

서울시 교통사고 위험요인에 대한 공간모형 개발 및 위험도 추정에 대한 연구*

정보미** · 강일석** · 허태영***

A Study on Estimation of Accident Hazard and Development of Spatial Model of Accident Risk Factor in Seoul Metropolitan Area*

Bo mi Jeong** · Il Suk Kang** · Tae-Young Heo***

요약 : 본 연구에서는 2010년 서울시의 416개 행정동 교통사고 자료를 이용하여 교통사고 발생건수에 영향을 미치는 위험요인을 확인하고 행정동별 교통사고 발생에 대한 위험도를 평가하기 위해 두 종류의 공간모형을 제안하였다. 본 연구의 모형을 위해 사용된 자료의 종속변수는 행정동 단위로 집계된 교통사고 발생건수이고, 설명변수는 도로 총연장, 버스 전용차로 설치비율, 도로 총 연면적, 교차로 및 횡단보도와 과속방지턱 설치 지점 수, 횡단보도 예고표시 설치율을 사용하였다. 공간모형 개발에 대한 여부를 판단하기 위해 서울시 각 행정동별 교통사고 발생건수의 공간적 상관성 여부를 모란지수(Moran Index)를 통해 확인하였다. 본 연구에서는 정규분포 기반의 회귀모형과 공간상관성이 반영된 공간자기회귀모형을 비교하였다. 또한, 포아송모형의 한계를 극복하기 위해 일반화 포아송모형과 공간적 상관성이 반영된 공간일반화 포아송모형을 비교하여 최적모형을 선택하였다. 공간일반화 포아송모형을 통해 각 행정동별 교통사고 발생에 대한 상대위험도를 추정하고 지도 위에 표출하여 위험지도를 구축하였다. 서울시 행정동별 교통사고 발생 위험도 평가 결과 416개 중 50개의 행정동이 상대적으로 위험지역으로 도출되었다. 특히, 서울시에서는 사고발생 위험도가 높은 상위 5개 동은 잠실 2동, 상계 10동, 잠실 4동, 반포 3동, 여의도 동 순으로 나타나 교통사고 안전관리를 위해 우선으로 교통안전대책을 세우는 것이 필요함을 알 수 있다.

주제어 : 모란지수, 공간모형, 일반화 포아송모형, 조건부자기회귀모형, 과대산포, 위험지도

ABSTRACT : The purpose of this research is to suggest two kinds of spatial model using traffic accident data observed at 416 administrative divisions in Seoul for 2010 year in order to observe explanatory variables affecting the number of accidents and to appraise relative risk of accidents occurred at each division. In this study, the number of traffic accidents collected at the level of administrative division is considered as a response variable and explanatory variables such as a road length, the total area, the number of crossroads, crosswalks and speed bumps, the rate of bus-only lane and that of sign for crosswalk are used. Significance test for Moran Index is conducted to determine whether spatial correlation exists in research region. After performing significance test, spatial autoregressive model reflecting spatial effect is compared to ordinary regression model based

* 본 논문은 서울연구원이 주최한 「공공데이터를 활용한 2014 서울연구논문 공모전」에서 제공한 데이터를 활용해 작성하였으며, 이 공모전에
서 입선논문으로 선정되었습니다. 또한 본 논문은 2014년도 충북대학교 학술연구지원사업의 연구비 지원에 의하여 연구되었습니다.

** 충북대학교 통계학과 석사과정(The Master's Course, Department of Information and Statistics, Chungbuk National University)

*** 충북대학교 통계학과 부교수(Associate Professor, Department of Information and Statistics, Chungbuk National University),

교신저자(E-mail: theo@chungbuk.ac.kr, Tel: 043-261-3741)

on gaussian distribution. Generalized poisson model is accepted to overcome over-dispersion that a limitation of poisson regression model. Moreover, to select optimal model, spatial generalized poisson regression model is compared to generalized poisson regression model. Through the result of analysis using spatial generalized poisson regression model, relative risk is estimated at the level of administrative division and represented on the map to exhibit the degree of the risk. As a result, 50 divisions are classified as hazard regions among 416 administrative divisions. To be specific, Jamsil 2-dong, Sanggye 10-dong, Jamsil 4-dong, Banpo 3-dong, Yeouido-dong are in top 5 risk areas, and these areas ought to be considered in priority when traffic safety measures are established.

Key Words : Moran Index, Spatial Model, Generalized Poisson Model, Relative Risk

I. 서론

서울시를 비롯한 각 지방자치단체에서는 교통사고 경감을 위해 교통사고 발생 위험지역에 대해 운전자 경각심 제고와 안전운행을 위해 다양한 정책적 노력을 기울이고 있다. 예를 들어, 서울시의 경우 교통사고 유형별, 사고자 연령별, 시간대별로 교통사고 다발지역을 분석하여 교통사고를 경감시키면서 시민들에게 경각심을 불러일으킬 수 있도록 하는 노력을 하고 있으며, 경남 김해의 경우 내비게이션을 통해 교통사고 위험구간 알리미 서비스를 제공하고 있다.

부산에서는 광안대교 운전 시 운전자에게 내비게이션을 통해 교통사고 위험구간에 대한 안내 서비스를 제공하고 있으며, 2003년 개통 이후 6년간 교통사고 분석을 토대로 과속위험, 사고다발지역에 도달하면 음성으로 알려주는 서비스를 시행하고 있다.

