든 독립변수의 공차값은 0.1보다 크고, VIF값은 10보다 작아 독립변수 간의 상호 연관성은 매우 약한 것으로 분석되었다.

3.3. 교통사고 모형개발

본 연구에서 대전광역시 CBD 내 교차로의 교통사고모형을 개발하기 위하여 종속변수로는 총 교통사고발생빈도(TOTAL)와 부상 교통사고발생빈도(IAB)로 하였으며, 독립변수로는 교통량정보, 교차로 기하구조 정보 등으로 하여 통계분석 software인 NLOGIT 4.0 econometrics를 이용하여 모형을 개발하였다. CBD 내 교차로 교통사고모형을 개발하기 위하여 교통사고모형개발에 많이 이용되고 있는 음이항 회귀모형을 적용하였으며, 특히 영과잉을 고려한 음이항 회귀모형을 적용하여 기존의 음이항 회귀모형과 비교하여 모형의 적합성을 분석하였다.

모형개발 방법은 모형개발 시 독립변수의 수와 종류를 단계적으로 대입하여 모형 적합도가 가장 높은 모형을 선정하는 방식인 단계적 방식(stepwise manner)으로 모형을 개발하였다. 단계적 방식은 F-통계량의 유의확률을 가장 낮게 하는 독립변수를 하나씩 선택하여 모형에 대입하는 방법으로, 본 연구에서는 CBD 내 교차로의 교통사고발생빈도에 영향을 많이 미친다고 판단되는 독립변수를 우선적으로 적용하여 모형을 개발하였다(Lee et al., 2015). 또한 CBD 내 교차로의 교통사고발생원인 분석과 영과잉 모형과 그렇지 않은 모형의 모형 적합도를 비교·분석하기 위하여 가능한 많은 독립변수를 모형개발에 적용하였으며, 또한 각 독립변수의 단위 변화(changing units)가 교통사고발생빈도에 미치는 영향을 분석하고자 각 산정된 계수에 대한 한계효과(marginal effects)를 산정하였다.

Table 3과 Table 4에서와 같이 대전광역시 CBD 내 교차로 교통사고모형 개발 결과, 총 교통사고발생빈도와 부상 교통사고발생빈도에 대한 교통사고모형에서 교차로 연평균일교통량(LAADT), 차로수(NLANE), 중앙분리대(BAR), 우회전전용차로(RP), 차량유도선(GL), 그리고 전방신호등(FS)이 교통사고발생빈도에 영향을 미치는 독립변수로 분석되었다. 개발된 각 교통사고모형에서 산정된 과포화계수 α 는 총 교통사고발생빈도에 대한 교통사고모형에서 음이항 회귀모형과 영과잉을 고려한 음이항 회귀모형이 각각 0.206(t-statistics: 8.123)과 0.171(t-statistics: 6.442)로 나타났으며, 부

상교통사고발생빈도에 대한 교통사고모형에서 음이항 회귀모형과 영과잉을 고려한 음이항 회귀모형은 각각 0.214(t-statistics:8.099)와 0.177(t-statistics:6.499)로 나타나 포아송 회귀모형보다는 음이항 회귀모형이 통계적으로 유의한 것으로 분석되었다. 또한, 영과잉 모형의 적합성을 검증하는 Vuong test의 경우 두 영과잉 모형 각각 2.390과 2.339로 나타나 영과잉 모형이 통계적으로 유의한 것으로 분석되었다.

4. 모형개발 결과 및 분석

본 연구에서는 산정된 두 교통사고모형(음이항 회귀모형 vs. 영과잉을 고려한 음이항 회귀모형)의 모형 적합도를 분석하기 위하여 로그-우도함수값(log-likelihood at convergence)과 우도비(likelihood ratio)를 산정하여 비교하였다. 로그-우도함수는 표본크기의 함수로 단독으로 모형의 적합도로 활용할 수 없지만, 여러 계수의

적합치를 비교하는데 이용할 수 있는 검증 방법이다. 로그-우도함수값은 0(zero)일 때를 제외하고 항상 음의 수이며, 0(zero)에 가까운 값일수록 모형적합도가 좋다고 판단할 수 있다. 또한, 우도비는 어떤 모형이 표본 데이터에 더 나은 적합도를 가지는지 확인하기 위해 모든 모수가 자유인 제약이 없는 모형과 귀무가설에 의해 더 작은 수의 모수로 제약되는 모형 등과 같은 두 가지 모형의 적합도를 비교하는 가설검증이다.

