

범죄환경 위험도 지수 개발 연구 II

-성범죄 발생의 환경요인 분석-

윤 상 연

목 차

제1장 서론	1
제1절 연구배경 및 목적	1
1. 연구배경	1
2. 연구목적	3
제2절 연구내용 및 방법	6
1. 연구내용	6
2. 연구방법	9
제2장 범죄예측 방법의 이해	13
제1절 예측적 경찰활동과 범죄예측	13
1. 범죄예측 개요	13
2. 범죄예측의 기본 원리와 기법	15
제2절 범죄예측 기법의 종류	17
1. 핫스팟(HotSpot)	17
2. 데이터 마이닝	19
3. 인근반복 모형	21
4. 위험영역 모형	23
5. 시공간 분석	25
제3절 회귀분석 기법의 검토	28
1. 회귀분석 기법	28
2. 범죄예측 기법의 비교	31

제3장 성범죄 발생 특성 33

제1절 성범죄의 심리적 원인 33

- 1. 성범죄 원인에 관한 심리적 이론 33
- 2. 성범죄자의 심리적 특성 34
- 3. 심리적 요인의 적용 가능성 36

제2절 성범죄의 발생 특성 37

- 1. 통계분석 개요 37
- 2. 범죄통계 상 성범죄의 특성 38
- 3. 범죄통계의 해석 43

제3절 현장조사 결과 45

- 1. 현장조사 개요 45
- 2. 범죄발생 장소의 특성 46
- 3. 현장조사 결과 해석 49

제4장 범죄위험 지수 연구 53

제1절 연구개요 53

- 1. 연구목적 53
- 2. 분석방법 56

제2절 위험요인 분석: 유동인구를 중심으로 57

- 1. 유동인구의 효과 57
- 2. 환경변인과의 관계 분석 62

제3절 회귀분석 결과와 고려사항 65

- 1. 회귀분석 결과(서울지역) 65
- 2. 조절효과의 확인 68
- 3. 지역별 차이 확인 71
- 4. 시간적 요인의 고려 78

제5장 결론	81
제1절 연구 의의	81
제2절 범죄위험도 개선을 위한 제언	86
제3절 범죄예측 기술에 대한 제언	87
참고문헌	91
부록1	95
부록2	97

표 목 차

〈표-1〉	현행 범죄위험도 지수의 문제점과 대안	7
〈표-2〉	법집행 기관의 범죄(시간·장소)예측 기술	15
〈표-3〉	5대 범죄 발생률에 대한 위계적 회귀분석 결과에서 확인된 변인의 효과	44
〈표-4〉	성범죄 피해자의 유형별 피해 경험	51
〈표-5〉	조사결과별 예상되는 예측 변인	52
〈표-6〉	인구사회적 변인 목록	56
〈표-7〉	서울시 행정동별 유동인구 기초분석	57
〈표-8〉	유동인구 상위지역(20위까지) 범죄유형별 10년간 발생건수	58
〈표-9〉	유동인구 하위지역(20위까지) 범죄유형별 10년간 발생건수	59
〈표-10〉	유동인구 및 5대 범죄 발생률 간의 상관관계 분석 결과	63
〈표-11〉	유동인구 및 5대 범죄 발생건수와 인구특성 간 상관관계 분석 결과	63
〈표-12〉	유동인구 및 5대 범죄 발생건수와 사회특성 간 상관관계 분석 결과	64
〈표-13〉	유동인구 및 5대 범죄 발생건수와 거처유형 간 상관관계 분석 결과	64
〈표-14〉	유동인구 및 5대 범죄 발생건수와 주거환경 간 상관관계 분석 결과	65
〈표-15〉	회귀모형에 대한 단계별 분산분석표	66
〈표-16〉	강간·강제추행발생률에 대한 위계적 회귀분석 결과	67

〈표-17〉 회귀모형에 대한 단계별 분산분석표	69
〈표-18〉 유동인구와 범죄경력자수가 강간·강제추행을 예측하는 회귀분석 결과	69
〈표-19〉 지역별 환경요인 설명량과 이에 따른 군집 분류	72
〈표-20〉 강간·강제추행발생률에 대한 위계적 회귀분석 결과(전라북도)	74
〈표-21〉 강간·강제추행발생률에 대한 위계적 회귀분석 결과(충청남도)	76
〈표-22〉 강간·강제추행발생률에 대한 위계적 회귀분석 결과(서울특별시)	78
〈표-23〉 시간적 요인 유형과 관련 현상	79

그림목차

〈그림-1〉 GeoPros 기능별 이용현황(2017년 기준)	2
〈그림-2〉 범죄예측 알고리즘 개발 프로세스	5
〈그림-3〉 연차별 연구내용	8
〈그림-4〉 연구 방법의 도식화	11
〈그림-5〉 엘립스 기법과 커널밀도 추정법(런던)	19
〈그림-6〉 거리에 따른 강도 발생 빈도	21
〈그림-7〉 인근반복모형을 사용한 범죄예측 예시	22
〈그림-8〉 CIPM을 사용한 범죄예측 모형 예시	24
〈그림-9〉 시공간 클러스터링의 예시	27
〈그림-10〉 성범죄의 발생경로	37
〈그림-11〉 성범죄 유형별 발생장소	39
〈그림-12〉 성범죄 유형별 월별 발생수준	40
〈그림-13〉 성범죄 유형별 요일별 발생수준	40
〈그림-14〉 성범죄 유형별 시간대별 발생수준	41
〈그림-15〉 성범죄 유형별 피의자 연령대	42
〈그림-16〉 성범죄 유형별 피의자 학력	42
〈그림-17〉 성범죄 유형별 피의자 전과	43
〈그림-18〉 성범죄 유형별 피의자 전과	44
〈그림-19〉 관악경찰서 관할 강제추행 발생장소 예시	48
〈그림-20〉 안양동안경찰서 관할 강제추행 발생장소 예시	48
〈그림-21〉 관악경찰서 관할 노상 성범죄 발생분포(2017년)	49

〈그림-22〉 안양동안 경찰서 관할 노상 성범죄 발생분포(2014~2017) ……	50
〈그림-23〉 서울특별시 경찰서관할구역별 총범죄건수와 범죄율 비교 ……	57
〈그림-24〉 유동인구와 성범죄 발생수준 관계의 산포도 ……	62
〈그림-25〉 행정동별 유동인구수준 및 성폭력 발생수준의 분포 ……	63
〈그림-26〉 유동인구와 범죄경력자수의 상호작용 ……	73
〈그림-27〉 군집 3의 요인별 설명량 비율 ……	75
〈그림-28〉 군집 2의 요인별 설명량 비율 ……	77
〈그림-29〉 군집 1의 요인별 설명량 비율 ……	79
〈그림-30〉 요일 및 시간별 사건발생 수에 관한 예시 ……	82
〈그림-31〉 요일 및 시간별 리듬 예시 ……	82

제1장 서론

제1절 연구배경 및 목적

1. 연구배경

경찰청은 증가하는 치안수요에 부응하기 위해서 많은 노력을 기울이고 있다. 탄력순찰제 운영 등, 국민의 구체적인 요구에 부합하는 치안서비스를 제공하기 위한 방안들을 도입하고 있다. 또한 도입한지 3년차를 맞은 범죄예방 경찰관(Crime Prevention Officer)들도 범죄예방 업무를 위한 역할들을 수행하고 있다.

그러나 범죄예방의 중요성이 커지고 있음에도 불구하고 범죄예방의 기초가 되는 범죄예측 분야의 활동이 충분히 활성화되고 있지 않다. 한쪽에서는 새로운 범죄예측 시스템이 도입되고 있지만¹⁾ 범죄예방 관련 전문인력의 부족과 조직내 인식 부족으로 인해 여전히 발생한 범죄에 대한 사후적 대응에 초점이 맞춰져 있는 것이 현실이다. 치안 수요의 증가에도 불구하고 지역 치안의 주요 담당자인 지구대나 파출소 역시 사건·사고 대응에 바빠 예방차원의 활동은 거의 못하고 있는 실정이다.

예방 활동에서 중요한 정보를 제공하는 범죄예측 기능의 경우 여전히 의지할 수 있는 정확하고 일관성 있는 예측자료가 없기 때문에 상급자의 개인적 경험이나 지역의 순찰 관행에 의존하는 경우가 많다. 범죄예측 기능을 가지고 있는 경찰청 시스템인 GeoPros의 기능별 이용건수 비교를 보면 경찰 내에서 범죄예

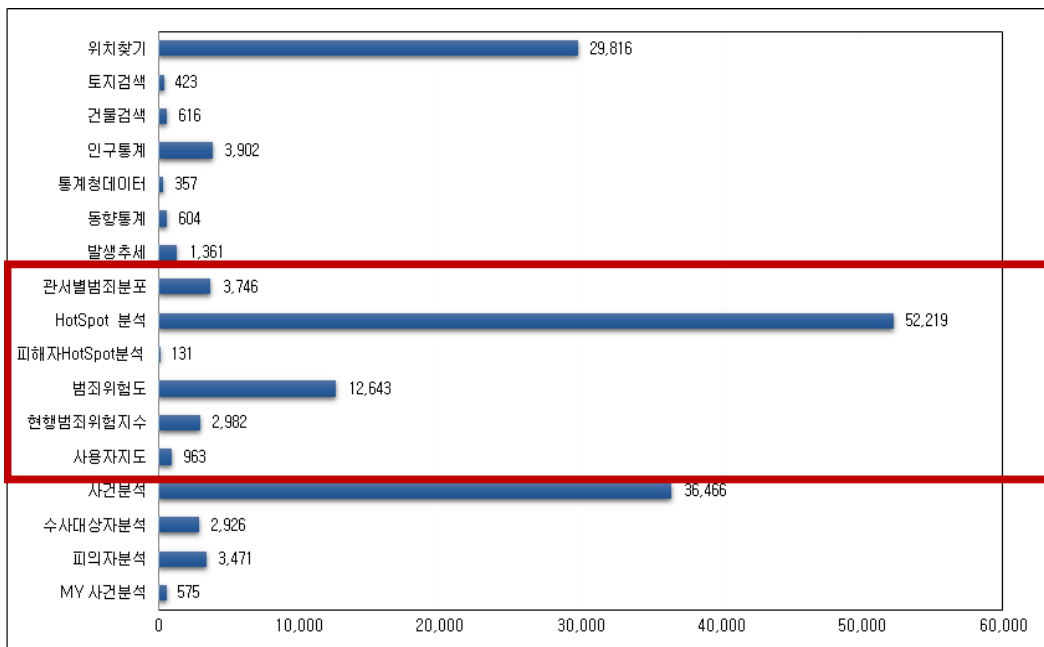
1) 경찰청에서 개발 중인 범죄분포 이해도구(Crime Layout Understanding Engine; CLUE. 2016~2018)는 빅데이터를 이용하여 범죄예측을 하는 기능을 탑재하고 있다. 또한 영등포구와 KT는 공동으로 범죄예방을 위한 도시환경 설계 플랫폼을 구축하였다.

측 기능에 대한 활용도, 그리고 그 기초가 되는 범죄위험도에 대한 신뢰 수준을 알 수 있다.

기능별 활용 수준을 보면, 범죄발생 지점에 대한 확인을 과거에 발생한 범죄의 기록인 HotSpot 분석을 통해 하는 경우가 대부분이라는 점을 알 수 있다.

<그림 1> GeoPros 기능별 이용현황(2017년 기준)

(단위: 회)



그 이유는 경찰관들이 지역의 사회적 요인 또는 타범죄발생에 의해 예측되는 범죄위험도(범죄위험지수)에 대한 인식(홍보) 부족이 원인일 수도 있지만 결과에 대한 정확성을 신뢰하지 않기 때문으로 보인다. 시스템 관리자들에 따르면 실제 자신이 경험한 위험도와 예측된 위험도가 일치하지 않는 것이 가장 큰 이유로 꼽히고 있다고 한다.

또한 편의성 측면에서도 범죄위험도는 단위 구역(치안블록 또는 기초경계구역) 별로 지수(색상) 형태로 동일하게 제시되는 반면, 핫스팟의 경우 발생 지점이 비교적 구체적으로 표시되기 때문에 지역 내에서의 순찰 대상지를 선정하는

데 더 상세한 정보를 제공해준다는 차이도 있다.

범죄예방 활동이 체계적으로 수행되기 위해서는 일관성 있는 범죄예측 및 범죄위험 요인에 대한 평가가 선행될 필요가 있다. 이를 위해서는 범죄위험도(지수)에 대한 지속적인 연구가 중요하다. 현재의 기술적 수준 및 제도적인 제한 내에서는 범죄예측의 정확도가 높지 않을 수 있다. 그러나 모든 과학적인 성과들은 하루아침에 이루어지지 않는다. 범죄예측이 중요하다면 장기적이고 체계적인 계획을 수립하고 시행착오를 통해서 발전시켜 나갈 필요가 있다.

특히 최근 주요 범죄 감소 추세에도 불구하고, 성범죄는 여전히 증가추세에 있다(치안정책연구소, 2018). 최근 ‘미투(#Me Too) 운동’으로 인해 촉발된 우리 사회의 성범죄에 대한 국민적인 관심도 증가했다. 그러나 기존의 성범죄 발생 원인과 발생 실태에 대한 연구는 범죄를 예측하고 예방활동을 벌이기에는 여전히 부족한 부분이 많다. 성범죄는 세부 유형별로 발생형태가 다르기 때문에 구체적인 연구가 필요한 상황이다. 또한 성범죄는 암수범죄가 많은 대표적인 범죄 유형이기 때문에 연구의 필요성은 더욱 크다고 하겠다.

이런 상황에서 지자체나 민간에서는 안전에 대한 주민의 수요에 대응하기 위해서 독자적으로 범죄예측 및 예방 시스템을 앞다투어 개발·적용하고 있다. 예를 들어 KT에서는 최근 발생이 증가하고 있는, 또는 두려움이 증가하고 있는 성범죄에 대한 대응을 위한 범죄예측시스템을 새롭게 구축하였다.²⁾ 치안활동의 중심에 있는 경찰이 범죄예측을 제대로 못 할 경우에는 효과적인 범죄예방 활동으로 이어질 수 없는 문제가 있다.

2. 연구목적

본 연구는 범죄환경 위험도 지수 개발을 위한 2차년도 연구로서, 위험도 지수는 매우 실증적인 문제임에도 불구하고 1차년도에는 문헌연구 및 국내외 사례연구, 그리고 현행 범죄위험도 기능에 대한 분석, 가능한 범죄환경 지표를 제안 하였으나 실증적인 검증과정이 없었기 때문에 변인의 효과를 보장할 수 없었다.

2) 한국일보, “빅데이터로 범죄 예측... 한국판 ‘마이내리티 리포트’ 구축한다.”, 2018. 09. 07.

따라서 실증적인 자료를 통해서 범죄환경 위험도 지수의 정확성을 확인하고 개선방안을 찾을 필요가 있다. 본 연구에서는 1차년도 연구에서 도출된 두 가지 연구방법을 사용하여 현재 범죄위험도를 개선할 수 있는 방법을 찾고자 한다. 첫째는 현장 조사를 통해서 기존에 제안된 범죄환경 요인들이 실제로 범죄 발생과 관련성이 있는지를 확인하는 것이다. 둘째는 통계자료를 바탕으로 지역별 범죄발생과의 관계를 통계적으로 분석하는 것이다.

범죄위험도지수의 정의는 다음과 같다.

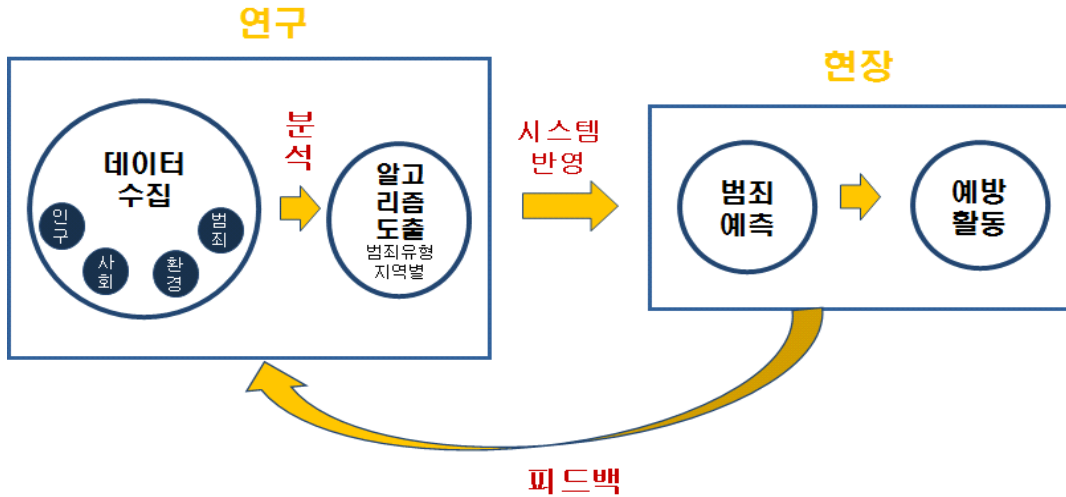
범죄위험지수는 전국을 36만 7천여개의 유사 블록으로 세분화하고 최종별, 시간대별 범죄위험도 예측 결과를 확인할 수 있는 기능이다. 범죄위험지수는 유동인구, 연령대별 인구 구성, 전과자수, 과거 3년간 범죄 동향, 최근 1주일 이내 주변 범죄 동향, CCTV, 경찰관서 접근성, 기상 정보 등 약 42개의 범죄유관변인들을 통해 범죄 위험도를 예측, 수치화한 정보이다. 범죄위험지수는 매일 시간대별(6개 시간대)로 산출된다.

출처: 지리적 프로파일링 시스템(GeoPros) 사용자 매뉴얼(2016)

이처럼 범죄위험지수는 매우 다양한 변인을 사용하여 위험도를 산정하고 있으며, 통계적으로 정확한 결과 산출을 위해서는 많은 과정이 요구되고, 다수의 조건이 충족될 필요가 있다.

완벽한 수준의 범죄위험도 개발을 위해서는 여러 가지 조건이 충족될 필요가 있다. 크게 데이터, 연구-분석, 시스템구축이라고 하는 세 개 단계로 구분할 수 있고, 각각의 단계에 대해서 복잡한 절차가 필요하다. 현재에는 세 개 단계 모두에서 한계가 확인되고 있어 각각에 대한 실태 분석 및 개선방안에 대한 모색이 필요하다.

<그림 2> 범죠평측 알고리즘 개발 프로세스



범죠평측은 일반 이론이나 외국 환경에 맞게 개발된 시스템 도입만으로는 부족하기 때문에 국내 환경에 맞는 알고리즘과 시스템 구축이 중요하므로 타당성 있는 위험도의 산출을 위한 시스템 구축을 위해서는 실제 적용대상 지역의 실증 데이터 기반 연구가 반드시 필요하다.

본 연구의 목적을 정리하면 다음과 같다.

첫째, 거시/미시 요인과 유발/억제 요인 등 개념적인 구분을 통해 예측요인의 작용방식을 고려하여 실질적인 예측요인의 발굴 및 예측 모델 수립을 위한 기초 자료를 제공한다.

둘째, 환경요인에 의한 범죄발생 예측의 정확성을 제고하여 범죠평방 정책 수립 및 활동을 위한 근거를 마련한다.

셋째, 현행시스템의 개선 또는 범죄위험도를 예측할 수 있는 신규 시스템의 개발 지원과 연계한다.

본 연구에서는 먼저 강간, 강제추행 등 성범죄(카메라등 이용 불법촬영 포함)의 발생원인을 환경적인 기회요인의 관점에서 분석하여 성범죄유형별 발생특성

을 검증하는 등 성범죄 발생 기제를 파악하고자 한다.

그리고 범죄예측에 실효성 있는 범죄환경 지표를 선정하고 데이터 분석 결과를 제시함으로써 성범죄예측을 위한 지수 개발을 위한 근거자료 제공할 것이다.

이러한 연구를 통해서 얻을 수 있는 기대 효과는 다음과 같다.

첫째, 세부 지역 단위 검토, 치안 활동 단위를 구체화, 개별 지역 블록 별 범죄 위험도의 개발을 통해서 경찰관의 현장 활동에 효과적으로 적용될 수 있도록 시스템을 개선할 수 있다.

둘째, 물리적 환경 중심의 범죄원인을 확인함으로써 범죄예방활동을 위한 실질적인 전략 수립이 가능하다.

셋째, 사회적으로 문제가 되고 있는 성범죄 대책에 연계하여 범죄취약계층인 여성에 대한 보호정책에 기여할 수 있다.

제2절 연구내용 및 방법

1. 연구내용

2017년 수행된 1차년도에는 다음과 같은 성과들을 얻을 수 있었다. 첫째, 외국의 범죄예측 시스템은 다양한 예측방법을 사용하고 있으며, 범죄 발생량이 많은 대도시를 중심으로 운용 중이다(ex. PredPol, Crime Insight & Prevention, HunchLab, PCA). 둘째, 현행 시스템(GeoPros)을 여러 차례 고도화 사업을 통해 개선해 왔지만 범죄위험도의 정확성 문제(예측 알고리즘) 개선에는 소극적이다. 셋째, 문제 해결 방안으로 데이터 수집 방식의 개선(CPO 활용), 알고리즘 개선이 필요하다.

본 연구는 실증적인 연구를 중심으로 하는 만큼 연구의 범위 설정이 중요하다. 연구목적에서 밝힌 바와 같이 본 연구는 범죄예측에 대한 일반적 논의를 위해 직접적으로 GeoPros의 범죄예측 기능에 탑재되어 있는 범죄위험도 지수를 개선하기 위한 연구이며, 2차년도 연구에서는 현재 GeoPros에 적용되고 있는 범죄위험도의 수정을 위한 실증적인 연구를 수행하고자 한다.

