

# GIS와 상관·회귀분석을 활용한 5대 범죄의 특성분석

## Analysis of the Five Major Crime Utilizing the Correlation-Regression Analysis with GIS

김창규\* · 강인준\*\* · 박동현\*\*\* · 김상석\*\*\*\*

Kim, Chang Kuy · Kang, In Joon · Park, Dong Hyun · Kim, Sang Seok

### 要 旨

현대사회의 시민들은 쾌적하고 안전한 생활공간과 환경을 희망한다. 그러나 현대사회로 접어들수록 급격하고 복잡한 변화를 맞이함으로써 지능적이고 고도화된 범죄가 빈번히 발생하고 있으며, 그 범죄들에 대응하기 위하여 다각적으로 노력이 행해지고 있다. 특히, 실생활에서 가장 많이 발생하고 두려움을 갖게 되는 5대 범죄(살인, 강도, 강간, 폭력, 절도)는 매우 위협적인 요소로 과학적인 대응방안이 필요하다. 이에 본 연구에서는 부산광역시 금정구 일대의 범죄 자료들을 바탕으로 5대 범죄에 대응하기 위하여 범죄유형별 빈도수를 분석하고 상관분석과 회귀분석을 이용하여 범죄 요인과 범죄 사이의 공간적 특성을 분석하였다.

핵심용어 : GIS, 5대 범죄, 상관분석, 회귀분석

### Abstract

People in the modern society want to live under safe and comfortable circumstances. As our society, however, is sharply developing, crimes are getting smarter and more difficult to treat. Above all, they often take place around us, and we are trying to cope with them variously in order to make our lives more comfortable and safer. In particular, five major crimes(Murde, Robber, Rape, Violence, Theft ) that most frequently occur in the real life are very threatening and fearful so it is necessary to deal with them with “the scientific method.” In this study, therefore, we searched the frequency of crime by its type and analyzed spatial characteristics between crimes and criminal factors by using regression analysis and correlation analysis based on the crime data that has occurred around Geumjeng-gu, Busan so that we can confront five major crimes.

Keywords : GIS, Five Major Crime, Correlation Analysis, Regression Analysis

## 1. 서 론

### 1.1 연구배경 및 목적

현대 사회의 시민들은 쾌적하고 안전한 생활공간과 환경을 희망한다. 그러나 급격한 발전으로 인해 여러 가지 문제점들이 대두되고 있다. 그 중 범죄는 시민들의 생활환경을 위협하고 직접적으로 영향을 미치는 매우 위험한 요소이다.

범죄 중 실생활에서 가장 많이 발생하고 두려움을 갖

게 되는 5대 범죄(살인, 강도, 강간, 폭력, 절도)는 매우 위협적이고 중요한 요소로 여겨지고 있으며, 5대 범죄의 문제점과 심각성을 해결하고자 최근 범죄의 실태, 원인, 범죄행동의 통제를 연구하는 많은 과학적인 방법들이 활용되고 있다.

본 연구의 목적은 5대 범죄발생에 대한 GIS 기법과 상관·회귀분석을 실시하여 범죄와 시공간 사이의 관계를 규명하고, 범죄 발생지에 대한 특성을 파악 하여, 범죄발생을 예측 하고자 한다.

Received: 2014.07.24, accepted: 2014.09.18

\* 정희원 · 부산대학교 사회환경시스템공학과 석사과정(Member, Department of Civil & Environmental Engineering, Pusan National University, kckya@naver.com)

\*\* 교신저자 · 정희원 · 부산대학교 사회환경시스템공학부 교수(Corresponding author, Member, Department of Civil & Environmental Engineering, Pusan National University, ijkang@pusan.ac.kr)

\*\*\* 정희원 · 부산대학교 사회환경시스템공학과 박사과정(Member, Department of Civil & Environmental Engineering, Pusan National University, sogone@pusna.ac.kr)

\*\*\*\* 정희원 · 동의과학대학교 산학협력단(Member, Ding-Eui Institute of Technology, kimss@dit.ac.kr)

1.2 연구동향

국내에서는 GIS를 이용한 범죄의 시공간적 패턴 연구(Shin and Hwang, 2005)를 비롯하여 GIS를 이용한 도시 범죄의 공간패턴 분석(Hwang and Hwang, 2003), 공간학적인 개념에서 범죄현상에 대하여 인식하고 원인을 규명하는 연구들이 꾸준히 진행되고 있다. 또한, 범죄발생의 공간적 특성을 파악하고 공간적 패턴과 범죄발생 요인간의 관계를 공간통계학적 관점에서 밝혀내고자 한 연구(kim et al. 2007, (Lockwood, 2007)등도 있다.