또한 강화도에서는 사고다발 도로구간에 대해 “교통사고 주의 안내도”를 제작하여 교통사고 경감 노력을 하고 있다. 서울시 송파구에서는 “정보가 있는 교통사고지도”를 발간하여 교통정책 개선 및 교통시설 투자 등에 활용할 예정이며, 사고다발지역에 대한 표지판 설치 및 주민 홍보를 통

해 안전사고 예방 노력을 병행하고 있다. 인천시 부평구는 교통사고발생지점에 대한 정확한 분석을 위해 교통사고 다발지역의 집중관리 및 시설물 보안을 하였으며 향후 교통사고를 미연에 방지하고자 “교통사고다발지점 현황 분석지도”를 제작하고 있다.

기존의 교통사고 분석을 위한 시스템에서는 각 구간별/지역별로 교통사고 건수 또는 사망사고 건수 등과 같이 빈도 또는 평균 사고건수 등과 같은 기술통계만을 교통정보시스템에 이식하고 있으며, 그래픽 요소를 가미하지 않고 숫자로 각 구간별 또는 지점별 정보를 제공해 주고 있다.

그러나 교통사고 경감을 위한 교통안전 분야의 주요한 목적은 교통사고 발생건수에 대한 영향요인을 찾아내고 각 지역별 위험도를 계산하여 패턴을 예측하는 것이다.

따라서 본 연구에서 제안된 최적의 교통사고모형을 통해 교통사고 발생에 영향을 주는 영향요인을 찾고, 서울시 각 행정동별 교통사고 위험도 순위를 평가할 수 있도록 위험도를 추정하고 이를 통해 서울시 행정동별 교통사고 위험지도를 생성할 수 있다(김학열·허태영, 2012).

본 연구에서는 교통사고 모형 개발을 위해 두 가지 형태의 모형을 제시하였다. 첫 번째는 교통

사고 발생 건수에 로그변환을 통해 종속변수를 정규분포화시켜 기존의 회귀모형에 공간영향력을 반영한 공간자기회귀모형(Spatial regression model)을 기존 회귀모형과 비교하여 공간자기회귀모형이 더 최적모형임을 확인하였으며, 두 번째는 교통사고 자료의 특성인 이산적 분포를 반영한 포아송회귀모형(Poisson regression model)과 일반화 포아송회귀모형(Generalized Poisson regression model)을 이용하여 자료의 공간상관성을 반영한 공간가산회귀모형(Spatial count regression model)을 제안하고(박병호·김윤환, 2007) 기존 일반화 포아송회귀모형과 비교하여 공간영향력을 반영한 공간가산회귀모형이 더 최적모형임을 확인하였다.

따라서 본 연구에서는 통계적 가정에 위배되지 않는 최적의 교통사고 모형을 개발하고 통계적 분석을 수행하였으며, 교통사고의 특성이 지역별로 서로 상호 연관관계가 있는지를 통계적으로 검정하여 공간상관성을 반영한 공간가산회귀모형을 제안하고 공간상관성을 반영하지 않은 기존 모형과의 비교를 통해 공간가산회귀모형의 우수성을 확인하고자 한다.

II. 교통사고 상대위험도 평가

본 연구에서는 교통사고 발생에 미치는 영향을 분석하기 위해 공간상관성을 반영한 공간자기회귀모형인 공간오차모형과 공간가산회귀모형인 공간일반화 포아송모형을 개발하고 평가하였다. 개발된 모형은 각 행정동별 사고위험도를 계산하고 위험구간을 설정할 수 있다.

교통사고 모형을 통해 종속변수인 교통사고 발생 건수에 대한 설명변수의 영향력을 나타내는 계수 추정값의 의미도 중요하지만 각 행정동별 교통

사고의 위험도를 파악하는 것도 매우 중요하다. 본 연구에서는 설명변수의 영향력을 고려하여 각 행정동별 교통사고 위험의 정도를 나타내었다.

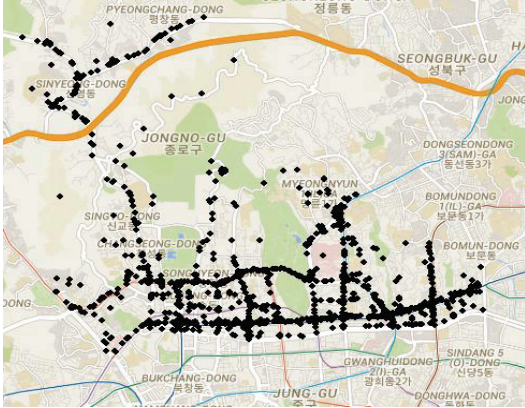
각 행정동별 상대위험도를 계산하기 위한 간략한 설명은 다음과 같다. 교통사고의 위험도를 나타내는 방법으로 많이 활용되고 있는 관측된 교통사고 건수(Y_i)에 대한 기대되는 교통사고 건수(E_i)인 사고위험은 $\hat{\theta}_i = Y_i/E_i$ 로 정의될 수 있다. 전통적인 추정방법보다는 베이지안 기법의 모수 추정 방법이 많이 활용되고 있으며, 포아송모형 등과 같은 가산모형을 이용해 다양한 사고위험도를 계산할 수 있다.

