

이변량 조건부자기회귀모형을 이용한 강력범죄자료분석

최정순^a, 박만식^b, 원유복^c, 김학열^d, 허태영^{1,e}

^aDivision of Biostatistics and Epidemiology, Medical University of South Carolina
^b성신여자대학교 통계학과, ^c서울시청 정보화기획담당관, ^d서경대학교 도시공학과
^e한국해양대학교 데이터정보학과

요약

본 연구에서는 5대 범죄 중 사람의 생명과 신체에 심각한 위해를 가하는 강력범죄인 살인과 강도 범죄의 이변량 가산자료에 대해 이변량조건부자기회귀모형을 사용하여 공간상관성을 반영한 강력범죄모형을 제안하였다. 범죄자료와 같은 가산자료에 대한 과대산포 검정을 위해 우도비 검정 실시하였으며, 그 결과 과대산포가 유의하지 않음에 따라 공간포아송모형을 이용하였다. 실증예제로 2007년 서울시에서 제공하는 25개 자치구별 강력범죄자료를 지리정보시스템을 이용하여 강력범죄 발생상태를 시각화하였으며 강력범죄에 영향을 주는 다양한 요인들에 대하여 분석을 실시하였다.

주요용어: 강력범죄, 공간포아송모형, 이변량 조건부자기회귀모형.

1. 서론

경제의 급속한 발전과 도시화에 따라 우리나라는 최근 10여년 간 총범죄 발생 건수가 2.5배 이상 증가하였으며, 특히 지역적 특성으로 농촌보다 도시지역에서 범죄의 발생이 상대적으로 크게 증가하는 것으로 나타났다. 행정구역 면적, 인구규모, 주택지 면적 등을 고려하지 않은 상태에서 범죄발생건수만으로 판단해 볼 때 서울지역 전체 강력범죄는 2005년 11만 2315건에서 2006년 10만 3487건으로 다소 감소했으나 2007년 10만 7380건으로 4,000건 이상 증가하였으며 2008년은 8월말 기준 6만 8950건에 이르는 것으로 나타났다. 2007년에는 서울에서 살인과 성폭행, 강도, 절도, 폭행 등 주요 강력범죄가 가장 많이 발생한 지역은 강남구로 모두 3만 2천 2백건에 달했으며 서초구는 2만 3천 7백건, 송파구는 2만 5백건 순으로 소위 강남 3구의 범죄발생건수가 가장 높은 수준으로 기록된 반면에 금천구에서는 9천 백건으로 범죄발생이 가장 적은 지역으로 집계되었고, 그 다음으로 도봉구(9,300건), 용산구(10,300건) 순으로 나타났다. 강도사건의 경우 2008년 4천 796건이 발생, 2004년에 비해 18.0% 줄었으나 2007년(4,439건)보다는 357건 늘었다. 2008년 기준으로 범죄 유형별 최다발생 지역을 살펴보면 구로구가 살인(16건) 및 강도(42건), 송파구는 강간(90건), 광진구는 절도(1,347건) 그리고 영등포구는 폭력(2,548건)이 가장 많이 발생하는 것으로 나타났다. 특히 구로구는 2006년 이래로 매년 살인 사건 다발지역 1위(2006년 16건, 2007년 19건, 2008년 16건), 송파구는 2007년 이래 강간사건 다발지역 1위(2007년 123건, 2008년 90건), 광진구는 2006년 이래 절도사건 다발지역 1위(2006년 1,859건, 2007년 1,862건, 2008년 1,347건), 영등포구는 2007년 이래 폭력사건 다발지역 1위(2007년 4,026건, 2008년 2,548건)로 나타났다.

이성우 (2004)는 총 범죄와 5대 범죄 자료와 관련하여 다양한 설명변수를 통하여 범죄모형을 구축하였으며, 특히 공간상관성을 반영한 모형인 공간계량경제모형을 이용하였으며, 이성우와 조중구

¹ 교신저자: (606-791) 부산시 영도구 동산동 1, 한국해양대학교 데이터정보학과, 조교수. E-mail: heoty@hhu.ac.kr

(2006)는 동일 자료에 대하여 공간 및 환경변수가 범죄피해에 미치는 영향을 연구하였다. 그러나 기존의 선행연구는 5대 범죄 중 특정범죄 사이의 상관성에 대한 고려가 거의 전무한 상태이다. 이에 본 연구는 높은 상관성을 나타내는 범죄간의 특성을 반영하여 모형을 구축하고 그 유용성을 평가하고자 하며 5대 범죄 중 살인, 강도와 같은 강력범죄는 서로간의 상관성이 높아 개별 모형보다는 두 강력범죄의 상관성을 고려한 이변량 모형(bivariate model)이 더 적합할 것으로 판단된다.