국외에서는 미국은 GIS를 이용한 범죄 지도를 제작하여 현업에 활용하고 있으며, 일본 및 유럽 등 선진국에서도 범죄현상에 대하여 GIS를 활용하여 공간적 측면에서 접근 하고 있다.

그러나 선행 연구들은 범죄의 공간적 발생패턴에 대한 접근은 부족하며, 선행 논문들에 제시된 대안들은 이론적 내용만을 포함하고 제시하는 수준에 머무르고 있다. 또한 통계분석기법도 공간 요인을 고려하지 않고 있어 분석결과의 신뢰성이 떨어진다고 판단된다.

1.3 연구방법 및 범위

본 연구의 방법은 먼저, 선행연구 및 문헌을 고찰하고, 연구의 방법과 범위를 선정하였다. 연구의 공간적 범위로는 부산광역시 금정구 일대를 선정하였으며, 연구자료는 2006년부터 2013년까지의 5대 범죄(살인, 강간·강제추행, 강도, 절도, 폭력은 제외)의 발생횟수, 범죄의 종류, 발생요일, 발생일시, 발생년도 등을 사용하였다. 마지막으로, 연구자료를 각각 분류한 후 사례지역의 범죄발생의 특성을 파악하고 GIS와 상관·회귀분석을 통한 통계적 분석을 실시하였다.

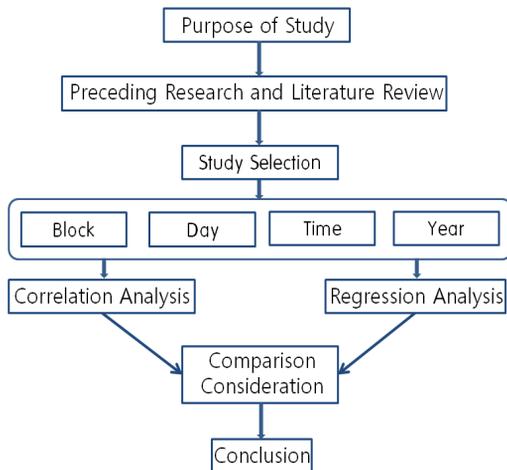


Figure 1. Study Flow Chart

석을 통한 통계적 분석을 실시하였다.

공간자료의 구축 및 분석은 ESRI사의 ArcGIS10.1을 사용하였고 통계분석은 SPSS21을 사용하였다.

Fig. 1은 전체적인 연구흐름도이다.

2. 범죄분석

2.1 연도별 범죄발생 분포

2006년부터 2013년도까지의 범죄의 종류와 범죄발생 횟수를 살펴보면, 2008년의 범죄 횟수가 2,414건으로 가장 많이 발생하였고, 2010년이 1781건으로 가장 작은 것으로 나타났다(Table 1).

2.2 시간대별 범죄발생 분포

금정구 일대의 시간대별 범죄발생 현황을 살펴보면, Table 2와 같이 모든 범죄 유형들이 야간에 집중하여 발생하고 있다. 범죄발생 건수는 절도, 강간·강제추행, 강도, 살인 순으로 나타남을 알 수 있다.

절도의 경우 04~07시를 제외하면 비슷한 추이를 보이고 있다. 강도와 강간·강제추행은 자정을 전후로 많이 발생하는 패턴을 나타냈다. 또한, 살인의 경우 자정에 발생횟수가 가장 많지만 우발적 상황에서 발생하는 경우가 많기 때문에 발생 패턴이 없는 것으로 나타난다.

2.3 요일별 범죄발생 분포

요일별 범죄발생의 경우, 평일보다는 주말에 더 많이 발생한다는 것을 알 수 있다. 절도의 경우 금요일과 토요일과 일요일에 많이 발생하며, 강간·강제추행도 평일보다는 주말에 많이 발생하는 패턴을 가지고 있다. 강도의 경우는 주말보다는 주중에 발생횟수가 더 많이 나타났다(Table 3).