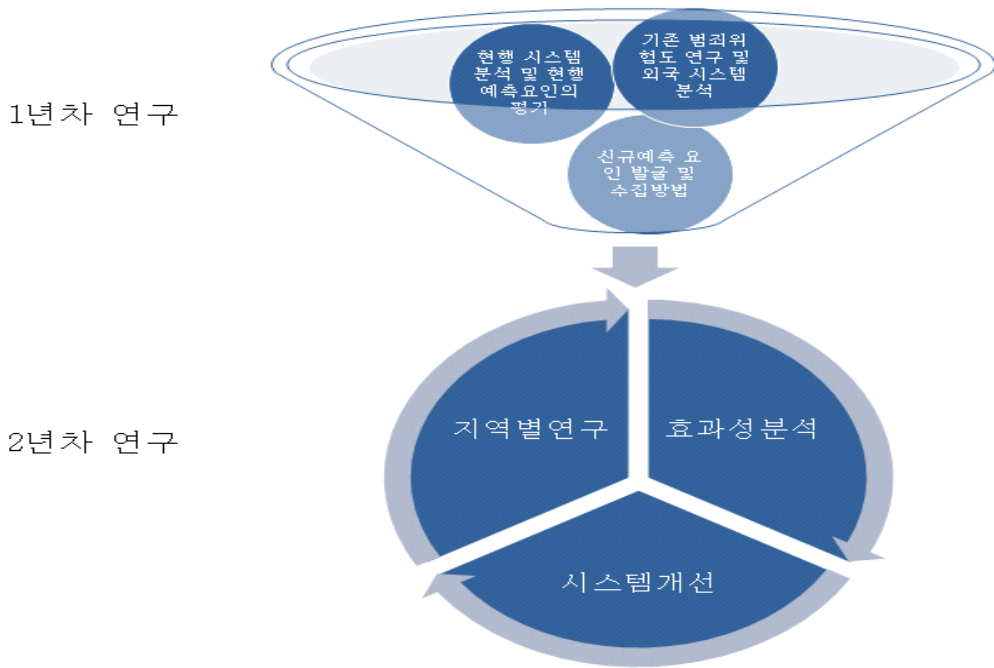
1차년도 연구에서는 기존 범죄위험도의 한계를 분석하고 이에 대한 대안으로 알고리즘 개선 방안 및 신규 변인의 수집을 위한 방안을 제안하였다. 2차년도 연구 수행과 관련된 1차년도 연구의 주요 성과를 발견된 문제(한계)와 대안 중심으로 구체적으로 제시하면 다음과 같다.

<표 1> 현행 범죄위험도 지수의 문제점과 대안

유형	문제	대안
공간 특성	데이터와 단위지역의 공간적 범위가 불일치하여 구역별 차이 반영 안 됨	적절한 범죄예측 단위 구역 설정
시간 특성	주요 변인의 업데이트 주기 느리고 시간대, 날씨 변인에 의해 변화	유동인구, 112 데이터 등 실시간 정보 활용
지역 특성	시군구별로 별도의 알고리즘을 적용하고 있어 인과적 관계에서의 파악 곤란	도시-농촌, 토지용도 등 지역특성에 따른 구분/통합
변인 관계	영향 요인의 효과를 선형적 관계로 전제하여 회귀분석에만 의존	지역별 영향 요인 간 관계 차별적 반영

본 연구의 연차별 연구내용을 제시하면 아래의 그림과 같다.

<그림 3> 연차별 연구내용



구체적인 검증은 위해서는 범죄유형별로 심층적인 연구가 필요한데, 최근 성범죄 발생이 증가하고 있고 지속적으로 사회적 이슈화할 가능성이 높아 우선순위를 두어 연구가 필요하다. 본 연구에서는 범죄유형을 성범죄로 제한하여 연구하고자 한다.

성범죄 예방전략 수립을 위해서는 정확한 범죄원인의 파악과 범죄예측이 필수적이며, 물리적 환경 요인은 범죄의 유발과 억제 과정에 관한 중요한 단서를 제공할 것으로 기대된다.

2. 연구방법

본 연구는 범죄예측을 위한 예측요인의 발굴 및 확인을 그 목적으로 한다. 이를 위해 현장조사, 데이터 분석 및 전문가 자문 등을 통해서 연구결과의 타당성을 확보하였다. 각 연구 방법과 관련된 사항을 아래와 같이 개괄적으로 소개하였다. 각 장별로 서로 다른 연구내용과 연구방법을 포함하고 있기 때문에 구체적인 연구방법은 각 장별에서 별도로 소개할 예정이다.

가. 문헌조사

최근 범죄예측은 범죄예방 기조에 발맞추어 상당한 수준으로 발전하고 있다. 경찰청 내에서도 GeoPros뿐만 아니라 경쟁력을 가진 시스템(CLUE; Crime Layout Utility Engine)이 완성 단계에 있고,³⁾ 유사한 기능을 가진 시스템들의 연구개발이 한창이다. 각종 지자체나 민간 기업들도 사회 안전에 대한 수요가 증가함에 따라서 범죄예측 및 예방을 위한 서비스를 제공하고 있다. 또한 외국의 다양한 시스템들이 소개되고 있어 이에 대한 참고는 물론, 그 허와 실에 대한 검토를 통해서 GeoPros의 기능 개선에 참고할 필요가 있다.

- GeoPros 매뉴얼 분석 등을 통한 예측시스템 심층 분석
- 성범죄 발생특성에 대한 이론적 검토 및 범죄통계 분석
- 외국의 범죄예측 기법 및 사례에 대한 분석⁴⁾

나. 현장조사

먼저 현장조사를 통한 범죄발생 환경 분석을 위해서 실제 범죄발생 장소(위치)에 관한 자료를 확보한 후 주변의 물리적 환경을 비교분석하였다.

3) 중앙일보, “52억짜리 AI 수사관 ‘클루’가 ‘살인의 추억’ 재발 막는다.”, 2017. 12. 08.

4) 본 연구에서는 랜드 연구소(RAND Corporation)에서 발간하는 보고서(Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations, 2013)의 내용을 상당부분 소개하였다. 랜드 연구소는 1948년에 설립된 미국 국방·행정 분야의 대표적인 민간연구소이다.

- 범죄발생 현황 분석을 통해 범죄 빈발 지역 및 이와 인접한 중립지역을 파악하여 일정 지역을 샘플로 선정하여 환경 지표 수집 및 분석

다. 데이터 분석

- 확보 가능한 범죄 발생 데이터 및 유동인구 등 범죄발생 관련 환경특성 자료를 활용한 통계분석 실시
- 실제 시스템 운영 수준(치안블록, 기초경계구역)이 아니라 행정동 수준에서 변인의 효과, 지역적 범위의 효과 등에 관해서 원형적인 분석(개발업체와의 공유 통해서 연구 결과 적용) 실시
- 지역별 인구·사회·경제적 차이를 고려(효과 통제)하여 물리적 환경의 순수 효과 확인
- 공간분석기법(space analysis technique)을 이용한 환경 특성 분석

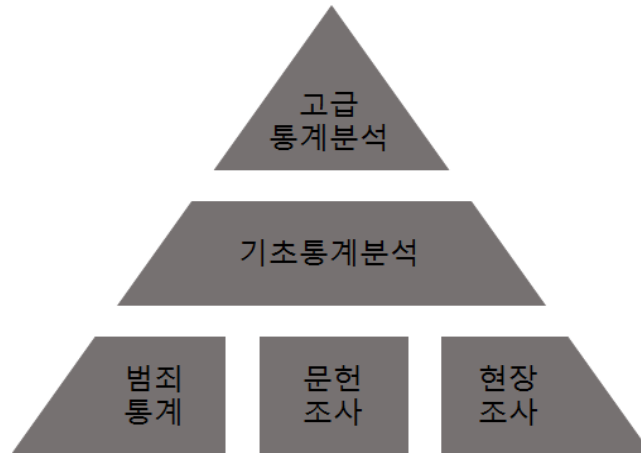
라. 전문가 자문

- 범죄학 전문가 및 지역 범죄예방경찰관(CPO)의 자문을 통해서 연구결과의 타당성 확보

■ 연구 프로세스

1. 문헌조사: 분석방법 도출
2. 범죄통계/현장조사: 요인 추출
3. 기초통계분석: 기초분석을 통한 유동인구, 범죄발생 현황에 대한 분석
4. 회귀분석 등 고급통계분석: 회귀분석을 통한 회귀식의 도출 및 환경요인의 상호작용 효과 분석

<그림 4> 연구 방법의 도식화



제2장 범죄예측 방법의 이해

제1절 예측적 경찰활동과 범죄예측

1. 범죄예측 개요

예측적 경찰활동은 통계적 예측을 통해 경찰 개입의 가능성이 있는 대상을 식별하고 범죄를 예방하거나 과거 범죄를 해결하기 위한 분석 기법(특히 양적 기법)을 적용하는 경찰활동 방법을 말한다.⁵⁾ 현재 미국 전역의 법집행 기관에서는 선도적으로 몇 가지 경찰활동 방법을 사용하고 있다.⁶⁾ 미국 이외에 영국, 중국 등에서도 발전된 컴퓨팅·IT 기술을 이용한 범죄예측 기술을 도입하고 있다.

예측 방법을 통해 경찰은 제한된 자원으로 보다 적극적인 치안 활동을 할 수 있다. 이 방법의 목적은 범죄를 예방하거나 수사를 보다 효과적으로 수행할 수 있는 전략을 개발하는 것이다. 그러나 예측적 경찰활동의 기초가 되는 범죄예측을 통해 영화 ‘반지의 제왕’에 등장하는 수정구슬(Palantir)처럼 원하는 것을 마음대로 들여다 볼 수 있는 것은 아니다.

전략이 효과적이라고 평가받기 위해서는 확실한 근거가 있어야 한다. 그래서 범죄예측에서는 치안과 관련된 데이터를 활용하여 통계적인 분석을 수행한다. 과거에는 범죄예측에 수학적으로 단순한 방법을 주로 사용하였기 때문에 데이터량이 적은 경우에도 사용이 가능했고, 초보 수준의 분석가가 사용할 수 있었다. 그러나 최근 증가하고 있는 예측 분석에는 대규모 데이터 세트와 정교한

5) Perry, W. L., (2013), “Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations.”, RAND Corporation, p.xiii.

6) 예측(prediction)과 예언(forecasting)은 비슷한 의미로 사용되기도 하지만 엄격한 의미에서는 서로 구분되는 개념이다. 예측은 객관적이고 과학적이며 재현 가능성이 있는 반면, 예언은 주관적이고 직관적이며 재현 불가능한 차이가 있다.

분석 방법이 사용된다. 따라서 범죄예측을 위해서 특별한 컴퓨터 프로그램이나 유사한 자원이 필요하다.

법집행 기관이 사용하는 넓은 의미의 범죄예측 기법에는 대표적으로 ‘범죄’, ‘범죄자’, ‘범죄자 신원’, 그리고 ‘피해자’를 각각 예측하는 네 가지 접근법이 있다.⁷⁾

- 범죄예측: 범죄의 위험이 증가하는 장소와 시간을 예측하는 접근법
- 범죄자 예측: 장래에 범죄 위험이 있는 개인을 식별하는 접근법
- 범죄자 신원 예측: 과거의 특정 범죄와 일치하는 용의자를 정확히 찾아내는 프로파일링 기법
- 범죄 피해자 예측: 범죄의 희생자가 될 가능성이 있는 집단 또는 개인 식별

본 연구에서 범죄예측은 범죄발생의 시간 및 장소를 예측하는 좁은 의미의 범죄예측을 의미하며, 표 2는 범죄 예측과 관련된 다양한 방법을 요약한 것이다.

범죄예측 방식을 전통적인 접근법과 최근의 접근법으로 구분했을 때, 표에서 볼 수 있듯이 기존의 접근법은 범죄 위치를 지도화하고 범죄가 집중되어 있는 우범지역(hotspot)을 인간의 판단에 의해 결정하는 것에서 시작되었다. 그러나 최근의 예측 분석방법은 가장 기본적인 수준에서 회귀 분석으로 시작하여 활발한 연구의 주제인 최첨단 수학 모델로 확장되었다.

7) Perry, W. L., (2013), “Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations.”, RAND Corporation, p.xiv.

<표 2> 법집행 기관의 범죄(시간, 장소)예측 기술

문제 : 위험이 높은 지역 식별	전통적 범죄 분석 (데이터양과 복잡성이 낮은 수준에서 중간 수준까지)	예측 분석 (대용량 데이터 및 높은 복잡성)
과거 범죄 데이터 사용	범죄 매핑 (핫스팟 식별)	고급 핫스팟 식별 모델; 위험 영역 분석
다양한 추가 데이터 (예 : 911 통화 기록, 경제 요인)	스프레드 시트 프로그램에서 생성 된 기본 회귀 모델	회귀, 분류 및 클러스터링 모델
최근 범죄로 인한 위험 증가에 대한 계산	최근 범죄 발생지 주변 지역의 위험 증가 가정	인근 반복 모델링
지역이 범죄의 위험에 가장 많이 노출 될 시기 결정	특정 지역의 범죄 빈도를 시간 / 날짜 (또는 특정 이벤트)별로 도표화 / 매핑	시공간 분석 방법
범죄의 위험을 증가시키는 지리적 특징 확인	범죄 발생 빈도가 가장 높은 지역 파악 및 추론 유도	위험 지형 분석

※ 출처: RAND Report(2013)

2. 범죄예측의 기본 원리와 기법

대부분의 예측 방법과 마찬가지로, 전술적 또는 전략적 관점에서 미래의 범죄 사건을 예측하는 것은 과거의 범죄 및 피해자에 대한 데이터를 분석하는 것에서 시작한다. 범죄예측을 위해서 과거를 연구하는 것은 법집행기관이 과거에 발생한 범죄에 더 큰 관심이 있기 때문은 아니다. 과거의 범죄발생 패턴을 연구함으로써 미래를 알 수 있기 때문이다.

범죄예측의 근본적인 가정은 과거가 현재를 알리는 단서가 된다는 것이다. 현재 또는 미래를 예측하기 위해서 과거를 연구하는 것은 범죄가 일정한 원칙 또는 원리에 의해서 발생한다는 것을 전제로 한다. 과거의 자료를 통해서 범죄 발생의 원리를 발견한다는 것은 범죄 유발·억제요인의 현황과 실태 파악을 통해서 범죄 예측을 할 수 있다는 것을 의미한다. 특히 최근(전술적 접근법의 경우 최근 며칠, 전략적 접근법의 경우 최근 몇 년 동안)에 발생하는 현상들은 현재의 범죄 발생을 예측하는데 더 중요한 의미를 갖는다.⁸⁾

주의할 것은 실무자가 “범죄의 원인은 무엇인가?”라는 질문에 답하기 위한 분석방법을 적용할 때 매우 신중해야 한다는 것이다. 특히 분석 결과에 대해 인과관계를 부여하지 않도록 주의할 필요가 있다. 특정 요인과 높은 범죄 위험 사이의 통계적 유의미성이 높다고 해서 반드시 그 요인이 범죄를 유발한다는 것을 의미하지는 않는다. 예를 들어 CCTV가 많이 설치되어 있는 지역에서 범죄가 많이 발생하는 경향이 있지만, 그것이 CCTV 설치의 결과는 아니다. 즉, CCTV 설치의 효과로 범죄가 많이 발생하는 것이 아니라 범죄가 많이 발생하는 지역에 범죄예방과 감시의 목적으로 CCTV를 설치하는 것이다.

범죄예측 기술은 네 가지 등급으로 나눌 수 있다.

- **고전적 기법:** 대부분의 회귀분석, 데이터 마이닝, 시계열 분석 및 계절적 조정과 같은 표준 통계 프로세스를 포함
- **단순한 기법:** 정교한 컴퓨팅이나 많은 양의 데이터를 요구하지 않는 대부분의 경험적 방법이 해당하며, 대용량 데이터보다 체크리스트와 색인 데이터를 주로 사용
- **복잡한 (응용) 기법:** 정교한 컴퓨팅 도구, 대용량 데이터가 필요한 새롭고 혁신적인 방법이나 기법(다수의 새로운 데이터 마이닝 기법과 일부 인공 반복 기법 포함)
- **맞춤형 기법:** 기존 기법들을 예측적 경찰활동 지원에 적합하도록 조정(ex. 고전 통계 방법을 사용하여 히트맵⁹⁾ 생성)

범죄예측을 위한 절대적인, 최고의 기법은 없다. 분석의 목적이 무엇인지, 분

8) 따라서 최근의 데이터에 대해서는 높은 가중치를 부여하는 방식으로 이러한 특성을 반영한다(Perry, W. L., (2013), “Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations.”, RAND Corporation, p.18.)

9) 히트맵(heat map)은 주어진 영역에서 범죄 활동의 강도를 묘사하는 단순한 색으로 구분된 그리드를 말한다.

석을 위해 사용할 수 있는 데이터가 어떤 것인지, 데이터 분석 프로그램과 컴퓨터 처리속도에 따라서 적합한 분석 기법이 달라진다. 현재 경찰에서 범죄예측을 위해서 활용하고 있는 다양한 기법들은 위의 기법들을 망라하고 있다. 분석이 필요한 경찰관이나 연구관 개인이 Excel이나 각종 통계프로그램을 이용하여 회귀분석, 시계열분석 등을 실시하기도 하고, GeoPros와 같은 시스템은 회귀분석 기반의 알고리즘에 기초하여 인공지능에 의해서 산출된다.

우리 경찰은 전국 범위의 예측시스템을 구축·운영하고 있으나 개별 단위(치안 블록, 기초경계구역)에 대한 예측결과의 정확성은 낮은 경향이 있다. 맞춤형 방법을 사용하여 더 많은 분류기준들을 수립하고 구체적인 타당성을 확보할 방법을 찾을 필요가 있다.

제2절 범죄예측 기법의 종류

범죄 위험을 예측하는 기법은 크게 핫스팟 분석, 회귀분석, 데이터 마이닝, 인근반복기법, 시공간 분석 및 위험영역 분석 등 크게 여섯 가지로 구분된다. 이 범주의 모든 기술이 복잡성 면에서 동등한 것은 아니다. 이 방법들은 과거 범죄 데이터 및 각종 관련 데이터에 기초하여 미래 범죄의 시간과 장소를 예측하기 위한 기법들이다.

1. 핫스팟(HotSpot)

핫스팟 기법은 범죄현상에 관한 정보를 지도상에 표현하는 기법으로, 범죄예측뿐만 아니라 범죄발생 현황을 확인하는데 가장 보편적으로 사용된다. 범죄가 균일하게 분포하지 않는다는 범죄의 편재성(偏在性) 가정을 기초로 범죄발생률이 가장 높은 지역을 파악하는 데 핫스팟 기법의 목적이 있다.

범죄지도의 제작은 19세기 프랑스에서 시작되어 1920년대 미국의 시카고

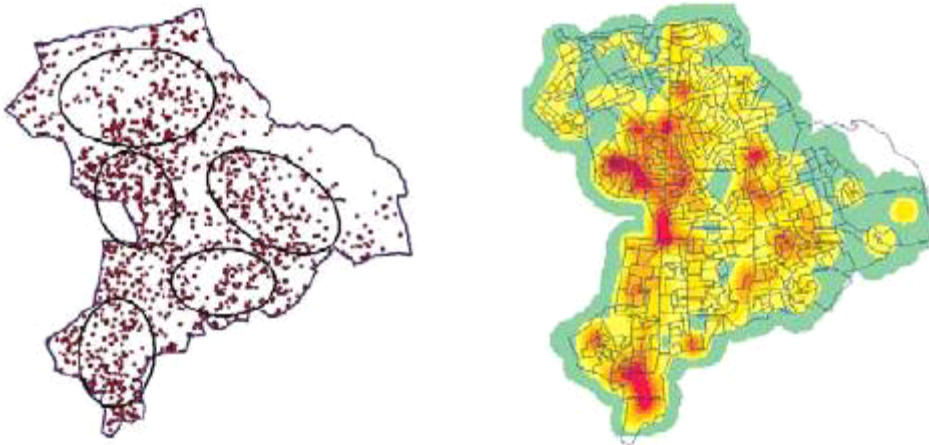
학파에서 본격적으로 사용하였다. 이후 일상활동이론, 환경범죄학 등을 기초로 하는 연구자들이 많이 사용하고 있으며, 범죄예방 분야에서도 주요 연구방법 중 하나로 활용되고 있다. 다양한 학문분야와 지리정보 기술의 결정체인 지리정보시스템(Geographic Information System: GIS)에서 핫스팟 기법은 중요한 위치를 차지하고 있다.

핫스팟 기법은 과거 범죄 데이터를 기반으로 범죄 위험이 증가한 영역을 예측한다. 기본적인 가정과 예측적 특징은 범죄가 이미 발생한 곳에서 범죄가 다시 발생할 가능성이 높다는 가정에 근거하고 있으며, 직관적으로 범죄의 발생 빈도를 확인하기 좋다는 장점 때문에 범죄지도 제작에서 빼놓기 힘든 기술이다.

그러나 핫스팟 기법에는 단점이 있다. 표현 단위영역과 표적 단위영역 간에 차이가 있을 경우 결과의 유용성이 반감된다. 우선 범위식별 단위 지역이 너무 작으면 결과에 관심 영역이 배제될 수 있다. 반대로 단위 지역이 너무 큰 경우에는 정확한 지점을 확인할 수 없어 치안 자원을 배치하는데 적합하지 않기 때문에 실행 가능성이 낮다.

핫스팟 기법에는 그리드 매핑 또는 주제 매핑과 같은 매우 간단한 방법도 있지만 수학적으로 엄격한 엘립스(ellipses) 기법과 커널 밀도 추정 방법도 있다. 이 두 가지 방법은 상대적으로 견고하고 지리적인 분할에 민감하지 않으며 통합성이 있어서 유리하다. CrimeStat 및 HotSpot Detective를 포함한 NIJ가 후원하는 분석도구도 이러한 방식에 포함된다. 모든 핫스팟 방법은 유사성이 높은 지역과 낮은 지역을 분류하는 클러스터링과 관련이 있다.

<그림 5> 엘립스 기법과 커널밀도 추정법(런던)



출처: Eck et al., Mapping Crime: Understanding Hot Spots, NIJ Special Report, 2005

핫스팟 기법에서 중요한 고려 사항은 정보, 즉 분석 결과로 무엇을 해야 할지를 명확히 하는 것이다. 핫스팟에서 범죄를 방지하기 위한 최적의 순찰 기간을 파악하기 위한 연구에서 Koper(1995)는 통계 분석을 통해 핫스팟에서 범죄가 발생할 확률이 다양한 기간의 경찰 순찰에 어떻게 영향을 받았는지 확인하기도 했다.¹⁰⁾

2. 데이터 마이닝(예측 분석)

데이터 마이닝(data mining)의 일반적인 정의는 대량의 데이터 가운데 숨겨져 있는 유용한 상관관계를 발견하여, 미래에 실행 가능한 정보를 추출해 내고 의사결정에 활용하는 과정을 말한다.¹¹⁾

전통적인 통계학에서도 데이터 마이닝과 동일한 목적으로 회귀분석, 판별분석, 주성분 분석 등을 사용했으나, 최근에 증가하고 있는 비정형 데이터,¹²⁾ 대

10) Koper, C. S. (1995). Just Enough Police Presence: Reducing Crime and Disorderly Behavior by Optimizing Patrol Time in Crime Hot Spots. *Justice Quarterly*, 12(4), 649-672.

11) 두산백과, [네이버 지식백과].