III. 공간모형

1. 자료 설명

본 연구에서 종속변수는 2010년 서울시 행정동 단위로 집계된 개별 교통사고 발생건수이다. 본 연구의 자료는 발생위치를 포함하고 있는 공간자료이며 집계된 교통사고 발생건수는 총 41,662건이다. <그림 1>은 서울시에서 발생한 교통사고 지점을 서울시 구글 지도상에 표출한 것 중 일부로 종로구의 사고지점을 나타낸 것이다.

설명변수는 각 행정동 단위로 교통안전시설 및 정책의 여건에 해당하는 도로의 총연장, 버스전용차로 설치비율, 도로의 총 연면적, 교차로·횡단보도·과속방지턱 설치 지점 수, 횡단보도 예고표시 설치율을 활용하였다.



〈그림 1〉 2010년 서울시 종로구 교통사고 지점

종속변수에 해당하는 교통사고 발생건수는 이 산형 자료로서 포아송분포를 따르므로 포아송모형을 활용하여 분석하는 것이 일반적이다. 그러나 가산자료 분석에 가장 많이 활용되고 있는 통계모형은 포아송모형이지만 포아송모형은 자료의 평균과 분산이 동일하다는 엄격한 가정을 가지고 있다.

따라서 본 연구에서는 기본적인 가산모형인 포아송모형을 확장한 과대산포와 과소산포 문제를 해결할 수 있는 모형을 이용하여 분석결과를 제시하고 비교하였다.

2. 가산자료 모형

1) 가산회귀모형

본 논문에서 종속변수로 사용할 교통사고 건수와 같이 정규분포를 따르지 않는 변수를 모형화할 때 쓰이는 간단한 모형 중 하나로 포아송모형이 많이 쓰인다. 본 연구에서 고려한 포아송 회귀 모형의 일반식은 식 (1)과 같다.

$$P(Y_i = y_i) = \frac{\lambda_i^{y_i} \exp(-\lambda_i)}{y_i!} \quad (1)$$

여기서, $P(Y_i = y_i)$ 는 교통사고 발생건수인 y_i 가 i 번째 행정구에서 발생할 확률을 의미하며, λ_i 는 평균 사고건수를 의미하며 포아송모형의 정준 연결함수(Canonical link function)를 로그함수를 이용하여 식 (2)와 같이 표현할 수 있다.

$$\log(\lambda_i) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} \quad (2)$$

여기서, x_i 는 i 번째 행정동에서 관측된 여러 속성변수들의 관측값을 의미하며, β_i 는 각 설명변수 계수값을 의미한다.

포아송모형을 각 행정동의 교통사고 건수에 적합시켜 얻는 여러 가지 합리적인 이유에도 불구하고 포아송모형은 평균과 분산이 같다는 강한 제약을 따른다. 현실적으로 얻어지는 교통사고 건수는 대부분 평균과 분산이 확연히 차이가 있음을 알 수 있으며, 본 논문의 자료에서도 마찬가지로 분산의 값이 평균보다 매우 큰 값을 가지는 과대산포(over-dispersion) 문제가 발생한다. 흔히 산포의 정도를 알아보는 데 산포지수(DI: dispersion index)가 사용된다. 산포지수는 식 (3)과 같이 주어진다.

$$DI = \frac{Var(Y)}{E(Y)} \quad (3)$$

산포지수의 값이 1 이상의 값을 가질 경우 과대산포가 발생하는 것을 짐작할 수 있으며 반대로 1 이하의 값을 가지는 경우에는 과소산포를 의미한다고 볼 수 있다.

본 논문에서 이용하고자 하는 교통사고 건수의 평균은 96.50이며 분산은 5,542.55를 가지므로 교통사고 건수의 산포지수는 57.44로 계산된다. 따라서 교통사고 건수는 과대산포의 문제점을 가진

다고 볼 수 있으며 이러한 경우에 포아송모형을 사용하기에는 적절하지 않을 수 있다. 이러한 문제점의 대안으로 일반적으로 음이항 모형과 일반화 포아송모형을 고려할 수 있으나 본 논문에서는 과대산포와 과소산포를 모두 고려할 수 있는 일반화 포아송모형을 이용하고자 한다.

2) 일반화 포아송모형

전통적인 포아송모형이 가지는 한계를 극복하기 위해 평균과 분산의 차이가 큰 경우에는 포아송모형의 대체모형으로 일반화 포아송모형을 많이 사용하고 있다(김학열·허태영, 2012). 일반화 포아송모형은 음이항 모형과 다르게 가산자료의 과대산포뿐만 아니라 과소산포의 문제점을 가지는 자료에도 적용할 수 있다는 여러 가지 장점이 있다(Consul and Jain, 1973; Consul and Famoye, 1992; Famoye et al., 2004; Bae et al., 2005; Famoye and Singh, 2006).

확률변수 Y_i 가 일반화 포아송분포를 따른다고 하면 식 (4)와 같이 표현이 가능하다.

$$Y_i \sim GP(\mu_i, a), i = 1, 2, \dots, n \quad (4)$$

여기에서 GP 는 일반화 포아송분포를 뜻하며, 일반화 포아송분포의 확률질량함수는 식 (5)와 같이 정의된다.

$$P(Y_i = y_i) = \left(\frac{\mu_i}{1 + a\mu_i} \right)^{y_i} \frac{(1 + ay_i)^{y_i - 1}}{y_i!} \times \exp\left(-\frac{\mu_i(1 + ay_i)}{1 + a\mu_i}\right), \quad y_i = 0, 1, 2, \dots \quad (5)$$

여기서 평균 $E(Y_i) = \mu_i$ 이고, 분산 $Var(Y_i) = \mu_i(1 + a\mu_i)^2$ 이다. 일반화 포아송분포는 표준 포아송분포의 자연스러운 확장으로 만약 $a = 0$ 이면 일반화 포아송분포는 평균과 분산이 동일한 표준 포아송분포로 자연스럽게 변환된다. 만약 $a > 0$ 이면 분산이 평균보다 큰 과대산포가 되며, $a < 0$ 이면 분산이 평균보다 작은 과소산포를 나타내는 분포를 의미한다.