본 연구에서는 서울시 자치구에서 발생한 5대 범죄 중 살인, 강도에 대해 서울시 자치구별 발생현황을 살펴보고 범죄발생과 영향을 주는 다양한 요인과의 관계를 분석해 보고자 한다. 또한 지역별 범죄 발생자료를 지리정보시스템을 이용하여 서울시의 자치구별 강력범죄 발생 실태를 시각적으로 보여 주었으며, 이변량 조건부자기회귀모형(Bivariate conditional autoregressive model)을 이용하여 지역 간 공간연관성을 반영한 공간포아송회귀모형과 일반적인 포아송회귀모형을 비교하고자 한다. 본 연구는 다음과 같이 구성되었다. 2장에서는 살인과 강도에 대한 공간 통계 모형을 제안하였으며, 본 연구에서 사용한 모형비교 기준을 소개하였다. 3장에서는 본 연구에서 사용하는 살인과 강도에 대한 서울시 자치구별로 수집된 자료를 설명하고 자료분석 결과에 대해 다루었으며, 4장에서는 결론 및 향후 연구를 제시하였다.

2. 통계 모형

2.1. 이변량 범죄 모형

일반적으로 빈도자료에 사용되는 통계 모형은 포아송 모형이며 이러한 포아송 모형은 자료의 평균과 분산이 동일하다는 가정을 가진다. 또한 자료의 평균이 분산보다 더 작은 과대산포의 문제가 야기 되면 이를 고려한 모형인 음이항 모형이나 일반화 포아송모형(Famoye, 1993; Choi 등, 2009)을 사용하는 것이 더 적합하다. 이러한 과대산포를 검정하기 위하여 우도비검정을 실시한 결과 강력범죄 자료의 경우 과대산포가 나타나지 않아 일반적인 포아송 모형에 이변량 공간모형을 적용하였다.

따라서 본 연구에서 이변량 범죄 모형을 구축하기 위하여 범죄에 영향을 미치는 변수들과 공간 영향을 동시에 고려한 공간자료모형을 개발하였다. 모든 지역 $\mathbf{s}_i (i = 1, \dots, n)$ 에서, $Y_1(\mathbf{s}_i)$ 을 지역 \mathbf{s}_i 에서의 살인건수라고 하고, $Y_2(\mathbf{s}_i)$ 을 지역 \mathbf{s}_i 에서의 강도건수라고 하면 통계 모형은 아래와 같다. $k = 1, 2$ 에 대해

$$Y_k(\mathbf{s}_i) \sim \text{Poisson}(\lambda_k(\mathbf{s}_i)), \quad i = 1, \dots, n,$$

$$\log(\lambda_k(\mathbf{s}_i)) = \mu_0(\mathbf{s}_i) + \mu_k(\mathbf{s}_i) + \phi_k(\mathbf{s}_i),$$

여기서, $\mu_0(\mathbf{s}_i)$ 는 $\log(\mathbf{s}_i$ 지역의 10만명당 상주 인구수)이며 $\mu_k(\mathbf{s}_i) = \beta_{0k} + \beta_{1k}x_{i1} + \beta_{2k}x_{i2} + \dots + \beta_{pk}x_{ip}$ 는 평균함수이고 $\phi_k(\mathbf{s}_i)$ 는 공간상관성을 반영한 공간모형으로 표현될 수 있다.