12) 특히 자연어에 해당하는 비정형데이터를 처리하여 정형화하고, 특징·정보를 추출하는 기술을 텍스트

용량의 데이터를 처리하기 위해서는 별도의 분석 기법이 요구되기 때문에 데이터 마이닝 이용이 대중화되었다.¹³⁾

데이터 마이닝은 다양한 통계적·수학적 기법, 인공지능 기반 패턴인식 기술 등을 이용하여 데이터 속에서 유의미한 관계, 유용한 패턴, 추세를 찾을 수 있다. 그래서 데이터 마이닝은 예측분석(predictive analysis)의 의미로 사용되기도 한다.

- **분류법(Classification):** 회귀 분석에서와 같이 연속되는 숫자(ex. 다음달에 평균 1.24건의 강도가 발생할 것으로 예측된다)가 아니라 결과의 범주를 예측한다(ex. 다음달에 강도가 발생할 가능성은 85%정도이다: 100% 이내에서의 범위).
- **클러스터링(Clustering):** 수치들을 수학적으로 유사한 그룹으로 세분한다. 이 모델은 미래의 상황이 이전 상황 클러스터와 유사 할 것이라고 예측한다.
- **특수한 회귀모델(Regression Model):** 회귀 모델 중에는 일반적인 것보다 입력 데이터와 출력 예측치 사이에 훨씬 더 복잡하거나 매우 다른 관계를 허용하는 방법이 있다. 회귀 모델은 입력 데이터와 출력 관계와 관련된 간단한 공식을 가정한다.
- **앙상블기법(Ensemble Method):** 여러 가지 간단한 예측 모델을 사용하여 투표 방식이나 출력의 가중 평균을 취하는 등의 방식으로 최종적으로 전체 예측을 산출한다.

데이터 마이닝을 위한 단일 핵심 기술은 없다. 해당 데이터에서 필요한 정보를 가장 잘 수행하는 모델링 방식을 선택하여 여러 가지 알고리즘과 매개 변수를 실험하는 것이 일반적이다. 전통적인 회귀 분석에서는 일반적으로 복잡한

트 마이닝(text mining)이라고 한다.

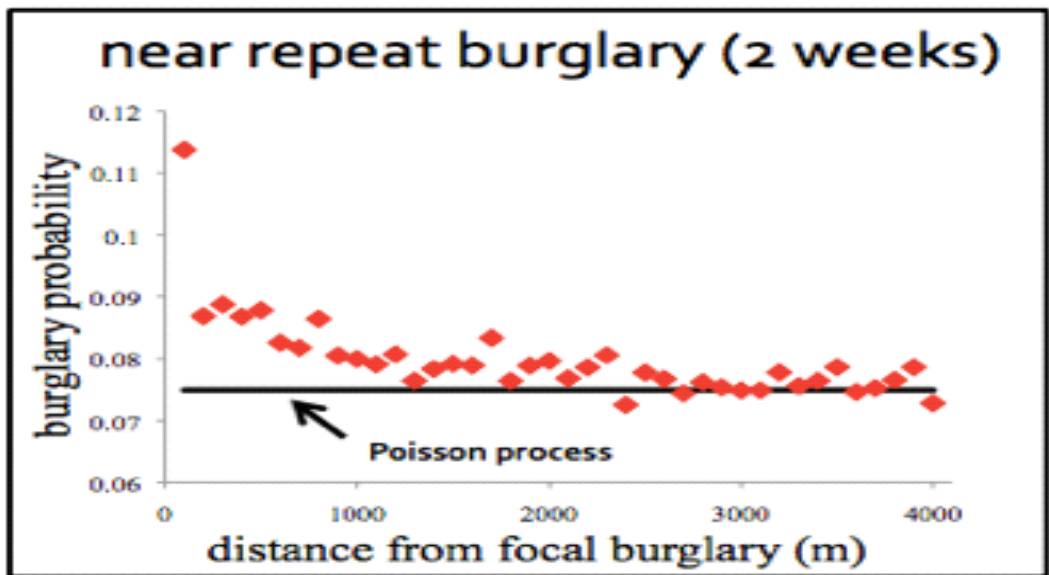
13) 최근 각광받고 있는 텍스트 마이닝도 일종의 데이터 마이닝 기술에 포함되는데, 텍스트 마이닝은 특히 자연어를 처리하여 정보를 추출하는 기술이다.

계산 문제를 회피하기 위해서 소수의 변인만 투입한다. 대안적으로 분석가들은 예상 모델을 수립하는 데 필요한 변수를 조합하여 새로운 변수를 생성한다. 예를 들어, 범죄 예측 모델을 구축할 때 이전 범죄의 여러 유형, 여러 유형의 장애 보고서, 여러 경제 및 인구 통계학적 변수, 날씨 및 데이터, 그리고 이들의 대푯값이나 결합에 의한 변수를 생성한다.

3. 인근반복 모형

인근반복 모형(Near Repeat Model)은 미래의 범죄가 현재의 범죄에 시공간적으로 매우 근접해서 발생한다는 가정을 전제로 한다. 즉, 최근에 범죄 수준이 높을수록 근접한 시간, 근접한 장소에서 범죄가 발생할 가능성이 높다는 것이다. 예를 들어 강도사건의 경우 범죄자들에게 지역 취약점이 잘 알려져 있기 때문에 영역 내의 인접 대상을 반복적으로 공략한다.¹⁴⁾

<그림 6> 거리에 따른 강도 발생 빈도



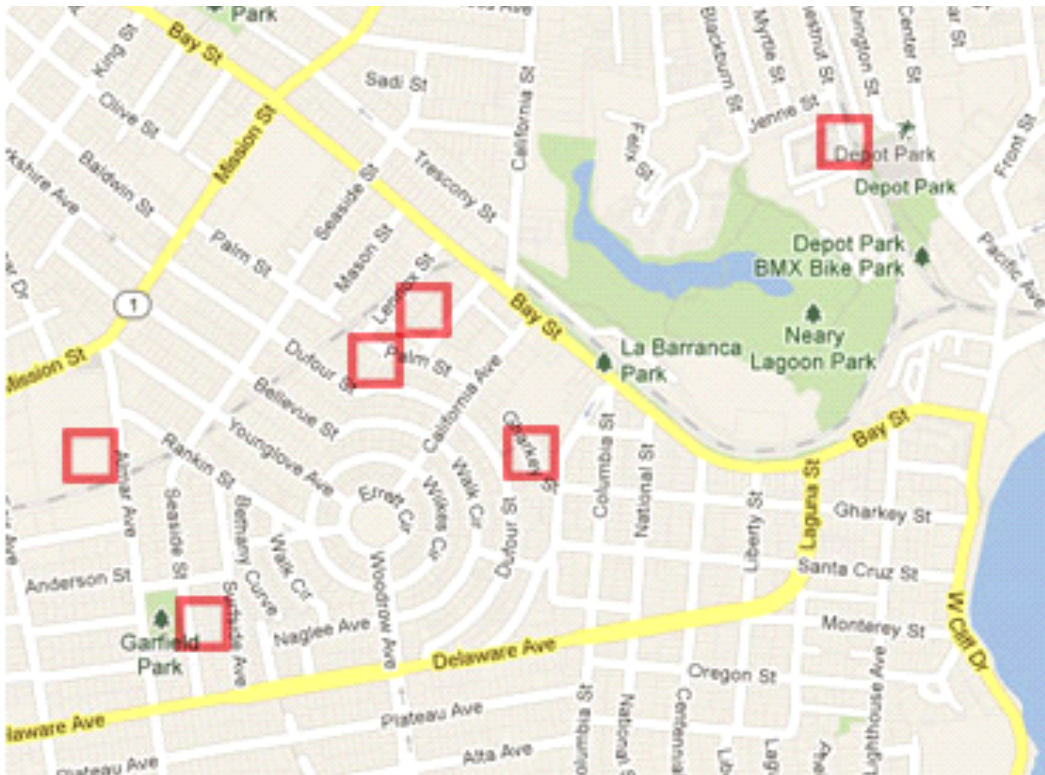
출처: PREDPOL 홈페이지(<https://www.predpol.com/predictive-policing-definition/>)

14) Mohler, G. O., Short, M. B., Brantingham, P. J., Schoenberg, F. P., & Tita, G. E. (2011). Self-Exciting Point Process Modeling of Crime. *Journal of the American Statistical Association*, 106(493), 100-108.

이들의 주장에 따르면 범죄는 전염병과 같이 지역 환경(미시 및 미세 장소)을 통해 확산된다고 한다. 범죄가 한 건 발생했다면 가까운 거리에서 범죄발생 위험이 증가한다는 것이다. 이 효과를 이용하기 위해 몰러 등은 자기자극과정(self-exciting process) 또는 지진 모델링(earthquake modeling) 알고리즘을 개발했다.¹⁵⁾

LA경찰서에서 사용하고 있는 PredPol은 인근반복모형을 사용하는 대표적인 범죄예측 시스템이다. LA지역의 경우에 총기사건 등 강력 범죄가 발생하는 지역과 아닌 지역의 구분이 비교적 뚜렷하고 다른 환경요인에 비해서 범죄발생 자체가 다음 범죄의 발생을 예측하는데 결정적인 요인으로서의 역할을 하는 특수성이 있다.

<그림 7> 인근반복모형을 사용한 범죄예측 예시



출처: PREDPOL 홈페이지(<https://www.predpol.com/predictive-policing-definition/>)

15) Mohler et al., 2011.

4. 위험영역 모형

위험영역 모형(Risk Terrain Model)은 첫째, 범죄 위험과 관련이 있는 지리적 특징(ex. 술집, 주류 판매점 등)을 식별하고, 둘째, 주어진 위험에 얼마나 근접하는지에 따라 범죄 위험을 예측하는 기술을 의미한다. 특정 위치는 위험을 유발하는 속성을 내포하고 있다고 가정한다.

경찰관의 관점에서 볼 때, 위험 영역 모델의 결과물은 핫스팟 방식의 결과와 질적으로 동일하다. 둘 다 가까운 장래에 높은 범죄를 당할 가능성이 있는 영역을 강조하여 표시하는 방식이기 때문이다. 그러나 핫스팟 기법은 근본적으로 범죄 자체가 발생한 지역을 표시하는 클러스터링 기법인 반면, 위험영역 모델링은 지역 특성을 기반으로 범죄 위험을 특성화하는 분류 접근법이다.

최근 국내에서는 허선영, 김주영, 문태현 등¹⁶⁾이 위험영역 모델을 적용하여 GIS 프로그램을 이용한 공간회귀분석에 의해서 물리적인 환경특성을 도출한 다음, 이 변수를 이용하여 베이저안 확률기반 범죄발생 위험성 예측모형(CIPM: Crime Incident Prediction Model)을 개발하였다. 이 연구에서는 사례지역의 경우 상업 및 유흥시설 밀집되거나 건물층수가 높은 지역, 상업 및 유흥시설이 혼재된 지역 등에서 범죄발생 확률이 높은 것으로 나타났다.

16) 허선영, 김주영, 문태현 (2017). 베이저안 확률 기반 범죄위험영역 예측 모델 개발, 2017, 89-101.

<그림 8> CIPM을 사용한 범죄예측 모형 예시



위험영역 모형과 유사하지만 다소 차이가 있는 중첩분석방법이 국내 지리통계학 분야 연구에서 사용되고 있다. 중첩분석은 2개 이상의 지도에서 공간 정보를 통합, 최적분석과정을 거쳐 결합된 속성에 대해 두 계층 이상의 관계를 분석하는 방법이다. 강선호 등은 범죄위험도 예측에 있어서도 중첩분석을 사용

하여 범죄예측이 가능한 것을 확인하였다.¹⁷⁾

노성훈은 상이한 범죄유형에 대해서 인근반복모형과 위험영역모형의 예측력을 비교하는 연구를 수행하였는데, 상대적으로 가중폭행의 경우 인근반복모형이, 성범죄의 경우 위험영역모형이 더 예측모형이 더 예측력이 높은 것으로 확인되기도 했다.¹⁸⁾

5. 시공간 분석

범죄 유형, 위치 및 시간, 또는 범죄의 현재 패턴은 일정한 관계가 있는 것으로 가정된다. 그러나 범죄 패턴이 시간이 지남에 따라 변할 수 있다는 것은 누구나 동의할 수 있을 것이다. 따라서 범죄 현장의 다양한 환경 및 시간적 특성을 포함한 예측 문제로 확장할 필요가 있다. 시간 정보는 범죄 사건 정보와 함께 향후 범죄의 위치와 시간을 예측하는 데 사용될 수 있다.

시공간 분석방법(spatiotemporal analysis)은 연쇄 범죄자의 석방 날짜 및 개인 정보를 수학적으로 엄격한 방법으로 결합하기 위해 개발되었다. 시공간적 분석에 일반적으로 사용되는 예측요인에는 다음과 같은 것들이 포함된다.

- 시간대, 요일, 시간 및 요일의 주기
- 이벤트(ex. 월급 날, 스포츠 이벤트, 콘서트) 및 일시적인 근접성
- 계절
- 날씨
- 연쇄 범죄의 범죄 간 간격
- 반복되는 위치
- 연쇄범죄에서 사건의 지리적 이동
- 사건의 공간적 배열
- 위치 유형(ex. 공원, 편의점, 공공주택)
- 지리적 관계(ex. 버스 정류장 근처, 주류를 판매하는 식당 근처)
- 환경 및 목표 요소 (ex. 가로등, 인접 조건, 교통량)
- 범죄 지역의 인구 통계 및 경제 데이터

※ 출처: RAND Report(2013)

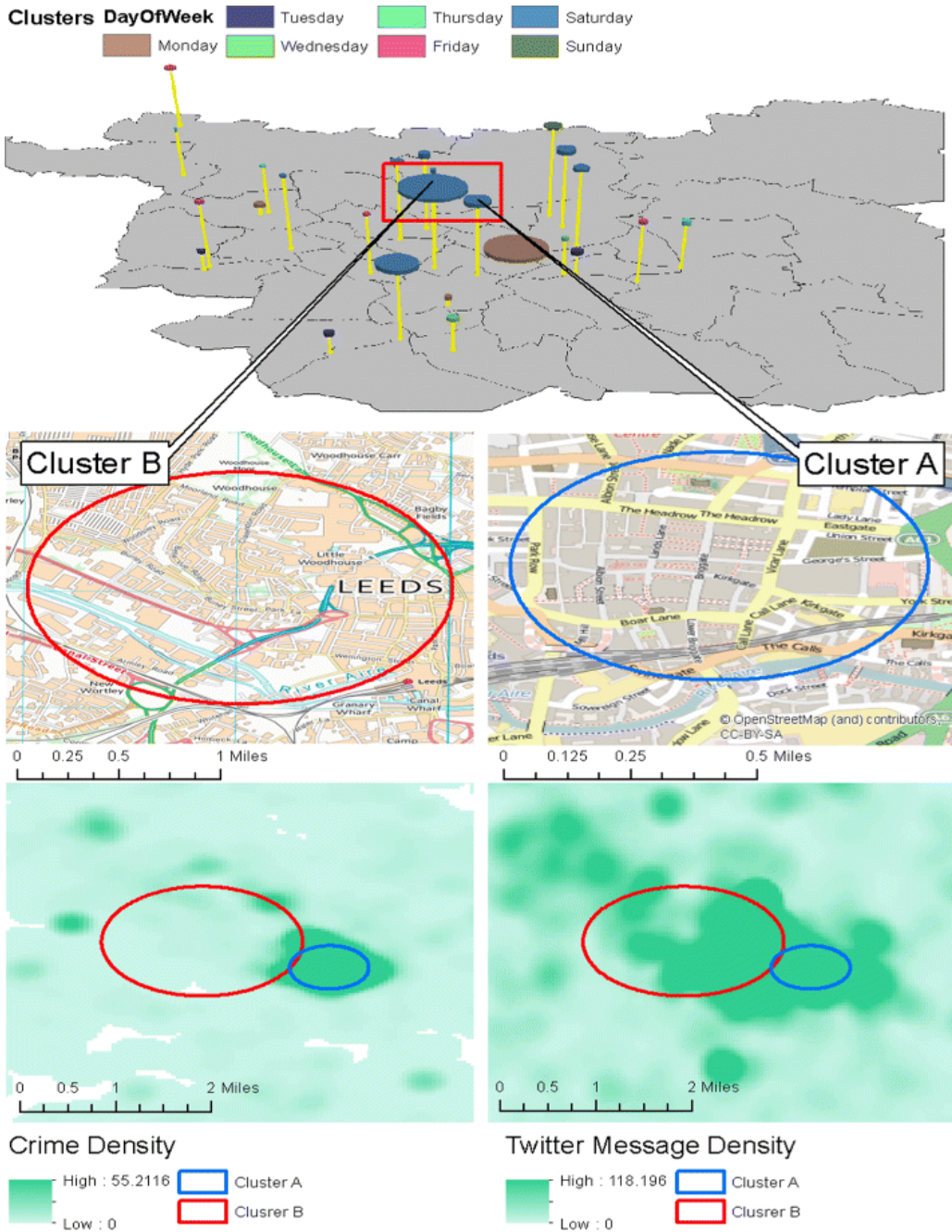
17) 강선호, 양진경, 한별, 이동희 (2017). 국가기초구역을 이용한 범죄예측모델: 안양시 사례를 중심으로. 한국경영과학회지, 42(4), 135-146.

18) 노성훈 (2015). 시공간 분석과 위험영역모델링을 활용한 범죄예측모형의 예측력 검증. 형사정책연구, 26(3), 239-266.

이러한 모든 기능(단독 또는 조합)은 단기 및 장기 문제 또는 핫스팟 모두를 분석할 때 예측적 가치가 있다. 특히 공간적인 요인들이 시간적인 요인에 따라서 범죄발생에 다르게 영향을 미치기 때문에 시공간적 요인을 동시에 고려하는 것은 중요한 의미를 가진다. 즉, 같은 물리적 환경을 가진 장소라고 하더라도 계절, 요일, 날씨 등에 따라서 범죄발생량은 달라진다. 공간적인 환경뿐만 아니라 시간적 요인을 고려하는 것은 짧은 주기를 가지는 순찰활동 등 범죄예방활동 필요성의 증가를 고려했을 때 매우 중요한 부분이다. <그림 9>는 특정 도시에서 요일이나 권역별로 범죄와 SNS 활동 분포의 차이가 있음을 보여주고 있다. A권역의 경우 SNS 활동과 범죄발생 모두 높게 나타나고 있지만 B권역의 경우 SNS 활동은 많지만 범죄발생은 낮은 것으로 나타났다. 이는 A권역과 B권역 모두 (유동)인구가 집중되는 지역이지만 B권역은 비교적 안전하다는 것을 보여준다.

시간과 공간에 대한 고려 사항은 범죄 현장(일반적으로 경찰 데이터 시스템에서 추출한 것)뿐만 아니라 범죄와 관련된 다른 장소에도 적용된다. 예를 들어, 성폭력이나 강도 사건은 범죄자가 실제 강도 또는 폭행이 발생한 장소와 무관한 피해자를 대상으로 한 범행 위치와 시간을 포함할 수 있다. 한 지역에서 도난당한 차량은 다른 지역에서 범행에 이용되고, 또 다른 지역에 버려질 수 있다. 이러한 사건에 대해 항목별로 단일 정보를 요구하는 일반적인 분석 기법에서는 각 요인의 특성이 특정되지 않고(상이한 3개의 시간과 장소) 충돌하여 정보 입력 및 분석에 한계가 발생한다. 그러나 시공간 분석을 활용할 경우 시공간적인 이동이 있는 범죄행위에 대해서도 유용한 정보와 의미를 도출할 수 있다.

<그림 9> 시공간 클러스터링의 예시



출처: Malleon, N., & Andresen, M. A. (2015). Spatio-temporal crime hotspots and the ambient population.

제3절 회귀분석 기법의 검토

지금까지 6가지 범죄예측 기법 중 5가지에 대해서 알아보았고, 나머지 한 가지인 회귀분석 기법은 범죄예측 기법에서 매우 중요한 위치를 차지하고 있다. 현재 국내의 대표적인 범죄예측시스템인 GeoPros에서도 주로 회귀분석 방법을 사용하여 범죄위험도를 계산하고 있다. 회귀분석은 다른 기법에 비해서 장점들을 가지고 있어서 본 연구에서도 별도로 다루게 되었다. 3절에서는 회귀분석을 중심으로 그 기법을 소개하고자 한다.

1. 회귀분석 기법

회귀 분석은 예측결과와 독립적인 예측(설명) 요인간의 수학적 관계에 적용하기에 적합하다. 핫스팟 매핑과는 달리 회귀 분석은 과거의 범죄뿐만 아니라 광범위한 범죄 관련 요인에 기초하여 미래의 범죄 위험을 예측한다.

회귀분석에서의 데이터 범위는 예를 들어, 미래의 강도에 대한 회귀 모델에는 범죄 유형에 대한 입력뿐만 아니라 다른 범죄 유형에 대한 계수, 기물 파손 및 기타 장애 유형에 대한 계수, 해당 지역의 주택수, 공가(空家) 수 등이 적용될 수 있다. 회귀는 상당 기간 범죄와 관련된 문제에 적용되었다. 그러나 데이터가 적은 경우에는 불안정할 수 있기 때문에 정교한 기술이 요구되기도 한다.

회귀분석 방법은 크게 종속변인에 해당하는 예측값과 독립적인 "설명"변수 사이의 세 가지 관계로 분류할 수 있다.¹⁹⁾

- **선형 회귀:** 선형 회귀는 관측된 데이터에 선형 방정식을 적용하여 두 개

19) Perry, W. L., (2013), Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations., RAND Corporation.

이상의 변수 간 기본 관계를 모델링한다. 기하학적으로 이 방법은 출력 변수와 입력 변수 사이의 관계에 선이나 평면을 맞추기 때문에 선형 모델이라고 한다. 선형 회귀 모형에서는 범죄에 대한 신뢰 구간을 추정 할 수 있다(ex. 2건에서 14건 사이에서 90 % 확률로 8건의 강도가 예상됨).

- **비선형 회귀**: 이 방법은 가중 평균보다 입력 변수와 회귀 변수 사이에 더 복잡한 수식을 허용한다. 실제로 비선형 관계를 맞추려면 더 복잡하고 시간이 많이 소요되는 통계 알고리즘이 필요하다.
- **회귀 스플라인(spline)**: 회귀 스플라인을 사용하면 종속 변수의 여러 영역에서 데이터를 모델링하는 여러 회귀 방법을 사용할 수 있다. 예를 들어, 특정 해에 한 지역에서 자동차 도난의 확률이 0.01% ~ 45%이고, 0.01%에서 5% 사이의 가장 좋은 설명 변수는 모델 유형, 차량 위치, 피의자 연령 등이지만 5%에서 45%까지의 도용에 대한 가장 유용한 변수는 연령과 상황인 경우가 있다. 회귀 스플라인을 사용하면 종속 변수의 전체 범위에 대해 구간별로 회귀 분석식을 만들 수 있다.²⁰⁾

회귀 모델에 포함할 입력 변수를 선택하는 것은 쉽지 않은 일이다. 많은 수의 후보 입력 변수가 있는 경우 특히 그렇다. 가능한 모든 변수를 하나의 모델에 넣을 수는 있지만 결과적으로 변수가 과잉된 모델이 도출될 가능성이 높다. 결과 공식은 입력 변수와 출력 변수 사이의 실제 관계라기보다는 입력 데이터의 일정한 노이즈를 반영한다.