Table 3에서 보는 바와 같이, 총 교통사고발생빈도에 대한 교통사고모형에서 음이항 회귀모형과 영과잉을 고려한 음이항 회귀모형의 모형 설명력을 비교해보면, 영과잉을 고려한 음이항 회귀모형의 로그-우도함수값이 -836.736으로 나타나 음이항 회귀모형의 -843.547보다 모형 설명력이 좋은 것으로 나타났다. 또한, 우도비(ρ^2)의 경우 영과잉을 고려한 음이항 회귀모형에서 0.193으로 나타나 일반 음이항 회귀모형의 0.187보다 개선된 것으로 분석되었다.

Table 3. Model Estimation Results of Total Accidents at Intersections in CBD Area

Variables	Negative binomial				Zero-inflated negative binomial			
	Coefficients	Standard error	t-statistics (P-value)	Average marginal effects	Coefficients	Standard error	t-statistics (P-value)	Average marginal effects
Constant	-6.193	1.728	-3.584 (0.000)	-	-6.101	1.515	-4.027 (0.000)	-
Ln(AADT)	0.672	0.153	4.391 (0.000)	10.723	0.684	0.135	5.071 (0.000)	11.137
NLANE	-0.112	0.047	-2.400 (0.016)	-1.793	-0.122	0.044	-2.791 (0.005)	-1.980
BAR	0.503	0.104	4.854 (0.000)	8.026	0.457	0.091	5.033 (0.000)	7.450
RP	-0.493	0.087	-5.678 (0.000)	-7.863	-0.441	0.817	-5.405 (0.000)	-7.191
GL	1.968	0.383	5.136 (0.000)	31.409	1.812	0.388	4.666 (0.000)	29.511
FS	-0.195	0.076	-2.574 (0.010)	-3.104	-0.219	0.070	-3.128 (0.002)	-3.564
Number of observations	248							
Dispersion parameter (t-Statistics/P-Value)	0.206 (8.123/0.000)				0.171 (6.442/0.000)			
AIC	6.867							
BIC	6.981							
Log-likelihood with constant only	-1037.250				-1037.250			
Log-likelihood at convergence	-843.547				-836.736			
ρ^2	0.187				0.193			
Vuong test	-				2.390			

대전광역시 CBD 내 교차로에서 발생한 총 교통사고발생빈도에 대한 교통사고모형(영과잉을 고려한 음이항 회귀분모형)을 분석해보면, 6개의 변수가 교통사고 발생빈도에 영향을 미친 것으로 분석되었다(t -statistics ≥ 1.96 과 P -value ≤ 0.050). 교통량(LAADT)의 경우 계수(coefficient)가 0.684로 교통량이 증가할수록 교통사고 발생빈도가 높아지는 것으로 나타났다. 교통량의 한계효과는 11.137로 교통량이 1단위 증가할 때마다 교통사고는 평균 11.137만큼 증가하는 것으로 나타났다. 이는 일반 음이항 회귀분모형으로 산정된 교통사고모형에서 교통량의 한계효과인 10.723과 비교하여 영과잉을 고려한 음이항 회귀분모형이 교통량에 더 민감한 것으로 나타났다.

또한, 차로수(NLANE), 우회전전용차로(RP), 그리고 전방신호등(FS)의 계수는 각각 -0.122, -0.441, -0.219로 나타나 차로수가 많을수록, 우회전전용차로와 전방신호등이 설치될수록 교통사고발생빈도가 줄어드는 것으로 분석되었다. 각 변수의 한계효과는 차로수, 우회전

전용차로, 전방신호등 각각 -1.980, -7.191, -3.564로 나타나 우회전전용차로 설치가 차로수 증가나 전방신호등 설치보다 교통사고발생빈도 감소에 더 큰 영향을 미치는 것으로 나타났다.

중앙분리대(BAR)와 차량유도선(GL) 설치의 경우 계수가 각각 0.457과 1.812로 산정되어 교통사고발생빈도에 부정적인 영향을 미치는 것으로 나타났다. 이는 기존 연구에서 밝힌 교차로 교통사고발생빈도 분석과는 상이한 결과이다. 하지만, 차량통행과 보행자통행이 다른 교차로에 비해 많은 도시지역 CBD의 경우, 사고심각도 감소를 위해 설치된 중앙분리대는 보행자 무단횡단을 방지하기 위해 설치된 안전시설이지만, 이러한 도로상에 설치된 교통안전시설과 많은 교통량으로 인한 상충빈도 증가로 충돌사고가 증가하는 것으로 판단된다. 또한, 차량유도선의 경우, CBD 특성상 많은 차로수, 넓은 교차로 면적, 다수의 회전차로 등으로 인해 차량간의 상충빈도 증가로 교통사고발생빈도가 증가할 확

Table 4. Model Estimation Results of Injury Accidents at Intersections in CBD Area