본 연구에서는 $\phi_i = (\phi_1(\mathbf{s}_i), \phi_2(\mathbf{s}_i))^T$ 을 다음에 소개되는 다변량 CAR 모형인 MCAR($\rho, \mathbf{\Omega}_0$)을 이용하여 공간상관성을 나타내었다(Banerjee 등, 2004; Gelfand와 Vounatsou, 2003).

$$\Phi \sim N_{2n} \left(\mathbf{0}, \left[(\mathbf{D} - \rho \mathbf{W}) \otimes \mathbf{\Omega}_0^{-1} \right]^{-1} \right), \quad (2.1)$$

여기서, $\Phi = (\phi_1^T, \dots, \phi_n^T)^T$ 이고, $j, l = 1, 2$ 에 대해 $\mathbf{\Omega}_0 = \{\omega_0^{jl}\}$ 는 2×2 대칭적 양정치행렬이다. 행렬 $\mathbf{W} = \{w_{ij}\}$ 의 원소는 만약 지역 i 와 지역 j 가 서로 공간적으로 인접하면 $w_{ij} = 1$ 이고, 그렇지 않으면 $w_{ij} = 0$ 을 갖는 공간 인접 정보를 포함한다. 행렬 \mathbf{D} 는 원소가 $d_{ii} = \sum_{i \sim j} w_{ij}$ 인 $n \times n$ 대각행렬이고, d_{ii} 는 지역 i 에 인접한 지역의 개수를 의미한다. 여기서, 모수 $\rho \in (-1, 1)$ 는 공간적 연관성을 나타내는 모수로서, 만약, $\rho = 0$ 이면 식 (2.1)은 공간 상관성이 없는 모형이 된다. 하지만, 모수 ρ 는 상관계수는 아니다. 본 연구에서는 살인과 강도건수의 공간상관성을 알아보도록 하였다.

2.2. 베이지안 추정

모수추정 방법으로는 베이지안 방법을 통해 모수를 추정하였으며, 본 연구에서는 베이지안 통계 패키지인 WinBUGS(<http://www.mrc-bsu.cam.ac.uk/bugs/>) 프로그램을 사용하였다. 모수추정을 위해 모수에 대한 사전분포를 정의하여야 한다 (Heo와 Hughes-Oliver, 2009). WinBUGS 프로그램에는 MCAR모형의 함수가 포함되어 있지 않아서, 두 모수 ρ 와 Ω_0 을 WinBUGS 프로그램을 이용하여 직접적으로 추정할 수가 없다. 하지만, 본 연구에서는 WinBUGS 프로그램에서 추정 가능한 모형인 일반화된 다변량 CAR 모형(GMCAR; Jin 등, 2005)을 이용하여 MCAR 모형의 모수를 추정하였다. 다변량 CAR 모형 중 일부는 GMCAR의 특별한 경우로 표현이 가능하다. 본 논문에서 사용하는 다변량 CAR 모형인 MCAR(ρ, Ω_0)은 GMCAR($\rho, \rho, \eta_0, 0, \tau_1, \tau_2$)로 표현이 가능하며, 모수들 간의 관계는 $\omega_0^{11} = \tau_1^{-1} + \eta_0^2 \tau_2^{-1}$, $\omega_0^{12} = \eta_0 \tau_2^{-1}$ 그리고 $\omega_0^{22} = \tau_2^{-1}$ 이다. 따라서 GMCAR모형의 모수인 $\rho, \eta_0, \tau_1, \tau_2$ 에 대해 사전분포를 정의함으로서, MCAR모형의 모수인 ρ 와 Ω_0 을 추정하고자 한다. 모수에 대한 사전분포를 정의하기에 앞서, 각 모수 $\beta = (\beta_{01}, \beta_{02}, \beta_1, \dots, \beta_p)^T$, $\rho, \eta_0, \tau_1, \tau_2$ 간에는 상호 독립임을 가정하자. 그러면, 모수 벡터 $\theta = (\beta, \rho, \eta_0, \tau_1, \tau_2)^T$ 의 사전분포는 다음과 같다.

$$p(\theta) = p(\beta)p(\rho)p(\eta_0)p(\tau_1)p(\tau_2).$$

평균함수의 회귀계수에 대한 사전분포는 $\beta_{01}, \beta_{02} \sim \text{Normal}(0, 10^6)$ 이고, 모든 $l = 1, \dots, p$ 에 대해 $\beta_l \sim \text{Normal}(0, 10^6)$ 이다. 연결 모수 $\eta_0 \sim \text{Normal}(0, 10^6)$ 이고, Kelsall과 Wakefield (1999)에 따라서 $\tau_1, \tau_2 \sim \text{Inverse Gamma}(0.5, 0.0005)$ 이다. 공간 연관성 모수 ρ 는 균등분포를 따르며 공분산 행렬 Ω_0 이 양정치 행렬을 만족하는 범위를 사용하였다 (Banerjee 등, 2004).