일반화 포아송모형은 과대산포 문제만을 해결할 수 있는 음이항모형보다 일반화된 모형으로 과대산포 또는 과소산포 문제를 해결할 수 있는 모형이 되는 장점을 가지고 있다(Wang and Famoye, 1997; Wulu et al., 2002).

3. 공간모형

1) 공간상관성 진단

본 연구에서는 서울시 각 행정동별로 교통사고 발생에 대한 공간적 상관성 여부를 판단하여 공간적 상관성이 존재하면 기존의 모형에 공간상관성을 반영한 공간모형을 제안하고자 한다.

공간상관성을 진단하기 위해 Moran's I 통계량을 사용하며, 식 (6)과 같이 계산된다.

$$I = \frac{N \sum_i \sum_j w_{ij} (Y_i - \bar{Y})(Y_j - \bar{Y})}{(\sum_i \sum_j w_{ij}) \sum_i (Y_i - \bar{Y})^2} \quad (6)$$

여기서, N 은 관측치의 수, Y_i 와 Y_j 는 각 지역의 관측치, \bar{Y} 는 관측값들의 평균값이고 w_{ij} 는 열의 합을 1로 하는 공간가중치 행렬(W)의 (i, j) 번째 원소값이다.

Moran's I 통계량에 대한 분포는 식 (7)과 같은

Z값으로 검정할 수 있다(Cliff and Ord, 1972).

$$Z_i = \frac{I - E(I)}{S.E(I)} \quad (7)$$

여기서 $E(I)$ 는 평균, $S.E(I)$ 는 공간상관성이 없다는 가정하에 계산된 표준오차이다.

Moran's I의 이론적인 범위는 일반적인 상관계수와 같이 -1과 1 사이에 위치하며, -1과 1에 가까울수록 공간상관성이 크고 0에 가까울수록 공간상관성이 작음을 의미한다.

2) 공간오차모형

본 연구의 첫 번째 제안 모형인 공간오차모형(SEM; Spatial error model)에 대한 설명은 식(8)과 같다(Wall, 2004).

$$Y = X\beta + e, e = \lambda W e + \epsilon, \epsilon \sim N(0, \sigma^2 I) \quad (8)$$

여기서 W 는 종속변수 Y 에 대한 공간가중행렬(spatial weight matrix)이며, λ 는 공간자기회귀계수로 공간상관성의 크기를 나타낸다. X 는 설명변수이며 β 는 설명변수에 대한 회귀계수를 의미한다.

식(8)의 모형에서 공간가중행렬 W 는 식(9)와 같이 정의된다.

$$W_{ij} = \frac{w_{ij}}{w_{..}}, w_{..} = \sum_{i=1}^n \sum_{j=1}^n w_{ij} \quad (9)$$

여기서, 행정동 i 와 j 가 이웃해 있으면 $w_{ij}=1$, 그렇지 않으면 $w_{ij}=0$ 이다.

3) 공간가산회귀모형

교통사고 건수 Y_i 가 기댓값 μ_i 를 가지는 포아송분포를 따를 때 $Y_i \sim \text{Poisson}(\theta_i E_i)$ 로 표현할 수 있으며, 여기에서 θ_i 는 i 번째 구간에서 교통사고가 발생할 위험도를 의미하며 기대 교통사고 건수 E_i 로 정의할 수 있다. 기대 교통사고 건수 E_i 는 관측된 교통사고 건수의 값에 따라 보정이 필요하며, 다음과 같은 식을 이용하여 보정하여 기대 교통사고 건수를 식(10)과 같이 계산할 수 있다(Jovanis and Chang, 1986).

$$E_i = \frac{\sum Y_i}{\sum T_i} \times T_i \quad (10)$$

여기에서 T_i 는 행정동 i 에서 관측된 도로의 총 연장(길이)을 의미한다. 전체 교통사고 건수는 해당 행정동의 도로 연장과 비례하기에 각 행정동별 기대 교통사고 건수를 산출해낼 수 있다.

위에서 설명한 포아송분포를 기반으로 한 상대 위험도에 대한 설명은 일반화 포아송분포에서 동일하게 적용할 수 있다. 포아송모형과 동일하게 기댓값을 로그연결함수를 이용하여 설명변수와 공간효과를 표현한 모형은 식(11)과 같다.

$$\begin{aligned} \log(\mu_i) &= \log(E_i) + \log(\theta_i) \\ \log(\theta_i) &= \beta_0 + \sum_{j=1}^k \beta_j x_{ij} + \delta_i + \epsilon_i, \\ i &= 1, 2, \dots, n \end{aligned} \quad (11)$$

여기서, θ_i 는 i 번째 행정동의 교통사고 위험도로 표현될 수 있으며, x_j 와 β_j 는 각각 j 번째 설명변수와 회귀계수를 의미한다.