총 범죄 수는 금요일과 토요일이 가장 많고, 일요일이 가장 적었으며, 월요일이 그 다음으로 범죄건수가 적게 나타났다.

Table 1. Status of Busan Geumjeong Crime

Division	Theft	Robbery	Murder	Rape	Total
2006	1904	26	4	38	1972
2007	1789	18	9	41	1857
2008	2337	33	4	40	2414
2009	2081	57	7	58	2203
2010	1664	21	5	91	1781
2011	1733	8	4	56	1801
2012	1829	16	2	59	1906
2013	1870	5	3	92	1970

Table 2. Distribution of Crime by Time Period

Division	Theft	Robbery	Murder	Rape
00	7811	132	25	208
01	289	7	0	15
02	317	3	0	22
03	237	5	0	18
04	176	10	0	23
05	114	5	0	7
06	121	2	0	10
07	155	0	0	9
08	264	1	0	4
09	328	0	8	1
10	370	1	0	13
11	313	1	0	7
12	322	1	1	3
13	350	1	0	4
14	352	2	1	6
15	369	1	1	1
16	357	1	0	7
17	368	1	0	6
18	463	2	2	11
19	421	2	0	9
20	435	1	1	7
21	399	1	3	17
22	463	1	1	17
23	412	3	2	34
계	15,206	184	45	459

Table 3. Distribution of Crime each Day

Division	Theft	Robbery	Murder	Rape	Total
Sunday	1950	26	4	70	2,050
Monday	2041	42	5	72	2,160
Tuesday	2159	24	9	66	2,258
Wednesday	2133	19	6	53	2,211
Thursday	2188	25	5	65	2,283
Friday	2551	27	5	76	2,659
Saturday	2245	21	4	70	2,340

Table 4. Distribution of Crime by Block

Division	Theft	Robbery	Murder	Rape	Total	CCTV	Average Population
Guseo	2129	20	6	88	2,243	34	56,402
Geumsa	523	4	0	14	541	12	10,764
Geumseong	32	0	0	2	34	4	1,204
Namsan	2040	9	3	32	2,084	29	35,711
Bugok	2722	26	7	76	2,831	46	60,773
Seo	2326	43	15	98	2,482	42	38,573
Sun · Dugu	298	4	0	9	311	8	2,877
Jangjeon	4277	63	3	124	4,467	36	34,330
Chungryoung · Nopo	638	14	3	28	683	16	10,259

2.4 행정동별 범죄발생 분포

Table 4는 금정구 행정동 별 범죄의 종류와 범죄건수를 나타내며, Fig. 2는 금정구 행정구역도를 추출하여 각 동의 범죄 발생 데이터를 백분율(%)로 나타낸다.

동별 범죄 발생 분포를 살펴보면, 절도는 장전동이 가장 많았으며, 강도, 살인, 강간강제추행은 서동이 가장 많이 발생하였다.

이는 기본적으로 평균 인구가 많을수록 범죄발생률이 높다는 것을 보여준다. 그러나 서동의 경우는 평균인구수를 고려하여도 다른 동에 비하여 범죄 발생률이 높은 것으로 나타났다. 장전동의 경우도 살인을 제외한 나머지 범죄가 인구수에 대비해서 구서동과 부곡동에 비해 높은 것을 알 수 있다. 범죄발생횟수가 가장 많은 지역은 장전동이며, 가장 적은 지역은 금성동이다.

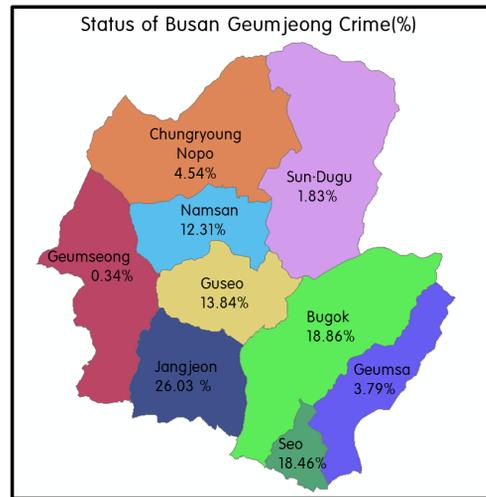


Figure 2. Status of Busan Geumjeong Crime(%)