모수 벡터 θ 에 대한 추론은 다음의 사후 분포를 기반으로 이루어진다.

$$p(\theta|\mathbf{Y}) \propto p(\mathbf{Y}|\theta) \times p(\theta).$$

주어진 자료에서 모수 추정은 마코프 연쇄 몬테카를로(Markov chain Monte Carlo; MCMC) 방법을 이용하였다.

2.3. 모형비교

공간상관성을 반영하지 않은 모형과 반영한 모형의 비교를 통해 서울시 자치구별 강력범죄모형에 가장 적합한 모형을 판단하기 위한 기준으로 베이지안적 모형비교 기준인 DIC(Deviance information criterion; Spiegelhalter 등, 2002)와 제곱근평균제곱예측오차(root mean squared prediction error; RM-SPE)를 이용하였다. 두 기준 모두 작은 값일수록 좋은 모형이 된다.

DIC는 자료에 대한 모형의 적합도(goodness of fit)와 모형의 복잡도(complexity)를 동시에 반영하는 척도로서 일반적인 적합도 기준인 AIC(Akaike information criteria)와 비슷한 척도이다. DIC는 이탈도(deviance)함수의 사후분포를 기반으로 하는데, 이탈도는 다음과 같다.

$$D(\theta) = -2\log[L(y|\theta)],$$

여기서, $L(y|\theta)$ 는 모수 θ 가 주어졌을 때 자료의 우도함수이다. DIC는 자료의 적합도를 나타내는 \bar{D} 와 모형의 복잡도를 나타내는 p_D 로 구성된다.

$$\text{DIC} = \bar{D} + p_D.$$

표 1: 모형에 사용된 자료 및 변수 설명

	변수	변수 설명(단위)	자료출처
종속변수	살인건수	살인 범죄건수	서울지방경찰청
	강도건수	강도 범죄건수	서울지방경찰청
설명변수	가구주 1인당 재산세	재산세/가구주 인구수	서울시 기본통계(통계연보)
	인구밀도	상주인구수/행정구역면적(km ²)	통계청 인구주택 총조사
	유동인구	명	통계청 인구주택 총조사
	청소년비율	15-24세	통계청 인구주택 총조사
	고학력비율	4년제 이상 대졸	통계청 인구주택 총조사
	개발제한구역비율	km ²	서울시 기본통계(통계연보)
	주택연상비율	주택연상면적(m ²)/총건물연상면적(m ²)	서울시 건축주택통계분석시스템
	숙박연상비율	숙박연상면적(m ²)/총건물연상면적(m ²)	서울시 건축주택통계분석시스템

이탈도의 사후 평균으로 정의되는 \bar{D} 는 다음과 같으며

$$\bar{D} = E_{\theta|y}[D(\theta)].$$

모형에서의 모수의 수를 의미하는 p_D 는 아래와 같이 정의된다.

$$p_D = E_{\theta|y}[D(\theta)] - D[E_{\theta|y}(\theta)].$$

모형을 평가하는데 있어서 DIC는 모형의 예측력을 반영하지 않고 있다. 하지만, 예측력 측면에서 모형을 평가하는 것도 상당히 유용하다. 따라서, 본 연구에서는 관측치와 예측치를 이용한 제공근평균 제공예측오차(RMSPE)를 이용하여 예측력 측면에서 공간 상관성을 반영한 모형과 반영하지 않은 모형을 비교하였다. 제공근평균제공예측오차는 아래의 식을 통하여 계산되었다.