Variables	Negative binomial				Zero-inflated negative binomial			
	Coefficients	Standard error	t-statistics (P-value)	Average marginal effects	Coefficients	Standard error	t-statistics (P-value)	Average marginal effects
Constant	-6.319	1.759	-3.593 (0.000)	-	-6.218	1.540	-4.038 (0.000)	-
Ln(AADT)	0.683	0.156	4.388 (0.000)	10.730	0.693	0.137	5.064 (0.000)	11.127
NLANE	-0.018	0.047	-2.280 (0.023)	-1.691	-0.116	0.044	-2.654 (0.008)	-1.868
BAR	0.500	0.106	4.725 (0.000)	7.866	0.456	0.093	4.915 (0.000)	7.313
RP	-0.492	0.089	-5.551 (0.000)	-7.735	-0.442	0.084	-5.283 (0.000)	-7.095
GL	1.944	0.388	5.014 (0.000)	30.557	1.787	0.391	4.573 (0.000)	28.689
FS	-0.209	0.077	-2.717 (0.007)	-3.279	-0.232	0.071	-3.260 (0.001)	-3.718
Number of observations	248							
Dispersion parameter (t-Statistics/P-Value)	0.214 (8.099/0.000)				0.177 (6.499/0.000)			
AIC	6.860							
BIC	6.973							
Log-likelihood with constant only	-1039.589				-1039.589			
Log-likelihood at convergence	-842.631				-836.415			
ρ^2	0.189				0.195			
Vuong test	-				2.339			

률이 높아지는 것으로 판단할 수 있다.

부상 교통사고발생빈도에 대한 교통사고모형은 Table 4에서 보는 바와 같이, 총 교통사고발생빈도에 대한 교통사고모형과 비슷한 결과를 도출하였다. 먼저 음이항 회귀모형과 영과잉을 고려한 음이항 회귀모형의 모형 설명력을 비교해보면 로그-우도함수값이 각각 -842.631과 -836.415로 영과잉을 고려한 음이항 회귀모형의 모형 설명력이 좋은 것으로 나타났다. 우도비(ρ^2)의 경우 음이항 회귀모형과 영과잉을 고려한 음이항 회귀모형 각각 0.189와 0.195로 나타나 영과잉을 고려한 음이항 회귀모형이 그렇지 않은 모형에 비해 모형이 개선된 것으로 분석되었다.

부상 교통사고발생빈도를 분석하기 위해 개발된 영과잉을 고려한 음이항 회귀모형을 살펴보면, 총 교통사고발생빈도에 대한 교통사고모형의 결과와 마찬가지로 교통량, 차로수, 중앙분리대, 우회전전용차로, 차량유도선, 전방신호등이 유효한 독립변수로 나타났다. 또한 개발된 모형의 독립변수의 계수와 한계효과는 총 교통사고발생빈도에 대한 교통사고모형의 결과와 비슷하게 나타났다.

5. 결론

본 연구는 대전광역시 CBD 내 교차로에서 발생하는 교통사고를 분석하기 위하여 음이항 회귀모형을 이용하여 교통사고모형을 개발하였으며, 교통사고모형의 모형 설명력을 높이기 위하여 영과잉을 고려하여 모형을 개발하였다. 또한, 영과잉을 고려한 교통사고모형의 모형 설명력의 개선효과를 분석하고자 일반적인 음이항 회귀모형의 모형 산정결과와 비교하여 모형 설명력을 분석하였다.

교통사고모형은 교통사고발생빈도에 영향을 미친다고 판단되는 설명변수를 단계적 방식(stepwise manner)을 이용하여 개발하였으며, 그 결과 총 교통사고발생빈도와 부상 교통사고발생빈도에 대한 교통사고모형의 독립변수로 교통량, 차로수, 중앙분리대, 우회전전용차로, 차량유도선, 전방신호등 등으로 나타났다. 또한, 교통사고모형의 모형 설명력을 분석해보면, 총 교통사고발생빈도와 부상 교통사고발생빈도에 대한 교통사고모형 모두에서 영과잉을 고려한 음이항 회귀모형의 로그-우도함수값이 각각 -836.736과 -836.415로 나타나 일반 음이항 회귀모형의 로그-우도함수값인 -843.547과 -842.631보다 좋은 것으로 나타났다. 또한 우도비(ρ^2)의 경우, 영과잉

을 고려한 음이항 회귀모형에서 각각 0.193과 0.195로 일반 음이항 회귀모형의 우도비(ρ^2) 0.187과 0.189보다 개선된 것으로 나타났다.

교통사고모형 개발 결과, 도시지역 CBD 내 교차로의 특성상 교통량과 보행량이 다른 지역에 비해 많아 일반적인 교차로에서 발생하는 교통사고특성과는 다른 경향을 보이고 있다. 이러한 CBD 내 교차로의 교통사고발생빈도를 줄이기 위하여 CBD 내 통과교통량을 줄이는 방안이 필요하며, 주간선도로와 보조간선도로 또는 이면도로와의 접속이 원활하도록 우회전전용차로 신설을 고려하여야 한다. 또한, 본 연구에서는 중앙분리대와 차량유도선이 교통사고발생빈도에 부정적인 영향을 미친다고 나왔는데 이는 교통안전시설의 부적절한 설치로 인한 차량간 상충으로 교통사고발생빈도가 높아지는 것으로 판단되어 교통안전설치 시 고려해야 할 것으로 판단된다.