### 3. 상관/회귀분석

#### 3.1 상관관계 분석

범죄분석 결과들을 SPSS를 이용하여 상관분석을 실시하였다. 변수간의 상관관계의 정도를 통계적으로 측정하는 방법 중에 가장 많이 사용하는 것이 “피어슨의 상관계수(Correlation Coefficient)이다. 상관계수는 r로 표기하며 r의 통계적 식은 아래의 식(1)과 같다.

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{(n-1)s_X s_Y} \quad (1)$$

단, n은 표본의 크기,  $s_X$ 와  $s_Y$ 는 각각 변수 X와 Y의 표준편차

상관계수 r은 -1부터 1까지의 값을 취할 수 있는데, 일반적으로 +0.1과 +1.0사이이면 양적 선형관계이며 -1.0과 -0.1사이이면 음적 선형관계를 가진다. r이 -0.1과 +0.1사이이면 거의 무시될 수 있는 선형관계이다.

Table 5는 각 동별 범죄발생수와 평균인구, 방범용 CCTV 운영 대수를 변수로 하여 상관분석을 실시한 결과이다. 상관분석 결과 모든 변수들이 양적 선형관계로 나타났다.

Table 6은 2006년부터 2013년까지의 전체 범죄건수

Table 5. Correlation Between the Number of CCTV

		Correlation Coefficient					
		Theft	Robbery	Murder	Rape	Average Population	CCTV
Theft	Pearson Correlation Coefficient	1	.894**	.483	.925**	.756*	.861**
	Significance Probability (Both Sides)		.001	.188	.000	.018	.003
	N	9	9	9	9	9	9
Robbery	Pearson Correlation Coefficient	.894**	1	.581	.937**	.528	.743*
	Significance Probability (Both Sides)	.001		.101	.000	.144	.022
	N	9	9	9	9	9	9
Murder	Pearson Correlation Coefficient	.483	.581	1	.667*	.638	.780*
	Significance Probability (Both Sides)	.188	.101		.050	.065	.013
	N	9	9	9	9	9	9
Rape	Pearson Correlation Coefficient	.925**	.937**	.667*	1	.759*	.868**
	Significance Probability (Both Sides)	.000	.000	.050		.018	.002
	N	9	9	9	9	9	9
Average Population	Pearson Correlation Coefficient	.756*	.528	.638	.759*	1	.929**
	Significance Probability (Both Sides)	.018	.144	.065	.018		.000
	N	9	9	9	9	9	9
CCTV	Pearson Correlation Coefficient	.861*	.743*	.780*	.868**	.929**	1
	Significance Probability (Both Sides)	.003	.022	.013	.002	.000	
	N	9	9	9	9	9	9

\*\* Correlation coefficient is significant at the 0.01 Masu level.

\* Correlation coefficient is significant at the 0.05 Masu level.

Table 6. Correlation of the Crime Count

		Correlation Coefficient		
		The number of crimes	Average Population	CCTV
The number of crimes	Pearson Correlation Coefficient	1	.756 <sup>a</sup>	.864 <sup>**</sup>
	Significance Probability (Both Sides)		.018	.003
	N	9	9	9
	Pearson Correlation Coefficient	.756 <sup>a</sup>	1	.929 <sup>**</sup>
Average Population	Significance Probability (Both Sides)	.018		.000
	N	9	9	9
	Pearson Correlation Coefficient	.864 <sup>**</sup>	.929 <sup>**</sup>	1
	Significance Probability (Both Sides)	.003	.000	
CCTV	N	9	9	9

\* Correlation coefficient is significant at the 0.05 Masu level.

\*\* Correlation coefficient is significant at the 0.01 Masu level.

와 평균인구, 방범용 CCTV 운영 대수를 이용하여 상관분석 하였다. ‘범죄건수’와 ‘평균인구’의 상관계수는 0.756, ‘범죄건수’와 ‘CCTV수’의 상관계수는 0.864로 나타났다. 상관분석 결과 범죄건수와 평균인구 수, 그리고 방범용 CCTV 수가 강한 양의 상관관계를 보였다.