식(11) 모형에서 ϵ_i 는 관측되지 않는 오차항, δ_i 는 지역별 공간적 상관성을 반영한 조건부자기

회귀모형(CAR: Conditional Autoregressive Model)을 의미하며 식 (12)와 같이 정의된다.

$$\delta_i | \delta_{-i} \sim N\left(\mu_i + \sum_{j=1}^n c_{ij}(\delta_j - \mu_j), \tau_i^2\right) \quad (12)$$

CAR모형에 대한 간략한 설명은 다음과 같다. $\{Z(A_i) : A_i \in (A_1, \dots, A_n)\}$ 이 격자 D 를 형성하는 정규확률과정(Gaussian random process)을 따르며 $Z(A_i)$ 가 지역 A_i 에서의 관측값이라고 하고 $\{A_1, \dots, A_n\}$ 이 격자 D 를 분할한다고 가정한다. 그러면 조건부 자기회귀모형은 식 (13)과 같이 표현할 수 있다(최정순 외, 2010a; 2010b).

$$Z(A_i) | Z(A_{(-i)}) \sim N\left(\mu_i + \sum_{j=1}^n c_{ij}(Z(A_j) - \mu_j), \tau_i^2\right) \quad (13)$$

여기서 $Z(A_{(-i)}) = \{Z(A_j) : j \neq i\}$ 를 뜻하며, $E(Z(A_i)) = \mu_i$, τ_i^2 는 조건부 분산을 나타낸다.

n 이 유한인 경우 행렬 $C = (c_{ij})$, $T = \text{diag}\{\tau_1^2, \dots, \tau_n^2\}$ 을 나타내며, 인수분해정리를 통해

$$\mathbf{Z} \sim N(\boldsymbol{\mu}, (I_n - C)^{-1}T)$$

와 같이 표현된다. $(I_n - C)$ 가 역행렬이 존재하기 위해서는 대칭행렬을 만족하여야 한다.

따라서 행렬 $(I_n - C)^{-1}T$ 는 대칭이며 양정치행렬이다. 행렬 C 는 가중행렬 W 를 통해 이루어진 행렬이며, 일반적으로 지역 i 와 j 가 인접하면 $w_{ij} = 1$ 을, 인접하지 않으면 $w_{ij} = 0$ 값을 사용한다.

또한 공간적 구조를 설정하고 인접성 기반의 이웃정보를 사용하기 위해 $c_{ij} = w_{ij}/w_{i+}$ 로 정의

하고, $\tau_i^2 = \sigma_c^2/w_{i+}$ 로 가정하며, 여기서 σ_c^2 는 공간 상관의 평활성을 나타내는 모수를 의미한다.

제안된 식 (11)의 모형에 대한 모수를 추정하기 위해 베이저안 방법을 이용하였으며, 이를 위해 아래와 같이 각 모수에 대한 사전분포(prior distribution)를 식 (14)와 같이 정의하였다.

$$\begin{aligned} \beta_j &\sim \text{Normal}(0, 10^6), \quad j=0,1,2,\dots,k \\ \epsilon_i &\sim N(0, \sigma^2) \\ \sigma^2 &\sim \Gamma(0.5, 0.005) \\ \sigma_c^2 &\sim \Gamma(0.5, 0.005) \end{aligned} \quad (14)$$

모수에 대한 추론은 사후분포를 통해 이루어지며 모형의 모수를 추정하기 위해 마코프 연쇄 몬테카를로(MCMC: Markov Chain Monte Carlo) 방법을 사용하였다. 모수추정결과는 세 개의 초기값(initial value)을 활용하여 100,000번 반복 후 처음 30,000번까지 제거한 나머지 결과값으로, 구체적인 모수 추정값은 결과에 제시하였다(최정순 외, 2010a; 2010b).

4. 모형 비교 척도

1) AIC

본 논문에서는 통계적 모형의 적합도 및 비교의 척도로서 AIC(Akaike Information Criterion)를 사용하였으며 식 (15)와 같이 표현한다(Akaike, 1974).

$$AIC = 2k - 2\ln(L) \quad (15)$$

여기서 k 는 모형의 모수 수를 의미하고 L 은 모형의 우도함수 최댓값을 의미한다. AIC는 데이터

로부터 모형을 생성하는 과정에서 손실되는 정보의 상대적인 추정치를 제공한다. 그러므로 주어진 데이터로부터 얻어진 모형들 중 AIC값이 가장 작은 모형이 가장 좋은 모형이라 할 수 있다.

2) DIC

베이저안 기법의 모형 비교 측도에 관해서는 DIC(Deviance information criterion)를 사용하였으며 식 (16)과 같이 표현한다(Spiegelhalter et al., 2002).

$$DIC = D(\bar{\theta}) + 2p_D = \bar{D} + p_D \quad (16)$$

여기서, $p_D = \bar{D} - D(\bar{\theta})$ 이며, $D(\bar{\theta})$ 은 로그가능도의 함수로 이루어진 이탈도(Deviance)를 의미하고 $\bar{\theta}$ 은 모수 θ 의 사후평균을 의미한다. \bar{D} 는 이탈도의 사후평균, p_D 는 모형을 이루는 효과적인 모수의 개수를 나타낸다.

DIC는 모형의 적합성과 복잡성을 동시에 고려할 수 있는 척도로서 AIC와 비슷하며, DIC의 값이 작은 모형일수록 잘 적합된 모형으로 고려할 수 있다. 한편, DIC는 이탈도의 사후평균이 중심값을 대표할 수 있는 척도임을 가정하고 있기 때문에 이탈도의 사후분포 모양이 양봉 분포(Bimodal distribution) 형태 또는 왜도가 높은 분포 형태를 가지고 있으면 모형 비교 척도에 있어서 DIC의 사용이 적절하지 않을 수 있다.