선행 지표(leading indicator)는 앞으로의 상황에 대한 힌트를 제공한다. 지평선의 천둥을 동반한 구름이 비가 몰려오고 있음을 암시하는 것과 같다. 변수는 가까운 미래에 범죄가 나타날 방향을 알려줄 수 있다. 이것은 더 심각한 범죄, 범죄의 위치를 이동시키는 지리적 변화(ex. 쇼핑몰, 스포츠 경기장 또는 대형건물 개장) 또는 범죄 활동의 위치 또는 심각성이 변화하는 사소한 범죄일 수 있다. 선행 지표로 분석 가능 대응책에서 사전 대책으로 이동하기 위한 정

20) Nyce, C. (2007). *Predictive Analytics White Paper*, Malvern, Pa.: American Institute for Chartered Property Casualty Underwriters/Insurance Institute of America.

책을 수립해야하지만 예측 모델이 관련성을 유지하도록 주의를 기울여야한다. 범죄의 유형이나 목표가 타당성을 잃게 되면 시스템은 변화에 맞게 수정될 필요가 있으며 새로운 주요 지표가 필요하다.

예를 들어, 특정 유형의 범죄가 증가하는 것을 예측할 수 있는 요인으로 날씨 요인이 알려져 있다. Bushman 등에 따르면 계절과 시간대와 같은 요인을 통제 할 때조차 높은 기온은 높은 범죄와 관련이 있음을 보여주었다.²¹⁾ 따라서 일기 예보는 매우 단기적인 관점에서 범죄 발생 장소를 예측할 수 있는 통계 모델을 제안해준다.

Gorr과 Olligschlaeger는 피츠버그의 데이터를 사용하여 일련의 범죄 범주를 예측하는 여러 회귀 접근법을 비교했다.²²⁾ 동일한 데이터 세트에서 서로 다른 복잡성의 회귀 분석을 수행하고 결과를 비교했다. 그들은 간단한 시계열이 더 복잡한 방법을 적용하는 경우보다 더 성능이 뛰어나다는 것을 발견했다. 특히 기본 데이터의 불안정성으로 인해 평활 계수(최근 데이터에 더 많은 가중치를 두는)를 사용하면 예측의 정확성이 향상되었다.

국내에서는 범죄예측의 선행지표로 1972년부터 2016년까지의 경제지표(국내총생산, 국내총소득, 국민총소득, 최종소비지출, 노동소득분배율, 총저축율, 가계순저축률, GDP디플레이터)의 효과를 확인한 연구가 수행되었다.²³⁾ 연구 결과에 따르면, 범죄 유형에 따라서 경제지표 별 선행성은 차이가 있었지만, 범죄학 분야에서 오랜 관심사인 경제과 범죄와의 관계를 경험적으로 확인했다는 점에서 의미가 있다. 다만, 범죄예측 연구는 국내에서의 지역별 범죄위험도를 예측하는데 관심이 있는 만큼 지역적 범위와 기간을 예측의 목적에 맞게 수집하여 활용할 필요가 있다.

21) Bushman, B. J., Wang, M. C., & Anderson, C. A. (2005). Is the Curve Relating Temperature to Aggression Linear or Curvilinear? Assaults and Temperature in Minneapolis Reexamined. *Journal of Personality and Social Psychology*, 89(1), 62-66.

22) Gorr, W., & Olligschlaeger, A. (2002). *Crime Hot Spot Forecasting: Modeling and Comparative Evaluation*, prepared for the U.S. Department of Justice, <https://www.ncjrs.gov/App/abstractDBDetails.aspx?id=195167>

23) 주일엽 (2018). 경제지표의 범죄발생에 대한 선행관계. *경찰학논총*, 13(3), 9-37.

2. 범죄예측 기법의 비교

범죄예측 기법들은 각각의 특성들을 가지고 있으며, 기법의 사용을 위한 조건이 다르고, 각 기법의 장단점도 다르다. 지금까지 소개된 총 6개의 범죄예측 기법들을 서로 비교하여 가장 적합한 기법을 선택할 필요가 있다.

- 핫스팟 분석, 회귀분석, 데이터 마이닝 및 인근반복 기법은 일반적으로 특정 시간대에 범죄가 발생하는 위치를 식별하는데 사용된다(방법 및 응용 프로그램에 따라 하루에서 1년까지 다양).
- 시간적 및 시공간적 방법을 사용하여 범죄 발생 가능성을 파악할 수 있는데, 이 방법은 주변인뿐만 아니라 지역특성에 대한 분석을 통해 피해자를 식별하기도 한다.
- 위험 영역 분석은 범죄위험을 야기하는 지형 공간적 요인을 식별하고 특정 유형의 범죄에 관련된 물리적 위치를 찾는 데 적합하다.

GeoPros 상에서 기존의 범죄환경 위험도는 범죄위험도지수와 범죄위험지수로 구분되어 있는데, 각각의 예측 방식은 똑같이 다른 회귀분석 방법에 의한 범죄예측 방식을 사용하였다.

기법별 장단점 비교 결과에 비추어 보더라도 데이터와 분석 프로그램의 가용성, 해석의 용이성 등을 고려했을 때 특정 시간·장소에서의 범죄발생 가능성을 예측하기 위해서는 회귀분석 방법을 통해서 범죄위험도 지수를 산출하는 것이 적합한 것으로 보인다. 다만, 회귀분석이 가지고 있는 한계를 고려한다면 회귀분석의 결과를 보완할 수 있는 다양한 방법을 적용할 필요는 있다.

현재 범죄예측에 현실적으로 적용 가능한 데이터는 각종 범죄발생 데이터, 공공기관의 행정데이터(전국), 지역 행정데이터(지자체), 기상데이터 등이다. 특히 우리나라의 경우 마약, 범죄집단 등 결정적인 요인이 다른 범죄발생에 영향을 미치는 경향이 뚜렷하지 않기 때문에 특정 범죄발생 자체로 다른 범죄를 예

측하기보다 환경적인 요인을 통해서 범죄를 예측하는 것이 더 적합한 것으로 보인다.²⁴⁾

범죄발생의 기본 전제인 잠재적 가해자와 피해자의 존재를 고려한다면 유동 인구 등의 요인을 고려하여 회귀분석 기법을 적용하는 것이 가장 적합할 것으로 보인다. 특히 단순한 선형회귀가 아니라 범죄결과에 미치는 위험요인들의 비선형적인 효과를 반영할 수 있는 방법을 추가로 고려할 필요가 있다.

24) 윤상연, 신상화 (2018). 범죄예측의 현재와 미래: 한국형 범죄예측 시스템 개선·도입에 대한 제언. 한국공안행정학회, 27(3), 11-38.

제3장 성범죄 발생 특성

범죄의 예측은 내용적으로는 범죄의 발생 기제 즉, 요인과 경로를 분석하는 과정이 필요하다. 범죄예측을 위해서는 발생 요인을 추출하고, 그 요인의 분포 수준을 파악해야 하는 것이다. 3장에서는 본 연구의 대상인 성범죄와 관련하여 문헌연구, 범죄통계 분석, 현장조사 결과를 통해서 발생특성과 예측 요인을 살펴보고자 한다.

제1절 성범죄의 심리적 원인

1. 성범죄 원인에 관한 심리적 이론

사회학습이론(Social Learning Theory)에서는 인간의 행위는 쾌락을 추구하고 고통을 피하도록 조직되어 있으며, 범죄를 포함한 인간의 모든 행위와 그 가치는 학습된다고 본다. 이 때문에 성범죄자들이 그들이 처한 사회적 환경 안에서 성범죄에 대한 관대한 신조를 배운다는 점을 강조한다. 성범죄자는 강간에 대한 전문지식을 상대적으로 더 효과적으로 배우고, 성범죄와 폭력성에 대해 우호적인 특징을 보인다.²⁵⁾

인지왜곡이론에 따르면 성범죄를 지향하는 태도나 행동은 다양한 대인관계 속에서 학습되고, 강압적이고 강제적인 남성-여성 관계를 나타내는 성범죄에 대한 신조는 사회화(학교, 가정, 직장 등)되고, 다양한 집단 내에서의 유대감, 미디어매체를 통해서 강화된다.²⁶⁾ 보편화된 ‘강간 신화’가 인지왜곡이론의 대표적인 사례이다.

25) Howitt, D. (2002). Forensic and criminal psychology. Harlow: Pearson Education.

26) Russell, D. E. H. (1984). Sexual exploitation-rape, child sexual abuse, and workplace harassment.

사회문화이론에 따르면 성범죄 발생률을 결정하는 사회문화적인 요인에는 크게 세 가지가 있는데, 그 중 ‘남녀불평등(경제적, 법적, 정치적 지위, 평균 수입 비교)’과 ‘사회적 혼란 요인(지리적 이동, 이혼·편부모 가정, 종교귀속 감소 등)’이 ‘문화적인 영향(학교 체벌, 사형 등 합법적인 폭력의 수용도)’에 비해 더 영향이 큰 것으로 알려져 있다.

페미니스트 이론은 성범죄가 사회적인 성구조 안에서 만들어진 것이라고 주장한다. 강간은 권력과 사회적 지위에서의 남녀불평등과 연관성이 있으며, 강간은 성적인 욕망보다 오히려 권력과 지배욕에 대한 욕망에 의해 동기화된다고 한다.

2. 성범죄자의 심리적 특성

가. 발달적 특성

성범죄자는 발달과정에서 정상적인 사람들과는 다른 문제들을 경험하는 것으로 알려져 있다. 성범죄자는 성장과정에서 성적 학대나 다른 폭력피해 경험이 있는 경향이 있고, 가족관계 문제가 발견된다.²⁷⁾ 부모와의 관계가 제대로 형성되어 있지 않아 타인과의 관계형성 어려움 등 초기 대인관계 장애를 보이는 것이 특징이다. 또한 성범죄자는 어린 시절 성폭력 피해에 노출,²⁸⁾ 신체적 학대 경험과 가정폭력 노출 경험²⁹⁾이 있는 것으로 알려져 있다.

나. 성격적 특성

성범죄자는 고립감, 사회성 및 정서성이 결여되어 있는 경향이 있다.³⁰⁾ 범행 당시 혼자 살고 있던 비율, 여성과의 친밀한 관계 경험의 결여, 정서적 상호작

27) Hall, G. K., & Barongan, C. (1997). Prevention of sexual aggression: sociocultural risk and protective factors. *American Psychologist*, 52(1), 5-14.

28) Becker, J. V., & Hunter, Jr. J. A. (1997). Understanding and treating child and adolescent sexual offenders. *Advances in Clinical Child Psychology*, 19, 177-197.

29) Ford, M. E., & Linney, J. A. (1995). Comparative analysis of juvenile sexual offenders, violent nonsexual offenders, and status offenders. *Journal of Interpersonal Violence*, 10(1), 56-70.

30) Dietz, P. E., Hazelwood, M. S., Warren, D. S. W. (1990). The sexually sadistic criminal and his offences. *Bulletin of American Academy of Psychiatry and the Law*. 16, 163-178.

용의 제한 등의 특성을 보인다. 다만, 이러한 결과는 성적 살인범과 단순 강간범의 비교 결과이기 때문에 제한적으로 해석할 필요가 있다.

성범죄자와 같이 정서공감력³¹⁾이 부족한 경우 타인의 고통이나 아픔을 이해하지 못하여 피해자의 고통상황에 대해 정상적으로 반응하지 못하는 등 공감능력 상의 문제를 동반한다. 성범죄자들은 유년시절 가정불안, 해체, 폭력이 일반적으로 나타나는 등 부적절한 사회적 기술, 취약한 또래관계, 사회적 고립 등 사회적 능력이 결여되어 있는 경향이 있다.³²⁾ 또한 성범죄자는 청소년기 행동장애 진단 및 반사회적 특질,³³⁾ 충동조절 문제와 생활충동성³⁴⁾ 등 정신적인 문제를 동반하기도 한다.

다. 인지적 특성

성범죄자는 성에 대한 왜곡된 인식, 올바른 의사소통 능력 부족 등으로 인해 자신의 행동, 타인의 반응, 상황 등을 적절하게 인식하지 못하는 경향이 있다.³⁵⁾ 행위의 정당화, 행위의 부정 및 최소화 등의 왜곡된 인지과정은 성범죄자가 일탈적인 범죄행위를 지속하는 핵심 원인으로 지목되고 있다.

예를 들어 한 청소년 성범죄자 대상 연구에서 참가자 3분의 1은 성관계를 타인에 대한 마음이나 사랑을 증명하기 위한 행위로 인식했고, 나머지는 권력과 지배력을 느끼기 위한 행위(23.5%), 분노를 해소하기 위한 행위(9.4%), 상처 입히고 비하하거나 벌주는 행위(8.4%)로 인식하였고 있다는 것이 확인되기도 했다.³⁶⁾

31) 정서공감력은 자기개념화와 감정에 대한 이해 및 타인에 대한 반응능력 등 인지적인 능력과 감정적인 능력으로 구성된다.

32) Bagley, C., & Shewchuk-Dann, D. (1991). Characteristics of 60 children and adolescents who have a history of sexual assault against others: evidence from controlled study. *Journal of Child and Youth Care, Special Issue*, 43-52.

33) Kavoussi, R. J., Kpalan, M., & Becker, J. V. (1988). Psychiatry diagnoses in adolescent sex offenders. *Journal of the American Academy of Child and Adolescent Psychiatry*, 27(2), 241-243.

34) Prentky, R. A., Harris, B., Frizzell, K., & Righthand, S. (2000). An actuarial procedure for assessing risk with juvenile sex offenders. *Sexual Abuse: a Journal of Research and Treatment*, 12(2), 71-93.

35) Carich, M. S., Newbauer, J. F., & Stone, M. H. (2001). Sexual offenders and contemporary treatment. *Journal of Individual Psychology*, 57(1), 3-17.

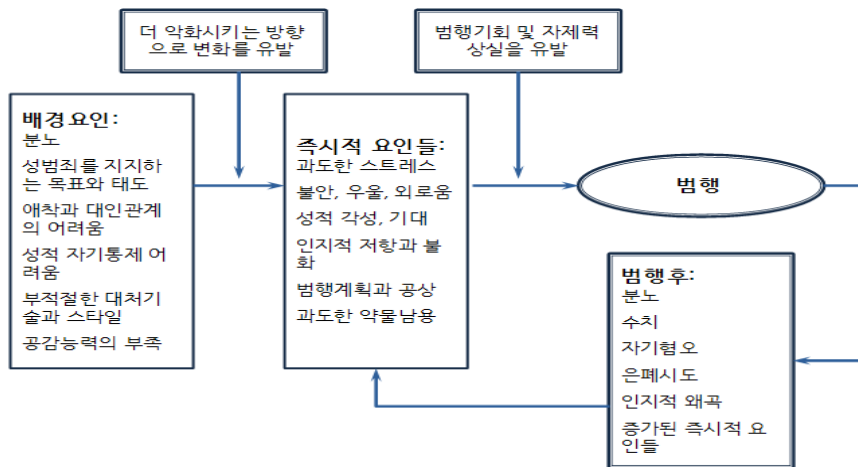
36) Ryan, G., Miyoshi, T. J., Metzner, J. L., Krugman, R. D., & Fryer, G. E. (1996). Trends in

가학적 성범죄자들은 범행 시 성적 환상상태에 놓여 있고, 여러 차례 성적 살인을 한 범죄자들의 경우 성적환상과의 관련성이 더욱 높았다.³⁷⁾ 특히 성적 환상은 관음증과 물품도착처럼 상상 속의 행위를 실현하는 범죄와 조직화된 범 죄현장을 남기는 습관을 가진 범죄자에게서 나타났다.

3. 심리적 요인의 적용 가능성

성범죄 발생의 심리적 원인들을 검토한 결과 성범죄가 발생하는데 있어서는 다양한 사회·문화적인 특성, 개인적 특성이 작용하고 있음을 확인할 수 있었다. 특히 개인적인 심리 특성에는 발달, 성격, 인지적인 특성들이 개입하고 있는 것으로 보인다. 윤정숙 등(2012)은 일반적인 이론적 논의보다 더 종합적으로 성범죄의 발생 경로를 구체적인 요인들을 적시하여 제시하기도 한다. 이에 따르면, 성범죄는 배경요인과 즉시적 요인들, 그리고 억제 요인의 부재와 같은 조건 하에서 발생한다고 제안하고 있다.

<그림 10> 성범죄의 범행경로



출처: 성범죄자를 위한 치료프로그램 개발 및 제도화방안(I). 윤정숙 등, 2012. p.247

a national sample of sexually abusive youths. Journal of the American Academy of Child & Adolescent Psychiatry, 35(1), 17-25.

37) Grubin, D. (1999). Actuarial and clinical assessment of risk in sex offenders. Journal of Interpersonal Violence, 14(3), 331-344.

개인적인 심리특성들은 성범죄의 발생을 직접적이고 심층적으로 이해하는데 도움을 주지만, 이러한 결과는 예측의 관점에서 활용하기는 쉽지 않을 것으로 보인다. 범죄예측은 개인 단위가 아닌 일정 범위의 지역 단위의 예측을 하고 있는데, 지역 수준에서 개인의 심리적 상태를 측정하고 수치화하는 것은 어렵다. 그리고 현실적으로 이러한 자료가 없는 실정이다.

제2절 성범죄의 발생 특성

1. 통계분석 개요

범죄예측은 범죄발생에 대한 원리 이해를 전제로 한다. 범죄발생에 있어서의 특성과 양상을 이해하는 것이 범죄예측이라는 결과를 정확하게 도출하는 것뿐만 아니라 의미 있는 범죄예측 요인을 도출할 수 있는 중요한 과정이다.

본 연구의 대상인 성범죄의 발생 및 피해 특성을 파악하여 범죄예측을 위한 변인을 도출하기 위해 성범죄에 관한 기초적인 자료를 분석하였다. 성범죄에 대한 기존의 통계적 결과들을 확인함으로써 성범죄가 가지고 있는 기본적인 특성들을 확인하고자 하였다.

성범죄의 발생 특성에 대한 연구들은 성범죄 발생의 시간적인 요인, 공간적인 요인들에 대해서 정보를 제공한다. 계절, 월별, 요일, 시간대별 성범죄 발생 특성, 그리고 성범죄 발생 장소 등에 대한 정보를 통해 성범죄가 발생하는 다양한 조건 또는 환경 등에 대한 통찰이 가능하다.

이를 위해 공식적인 범죄통계 자료인 ‘범죄통계’의 최근 3년간(2014~2016) 자료를 분석하여 성범죄의 발생추세 및 유형별 특징을 분석하였다. 성범죄 유형으로는 성폭력 범죄에 해당하는 강간, 강제추행³⁸⁾ 이외에 성풍속범죄에 해당하는 불법촬영(성폭력범죄의처벌등에관한특례법(카메라등이용촬영))을 포함하였다.

38) 범죄통계의 강간·강제추행 죄에서는 강간, 강제추행 이외에도 유사강간(2012.12 신설), 기타 강간·강제추행을 구분하고 있으나, 본 분석에서는 강간과 강제추행만 포함하였다.

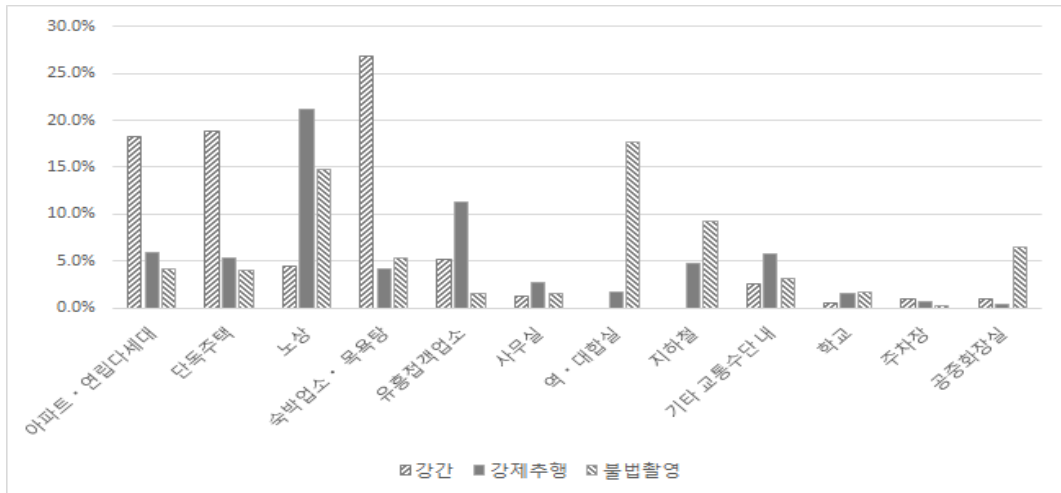
2. 범죄통계 상 성범죄의 특성

가. 성범죄 발생의 공간적 특성

성범죄 유형에 따라서 발생장소는 큰 차이를 보였다. 강간은 폐쇄된 공간, 강제추행과 불법촬영은 개방된 공간에서 많이 발생하며, 특히 불법촬영은 노상 또는 역·대합실, 지하철 등 공공장소발생률이 높았다.

범죄유형별로는 차이가 있지만 강간과 강제추행의 발생건수 차이가 워낙 크기 때문에 강간·강제추행을 묶어서 분석할 경우에는 강제추행의 발생 특성이 크게 반영된다. 전체 성폭력 발생 건수를 기준으로 분석해보면 대부분의 범죄가 개방된 공개적인 장소에서 발생하는 것으로 나타난다.