### 3.2 회귀분석

선형회귀분석을 통해 변수들 사이의 관계에 있어 한 변수를 원인으로, 다른 변수들을 결과로 하여 원인변수들과 결과변수 사이의 선형식을 구하고자 한다.

회귀분석을 통하여 Table 7과 같이 절도, 강도, 살인, 강간·강제추행에 대한 회귀 계수를 얻었다.

회귀분석을 통해 얻은 계수를 바탕으로 얻은 회귀식은 식(2), (3), (4), (5)와 같다.

$$\text{범죄종류(절도)} = 356.498 + (0.047 * \text{평균인구}) \quad (2)$$

$$\text{범죄종류(강도)} = 6.937 + (0.000 * \text{평균인구}) \quad (3)$$

Table 7. The Coefficients in the Regression Model

Coefficients <sup>a</sup>					
Model	Non-standardized coefficients		Standardized Path Coefficients	t	significance probability
	B	Standard Error	Beta		
1 (Constant)	6.937	10.495		.661	.530
Average Population	.000	.000	.518	1.604	.153

a. The Dpendent Variable : Robbery

Coefficients <sup>a</sup>					
Model	Non-standardized coefficients		Standardized Path Coefficients	t	significance probability
	B	Standard Error	Beta		
1 (Constant)	.315	2.179		.145	.889
Average Population	.000	.000	.638	2.189	.065

a. The Dpendent Variable : Murder

Coefficients<sup>ø</sup>

Model	Non-standardized coefficients		Standardized Path Coefficients	t	significance probability
	B	Standard Error	Beta		
1 (Constant)	356.498	538.297		.662	.529
Avrage Population	.047	.015	.756	3.055	.018

ø. The Dependent Variable : Theft

Coefficients<sup>ø</sup>

Model	Non-standardized coefficients		Standardized Path Coefficients	t	significance probability
	B	Standard Error	Beta		
1 (Constant)	10.389	17.108		.607	.563
Avrage Population	.002	.000	.759	3.081	.018

ø. The Dpendent Variable : Rape

Table 8. Summary of Regression Models

Model Summary

Model	R	R Square	The modified R-squared	The standard error of estimate
1	.756 <sup>ø</sup>	.571	.510	978.148

ø. Predictive value : (Constant), Avrage Population

Model Summary

Model	R	R Square	The modified R-squared	The standard error of estimate
1	.518 <sup>ø</sup>	.269	.164	19.071

ø. Predictive value : (Constant), Avrage Population

Model Summary

Model	R	R Square	The modified R-squared	The standard error of estimate
1	.638 <sup>ø</sup>	.406	.322	3.960

ø. Predictive value : (Constant), Avrage Population

Model Summary

Model	R	R Square	The modified R-squared	The standard error of estimate
1	.759 <sup>ø</sup>	.576	.515	31.087

ø. Predictive value : (Constant), Avrage Population

$$\text{범죄종류(살인)} = 0.315 + (0.000 * \text{평균인구}) \quad (4)$$

$$\text{범죄종류(강간·강제추행)} = 10.389 + (0.002 * \text{평균인구}) \quad (5)$$

회귀분석을 통해 얻은 회귀식의 정확도를 검증하기 위해 모형요약을 실시하였다. 모형요약으로 R제곱, 수정된 R값 얻을 수 있으며, 이 값이 1에 가까울수록 선형 회귀식의 정확하다고 할 수 있다. 그러나 R제곱과 R

값이 절대적으로 얼마 이상이어야 된다는 규칙은 없다. R제곱은 변수가 많을수록 자연히 증가하는 성질이 있다. 모형요약을 통해 얻은 R제곱값(결정계수)을 살펴 보면, 각각 0.571, 0.164, 0.406, 0.576의 결정계수가 나타났다.

### 3.3 비교고찰

회귀분석을 이용하여 얻은 식을 이용하여 2008년도 부터 2011년까지의 부산광역시 동래구의 절도범죄 건 수와 평균 인구를 이용하여 금정구 회귀분석에서 얻은

Table 9. Dongnae Crime Number and Average Population

Year	Theft	Average Population
2013	2511	40534
2012	2617	40397

식에 적용시켜보았다.