지금까지 교통사고모형은 단순히 지형 및 기하구조 정보와 교통운영정보 등을 독립변수로 하는 가산모형인 포아송 회귀모형이나 음이항 회귀모형 등을 이용하여 모형을 개발하고 교통사고발생에 대한 분석 및 예측을 시행하였다. 하지만 도시지역 교차로의 교통사고는 매년 발생하지 않을 수도 있어(zero accident) 교통사고가 발생하지 않은 교차로 또는 해당년도를 고려하여 교통사고모형을 산정하여야 교통사고발생빈도의 설명력을 높일 수 있다. 본 연구에서 산정된 교통사고모형을 비교해보면, 영과잉을 고려한 음이항 회귀모형으로 산정한 교통사고모형이 그렇지 않은 모형과 비교해 통계학적으로 모형의 설명력이 향상된 것으로 나타났다.

이러한 결과는 향후 교통사고발생 원인을 분석하고 교통사고예방을 위한 안전시설 설치 시 교통사고예방 효과가 높은 시설물 파악의 정확도를 높여줄 것으로 기대된다.

REFERENCES

- Carlin, B. P., and Louis, T. A., 2009. Bayesian Methods for Data Analysis, New York, NY, CRC Press.
- Chin, H. C., and Quddus M. A., 2003. Applying the Random Effect Negative Binomial Model to Examine Traffic Accident Occurrence at Signalized Intersection, Accident Analysis and Prevention, vol. 35, No. 2. 253-259.
- Jonsson, T., Lyon, J., Washington, S., Schlakwyk, I., and Lord, D., 2009. Differences in the Performance of Safety Performance Functions Estimated for Total Crash Count and for Crash Count by Crash Type, Transportation Research Record, 2102, 115-123.

- Jovanie, P. P., and Chang, H., 1986. Modeling the Relationship of Accidents to Miles Traveled, *Transportation Research Record*, 1068. 42-51.
- Korean National Police Agency, 2015. Traffic Accident Statistics.
- Kweon, Y., 2012. Disaggregate Safety Evaluation for Signalized Intersections and Evaluation Tool, *Journal of Advanced Transportation*, Vol. 46, No. 2. 151-160.
- Lee, S. H., Lee, Y. D., and Do, M. S., 2012. Safety Impacts of Red Light Enforcement on Signalized Intersections, *Journal of Korean Society of Transportation*, Vol. 30, No. 6. 93-102.
- Lee, S. H., Lee, Y. D., and Do, M. S., 2016. Analysis on Safety Impact of Red Light Cameras using the Empirical Bayesian Approach, *Transportation Letters*, Online Published.
- Lee, S. H., Park, M. H., and Woo, Y. H., 2015. A Study on Developing Crash Prediction Model for Urban Intersections Considering Random Effects, *Journal of the Korea Institute of Intelligent Transport Systems*, Vol. 14, No. 1. 85-93.
- Lord, D., Washington, S. P., and Ivan, J. N., 2005. Poisson, Poisson-Gamma and Zero-Inflated Regression Models of Motor Vehicles: Balancing Statistical Fit and Theory, *Accident Analysis and Prevention*, Vol. 37. 35-46.
- Miaou, S. P., and Lum, H. A., 1993. Modeling Vehicle Accidents and Highway Geometric Design Relationship, *Accident Analysis and Prevention*, Vol. 25, No. 6. 689-709.
- Mitra, S., and Washington, S., 2012. On the Significance of Omitted Variables in Intersection Crash Modeling, *Accident Analysis and Prevention*, Vol. 49, 439-448.
- Mouatassim, Y., and Ezzahid, E. H., 2012. Poisson Regression and Zero-inflated Poisson Regression: Application to Private Health Insurance Data, *European Actuarial Journal*, Vol. 2, No. 2, 187-204.
- Na, H., and Park, B. H., 2012. Accident Models of Circular Intersection by Cause Using ZAM, *International Journal of Highway Engineering*, Vol. 14, No. 2. 101-108.
- Retting, R. A., Ulmer, R. G., and Williams, A. F., 1999. Prevalence and Characteristics of Red Light Running Crashes in the United States, *Accident Analysis and Prevention*, Vol. 31. 687-694.
- Vuong, Q. H., 1989. Likelihood Ratio Tests for Model Selection and Non-nested Hypotheses, *Econometrica*, Vol. 57, No. 2, 307-333.