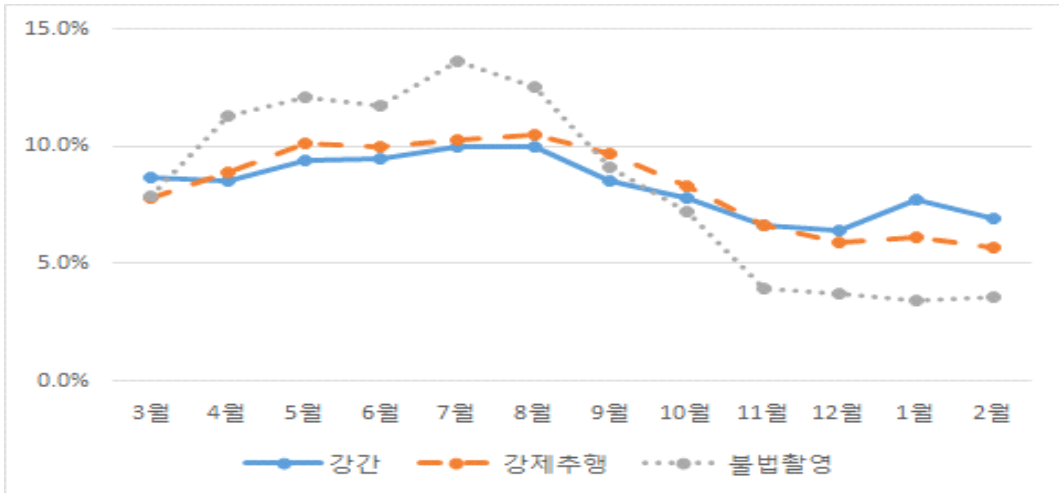
<그림 11> 성범죄 유형별 발생장소



나. 성범죄 발생의 시간적 특성

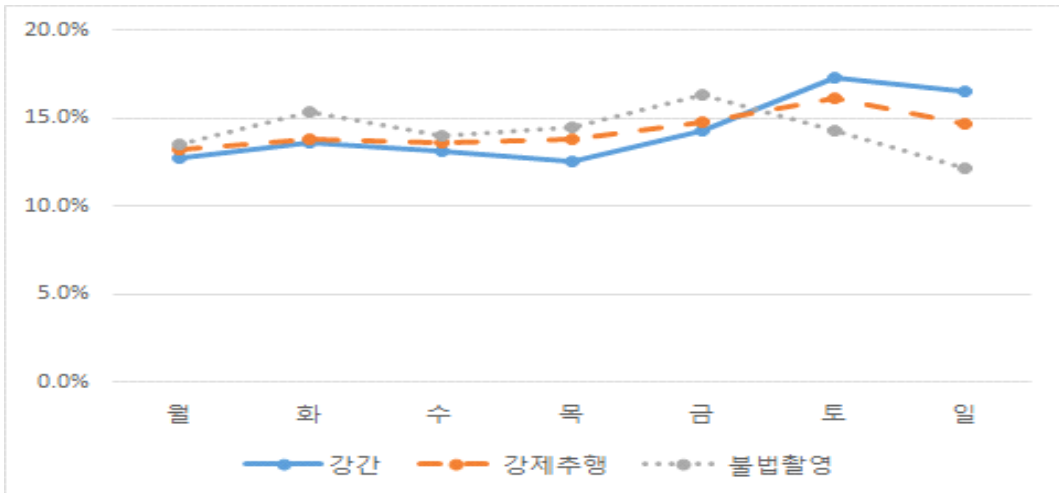
성범죄는 계절에 따른 발생량의 변화가 큰 범죄에 해당한다. 월별 발생비율을 보면 겨울철에는 감소하였던 발생량이 봄철이 되면서 증가하기 시작하여 여름철까지 높은 발생 비율을 유지하게 된다. 복장 등 특정 기회요인에 영향을 크게 받는 불법촬영의 변화가 더 극적으로 확인되었다.

<그림 12> 성범죄 유형별 월별 발생수준



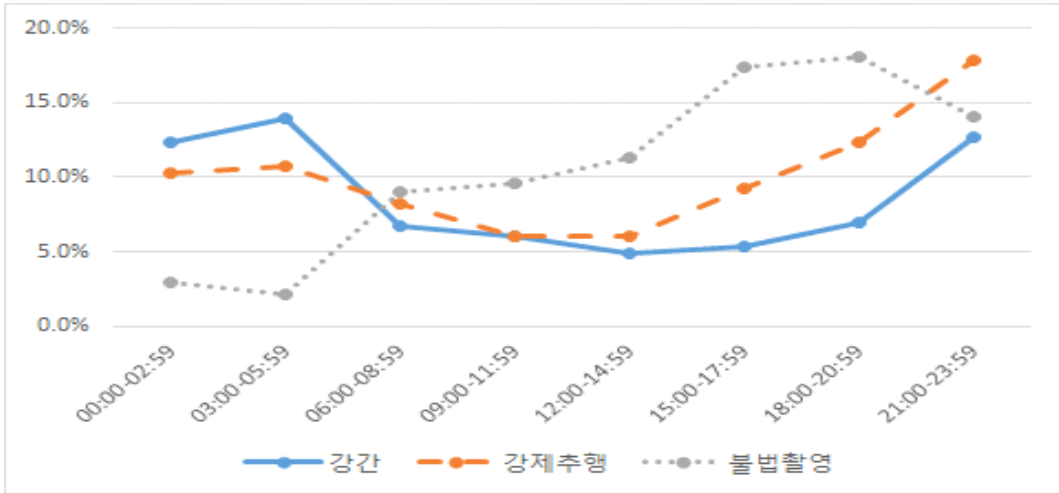
요일별 발생비율의 변화를 보면 요일간의 큰 차이는 없으나 강간과 강제추행은 평일에 비해서 주말에 발생량이 높은 것으로 나타났다. 반면 출퇴근 시간의 혼잡함을 주로 이용하는 불법촬영은 주말에 감소하였다.

<그림 13> 성범죄 유형별 요일별 발생수준



시간대별 발생비율을 보면 범죄유형간 뚜렷한 차이를 보였다. 일조량에 영향을 크게 받는 불법촬영은 06시를 기준으로 증가하다가 21시 이후로는 급격하게 감소, 반면 강간과 강제추행은 지속적으로 증가하여 새벽에도 발생이 지속되었다.

<그림 14> 성범죄 유형별 시간대별 발생수준

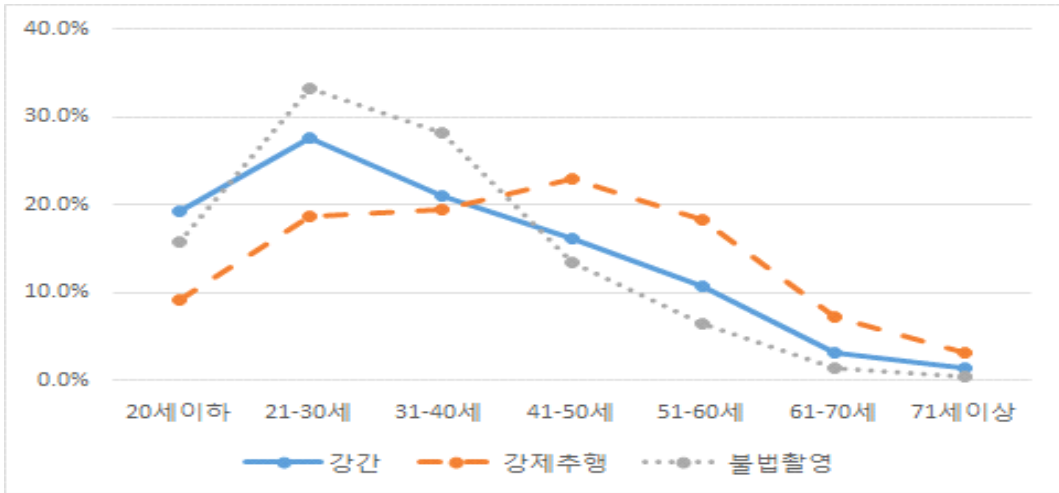


이상의 분석에서 볼 수 있는 것처럼 범죄발생은 시간적인 요인과도 밀접하게 관련되어 있으며, 계절(월), 요일, 시간대 등에 따른 범죄발생에서의 변화 폭이 크기 때문에 이러한 차이를 간과할 경우 이용자의 필요에 부합하는 구체적이고 정확한 예측이 어렵다는 것을 알 수 있다. 범죄예측 정보를 이용하는 경찰관이 범죄예방 활동에 활용하기 위해서는 시간적 정보에 대한 고려가 필요함을 알 수 있다.

다. 성범죄 피의자 특성

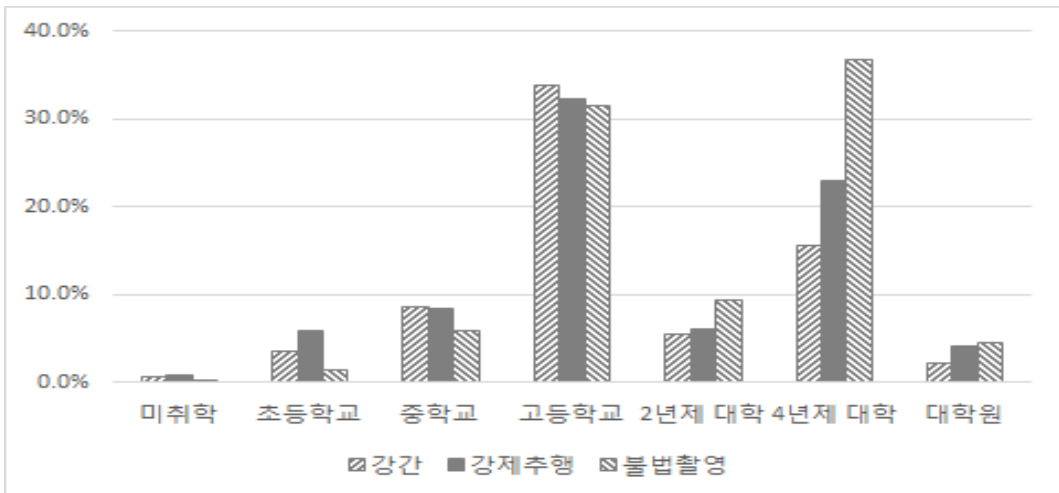
피의자 연령을 확인한 결과 범죄유형별로 피의자 연령대별 비율이 다소 차이가 있는 것으로 확인되었다. 20-30대 비율이 가장 높은 등, 강간과 불법촬영의 비율이 유사한 반면 강제추행의 경우 40대가 가장 높은 비중을 차지해 다른 두 유형의 범죄에 비해 피의자의 연령대가 높은 것으로 나타났다. 그러나 전체적으로는 20-30대가 가장 높은 비중을 차지하고 있어 피의자의 인구구성적인 특성상 특정 연령대에 집중되는 것을 알 수 있다.

<그림 15> 성범죄 유형별 피의자 연령대



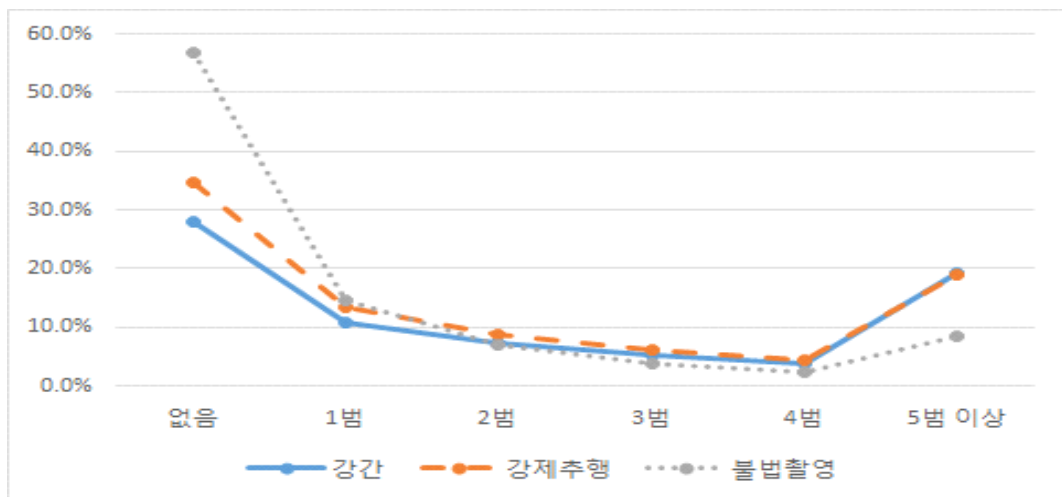
피의자의 학력은 심각한 성범죄(강간>강제추행>불법촬영)일수록 학력이 낮은 경향이 나타났다. 성범죄 전체적으로 봤을 때 성범죄자의 학력은 고졸이하의 비율이 대졸 이상의 비율보다 높았다. 학력이 미상인 경우를 포함한 전체 피의자 중 강간의 경우 대졸 이상의 학력자가 23.1%, 강제추행의 경우 33.1%에 그쳤다. 그러나 불법촬영의 경우 4년제 대학졸업 비율이 50.6%로 상대적으로 훨씬 높은 비율을 보였다.

<그림 16> 성범죄 유형별 피의자 학력



피의자의 범죄 전력을 전과수로 구분했을 때 초범인 경우가 가장 많았지만 (강간 28.0%, 강제추행 34.6%, 불법촬영 56.8%) 전과 유무로 분류할 경우 불법촬영을 제외하고는 전과가 있는 피의자의 비율이 더 높았다. 기존의 연구들에 따르면 다수 범죄는 기존의 범죄 전력이 중요한 위험요인으로 알려져 있다.³⁹⁾ 특히 성범죄의 경우에는 범죄 전력이 중요한 역할을 하는 것으로 알려져 있다.⁴⁰⁾

<그림 17> 성범죄 유형별 피의자 전과



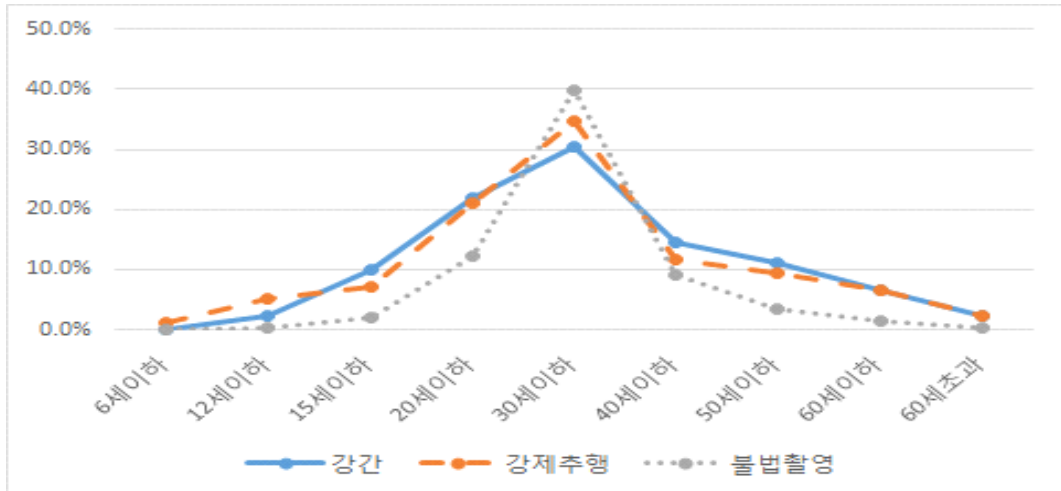
라. 피해자 특성

성범죄 피해는 범죄유형을 불문하고 10대와 20대에 집중되어 있었다. 20대의 비율이 뚜렷하게 높게 나타났지만 아래의 그림에서 볼 수 있듯이 10대의 경우에 여러 연령대(6세 이하, 12세 이하, 15세 이하, 20세 이하)로 구분되어 있기 때문에 실제로는 20대만큼 높은 수준의 피해자 비율을 보이는 것으로 나타났다.

39) Hanson, R. K., & Thornton, D. (1999). Static 99: Improving actuarial risk assessments for sex offenders (Vol.2). Ottawa, Ontario: Solicitor General Canada.

40) 이수정, 김경옥. (2005). 성범죄 재범율에 관한 바른 이해와 재범 방지 대안 모색. 한국심리학회지: 사회 및 성격, 19(3), 83-99.

<그림 18> 성범죄 유형별 피해자 연령대



3. 범죄통계의 해석

성범죄는 현재 살인, 강도, 폭력, 절도 등 주요범죄와 교통범죄 등 범죄발생 건수의 전반적인 감소에도 불구하고 증가추세에 있어 많은 관심이 필요한 시점이다.⁴¹⁾ 범죄통계 상의 증가가 실제 발생의 증가가 아닌 신고율의 증가라 하더라도 심각한 권력적 범죄인 성범죄에 대해 많은 관심이 필요하다. 경찰에서도 성범죄의 심각성을 인식하고 관련된 다양한 정책을 수립·시행하고 있다.

성범죄는 시간적으로는 계절적인 영향, 시간대별 발생량의 차이를 보이고, 장소에서도 발생특성이 뚜렷하여 환경의 영향이 큰 범죄로 분석되었다. 또한 피해자는 학력 수준이 낮은 편이고, 전과가 있는 등의 특성을 보였다. 피해자의 경우 특히 취약한 연령대인 10대와 20대가 압도적으로 많았다.

기존의 연구들을 통해서 성범죄에 영향을 미치는 요인들의 효과가 확인되었다. 성범죄의 발생수준은 지역의 인구, 사회, 주거환경 등 다양한 환경 요인들의 영향을 받는 것으로 나타났다. 아래는 다양한 환경적 요인을 사용하여 회귀 분석을 한 결과를 요약한 표이다.⁴²⁾

41) 치안정책연구소 (2018). 치안전망. p.89.

<표 3> 5대 범죄 발생률에 대한 위계적 회귀분석 결과에서 확인된 변인의 효과

	살인발생률	강도발생률	강간강제추행 발생률	절도발생률	폭력발생률
성비		☆	☆☆☆	☆☆	☆☆☆
20-30대비율	☆☆☆	★★★	★★★	★	★★★
대졸이상비율		★★★	★★★	★★★	★★★
1인가구비율			★★		
범죄경력자비율	★★★	★★★	★★★	★★★	★★★
비거주용건물내 주택비율	★★★	★★★	★★★	★★★	★★★
주택이외거처 비율	★★★	★★★	★★★	★★★	★★★
소형주택비율	★★★	★★★	★★★	★★★	★★★
숙박업소비율	★★★	★★★	★★★	★★★	★★★
주점비율	★		★★		★★

★: 정적(+)효과, ☆: 부적(-)효과, 별의 개수는 유의미도 기준과 동일

이 표에 따르면 강간과 강제추행은 다른 범죄유형에 비해서 다양한 환경요인에 영향을 받는 것으로 확인되었다. 상대적으로 주점비율과 같은 주거환경 요인이나 1인가구비율과 같은 인구·사회적 특성 요인의 영향이 큰 것으로 나타나 분석이 필요할 것으로 보인다. 다른 범죄들과 비교를 하더라도 데이터 분석을 통한 성범죄의 원인분석에서는 범죄의 발생원인이 더 다양하다는 것이 특성으로 나타나고 있다.

42) 박준휘 등 (2017). 국민안전 보장을 위한 형사정책의 실효성 제고방안 연구(II): 1인가구 밀집지역의 안전실태와 개선방안 연구. 한국형사정책연구원, p.203.

제3절 현장조사 결과

1. 현장조사 개요

성범죄에 대한 범죄예측에는 다양한 환경적 요인들이 고려되어야 하기 때문에 성범죄 발생지의 물리적 환경을 현장조사를 통해서 확인함으로써 성범죄 발생에 영향을 미치는 물리적 환경을 분석하였다.

범죄발생에 있어서 물리적 환경의 중요성은 다수의 연구들을 통해서 확인되고 있다. 범죄에 영향을 미치는 유발요인과 억제요인은 다양하지만 성별, 연령대 등 인구 특성과 인구밀도, 노령화지수, 1인가구비율 등 사회적 지표뿐만 아니라 물리적 환경의 영향은 매우 큰 것으로 나타났다.⁴³⁾

그러나 이러한 연구들은 실제 범죄발생 장소를 둘러싼 주변의 물리적 환경이라기보다는 행정동 등 넓은 지역단위 전체의 환경 특성과 범죄발생 빈도(비율)와의 관계에 관한 연구이거나 또는 미시적인 물리적 환경을 확인한 연구의 경우에도 실제로 그러한 환경 하에서 범죄가 발생하였는지는 확인되지 않았기 때문에 기존연구에서의 범죄환경이 직접적인 범죄발생에 인과적인 영향을 미치는지는 확인되지 않았다.

본 연구에서는 이러한 한계를 극복하기 위해서 강간, 강제추행, 불법촬영 범죄에 대한 1년 또는 3년 간의 발생 기록에 대한 발생위치(주소), 발생시간 등 자료를 제공받아 현장의 환경요인(CCTV, 가로등, 골목길 등)에 대한 조사를 실시함으로써 실제 범죄발생지의 물리적 환경을 확인하는 연구를 수행하였다.

현장조사를 위해서 서울과 경기도에서 각각 1개 경찰서로부터 최근의 성범죄 발생 기록을 제공받았고, 온라인 지도를 통해서 건물·도로의 분포 등 전반적인 물리적 환경을 확인하고 현장조사를 실시하여 물리적 환경들을 측정하였다.

43) 박준휘 등 (2017). 국민안전 보장을 위한 형사정책의 실효성 제고방안 연구(II): 1인가구 밀집지역의 안전실태와 개선방안 연구. 한국형사정책연구원, pp.188-200.

2. 범죄발생 장소의 특성

관악경찰서와 안양동안경찰서로부터 제공받은 성범죄 유형과 주소 정보를 활용하였다. 관악경찰서의 경우 2017년 한 해 동안 발생한 노상 강제추행 24건을 대상으로 현장조사를 하였고, 안양동안경찰서의 경우 2015년부터 2017년 3년간 발생한 성범죄 436건 중 노상에서 발생한 87건에 대해서 분석을 하였다.

가. 물리적 환경(미시적 환경)

현장조사에서 확인된 결과에 따르면 노상에서 강간 또는 강제추행이 발생한 시간은 대부분 야간이었다. 이러한 점을 고려한다면 일반적으로 생각해볼 수 있는 물리적 환경은 CCTV나 가로등이 없거나 작동이 제대로 되지 않는 물리적 환경을 떠올리기 쉽다.

그러나 실제로 성범죄 발생지들을 방문했을 때의 실제 물리적 환경은 대부분 CCTV와 가로등이 발생장소의 상단에 설치되어 있었다. 그렇지 않은 경우라고 하더라도 CCTV의 존재를 확인하거나 가로등의 불빛이 비치는 위치에 설치되어 있었다. 또한 도로의 길이 또는 골목의 폭 역시 좁지 않은 곳이 대부분이었다.

아래는 관악경찰서 관할의 신림역 주변의 한 강제추행 발생 장소를 지도상에 표현한 것이다. 24시간 운영 중인 편의점 바로 옆에서 해당 사건이 발생하였음을 확인할 수 있었다.

<그림 19> 관악경찰서 관할 강제추행 발생장소 예시



※ ●: 범행발생장소

또한 아래의 그림은 안양동안경찰서 관할의 범계역 근처에서 발생한 강제추행 사건의 발생장소를 표현했다.⁴⁴⁾ 이 사건에서도 발생장소는 늦은 시간까지 주민들의 이동량이 많은 지하철역 근처이고 바로 근처에 과출소와 119구조대 건물이 있는 것을 확인할 수 있다.