$$RMSPE = \sqrt{\frac{1}{n} \sum_{i=1}^n [Y(s_i) - \hat{Y}(s_i)]^2}.$$

3. 실증예제

3.1. 자료설명

서울시의 자치구별 범죄발생과 다양한 요인들과의 연관성을 확인하기 위하여 서울시에서 제공한 범죄, 사회 경제학적 변수 그리고 도시계획 변수 자료를 통합하여 분석에 필요한 단일 자료를 구축하였다. 서울시의 자치구별 범죄발생에 대한 자료는 2007년 서울시의 자치구별 5대 범죄인 살인, 강도, 강간, 폭력, 절도 중에서 사람의 생명과 신체에 심각한 피해를 가하는 강력범죄인 살인과 강도 발생 빈도에 대하여 공간상관성을 반영한 모형과 반영하지 않은 모형을 비교하고, 공간모형으로는 이변량조건부자기회귀모형을 제안하였다.

본 연구에 사용된 범죄모형의 구축을 위하여 사용된 변수들에 대한 간략한 설명은 아래와 같다. 서울시에서 제공한 25개 자치구별 살인과 강도 범죄에 대한 빈도수에 대하여 개별 자치구와 일치하는 재산세, 인구밀도, 유동인구, 청소년비율, 고학력비율, 개발제한구역비율, 주택연상비율, 숙박연상비율을 설명변수로 자치구별 자료로 재구성하여 단일 자료를 구축하였다(서울시, 통계청). 모형에 사용된 자료 및 자세한 설명변인에 관한 설명은 표 1에 있다.

3.2. 결과

모수 추정 결과는 두 개의 초기값을 활용하여 20,000번 반복후 10,000번까지 제거하여 10,000번을 통한 결과값이며 구체적인 모수 추정값은 표 2에 제시하였다.

표 2: 모형별 모수 추정값과 베이지안 신뢰구간.

	Without Spatial Effects					With Spatial Effects				
	Mean	Std	2.5%	50.0%	97.5%	Mean	Std	2.5%	50.0%	97.5%
β_{01}	3.727	2.618	-1.283	3.757	8.992	6.743	5.271	-3.681	6.963	16.440
β_{02}	5.323	2.617	0.282	5.353	10.630	4.162	4.571	-4.808	4.169	13.100
β_1	-0.057	0.103	-0.256	-0.059	0.147	-0.017	0.178	-0.357	-0.018	0.338
β_2	-0.343	0.147	-0.626	-0.344	-0.054	-0.413	0.246	-0.942	-0.405	0.057
β_3	0.268	0.184	-0.091	0.264	0.629	0.226	0.340	-0.400	0.217	0.970
β_4	0.001	0.671	-1.283	-0.011	1.352	0.509	1.107	-1.659	0.482	2.711
β_5	-0.277	0.248	-0.758	-0.276	0.227	-0.219	0.464	-1.241	-0.188	0.653
β_6	0.005	0.006	-0.006	0.005	0.017	0.005	0.009	-0.013	0.005	0.023
β_7	-0.329	0.234	-0.785	-0.330	0.130	-0.254	0.418	-1.083	-0.267	0.554
β_8	0.158	0.041	0.078	0.159	0.237	0.180	0.073	0.031	0.182	0.323
ρ						0.411	0.478	-0.749	0.541	0.978
ω_0^{11}						0.608	0.401	0.045	0.538	1.598
ω_0^{12}						0.268	0.163	0.048	0.233	0.680
ω_0^{22}						0.145	0.096	0.030	0.125	0.384
γ						0.959	0.133	0.521	0.998	1.000

표 3: 모형 비교

Model	DIC	p_D	RMSPE
Without Spatial Effects	355.0	9.991	7.815
With Spatial Effects	319.3	25.617	4.064