Table 9는 동래구의 2012년과 2013년의 절도범죄건수와 평균인구이다.

$$\text{범죄종류(절도)}=356.498 \\ +(0.047*\text{평균인구}) \quad (2)$$

식(2)를 이용하여 절도 건수를 구하면 2012년은 2255건 2013년은 2261건으로 나타났으며 2012년은 13.8% 2013년은 9.9%의 차이가 났다.

비교고찰을 통하여 GIS의 공간분석 기능과 상관·회귀분석방법을 이용하여 범죄 현상의 공간적 특성을 밝혔다는 점에서 연구의 의의를 들 수 있다. 연구결과를 통하여 범죄예방을 위한 치안 활동과 더불어 정보를 제공하고 범죄 예방과 관련 등에 기초적인 자료로 이용될 수 있을 것으로 기대된다.

#### 4. 결론

본 연구를 통하여 부산광역시 금정구를 대상으로 하여 범죄발생의 분포 패턴을 파악하여 다음과 같은 결론을 얻었다.

첫째, 2006년부터 2013년까지 8년간 절도범죄가 가장 많이 일어났으며, 연도별로는 2009년이 가장 많았다. 발생시간대 별로는 심야시간대에 범죄 발생수가 집중되어 있고 가장 많은 범죄가 발생한 시간대는 00시로 나타났다. 동별 범죄 건수는 인구수와 CCTV가 많을수록 대체적으로 범죄 발생율이 높았다.

둘째, 상관분석을 통해 r값이 강한 양적 선형관계를 도출하고 금정구의 범죄발생 현황을 회귀분석을 통해 R값과 R제곱의 값이 강도는 0.2이상, 살인은 0.4이상, 절도와 강간·강간추행은 0.5이상의 선형관계가 있다고 나타났다.

셋째, 회귀식의 유효함을 알아보기 위하여 부산광역시 동래구의 2012년과 2013년의 절도범죄 횟수와 평균인구를 이용하여 금정구를 대상으로 얻은 회귀식에 적용시켜보았다. 2012년은 13.8% 2013년은 9.9%의 오차 범위가 나타났다.

본 논문에서는 평균인구수와 범죄발생수를 이용하여 회귀분석을 실시한 결과 평균 11%의 오차범위가 나타

났다. 향후 연구를 통해 독립변수가 되는 여러 요인들을 지역에서의 실정에 맞게 효과적으로 수집하고 분석할 수 있는 방법이 개발되어야 한다.

#### 감사의 글

이 논문은 부산대학교 자유과제 학술연구비(2년)에 의하여 연구되었음

#### References

1. Kim, D. G., Yun, Y. J. and Ahn, G. H., 2007, A study on urban crime in relation to land use patterns, Korea Planners Association, Vol. 42, No. 3, pp. 155-168.
2. Shin, W. R., and Hwang, C. Soo., 2005, The analysis of spatio-temporal pattern for urban crime: using gis methodology, The Korean Geographical Society 2005Year Spring Joint Meeting, The Korean Geographical Society, pp. 35-36.
3. Yang, B. S., 2010, Five major crime spatial characteristics analysis using gis, Master's Thesis, pp. 25-43.
4. Lee, H. H., 2000, Crime analysis using gis, Korean Institute of Criminology, Vol, 43, pp. 287-305.
5. Jeon, J. H., Kwon, J. H., Yang, H. J., 2007, Spatial crime analysis using gis, Korean Society for Geospatial Information System, Vol, 15, No. 1, pp. 3-7.
6. Choi, J. H., Kang, I. J., Kim, S. K. and Soel, M. H., 2011, Spatial features analysis of crime occurrence place using gis, Korean Society for Geospatial Information System, pp. 300-302.
7. Hwang, S. Y. and Hwang, C. S., Spatial pattern analysis of urban crime using gis, Korea Planners Association, Vol. 38, No. 1, pp. 53-66.
8. Hwang, C. H. and Suh, Y. C., 2008, A study on demonstrating the propriety of the location for police for jigudea(police box) by crime gis analysis, Korean Society of Hazard Mitigation, pp. 357-340.
9. Lockwood, d., 2007, Mapping crime in savannah, Social Science computer Review, Vol. 25, No. 2, pp. 194-209.