<그림 20> 안양동안경찰서 관할 강제추행 발생장소 예시



※ ●: 범행발생장소

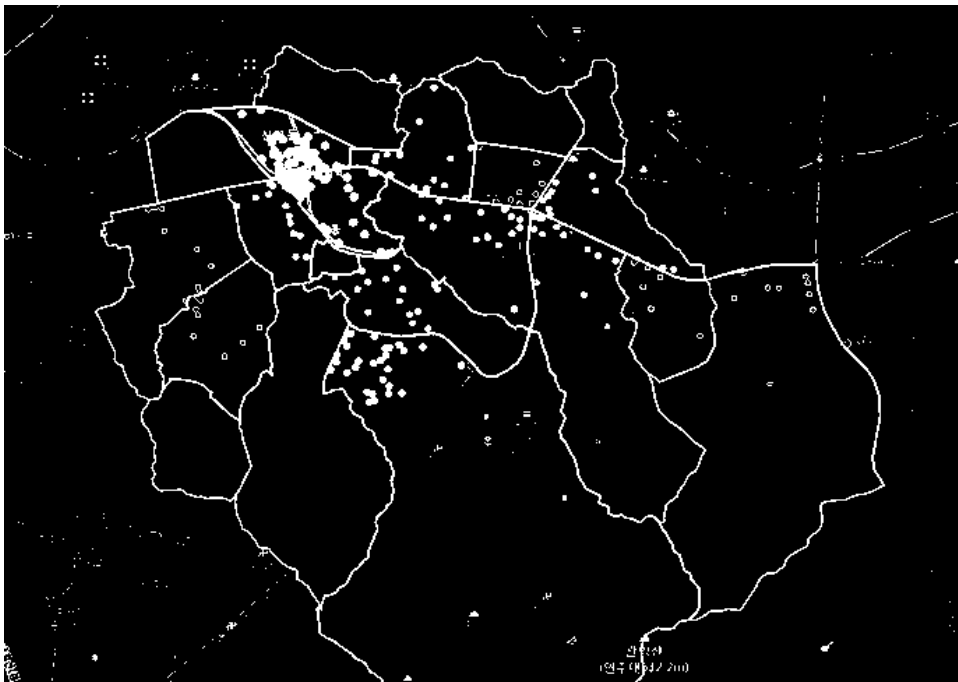
44) 발생 지점이 건물 내인 것처럼 표현되어 있지만 주소를 입력해서 발생장소를 검색한 결과를 현출한 것이기 때문에 이러한 결과가 발생한 것이며, 실제로는 건물 밖 노상에서 범죄가 발생하였다.

나. 지형적 특성

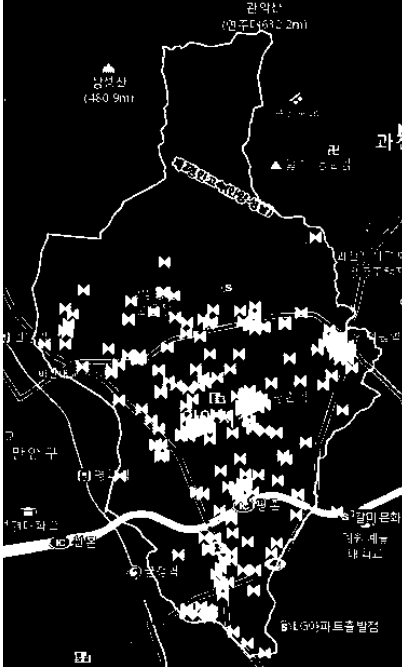
미시적 환경의 특성 확인에서 특이점을 발견하기 어려웠기 때문에 보다 거시적인 관점에서 공간적인 환경특성을 분석해 보았다. 노상범죄를 포함, 장소유형을 구분하지 않고 성범죄 발생장소를 확인하였다. 그 결과 관악경찰서와 안양동안경찰서 모두 성범죄 발생에서 인구의 이동이 많은 지하철역, 유흥가 밀집지역, 하천 주변 등을 중심으로 범죄발생이 집중해 있음을 확인할 수 있었다. 행정동별 경계에 집중되어 있는 것으로 확인되는데, 이는 경계가 도로를 중심으로 구축되어 있기 때문이다.

반대로 도심에서는 드문 토지용도이기는 하지만 관악구의 경우에는 지형적으로 다양한 형태를 포함하고 있는데, 인구의 이동이 거의 없는 농지, 산지 등의 지형에서는 성범죄 역시 거의 발생하지 않는 것을 알 수 있다.

<그림 21> 관악경찰서 관할 노상 성범죄 발생분포(2017년)



<그림 22> 안양동안경찰서 관할
노상 성범죄 발생분포(2015~2017년)



그러나 이러한 결과의 분석에는 여러 가지 한계들이 있었다. 우선 공간특성을 분석할 수 있는 표준화된 방식이 존재하지 않고, 표준화된 방식을 사용하더라도 지역전체의 공간적인 특성을 수치화하는 것은 시간·비용 면에서 매우 어려운 작업이다. 또한 대부분 발생 시간이 야간임에도 불구하고 유사한 시간에 환경을 측정하지 못하고 주간에 현장조사를 한 것도 한계로 볼 수 있다. 그러나 시간적인 조건의 차이를 고려하더라도 발생지의 물리적 환경은 다른 장소와의 차이를 확인하기 힘들었다.

3. 현장조사 결과 해석

노상 성범죄는 주로 심야 시간에 발생하였으나 발생지의 물리적인 환경은 예상했던 것보다 취약하지 않았고, 대부분의 위치에 CCTV 및 가로등이 설치되어 있었다. 미시적인 물리적 환경에서 범죄발생 원인에 대한 단서를 확인하기가 어려웠기 때문에 관할 구역 전체에 대한 거시적인 분석이 필요했다.

지도를 통해 확인한 범죄발생 장소는 주로 유흥지역이나 지하철역 근처로 확인되어 유동인구와 음주상태 등과 관련이 있는 것으로 추정되었다. 또한 노상 성범죄의 경우 범죄발생의 전제인 유동인구가 중요한 의미를 가지는 것으로 추정할 수 있다. 성범죄는 대표적인 대인범죄로, 대부분의 경우 신체적 접촉 상황에서 발생한다.⁴⁵⁾ 따라서 개인들 간의 접촉 빈도가 높은 지역에서 상대적으로

45) 법률적 의미의 성범죄는 형법이나 특별법 상의 강간, 유사강간, 강제추행, 기타 강제추행 등을 의미한다. 그러나 최근 성범죄가 여성에 대한 범죄로 확대 인식되는 경향이 있는데, 여성에 대한 납치,

많은 비면식 관계에서의 성범죄가 발생하는 것은 당연한 일일 수 있다.

그 외에도 기존의 범죄예측에서는 범죄발생이 어려운 산지, 농지 등에 대한 고려가 부족하였는데, 유동인구의 효과를 고려할 때는 토지용도 등을 반영할 필요가 있다. 즉, 산지, 농지의 경우 인구의 이동 자체가 적고 일정한 시간에 주로 이용하기 때문에 범죄발생 가능성이 매우 낮다. 따라서 구역 내에 이러한 지형을 포함하고 있는 지역의 경우에는 범죄위험도 예측에 있어서 고려가 되어야 할 것이다.

범죄발생에 있어서 물리적 환경을 고려하는 것은 물리적 특성이라고 하는 객관적인 특성에 대해서 측정 가능한 지표를 이용할 수 있다는 점에서 장점을 가진다. 그러나 조사 대상이 지나치게 미시적이어서 지역 범위를 확장하여 적용하는 데 한계가 있음을 확인하였다. 물리적 환경의 효과를 확인하기 위해서는 범죄가 발생한 장소뿐만 아니라 범죄가 발생하지 않은 장소 간 비교가 필요한데, 지역 범위가 광범위하여 현장조사 수행이 어려웠다.

현장조사 결과와 관련하여 경찰에서 인지된 사건의 경우 CCTV와 가로등에 의해서 가해자에 대한 인식이 용이하고 증거물의 확보가 용이하다는 점을 고려한다면, 신고 되지 않는 범죄의 발생이 매우 높음에도 불구하고 파악이 되고 있지 않을 가능성이 높다. 그나마 피의자에 대한 지각, 그리고 신고로 이어지기 위해서는 야간에 가로등과 CCTV가 있어야 가능할 것이다. 이후의 분석 내용들은 상당부분 이러한 한계를 가지고 있음을 밝힐 필요가 있다.

성범죄 피해에 관한 한 연구 결과⁴⁶⁾에서도 온라인 설문 응답자 1,483명 중 5년 이내에 비면식 성범죄 피해 경험이 한 번이라도 있는 피해자는 508명이었

감금 등을 포함하기도 하고 성희롱을 포함하기도 한다. 본 연구에서는 연구자료의 한계, 개념의 명확성을 위해 형법(특별법 포함) 상 성범죄만을 연구대상으로 할 예정이며, 성범죄 개념이 확장되어 사용된 경우(성범죄 피해자 분석)에는 개별적으로 명시하였다.

46) 산업통상자원부에서 발주한 ‘사회안전망 기반 성범죄 예방 및 불안해소를 위한 서비스 디자인’ 과제 수행을 위해서 팀인터페이스의 의뢰로 리서치 회사 Data Spring에서는 패널로 보유하고 있는 여성을 대상으로 2018년 상반기에 성범죄 피해 경험에 대한 온라인 설문조사를 실시하여 결과를 얻었다.

- 전체 메일 발송량: 18,461 명
- 참여시도자: 1,580 명
- 탈락자: 1,009 명(비면식범 성범죄 비경험자 975 명)
- 참여 거부자(비동의자): 8명

으며(34.3%), 성범죄 유형별로 경험 비율에는 차이가 있었지만, 공식적인 범죄 통계에서 확인되는 결과와는 큰 차이가 있어서 실제 발생하는 성범죄들이 많은 경우에 신고가 되지 않는 것으로 추정할 수 있다.

<표 4> 성범죄 피해자의 유형별 피해 경험

성범죄 유형	경험자(중복응답)	비율(5년)
강간	107	7.2%
몰카(불법촬영)	69	4.7%
강제추행	394	26.6%

위의 표는 설문조사 결과를 바탕으로 성범죄 유형별 피해경험 비율을 추정한 결과이다. 공식 통계 상 10대~40대 여성 성범죄 피해율을 확인한 결과 강간 0.043%, 강제추행 0.123%, 몰카 0.043%로, 설문조사 결과에서 확인된 비공식 피해율에 비해 낮았다. 이처럼 공식통계와 피해조사 결과가 커다란 차이를 보이는 이유는 우선, 실제 발생건수와 보고 건수가 다르기 때문이며, 대부분의 여성은 성범죄 피해를 경험하고도 이를 보고하지 않는 경향이 있다는 점(온라인 조사의 결과에서도 신고비율은 강간 26.2%, 강제추행 11.1%에 불과)은 기존의 많은 연구들에 의해서 지지되는 부분이다.

특히 미시적인 환경 분석은 발생보고(신고)가 선택적인 경우 더욱 큰 영향을 받을 수밖에 없다. 행정동 단위의 분석에서도 신고율이 낮은 것은 마찬가지로 큰 한계이기는 하지만 더 큰 지역 범위로 수치가 합쳐지기 때문에 전체적인 오차는 감소한다고 볼 수 있다. 따라서 성범죄에 대한 신고비율이 지역적으로 큰 차이가 없다면, 신고된 사건들을 토대로 상대적인 범죄발생률을 예측할 수 있다는 것을 의미한다.

마지막으로 3장에서의 연구를 통해서 얻은 결과를 바탕으로 성범죄와 관련된 원인 또는 예측요인을 아래와 같이 분류해 보았다. 그 결과 심리학 이론은 성범죄 발생의 근원적인 이유를 설명해주기는 하지만 개인의 내면적인 특성들이 대부분이라 이를 사회적 지표에 의해서 확인하기 어렵다는 한계가 있었다. 현장조사의 경우에는 가장 직접적인 물리적 환경을 포함하고 있기 때문에 좋은 자료가 될 수 있으나 전국적인 자료를 수집하기도 어려울 뿐만 아니라, 본 연

구를 통해서도 기존의 범죄환경에 대한 연구결과 또는 일반적 인식과도 차이가 있어 대안으로 적용하기 어려운 것으로 판단된다. 마지막으로 범죄통계에 의해서 확인된 지표들은 공공데이터 등을 통해서 수집이 가능하고 통계적인 검증이 용이하다는 측면에서 범죄위험도 지수를 개선하는데 있어서 가장 적합한 변수로 사용될 것으로 기대된다.

<표 5> 조사결과별 예상되는 예측 변인

조사	심리학이론	범죄통계	현장조사
변인	남녀불평등, 사회적 혼란(지리적 이동, 이혼·편부모가정), 폭력의 수용도, 성폭력·가정폭력 피해경험, 성에 대한 왜곡된 인식	공간·시간적 편재성, 피의자 연령대, 학력·전과, 피해자 연령대	유동인구, 지하철버스 정류장, 숙박업소, 주점 등 유흥업소
비고	공식적인 통계자료로 수치화하기 어려움	공식적인 통계자료를 변인으로 활용하기에 적합	CCTV, 가로등, 가로의 폭 등은 효과가 확인되지 않음

4장 범죄위험 지수 연구

제1절 연구개요

1. 연구목적

앞서 이론적 검토를 통해서 범죄예측 알고리즘을 구성하기 위해서는 많은 기초연구가 필요함을 주장하였다. GeoPros의 경우에도 개발 및 고도화 과정에서 연구를 하고 있지만 가용 데이터의 한계뿐만 아니라 개발자들의 범죄 분야에 대한 기초 지식의 부족으로 인해서 충분한 심층 연구가 부족한 상황이다.

범죄예측은 투입되는 데이터에 따라서 결과가 상이한, 즉 데이터 의존적인 분석방법이기 때문에 국내와는 범죄발생 조건 및 발생 형태가 다른 외국에서의 기존연구 결과에 기초하여 범죄예측의 방향과 방법을 구체적으로 도출하기 어렵다. 따라서 기존 위험도 지수의 문제점을 재차 확인하고 이러한 문제점을 개선할 수 있는 방법을 찾는 방식으로 연구를 수행하였다.

범죄예측의 정확성을 높일 수 있기 위해서는 주요 변인들의 효과를 다양한 방식으로 검증할 필요가 있다. 다양한 행정 자료를 이용하여 그 관계를 검토하여야할 뿐만 아니라 다른 환경과의 상호작용 하에서 개별 변인의 효과가 어떻게 영향을 받는지도 확인할 필요가 있다.

4장의 분석에서는 특히 유동인구라고 하는 범죄발생의 기본적인 조건을 심층적으로 분석하고자 한다. 일상활동 이론 등의 범죄학 이론에 기초해서 본다면 특정 지역의 인구는 범죄발생의 기본적인 조건을 형성한다. 현재의 GeoPros 알고리즘에서 CCTV, 가로등 등의 요인들이 범죄발생을 유발하는 것으로 예측하는 현실을 고려했을 때, 그 원인을 확인할 필요가 있다. 그리고 이러한 요인들은 유동인구와 밀접한 관계가 있다는 기존연구를 고려한다면 범죄

발생에 있어서 유동인구의 효과에 대한 심층적 연구는 반드시 필요할 것으로 보인다.

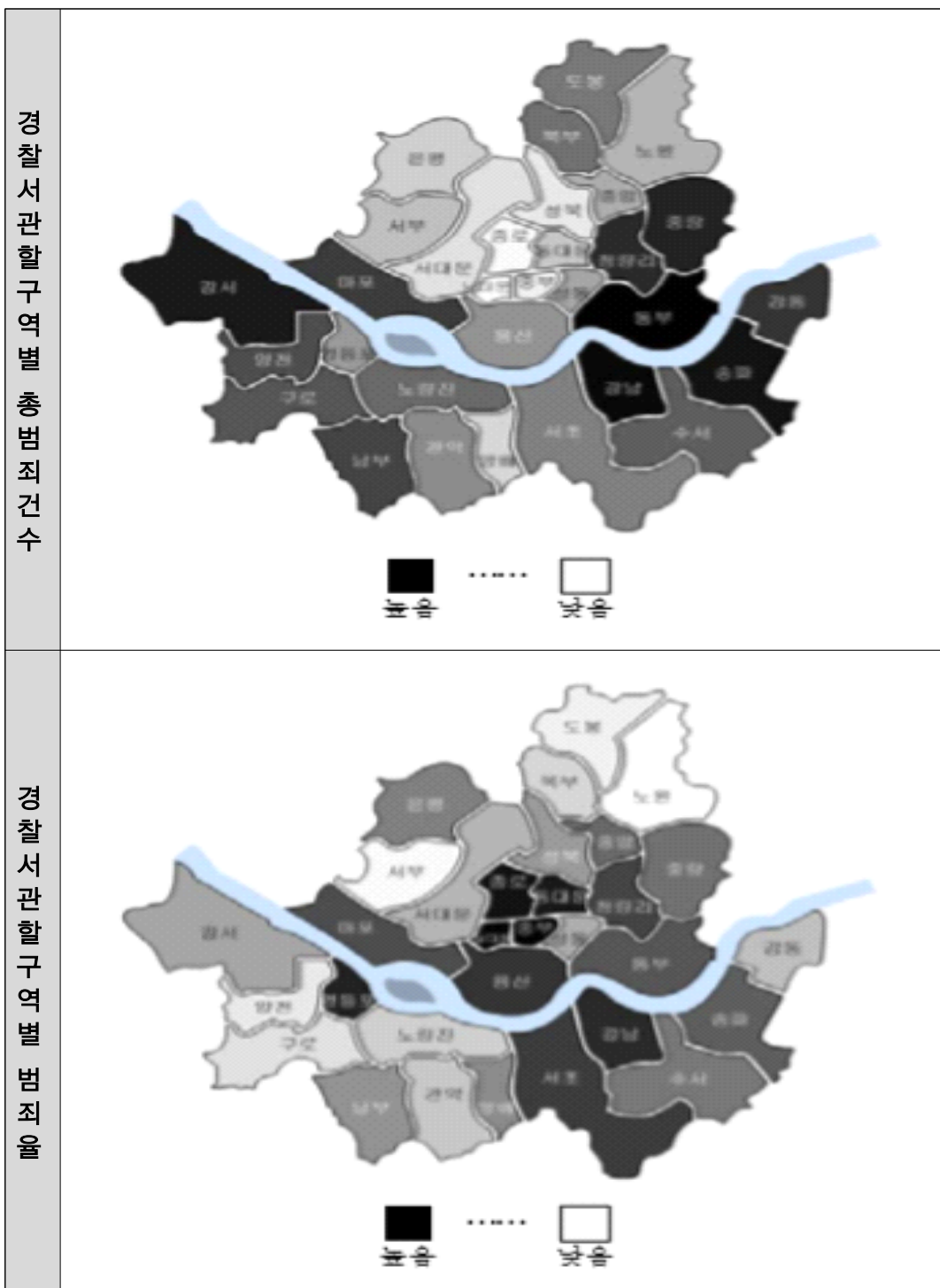
유동인구가 중요한 이유는 본 연구에서의 주요 관심사인 범죄위험도가 인구 대비 상대적인 범죄발생 수준인 범죄발생률에 기초한 것이 아니라 단위 지역별 발생건수에 기초한 것이기 때문이다. 범죄위험도는 결국 단위지역들의 상대적인 범죄발생수준을 보여주는 지수의 개념이다.

그러나 그동안 대부분의 연구들이 지역 내에서의 인구대비 범죄발생률에 주로 관심을 가지고 연구를 해왔다. 기본적으로 인구가 많은 지역에서 범죄발생건수가 많은 것은 어찌보면 당연한 것이니 인구수준을 고려하기 위함이다. 그러나 범죄예방 활동의 측면에서는 절대적인 발생건수 자체가 더 중요한 의미를 가진다. 원인을 불문하고 범죄발생량이 많은 지역에 더 많은 경찰력을 투입하는 것이 더 효과적이기 때문이다. 그렇기 때문에 범죄발생의 기본 조건인 인구는 범죄위험도에서 중요한 의미를 가진다. 다만, 거주인구가 중요할지 유동인구가 중요할지는 분석을 통해서 확인이 필요한 부분이다.

실제로 기존의 연구 결과에 따르면 범죄건수와 범죄율에 따라서 범죄의 분포 양상은 전혀 다르게 나타난다.⁴⁷⁾ 서울의 경우를 예를 들어볼 때 범죄발생 총량이 많은 경찰서 관할지역의 경우에 거주인구 역시 많은 경향이 있기 때문에 인구대비 범죄발생률의 경우에는 아래의 그림과 같이 오히려 반대의 결과가 나타나고 있다. 따라서 범죄발생건수와 범죄발생률은 별개이며, 범죄발생건수에 영향을 미치는 인구 변인은 거주인구보다는 유동인구라고 볼 수 있다.

47) 이성우 (2004). 서울시 범죄발생의 도시계획적 함의. 서울시정개발연구원.

<그림 23> 서울특별시 경찰서관할구역별 총범죄건수와 범죄율 비교



2. 분석방법

현 시스템에서 적용되고 있는 변인들을 치안블록(36만여 개) 및 기초경계구역(3만여 개) 그대로 사용하지 않고 이용 가능한 데이터의 지역적 단위를 고려하여 행정동 단위의 분석을 수행하였다.

전국의 행정동(3,470개)에 대한 인구 특성, 주거 특성 등 이외에도 유동인구 등의 자료를 활용하여 범죄에 미치는 영향 요인, 그리고 영향을 미치는 방식 등에 대해서 연구하였다.

<표 6> 인구사회적 변인 목록

	구분	실제 사용형태	변수 유형	측정 연도
인구 구성 변인	성별	성비 (남성인구/여성인구)	연속	2015
	연령대	연령대비율 (연령대별인구/전체인구)	연속	2015
	학력	학력수준별비율 (고등학교졸업이하인구/전체인구)	연속	2015
사회 특성 변인	인구	인구밀도	연속	2015
	노령인구	노령화지수	연속	2015
	범죄경력자	범죄경력자비율	연속	2015
주거 환경 변인	거처유형	거처유형별 비율 (유형별 거처수/전체 거처수)	연속	2015
	숙박업소	숙박업소 비율 (숙박업소수/전체인구)	연속	2014
	주점	주점 비율 (주점수/전체인구)	연속	2014
	노후주택	30년 이상 주택 비율 (30년이상주택 수/전체주택 수)	연속	2015
	소형주택	20m ² 이하 주택 비율 (20m ² 이하 주택 수/전체주택 수)	연속	2015

특히 서울지역의 경우에는 서울시에서 제공하는 공공데이터를 활용하여 전출입 인구, 이혼건수 등 인구특성과 도소매업, 공공행정업, 숙박및음식점업 등의

사업체와 종사자, 단란주점, 유흥주점, 주차장 수와 같은 성범죄 관련 요인들의 효과도 확인하였다.