표 2에서 보는 것과 같이 $MCAR(\rho, \Omega_0)$ 에서 공간상관계수 ρ 가 통계적으로 유의하지 않기에 강력범죄자료의 경우 공간상관이 존재하고 있지 않음을 알 수 있다. 공간상관성을 고려한 모형에서 살인과 강도의 상관계수, $\gamma = \omega_0^{12} / \sqrt{\omega_0^{11}\omega_0^{22}}$ 는 0.959이므로, 두 강력범죄간에는 서로 양의 상관관계가 높음을 알 수 있다. 공간상관성을 반영하지 않은 이변량포아송모형의 추정결과 강력범죄에 영향을 미치는 변수로서 인구밀도와 숙박연상비율이며, 재산세, 유동인구, 청소년비율, 고학력비율, 개발제한구역비율, 주택연상비율은 범죄율에 영향을 미치지 않는 것으로 나타났다. 특히, 인구밀도는 강력범죄에 부(-)의 영향을 나타내며, 숙박연상비율은 정(+)의 영향을 나타냈다. 인구밀도가 부의 영향을 미치는 이유는 강력범죄가 일반적으로 사람의 눈을 피해 인적이 드문 장소에서 발생하는 경향이 반영된 결과라고 판단된다. 또한 숙박연상비율의 경우, 숙박연상비율이 유흥가 및 상업지역의 특성을 대표하는 대리변수로 작용함으로써 강력범죄와의 상관성을 나타낸 것으로 추측할 수 있다. 이와는 대조적으로 공간상관성을 반영한 이변량포아송모형의 경우 인구밀도는 통계적으로 유의하지 않고, 숙박연상비율만이 정(+)의 영향을 주는 것으로 나타났다. 이러한 결과는 이성우 (2004)가 연구한 결과와 다른 양상의 결과를 보여주고 있다. 이는 각 년도별 범죄의 행태가 변화하고 있으며 이러한 변화를 인지하지 못한 개별모형 보다는 시간(년도별)과 공간(자치구별)의 변동성을 동시에 고려할 수 있는 패널자료분석 또는 시공간모형 등의 개발이 필요하다고 보여진다.

표 3에서는 DIC와 RMSPE를 이용하여 공간영향력을 고려한 모형과 그렇지 않은 모형을 비교하였다. 공간영향력을 고려한 모형이 모형 적합도 측면에서 비교하는 DIC값과 모형의 예측력 측면에서 비교하는 RMSPE값 모두 공간영향력을 고려하지 않은 모형보다 더 작은 값을 가짐을 알 수 있다. 즉, 강력범죄자료는 DIC와 RMSPE를 이용한 모형 비교에서 공간상관성을 반영한 모형이 잘 설명하고 있음을 알 수 있다. 따라서, 공간상관계수가 비록 통계적으로 유의하지 않다 하더라도 공간 효과가 반영된 모형이 모형의 적합도(DIC)와 예측력(RMSPE) 측면에서 더 적합함을 알 수 있다.

표 4: 서울시 25개 자치구별 강력범죄 관측값과 공간상관성을 반영하지 않은 모형과 반영한 모형과의 적합값 비교

자치구	Observation		without spatial effects		with spatial effects	
	살인	강도	살인	강도	살인	강도
종로구	8	41	7	35	8	39
중구	4	29	6	31	5	28
용산구	9	24	6	29	6	30
성동구	5	24	5	22	5	22
광진구	8	53	8	40	10	45
동대문	13	40	8	39	10	42
중랑구	13	40	8	38	10	41
성북구	12	40	9	46	9	44
강북구	12	47	9	43	11	46
도봉구	4	27	4	22	5	23
노원구	11	40	11	54	9	47
은평구	11	37	9	42	9	40
서대문	4	21	8	40	5	30
마포구	5	28	8	40	6	33
양천구	8	24	6	29	6	28
강서구	4	53	15	72	8	54
구로구	19	52	9	44	14	52
금천구	6	31	6	31	7	31
영등포	13	50	10	50	11	50
동작구	3	23	5	23	4	22
관악구	13	88	12	60	17	74
서초구	7	70	11	53	12	61
강남구	12	95	18	90	16	92
송파구	8	65	12	61	11	61
강동구	7	35	9	45	7	40

표 4는 서울시 25개 자치구별 강력범죄 관측값과 모형 적합값을 나타내었다. 두 강력범죄의 관측값과 모형적합값을 비교한 결과 역시 공간상관성을 반영한 모형이 관측값에 더 근접한 적합값을 제공함을 알 수 있으며, 그림 1에서는 지리정보시스템을 이용하여 관측값과 모형적합값의 차이를 시각적으로 표현하였다. 살인과 강도의 잔차 절대값이 공간상관성을 반영한 모형이 공간상관성을 반영하지 않은 모형보다 더 작은 값을 가진다는 사실을 지도를 통해서도 알 수 있다.