IV. 분석 및 결과

본 연구에서는 서울시 416개 행정동에 대해 교통사고 발생건수에 영향을 미치는 다양한 요인들의 영향력을 파악하고 그에 따른 행정동별 상대위

험도를 계산하여 표출하였다.

각 행정동에 대해 공간적 상관성을 진단하기 위한 Moran's I 통계량은 0.1880으로 P-값이 <.0001로 매우 작아 유의수준 5%에서 통계적으로 유의하기에 공간상관성이 존재함을 알 수 있다. 또한 Moran's I 통계량이 양의 값을 가지기 때문에 양의 상관성이 존재하여 기존의 공간적 상관성을 반영하지 않은 모형보다 공간적 상관성을 반영한 공간모형이 더 적합함을 알 수 있다.

<표 1>의 결과는 정규분포 기반의 모형 결과이며, <표 2>의 결과는 포이송분포 기반의 모형 결과로서 자세한 설명은 다음과 같다.

교통사고 발생건수에 미치는 영향을 교통안전시설에 해당하는 요인들의 영향력을 파악하기 위하여 정규분포 기반의 SEM모형을 구축하였다. <표 1>은 공간상관성을 반영하지 않은 회귀모형과 공간상관성을 반영한 SEM모형의 결과를 나타낸다.

기존회귀모형과 SEM모형의 분석결과에서는 두 모형 모두 모든 변수들이 유의수준 5%에서 유의하게 판단되어 해당 요인들이 교통사고 발생건수에 영향을 준다고 할 수 있다. 여기서 도로 길이와 버스전용차로 설치율, 도로 총 연면적, 교차로 수, 횡단보도 수는 교통사고 발생건수에 양으로 영향을 미치며 횡단보도 예고표시율과 과속방지턱의 수는 그 수와 비율이 클수록 교통사고 발생건수에 음으로 영향을 미치는 것으로 나타났다. 또한, 공간자기회귀계수의 추정값은 0.332이고 통계적으로 유의하게 판단되었다.

두 모형의 AIC값을 비교해 보면, 기존회귀모형이 595.74, SEM모형이 578.65로 SEM모형의 AIC값이 더 작게 나온다. 이것은 SEM의 공간영향력에 해당하는 부분이 기존회귀모형에서는 오차항으로 포함되기 때문에 공간상관성이 존재하는 경

〈표 1〉 회귀모형과 SEM모형의 모수추정 결과

모수	회귀모형				SEM			
	추정값	표준오차	Z값	P-값	추정값	표준오차	Z값	P-값
절편	3.43E+00	7.93E-02	43.2330	<.0001	3.42E+00	8.06E-02	42.4598	<.0001
도로 총연장	1.46E-02	2.63E-03	5.5530	<.0001	1.46E-02	2.64E-03	5.5399	<.0001
버스전용차로 설치율	4.82E+00	6.12E-01	7.8760	<.0001	4.71E+00	6.02E-01	7.8167	<.0001
총면적	1.66E-07	3.40E-08	4.8650	<.0001	1.19E-07	3.69E-08	3.2186	0.0013
교차로 수	2.98E-02	3.63E-03	8.2070	<.0001	3.18E-02	3.60E-03	8.8153	<.0001
횡단보도 수	7.70E-05	2.54E-05	3.0390	0.0025	6.98E-05	2.44E-05	2.8547	0.0043
횡단보도 예고표시율	-6.50E-01	2.22E-01	-2.9290	0.0036	-5.06E-01	2.19E-01	-2.3103	0.0209
과속방지턱 설치 수	-1.48E-03	3.76E-04	-3.9340	<.0001	-1.36E-03	3.84E-04	-3.5327	0.0004
공간자기회귀계수	-				3.32E-01	7.22E-02	19.0850	<.0001
	AIC : 595.74				AIC : 578.65			

〈표 2〉 일반화 포아송모형과 공간일반화 포아송모형의 모수 추정값과 신뢰구간

모수	일반화 포아송모형				공간일반화 포아송모형			
	평균	표준편차	2.50%	97.50%	평균	표준편차	2.50%	97.50%
절편	-0.0987	0.0254	-0.1487	-0.0486	-0.0985	0.0223	-0.1423	-0.0541
버스전용차로 설치율	0.3279	0.0240	0.2805	0.3742	0.3240	0.0236	0.2781	0.3705
총면적	0.0766	0.0247	0.0286	0.1249	0.0605	0.0288	0.0022	0.1169
교차로 수	0.1125	0.0306	0.0532	0.1733	0.1236	0.0324	0.0608	0.1875
횡단보도 수	0.0614	0.0299	0.0030	0.1196	0.0443	0.0280	-0.0089	0.1013
횡단보도 예고표시율	-0.0438	0.0284	-0.1004	0.0119	-0.0488	0.0294	-0.1086	0.0074
과속방지턱 설치 수	-0.0693	0.0296	-0.1271	-0.0120	-0.0546	0.0307	-0.1160	0.0048
산포모수	0.6889	0.0342	0.6070	0.7411	0.6827	0.03592	0.5987	0.7377
	DIC : 4,146.38				DIC : 4,128.37			

우 공간영향력을 반영한 모형이 더 적합한 모형이라 할 수 있다는 것이다.