제2절 위험요인 분석: 유동인구를 중심으로

1. 유동인구의 효과

가. 기초 분석

서울시 내에서 1일 평균 유동인구⁴⁸⁾는 다음과 같은 기초적인 속성이 확인되었다. 행정동 단위로 최소 6,064명, 최대 2,005,679명의 인구가 이용하는 것으로 확인되었다. 서울시에 해당하는 행정동임에도 불구하고 최소지역과 최대지역 간에는 무려 331배의 차이가 있었다. 평균적으로는 1일 평균 29만 명 정도가 각 행정동을 이용하는 것으로 나타났다.

<표 7> 서울시 행정동별 유동인구 기초분석

N	최소값	최대값	평균	표준편차
424	6064	2005679	285238.74	194197.353

나. 유동인구와 범죄발생 간 관계

서울시에서 유동인구가 가장 많은 지역 상위 20위까지의 행정동에 대해서 범죄유형별 발생수준을 확인하였다. 유동인구가 가장 많은 ○동의 경우에는 범죄발생 건수가 유동인구에 비례해서 높은 것은 아니지만 유동인구가 많은 행정동의 범죄발생 수준은 서울시 행정동 평균에 비해서 높은 수준이다.

48) 유동인구의 추정은 빅데이터 분석 전문업체인 OPENmate에서 독자적으로 개발한 방법으로 도출되었으며, 다음과 같은 절차를 통해서 이뤄졌다. 1) 도보가능 도로 위 10m 간격으로 유동인구를 추정할 기본 포인트 구축 2) 유동인구 추정을 위한 변수 설정(인구유발원단위, 버스정류장, 지하철, 토지용도, 최인접 도로차선 수, 핵심상권여부), 3) 회귀모형을 활용한 1일 평균 유동인구수 추정, 4) 관서영역 내에 포함되는 도로 위 10m 포인트를 공간연산 5) 관서영역별로 일평균유동인구수 평균 집계

<표 8> 유동인구 상위지역(20위까지) 범죄유형별 10년간 발생건수

행정동*	유동인구	범죄 경력자수	살인 건수	강도 건수	강간 강제 추행 건수	절도 건수	폭력 건수
중구 〇동	2005679	372	36	27	263	5225	2956
강남구 〇〇〇동	1338144	3655	21	98	885	7281	8240
영등포구 〇〇〇동	1286920	2846	38	108	668	6416	8881
마포구 〇〇동	1030581	2011	10	78	1153	9127	8592
서대문구 〇〇동	1021195	1231	9	50	902	7255	4800
종로구 〇〇〇〇〇〇〇동	962969	1448	46	129	965	10149	10924
서초구 〇〇〇동	902022	1643	10	25	407	2307	2708
동대문구 〇〇동	794907	3892	19	73	194	3331	3635
강남구 〇〇〇동	779185	1724	18	83	369	3735	3628
강남구 〇〇동	736724	1697	8	50	304	2952	2980
용산구 〇〇〇동	705680	1246	5	41	242	3623	2300
강남구 〇〇〇동	705288	1098	29	74	312	3453	2922
강동구 〇동	700308	4448	9	66	272	3212	4697
동작구 〇〇동	682221	2542	5	13	121	1753	1529
동대문구 〇〇동	678380	2999	21	57	158	2562	2668
영등포구 〇〇동	657858	1262	20	30	403	3191	3536
중구 〇〇〇동	647158	366	3	13	85	1572	1582
노원구 〇〇〇동	646800	1898	8	19	168	2563	2973
구로구 〇〇〇동	642105	3544	26	98	399	4228	4281
관악구 〇〇동	597534	1719	6	71	448	2910	4681

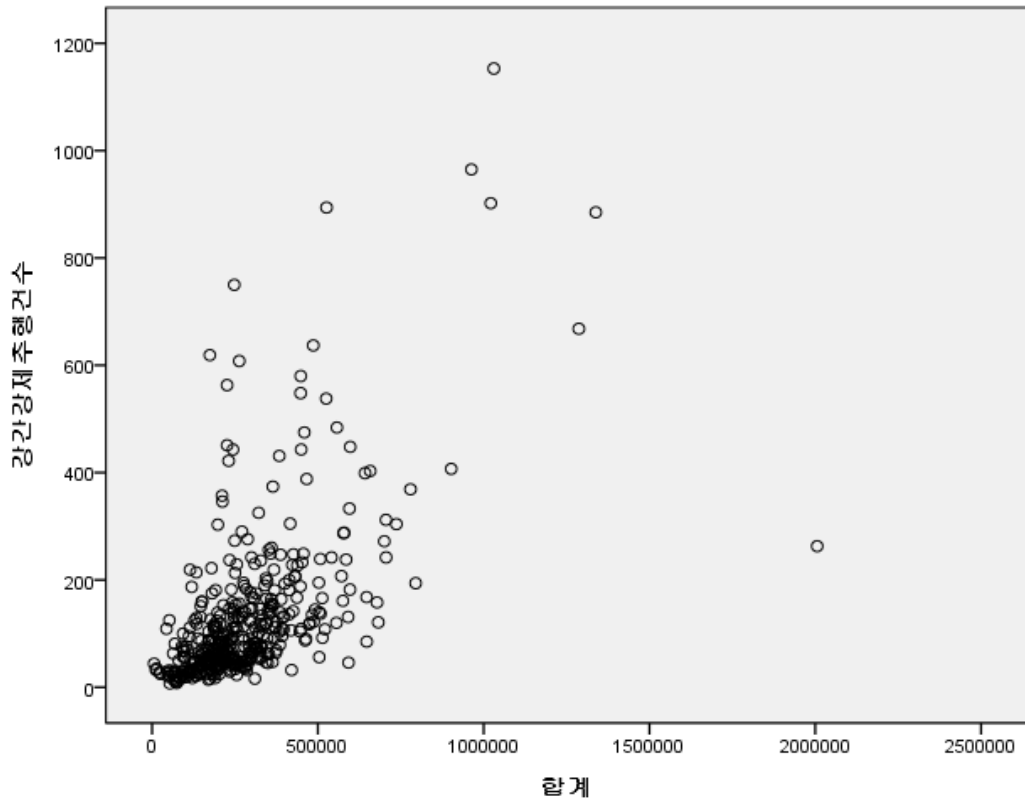
<표 9> 유동인구 하위지역(20위까지) 범죄유형별 10년간 발생건수

	유동인구	범죄 경력자수	살인 건수	강도 건수	강간 강제 추행 건수	절도 건수	폭력 건수
강남구 ○○동	6064	1393	4	5	44	581	748
송파구 ○○동	11651	234	0	0	32	419	370
서초구 ○○동	13294	541	3	6	35	426	328
종로구 ○○동	23968	234	2	6	28	425	282
송파구 ○○○동	24878	295	0	6	25	206	157
송파구 ○○○동	43205	1168	3	2	109	1372	909
종로구 ○○동	46919	699	4	4	26	449	272
종로구 ○○동	48169	536	0	3	20	243	254
송파구 ○○동	48600	561	6	7	31	679	426
광진구 ○동	52168	909	4	11	125	1022	911
용산구 ○○○동	52710	789	3	3	21	391	386
종로구 ○○○동	53746	713	1	4	7	237	204
강서구 ○○○동	56302	1191	1	0	31	455	492
용산구 ○○○동	63645	1167	0	8	63	627	503
관악구 ○○동	65349	1413	1	7	21	279	440
구로구 ○○○동	66469	1151	3	2	27	403	403
양천구 ○○○동	66981	850	3	5	30	574	663
강동구 ○○○동	67321	1002	1	2	12	375	332
강서구 ○○○동	68647	2033	9	8	81	605	762
종로구 ○○동	71587	378	1	4	14	353	222

* 실제 행정동명을 적시하였을 때 발생할 수 있는 부작용을 고려하여 구체적인 행정동명을 생략함

강간·강제추행 범죄에 한해서 산포도를 보면 유동인구가 많은 지역에서 성범죄가 많이 발생하고 있다는 것을 시각적으로 확인할 수 있다. 명동과 같은 초고도 유동인구 지역에서는 오히려 발생 수준이 크게 높지 않지만, 그 이외의 유동인구 과밀 지역에서의 성범죄 발생 수준은 대체로 높은 것으로 나타났다.

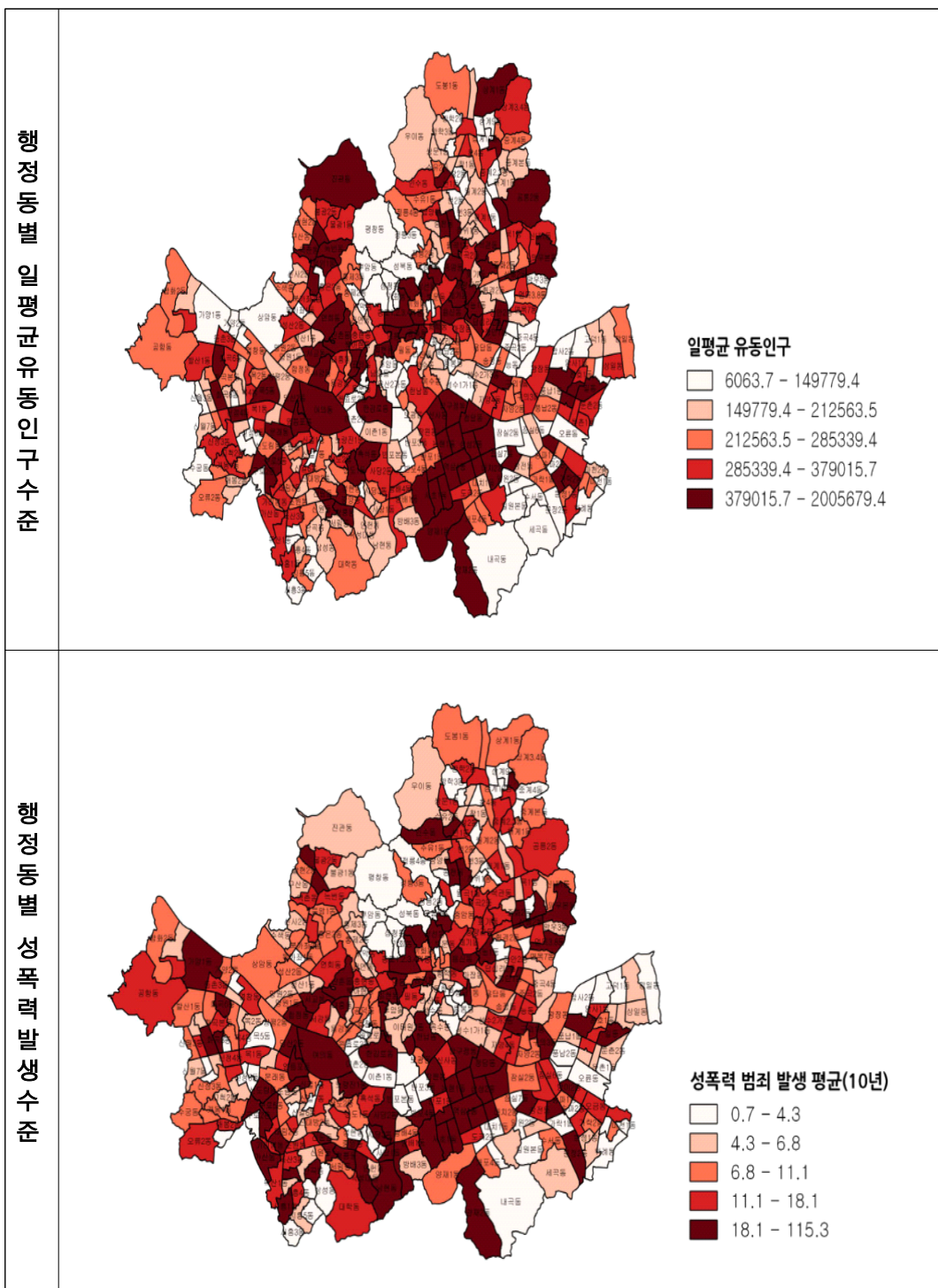
<그림 24> 유동인구와 성범죄 발생수준 관계의 산포도



다. 유동인구와 성범죄 발생의 공간분석

유동인구는 경찰청에서 사용하는 2017년 기준 일평균 유동인구를 사용하였으며, 범죄 발생율은 경찰청 KICS 데이터 접수사건 기준 최근 10년(2007~2016년) 자료 평균을 사용하였다. 이외 유흥업소, 숙박업소 비율, 1인가구 비율, 여성인구 거주 비율, 20~30대 거주 비율, 1인가구 비율 등은 공간정보포털 및 통계청 데이터를 수집, 가공하였다. 그리고 이를 QGIS 프로그램을 사용하여 공간분석을 실시하였다. 그 결과로 확인된 결과는 아래의 그림과 같다.

<그림 25> 행정동별 유동인구수준 및 성폭력발생수준의 분포



유동인구와 성폭력 발생 수준에 대한 행정동별 공간분석을 실시한 결과, 그림 상에서 확인할 수 있는 것처럼 대체로 유동인구가 많은 지역에서 성범죄 발생 수준 역시 높은 것으로 확인되었다. 서울특별시에서 유동인구와 성범죄 발생량은 편재하며, 두 변인 간의 이변량(bivariate) 자기상관계수를 통해서 두 변인 간의 유의미한 관련성이 있는 것으로 확인되었다(Moran's I = 0.127, $p < .001$).

2. 환경변인과의 관계 분석

유동인구와 범죄와의 관계를 이해하기 위해서 유동인구의 의미를 입체적으로 이해하기 위해서 유동인구, 범죄유형별 발생건수, 그리고 인구, 사회, 주거환경 변인들 간의 관계를 상관관계 분석을 통해서 확인해 보았다. 유동인구는 범죄 발생의 전제일 뿐, 직접적인 범죄발생의 조건을 형성하는 것은 다른 환경적인 요인들을 확인함으로써 알 수 있다.

우선, 유동인구는 범죄발생과 유의미한 상관관계가 있는 것으로 확인되었다. 그리고 .502~.740까지 높은 수준의 상관관계⁴⁹⁾가 있는 것으로 나타나서 범죄 발생건수 간의 상관 수준만큼 높은 상관관계를 보였다. 이러한 결과는 유동인구가 범죄발생 수준을 가늠할 수 있는 매우 중요한 요인이라는 것을 확인시켜 주는 결과이다.

49) 상관관계의 해석 기준(상태제, 알기쉬운 통계분석: 기술통계에서 구조방정식 모형까지, 학지사, 2010, 247쪽)

상관계수의 범위	상관관계의 해석
± .00 ~ .20	상관이 매우 낮다
± .20 ~ .40	상관이 낮다
± .40 ~ .60	상관이 있다.
± .60 ~ .80	상관이 높다
± .80 ~ 1.00	상관이 매우 높다

<표 10> 유동인구 및 5대 범죄 발생률 간의 상관관계 분석 결과

	유동인구	살인건수	강도건수	강간강제추행건수	절도건수	폭력건수
유동인구	1					
살인건수	.502***	1				
강도건수	.562***	.656***	1			
강간강제추행건수	.592***	.509***	.770***	1		
절도건수	.740***	.574***	.798***	.885***	1	
폭력건수	.694***	.639***	.832***	.849***	.910***	1

유동인구는 지역의 인구특성과도 상관관계가 있는 것으로 나타났다. 유동인구는 기본적으로 지역의 인구(주민수)나 가구수와 상관관계가 있는 것으로 나타났다, 특히 유동인구가 많은 지역은 20-30대 비율이 높은 반면, 40-50대 비율과 고졸이하 비율은 낮았다. 이는 유동인구가 많은 지역의 특성을 설명해주는 부분이다.

또한 강간·강제추행 역시 20-30대 비율과 정적인 상관, 40-50대 비율, 고졸이하 비율과는 부적의 상관을 나타내어 앞서 확인했던 통계분석 결과에서 피의자의 연령대가 20-30대에 집중되었던 것과 일치하는 결과로 해석할 수 있다. 또한 강간강제추행 발생건수는 높은 수준은 아니지만 기본적으로 인구 및 가구수와 상관이 있지만, 유동인구와의 상관관계 수준(표 21 참조)보다는 낮은 것으로 나타나, 지역의 상주인구나 가구수보다 유동인구가 강간·강제추행의 발생량과 더 관련이 있는 것으로 나타났다.

<표 11> 유동인구 및 5대 범죄 발생건수와 인구특성 간 상관관계 분석 결과

	유동인구	살인건수	강도건수	강간강제추행건수	절도건수	폭력건수
인구(주민수)	.240	.064	.163	.122	.157	.164
전체가구수	.351	.163	.311	.284	.320	.338
성비	.009	.127	.116	.017	.068	.157
20-30대비율	.268	.102	.252	.454	.394	.339
40-50대비율	-.165	-.030	-.130	-.352	-.263	-.199
60대이상비율	.015	.203	.092	-.050	-.031	.080
고졸이하비율	-.160	.039	.039	-.205	-.146	.013

유동인구는 지역의 사회적 특성과도 관련이 있는 것으로 나타났다. 노령화지수와 1인가구수, 그리고 범죄경력자수와 정적인 상관관계가 확인되었다. 이는 유동인구가 많은 지역의 사회적 특성을 보여주는 결과이다. 즉, 유동인구가 많은 지역은 노령화가 많이 진행되어 있고, 1인가구와 범죄경력자가 많은 등의 특성을 보이는 것으로 해석할 수 있다. 강간·강제추행 발생건수도 이와 유사한 보이고 있어, 강간·강제추행의 발생이 유동인구와 공통의 환경적인 특성에 영향을 받을 가능성을 다시 확인할 수 있었다.

<표 12> 유동인구 및 5대 범죄 발생건수와 사회특성 간 상관관계 분석 결과

	유동인구	살인건수	강도건수	강간강제추행건수	절도건수	폭력건수
인구밀도	-.075	-.118	.019	-.087	-.098	-.046
노령화지수	.232	.292	.287	.246	.274	.344
1인가구수	.445	.282	.468	.517	.532	.548
범죄경력자수	.264	.243	.363	.167	.251	.372
수급가구수	.052	.179	.147	.019	.056	.176
수급자수	.005	.127	.093	-.020	.010	.117

유동인구 수준에 따라서 지역의 주거환경 중 거처 유형의 분포가 다르게 나타났다. 아파트, 연립주택 등 일반적인 주거 형태는 유동인구와 크게 관계가 없는 것으로 나타났지만, 비거주용건물내주택, 주택이외거처 등 주택 이외의 거처는 유동인구와 관계가 높은 것으로 나타났다. 범죄와의 관계에 있어서도 매우 유사한 형태가 나타났는데, 성범죄의 경우 단독주택이나 주택 이외의 거처가 많은 지역에서 더 많이 발생하는 것으로 나타났다.

<표 13> 유동인구 및 5대 범죄 발생건수와 거처유형 간 상관관계 분석 결과

	유동인구	살인건수	강도건수	강간강제추행건수	절도건수	폭력건수
단독주택	.252	.233	.417	.302	.342	.395
아파트	.040	-.093	-.174	-.100	-.132	-.133
연립주택	.062	.058	.114	.075	.078	.079
다세대주택	.135	.091	.273	.176	.254	.225
비거주용건물내주택	.362	.180	.365	.417	.449	.438
주택이외거처	.403	.259	.323	.434	.456	.457

유동인구가 많은 지역은 숙박업소, 주점과 소형주택이 많은 것으로 나타났다. 이는 유동인구가 많은 지역에 유흥업소가 밀집되고 부동산 수요와 가격이 높아 소형 주거형태가 많은 일반적인 현실이 반영된 결과이다. 이런 지역은 거주민들의 주거 이동이 잦고, 이웃들 간의 공동체 의식이 부족한 등 사회해체 수준이 높은 것으로 볼 수 있다. 주거환경 특성은 성범죄와도 밀접한 관계를 가지고 있었으며, 숙박업소수, 주점수, 소형주택수는 성범죄 발생수준과 높은 상관관을 보였다.

<표 14> 유동인구 및 5대 범죄 발생건수와 주거환경 간 상관관계 분석 결과

	유동인구	살인건수	강도건수	강간강제추행건수	절도건수	폭력건수
숙박업소수	.668	.419	.606	.754	.798	.838
주점수	.597	.453	.546	.561	.642	.704
노후주택수	.152	.054	.071	.009	.050	.025
소형주택수 (연면적6평미만)	.443	.266	.424	.452	.496	.510

제3절 회귀분석 결과와 고려사항

1. 회귀분석 결과(서울지역)

강간·강제추행 발생률에 대한 3단계 위계적 회귀분석을 실시한 결과 3단계 모형 모두 유의미한 모형인 것으로 확인되었다. 인구구성변인을 투입한 1단계 모형($F(5,418)=28.266$, $p<.001$), 사회특성변인을 추가로 투입한 2단계 모형($F(12,411)=32.972$, $p<.001$), 주거환경변인을 추가로 투입한 3단계 모형($F(22,401)=33.111$, $p<.001$) 모두 유의미하였다.

<표 15> 회귀모형에 대한 단계별 분산분석표

모형	제곱합	자유도	평균 제곱	F	
1	회귀 모형	2327034.779	5	465406.956	28.266***
	잔차	6882415.060	418	16465.108	
	합계	9209449.840	423		
2	회귀 모형	4517208.112	12	376434.009	32.972***
	잔차	4692241.727	411	11416.647	
	합계	9209449.840	423		
3	회귀 모형	5939722.717	22	269987.396	33.111***
	잔차	3269727.122	401	8153.933	
	합계	9209449.840	423		

*** p < .001

인구구성변인을 투입한 1단계 모델($R^2=.253$, adj. $R^2=.244$)에 비해서 사회특성변인을 추가로 투입한 2단계 모델($R^2=.490$, adj. $R^2=.476$)에서 유의미하게 설명량이 증가하였다($\Delta R^2=.238$, $\Delta F=27.406$, $p<.001$). 주거환경변인을 추가로 투입한 3단계 모델($R^2=.645$, adj. $R^2=.625$)에서도 유의미한 설명량의 증가가 있었기 때문에($\Delta R^2=.154$, $\Delta F=17.446$, $p<.001$) 3단계 모델이 적합하다는 것을 확인할 수 있었다.

3단계 모델을 최적 모델로 보고 강간·강제추행발생률에 대한 위계적 회귀분석 결과를 분석하면, 인구구성 변인 중에서는 고졸이하비율이 부적인 영향을 미쳤고($\beta=-.264$, $t=-3.912$, $p<.001$), 사회특성 변인 중에서는 노령화지수가 정적인 영향을 미쳤다($\beta=.199$, $t=2.279$, $p<.05$). 즉, 고졸이하의 비율이 높은 지역일수록 오히려 범죄발생률이 작았고, 노령화지수가 높은 지역에서는 범죄발생률이 큰 것으로 나타났다. 통상 고졸 이하의 학력이 범죄를 정적으로 예측하는 것으로 알려져 있다는 점을 고려하면 이례적인 결과로 볼 수 있으며, 서울 지역의 특수성을 고려하여 학력에 따른 분포를 살펴볼 필요가 있다.