4. 결론

본 연구에서는 서울시의 각 행정 구별 발생된 중대범죄인 살인과 강도 자료에 대하여 이변량 공간 포아송회귀모형을 적합시키고 공간상관성의 여부를 판단하였다. 중대범죄인 살인과 강도의 연관성은 매우 높게 나타났으며, 이러한 연관성을 반영한 이변량 모형이 적합함을 알 수 있다. 또한 서울시 자치구별에 따른 공간상관성이 존재하지 않는 것으로 나타나 중대범죄 자료의 경우 공간적 연속성 또는 연계성이 그리 크지 않다는 것을 알 수 있다.

모형판단 기준인 DIC와 RMSPE 기준은 그 값이 작을수록 선택된 모형이 자료를 잘 설명하고 있음을 나타냄에 따라 공간상관성을 반영한 모형이 공간상관성을 반영하지 않은 모형보다 모형적합에 있어 강력범죄를 더 잘 설명하고 있다고 말할 수 있다. 또한 공간상관성을 반영하지 않은 포아송모형으로 얻어진 모수 추정값과 공간상관성을 반영한 공간포아송모형으로 얻어진 모수 추정값이 다름을 보여줌에 따라 공간상관성을 반영하지 않은 모형의 모수 추정값들이 부정확한 통계적 추론에 도달 할 수 있으

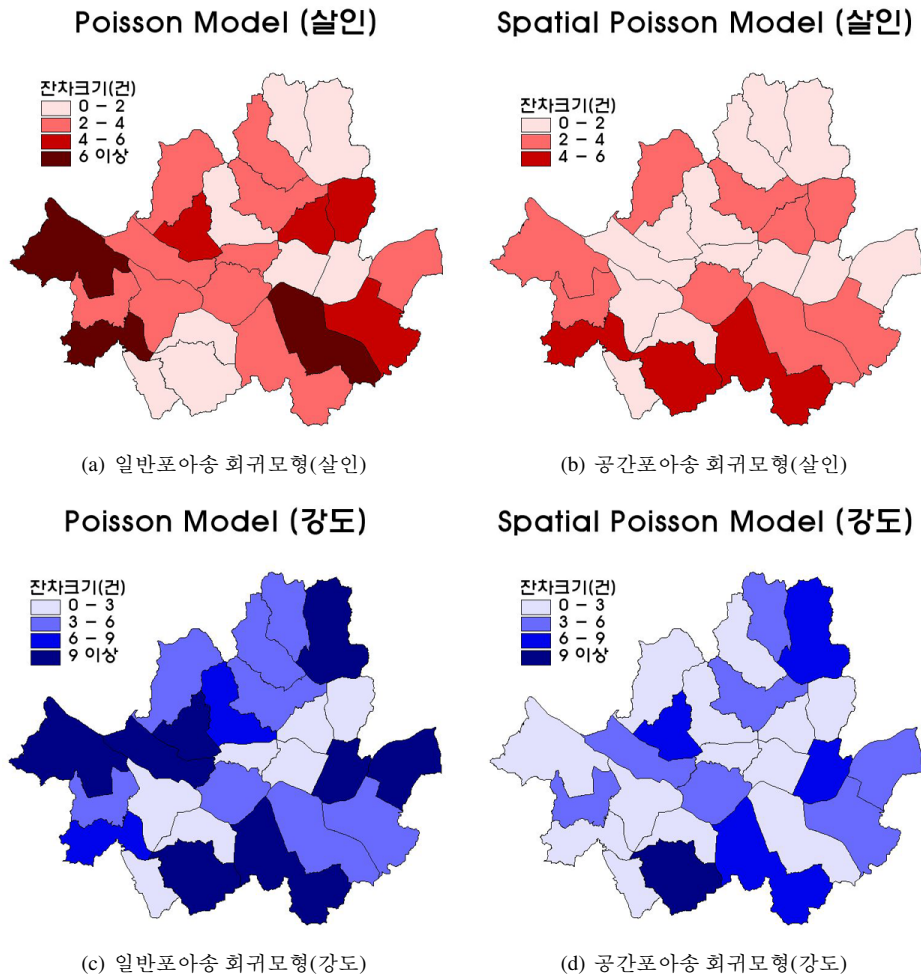


그림 1: 범죄건수의 잔차 절대값

며 이런 부정확한 추론으로부터 범죄모형에 대한 잘못된 결론을 도출할 위험이 있음을 알 수 있다.