교통사고 모형을 구축하기 위해 다양한 가산분포 기반의 교통사고 모형을 고려하였으나 앞서 산포지수를 통해 과대산포를 고려할 수 있는 일반화 포아송모형을 적합시켜 교통사고 발생건수와 다양한 요인의 관계를 모형으로 구축하였으며, 기대 사고 건수에 도로 총연장 변수가 사용되었으므로 일반화 포아송모형에서는 제거한다. 〈표 2〉는 일

반화 포아송모형의 결과를 나타내며, 〈표 1〉의 결과와 비교하면 모든 설명변수의 추정계수 방향이 같은 것을 알 수 있다.

일반화 포아송모형과 공간적 영향력이 반영된 공간일반화 포아송모형을 DIC값을 이용하여 비교해 보면 일반화 포아송모형이 4,146.38, 공간일반화 포아송모형이 4,128.37로 공간일반화 포아송의 DIC값이 더 작게 나온다. 〈표 1〉에서와 같이 공간일반화 포아송모형의 공간영향력에 해당하는 부분이

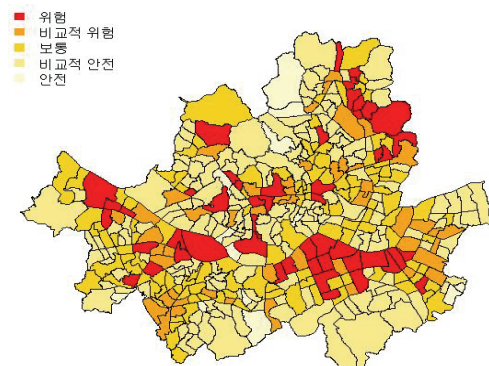
〈표 3〉 교통사고 위험도 분석결과

행정동	위험도	표준편차	2.50%	97.50%	위험 여부
청운효자동	0.814	0.169	0.511	1.171	×
사직동	1.370	0.274	0.872	1.945	×
삼청동	0.744	0.196	0.417	1.181	×
부암동	0.548	0.123	0.336	0.813	×
평창동	0.457	0.091	0.295	0.652	×
무악동	2.183	0.582	1.209	3.479	○
교남동	1.073	0.275	0.609	1.676	×
⋮	⋮	⋮	⋮	⋮	⋮
성내1동	0.835	0.205	0.479	1.275	×
성내2동	0.821	0.179	0.504	1.203	×
성내3동	1.252	0.269	0.770	1.830	×
길동	1.455	0.256	0.993	1.990	×
둔촌1동	1.050	0.425	0.453	2.085	×
둔촌2동	0.672	0.161	0.395	1.019	×
월곡1동	1.196	0.291	0.696	1.824	×

일반화 포아송모형에서는 오차항으로 포함되기 때문에 공간상관성이 존재하는 경우 공간영향력을 반영한 모형이 더 적합한 모형이라 할 수 있다.

각 행정동별 교통사고의 위험구간을 확인하기 위해 식 (11)에서 제시한 교통사고 위험도를 계산하여 나타내었다. 본 논문에서는 지면상 400여 개의 행정동에 대해 각 동별 사고위험도 제시가 불가능해 몇 개의 행정동에 대해서만 〈표 3〉에 일반화 포아송회귀모형을 통한 위험도를 제시하였으며, 〈그림 2〉와 같이 지도를 통해 구현하여 명암으로 표시하였다. 〈표 3〉의 분석결과를 보면 위험도 기준으로 사직동, 무악동, 교남동, 성내3동, 길동, 둔촌1동, 월곡1동의 위험도가 1보다 크게 나타났다. 그러나 무악동을 제외한 나머지 행정동은 95% 신뢰구간 안에 1이 포함되어 통계적으로 위험하지 않은 행정동으로 평가되었으며, 무악동은 95% 신뢰구간 안에 1이 포함되지 않아 위험한 행정동으로 평가되었다. 이와 같이 본 연구에서 추정된 상대위험도는 상대위험도의 신뢰구간을 제시할 수 있다는 장점도 가지고 있다.

〈그림 2〉에서와 같이 공간일반화 포아송모형을 이용하여 〈표 3〉의 위험도 값을 서울시 각 행정동별 지도 위에 표시하였으며, 행정동의 색이 진할수록 교통사고 위험도가 높은 행정동임을 의미한다. 결과적으로 서울시 행정동별 교통사고 발생 위험도 평가 결과 위험도가 높은 상위 5개 동은 잠실2동, 상계10동, 잠실4동, 반포3동, 여의도동 순으로 나타나 교통사고 안전관리를 위해 우선적으로 교통안전대책을 세우는 것이 필요함을 알 수 있다.



〈그림 2〉 서울시 행정구별 교통사고 발생 위험도 지도

V. 결론

본 연구에서는 서울시 각 행정동별 교통사고 발생건수에 영향을 주는 영향요인을 확인하고 행정동별 교통사고 발생에 대한 상대위험도를 추정할 수 있는 모형을 개발하였다. 특히 각 행정동별 교통사고 발생에 대한 공간적 상관성 존재 여부를 확인하고 공간적 상관성이 존재할 때 공간상관성을 모형에 반영한 두 종류의 공간모형을 제안하였다. 제안된 공간모형과 공간상관성을 반영하지 않은 기존 모형들과의 비교를 통해 공간모형의 우수성을 입증하였다.

본 연구는 포아송모형을 확장한 일반화 포아송모형을 소개하고 공간적 상관성을 반영한 공간모형을 통해 각 행정동별 상대위험도를 추정하여 교통사고 위험이 높은 지역을 평가할 수 있고, 상대위험도가 높은 지역에서의 교통사고 유발요인들을 역으로 파악하면 유발요인들이 교통사고 발생건수에 미치는 영향을 파악할 수 있다는 장점을 가지고 있다.

본 연구에서 제시한 두 가지 공간모형 결과 버스전용차선 설치율이 높고 교차로 수나 횡단보도 수가 많을수록 교통사고 발생건수가 증가하며, 횡단보도 예고표시율이 높고 과속방지턱이 많을수록 교통사고 발생건수가 감소한다는 것을 알 수 있어 교통사고 위험도가 높은 지역에 대해 우선적으로 횡단보도 예고표시와 과속방지시설을 설치하는 것이 필요하다고 할 수 있다.

이러한 모형의 결과는 교통사고 예방을 위한 환경 개선사업의 우선순위 결정, 순찰 강화 등의 교통안전정책 발굴, 지역적 교통사고 취약성과 위험도에 따른 교통사고 관리체계 확립 등을 추진하는 데 활용할 수 있다.

VI. 연구의 한계 및 향후 연구

본 연구에서는 2010년 서울시 교통사고 자료를 활용하여 종속변수 및 설명변수를 설정하였다. 그런데 교통사고 모형의 종속변수로 본 연구에서 사용한 교통사고 발생건수 외에 사망자 수, 부상자 수 등을 활용할 수 있다. 한 가지 예로 사망자 수는 0 값이 과대하게 관측되므로 영 과잉 포아송모형을 고려해 볼 수 있다.

영 과잉 포아송모형은 종속변수 값이 0 값을 과대하게 갖는 경우, 0 값이 관측되는 경우 등 두 가지로 나누어 베르누이분포와 포아송분포의 혼합된 분포 형태를 이용한 모형이다.

본 연구의 교통사고 모형을 구축하는 데 있어 설명변수로 설정한 행정동별 도로 연장이나 연면적보다 더 높은 영향을 줄 것으로 예상되는 행정동별 교통량 자료와 같은 주요 변수를 고려하여 추가적인 연구가 필요하다.

참고문헌

- 김학열 · 허태영, 2012, “일반화포아송회귀모형을 이용한 교통사고모형개발”, 『국토계획』, 47(1): 129~138.
- 박병호 · 김운환, 2007, “고속도로 교차점의 기하구조와 교통사고의 관계분석: 다중회귀분석, 포아송 회귀분석, 능형회귀분석 모형을 중심으로”, 『건설기술연구소 논문집』, 26(1): 43~58.
- 최정순 · 박만식 · 원유복 · 김학열 · 허태영, 2010a, “공간적 연관구조를 고려한 총범죄 자료분석”, 『응용통계연구』, 23: 335~344.
- _____, 2010b, “이변량 조건부자기회귀모형을 이용한 강력범죄자료분석”, 『한국통계학회논문집』, 17: 413~421.
- Akaike, H., 1974, “A new look at the statistical model identification”, 『IEEE Trans. Aut. Control』, 19: 716~723.

- Bae, S., Famoye, F., Wulu, J. T., Bartolucci, A. A. and Singh, K. P., 2005, "A Rich Family of Generalized Poisson Regression Models with Applications", *Mathematics and Computers in Simulation*, 69: 1~20.
- Cliff, A. and Ord, J., 1972, "Testing for spatial autocorrelation among regression residual", *Geographic Analysis*, 4: 267~284.
- Consul, P. C. and Famoye, F., 1992, "Generalized Poisson regression model", *Communications in Statistics, Theory and Methods*, 21: 89~109.
- Consul, P. C. and Jain, G. C. 1973, "A generalization of Poisson distribution", *Technometrics*, 15: 791~799.
- Famoye, F. and Singh, K. P., 2006, "Zero-Inflated Generalized Poisson Regression Model with an Application to Domestic Violence Data", *Journal of Data Science*, 4: 117~130.
- Famoye, F., Wulu, J. T. and Singh, K. P., 2004, "On the Generalized Poisson Regression Model with an Application to Accident Data", *Journal of Data Science*, 2: 287~295.
- Jovanis P. P. and Chang, H. L., 1986, "Modeling the relationship of accidents to miles traveled", *Transportation Research Record*, 1068, Transportation Research Board, Washington D.C.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A., 2002, "Bayesian Measures of Model Complexity and Fit" (with discussion), *Journal of the Royal Statistical Society, Ser. B*, 64: 583~639.
- Wall, M., 2004, "A close look at the spatial structure implied by the CAR and SAR models", *Journal of Statistical Planning and Inference*, 121: 311~324.
- Wang, W. and Famoye, F., 1997, "Modeling Household Fertility Decisions with Generalized Poisson Regression", *Journal of Population Economics*, 10: 273~283.
- Wulu, J. T., Singh, K. P., Famoye, F., Thomas, T. N. and McGwin, G., 2002, "Regression analysis of count data", *Journal of the Indian Society of Agricultural Statistics*, 55: 220~231.

원 고 접 수 일 : 2015년 3월 30일

1차심사완료일 : 2015년 5월 19일

최종원고채택일 : 2015년 7월 13일