본 연구의 경우 서울시의 범죄모형에 대해 고정된 년도(2007년)에 대하여 공간적 연관성(자치구 별)만을 고려한 한계가 있어 향후 공간과 시간을 동시에 고려한 모형의 확장 및 다양한 공간가중치 행렬을 통한 모형 개발을 연구 중에 있다. 또한 Stern과 Cressie (2000)이 제시한 공간 포아송 모형을 통한 각 지역별 질병에 대한 상대위험도(relative risk)를 추정한 연구와 같이 서울시 각 구별 범죄에 대한 지역별 상대위험도 추정 모형을 개발 중에 있다.

참고 문헌

이성우 (2004). 서울시 범죄발생의 도시계획적 함의, <서울시정개발연구원 2004-PR-13>.
 이성우, 조중구 (2006). 공간적, 환경적 요인이 범죄피해에 미치는 영향, <서울도시연구>, 7, 57-76.
 Banerjee, S., Carlin, B. P. and Gelfand, A. E. (2004). *Hierarchical Modeling and Analysis for Spatial Data*, Chapman & Hall, New York.

- Choi, J., Fuentes, M. and Reich, B. J. (2009). Spatial-temporal association between fine particulate matter and daily mortality, *Computational Statistics and Data Analysis*, **53**, 2989–3000.
- Famoye, F. (1993). Restricted generalized Poisson regression model, *Communication in Statistics-Theory and Methods*, **22**, 1335–1354.
- Gelfand, A. and Vounatsou, P. (2003). Proper multivariate conditional autoregressive models for spatial data analysis, *Biostatistics*, **4**, 11–25.
- Heo, T-Y. and Hughes-Oliver, J. (2009). Uncertainty adjustments to determine atmospheric dispersion models, *International Journal of Environmental Pollution*, In press.
- Jin, X., Carlin, B. P. and Banerjee, S. (2005). Generalized hierarchical multivariate CAR models for areal data, *Biometrics*, **61**, 950–961.
- Kelsall, J. E. and Wakefield, J. C. (1999). Discussion of “Bayesian models for spatially correlated disease and exposure data”, In *Bayesian Statistics 6*, eds. Bernardo, J. M., Berger, J. O., Dawid, A. P., and Smith, A. F. M. Oxford University Press, Oxford.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and Van der Linde A. (2002). Bayesian measures of model complexity and fit, *Journal of the Royal Statistical Society, Series B*, **64**, 583–639.
- Stern, H. S. and Cressie, N. (2000). Posterior predictive model checks for disease mapping models, *Statistics in Medicine*, **19**, 2377–2397.

2009년 11월 접수; 2010년 2월 채택

Analysis of Violent Crime Count Data Based on Bivariate Conditional Auto-Regressive Model

Jungsoon Choi^a, Man Sik Park^b, Yu-Bok Won^c, Hag-Yeol Kim^d, Tae-Young Heo^{1,e}

^aDivision of Biostatistics and Epidemiology, Medical University of South Carolina

^bDepartment of Statistics, Sungshin Women's University

^cDepartment of Information System Planning, Seoul Metropolitan Government

^dDepartment of Urban Engineering, Seokyeong University

^eDepartment of Data Information, Korea Maritime University

Abstract

In this study, we considered bivariate conditional auto-regressive model taking into account spatial association as well as correlation between the two dependent variables, which are the counts of murder and burglary. We conducted likelihood ratio test for checking over-dispersion issues prior to applying spatial poisson models. For the real application, we used the annual counts of violent crimes at 25 districts of Seoul in 2007. The statistical results are visually illustrated by geographical information system.

Keywords: Violent crime, spatial poisson model, bivariate conditional auto-regressive model.

¹ Corresponding author: Assistant professor, Department of Data Information, Korea Maritime University, 1 Dongsan-Dong, Yeongdo-Gu, Pusan 606-791, Korea. E-mail: heoty@hhu.ac.kr