서울특별시에서 성범죄발생량과 관련하여 3단계 모델에서 가장 뚜렷한 효과

가 확인되는 것은 주거환경 변인들이었다. 비거주용건물내주택수($\beta=-.178$, $t=-2.486$, $p<.05$), 숙박업소수($\beta=.555$, $t=.11.344$ $p<.001$), 주점수($\beta=.102$, $t=2.096$, $p<.05$), 노후주택수($\beta=-.078$, $t=-2.301$, $p<.05$) 등의 변인이 유의미한 효과가 있는 것으로 나타났다. 즉, 예상한 바와 같이 숙박업소, 주점 등이 많은 지역에서 성범죄가 많이 발생하였다. 그러나 일반적으로 부정적인 주거환경으로 구분되는 비거주용건물내주택수와 노후주택의 경우에는 성범죄 발생량과 부적 관계가 나온 것은 예상 밖의 결과라고 할 수 있다. 이와 관련해서는 역시 서울 지역의 특수성을 고려할 필요가 있다.

<표 16> 강간·강제추행발생률에 대한 위계적 회귀분석 결과

모형	비표준화 계수		표준화 계수	t	p	공선성 통계량		
	B	표준오차	베타			공차	VIF	
1	(상수)	-872.926	206.172		-4.234	.000		
	성비	-.815	.895	-.043	-.911	.363	.794	1.260
	20-30대비율	1829.167	238.988	.734	7.654	.000	.195	5.138
	40-50대비율	1449.400	431.520	.316	3.359	.001	.202	4.952
	60대이상비율	875.200	217.751	.215	4.019	.000	.627	1.594
	고졸이하비율	-282.593	71.722	-.207	-3.940	.000	.649	1.540
2	(상수)	293.313	241.704		1.214	.226		
	성비	-1.422	.813	-.075	-1.749	.081	.667	1.500
	20-30대비율	173.051	304.200	.069	.569	.570	.083	12.005
	40-50대비율	52.203	398.205	.011	.131	.896	.164	6.082
	60대이상비율	-860.921	358.772	-.211	-2.400	.017	.160	6.241
	고졸이하비율	-373.756	101.513	-.273	-3.682	.000	.225	4.450
	인구_내국인	-.003	.001	-.164	-1.870	.062	.161	6.197
	인구밀도	.000	.001	-.027	-.671	.503	.740	1.352
	노령화지수	.980	.274	.347	3.582	.000	.132	7.577
	수급가구수	-.016	.023	-.036	-.694	.488	.469	2.133
	1인가구수	.012	.008	.139	1.522	.129	.148	6.742
	범죄경력지수	.050	.017	.305	2.852	.005	.108	9.243
	합계	.000	.000	.351	7.531	.000	.570	1.754
3	(상수)	51.885	219.849		.236	.814		
	성비	-1.044	.760	-.055	-1.373	.170	.545	1.835
	20-30대비율	338.782	277.004	.136	1.223	.222	.072	13.937
	40-50대비율	338.591	364.901	.074	.928	.354	.140	7.151
	60대이상비율	-353.872	351.686	-.087	-1.006	.315	.119	8.397
	고졸이하비율	-360.969	92.274	-.264	-3.912	.000	.194	5.149

모형	비표준화 계수		표준화 계수	t	p	공선성 통계량	
	B	표준오차	베타			공차	VIF
인구_내국인	.009	.008	.538	1.089	.277	.004	275.888
인구밀도	-8.337E-5	.000	-.006	-.169	.866	.626	1.596
노령화지수	.562	.246	.199	2.279	.023	.116	8.596
수급가구수	.003	.023	.006	.112	.911	.316	3.160
1인가구수	.031	.025	.346	1.253	.211	.012	86.342
범죄경력자수	.028	.017	.171	1.607	.109	.078	12.851
합계	4.199E-5	.000	.055	1.144	.253	.380	2.634
단독주택	-.031	.029	-.440	-1.064	.288	.005	192.906
아파트	-.032	.028	-.673	-1.170	.243	.003	373.562
연립주택	-.006	.034	-.012	-.165	.869	.179	5.601
다세대주택	-.033	.027	-.337	-1.240	.216	.012	83.585
비거주용건물내주택	-.123	.050	-.178	-2.486	.013	.173	5.796
주택이외거처	-.025	.030	-.105	-.833	.405	.056	17.920
숙박업소수수준	13.674	1.205	.555	11.344	.000	.370	2.699
주점수수준	10.884	5.192	.102	2.096	.037	.371	2.698
연면적6평미만주택수	-.047	.031	-.071	-1.530	.127	.413	2.423
노후주택수	-.011	.005	-.078	-2.301	.022	.775	1.290

* p < .05, ** p < .01, *** p < .001

2. 조절효과의 확인

유동인구가 성범죄 발생에 미치는 효과를 조절하는 변인이 있는지 확인하고자 하였다. 이를 위해서 2단계의 위계적 회귀분석을 통해 특정 변인의 조절효과를 분석하였다.

우선 유동인구가 성범죄 발생에 미치는 효과가 범죄경력자수에 따라서 달라지는지 확인하기 위해서 1단계에서는 유동인구와 범죄경력자수의 효과를 독립적으로 확인하였고, 2단계에서는 유동인구와 범죄경력자수의 상호작용항(유동인구*범죄경력자)을 생성하여 추가하였다.

1단계 모델이 강간·강제추행발생건수에 대한 효과 중 34.7%를 설명하는 것으로 확인되었으며, 유의미한 효과가 있는 것으로 나타났다($F(2,421)=113.576, p<.001$). 2단계에서는 35.7%를 설명하여 설명력이 다소 증가하였으며 ($F(3,420)=78.868, p<.001$), 증가량의 변화가 유의미한 것으로 나타났다. 따라

서 성범죄에 미치는 유동인구에 대해서 범죄경력자수가 조절효과가 있는 것으로 판단할 수 있다.

<표 17> 회귀모형에 대한 단계별 분산분석표

	모형	제곱합	자유도	평균 제곱	F	Adj. R ²	ΔR ²
1	회귀 모형	3227554.555	2	1613777.277	113.576***	.347	.350***
	잔차	5981895.285	421	14208.777			
	합계	9209449.840	423				
2	회귀 모형	3318575.884	3	1106191.961	78.868***	.356	.010*
	잔차	5890873.955	420	14025.890			
	합계	9209449.840	423				

1단계의 분석 결과에 따르면 유동인구가 기본적으로 성범죄 발생에 유의미한 영향을 미치는 것으로 나타났고($t = 14.459, p < .001$), 2단계에서 상호작용항을 투입하였을 때, 상호작용항의 효과도 유의미한 것으로 나타나($t = 2.547, p < .05$), 유동인구와 범죄경력자수의 상호작용 즉, 조절효과가 있는 것을 알 수 있다.

<표 18> 유동인구와 범죄경력자수가 강간·강제추행을 예측하는 회귀분석 결과

	모형	비표준화 계수		표준화 계수	t	p
		B	표준오차	베타		
1	(상수)	.734	14.822		.050	.961
	유동인구	.000	.000	.589	14.459	.000
	범죄경력자수	.002	.007	.012	.290	.772
2	(상수)	39.299	21.120		1.861	.063
	유동인구	.000	.000	.438	6.126	.000
	범죄경력자수	-.022	.011	-.133	-1.908	.057
	유동인구*범죄경력자수	6.669E-8	.000	.255	2.547	.011

상호작용효과를 시각적으로 제시하기 위해서 2단계 모델의 비표준화계수(B)와 두 독립변인에 의해서 구성하였다. 아래와 같은 식에 의해서 독립변인의 값

에 따른 강간·강제추행 발생수준을 도출할 수 있다.

X_1 = 유동인구*의 평균 ± 표준편차

X_2 = 범죄경력자수**의 평균 ± 표준편차

* 유동인구: M = 285238.74, SD = 194197.353

** 범죄경력자수: M = 1957.10, SD = 905.224

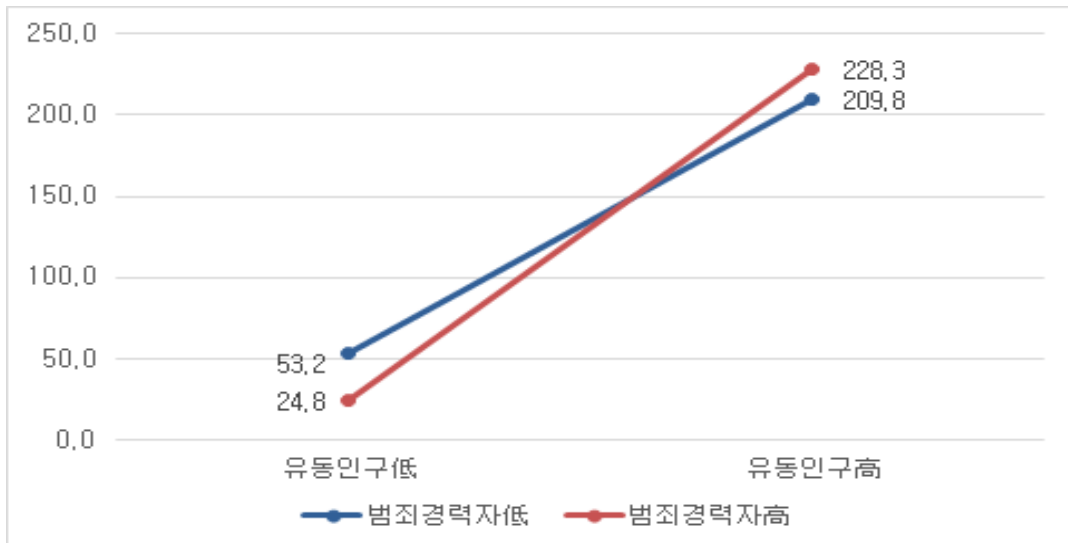
$$Y = 39.299 + 0.000333 X_1 + -0.0217295 X_2 + 6.669028E-08 X_1X_2$$

이 값들을 바탕으로 그림을 그리면 아래 그림과 같다. 이 그림에 따르면 범죄경력자 수준에 따라서 성범죄발생 수준은 차이가 있었다. 유동인구가 적은 지역에서는 범죄경력자가 적은 경우가 성범죄발생 수준이 더 높았지만, 유동인구가 많은 지역에서는 범죄경력자가 많은 지역에서 성범죄 발생 수준이 더 높았다.

언뜻 이해하기 어려운 결과지만, 인구(주민)대비 유동인구 비율 또는 인구(주민)대비 범죄경력자 비율이 아니라 절대적인 유동인구, 그리고 범죄경력자수를 독립변인으로 사용하였기 때문에 실제로 유동인구가 많으면서 범죄경력자수가 적거나 유동인구가 적으면서 범죄경력자가 많은 지역이 많지 않기 때문에 그 특수성이 반영된 것으로 보인다.

그림에서 볼 수 있듯이 실제로 성범죄 발생수준은 거의 유동인구에 의해서 결정되기 때문에 범죄경력자수와 상호작용의 크기는 작은 수준이다. 그러나 실제 회귀식에서 도출되는 범죄위험도지수의 크기는 작은 수준의 차이가 중요하기 때문에 이러한 변인들의 효과에 대해서도 간과할 수 없다.

<그림 26> 유동인구와 범죄경력자수의 상호작용



3. 지역별 차이 확인

현재 GeoPros 상에서는 시군구별로 별도의 회귀식을 사용하여 위험도지수를 산출하고 있다. 그러나 시군구별로 별도의 회귀식을 사용할 경우에는 범죄예측 결과는 확인할 수 있으나, 그 결과가 어떤 경로에 의해서 또는 어떤 요인들의 영향에 의해서 발생한 것인지를 인식하기가 어렵다는 단점이 있다.

따라서 범죄예측의 회귀식 도출 단위를 일종의 권역으로 분류하여 관리하는 것이 필요하다. 적정 단위가 어느 수준인지를 확인하기 위해서는 다양한 지역수준(시도 > 시군구 > 행정동 > 기초경계구역 > 집계구 > 치안블럭)에서의 분석이 필요하다. 이 과정을 통해서 가장 적합한 수준의 권역구분 작업을 하는 것이 중요하다.

우선, 시도별로 세 가지 환경변인군으로 분류하여 3단계 모델의 위계적 회귀분석을 실시하였다. 각 모델별로 추가되는 설명량의 크기는 지역별 환경변인군의 영향력을 의미한다. 3단계 모델의 위계적 회귀분석을 실시하여 분석한 결과 지역별로 세 가지 변인군(이후 요인)의 영향력이 다른 것으로 나타났다.⁵⁰⁾ 인구구성 변인(성비, 연령대비율, 학력 비율)의 영향력이 큰 지역도 있었고, 사회

특성변인(인구밀도, 노령화지수, 인구, 1인가구, 기초생활수급가구, 범죄경력자수)의 영향력이 큰 지역도 있었다. 그리고 상대적으로 주거환경(유형별 거처수, 숙박업소수, 주점수, 노후주택수, 소형주택수)의 영향력이 큰 지역도 있었다.

<표 19> 지역별 환경요인 설명량과 이에 따른 군집 분류

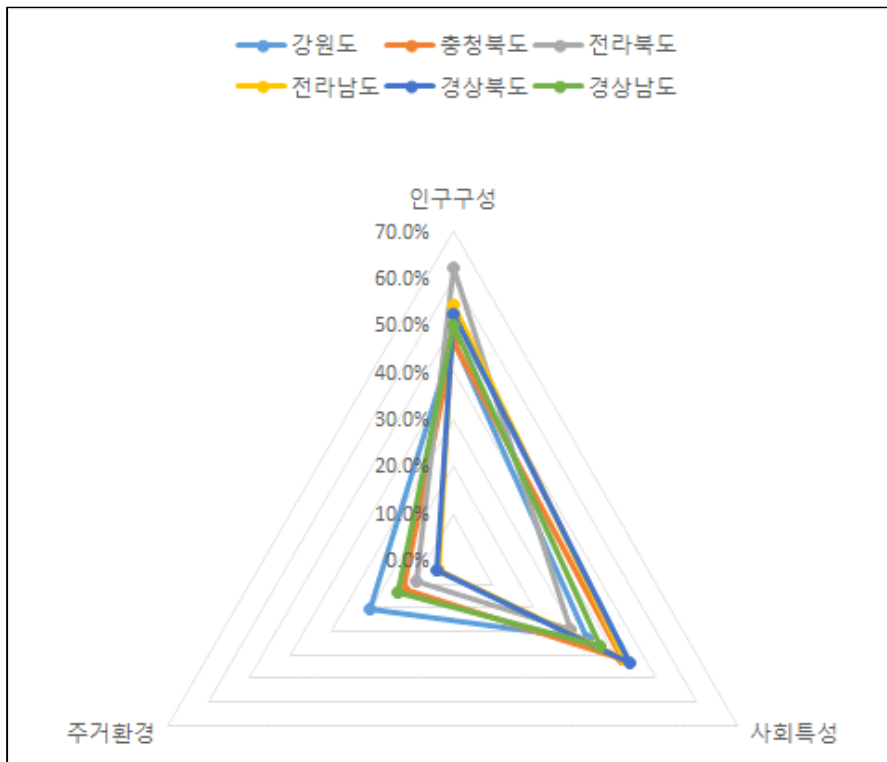
지역구분	설명량				군집
	인구구성	사회특성	주거환경	합계	
서울특별시	0.250 (38.9%)	0.166 (25.9%)	0.226 (35.2%)	0.642 (100.0%)	1
부산광역시	0.194 (30.4%)	0.125 (19.6%)	0.320 (50.1%)	0.639 (100.0%)	1
대구광역시	0.333 (54.7%)	0.127 (20.9%)	0.149 (24.5%)	0.609 (100.0%)	1
인천광역시	0.257 (37.2%)	0.188 (27.2%)	0.245 (35.5%)	0.690 (100.0%)	1
광주광역시	0.221 (25.3%)	0.369 (42.3%)	0.282 (32.3%)	0.872 (100.0%)	1
대전광역시	0.357 (45.4%)	0.144 (18.3%)	0.285 (36.3%)	0.786 (100.0%)	1
울산광역시	0.140 (14.9%)	0.574 (61.3%)	0.223 (23.8%)	0.937 (100.0%)	2
경기도	0.238 (34.4%)	0.241 (34.9%)	0.212 (30.7%)	0.691 (100.0%)	1
강원도	0.367 (46.7%)	0.259 (33.0%)	0.160 (20.4%)	0.786 (100.0%)	3
충청북도	0.366 (46.4%)	0.329 (41.7%)	0.094 (11.9%)	0.789 (100.0%)	3
충청남도	0.258 (30.2%)	0.544 (63.6%)	0.053 (6.2%)	0.855 (100.0%)	2
전라북도	0.511 (62.3%)	0.236 (28.8%)	0.073 (8.9%)	0.820 (100.0%)	3
전라남도	0.407 (54.5%)	0.313 (41.9%)	0.027 (3.6%)	0.747 (100.0%)	3
경상북도	0.454 (52.4%)	0.378 (43.6%)	0.035 (4.0%)	0.867 (100.0%)	3
경상남도	0.439 (50.2%)	0.318 (36.3%)	0.118 (13.5%)	0.875 (100.0%)	3
제주도	0.401 (41.1%)	0.530 (54.3%)	0.045 (4.6%)	0.976 (100.0%)	2

50) 각 모델(단계)별로 추가되는 설명력의 크기를 기준으로 하였다.

세 가지 요인의 영향력의 상대적 크기를 기준으로 하여 유사한 유형을 분류하기 위해서 군집분석을 실시하였다. 그 결과 크게 세 가지 유형의 지역으로 분류가 되었다.

설명 상의 편의를 위해서 3군집부터 설명을 하면, 다른 요인들에 비해서 인구구성 요인의 영향력이 가장 크고, 사회특성 요인의 영향력도 어느 정도 있으나 주거환경의 영향이 거의 없는 지역이 3군집으로 분류되었다. 이러한 지역들에는 강원도, 충청북도, 전라남·북도, 경상남북도 등 상대적으로 비도시적인 특성이 강한 지역이 이 유형에 포함되었다.

<그림 27> 군집 3의 요인별 설명량 비율



전라북도가 대표적인 사례로 볼 수 있으며, 전라북도 지역에서는 3단계 모델을 기준으로 40-50대 비율이 성폭력 발생에 영향을 미치는 가장 중요한 인구구성 변인이었다. 그리고 사회특성 변인 중에는 1인가구수가 유의미한 변인으로

확인되었다. 마지막으로 주거환경 변인 중에는 단독주택수, 숙박업소수, 노후주택수 등이 성범죄 발생수준에 영향을 미치는 유의미한 변인으로 확인되었다.

특이한 점은 3군집의 특성상 인구구성요인의 효과가 커야함에도 불구하고 3단계모델을 적용했을 때 유의미한 변인이 40-50대 비율밖에 남지 않았다. 이러한 결과는 3단계 모델에서 투입된 주거환경 변인의 효과가 인구구성요인의 효과를 간섭하여 독립적인 효과가 크지 않기 때문으로 볼 수 있다

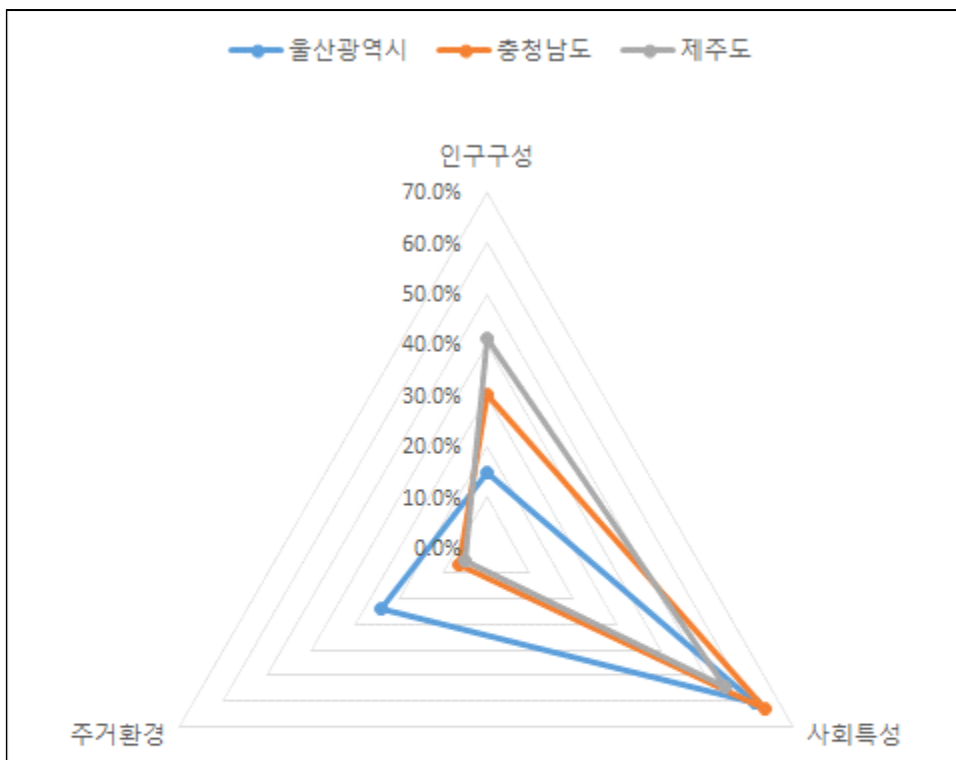
<표 20> 강간·강제추행발생률에 대한 위계적 회귀분석 결과(전라북도)

모형	비표준화 계수		표준화 계수	t	p	공선성 통계량		
	B	표준오차	베타			공차	VF	
3	(상수)	-137.955	73.735		-1.871	.063		
	성비	.341	.263	.058	1.296	.197	.485	2.061
	20-30대비율	75.205	99.128	.116	.759	.449	.041	24.501
	40-50대비율	212.428	84.822	.159	2.504	.013	.239	4.192
	60대이상비율	113.746	59.937	.394	1.898	.059	.022	44.960
	고졸이하비율	-35.702	43.934	-.101	-.813	.417	.063	15.970
	인구수	.005	.003	1.169	1.888	.061	.002	400.152
	인구밀도	.000	.000	-.020	-.416	.678	.424	2.357
	노령화지수	-.009	.015	-.056	-.638	.524	.123	8.151
	1인가구수	.014	.005	.381	2.840	.005	.053	18.841
	수급가구수	.005	.010	.024	.480	.632	.380	2.630
	범죄경력지수	-6.356E-5	.009	-.001	-.007	.994	.054	18.691
	단독주택수	-.015	.007	-.461	-2.165	.032	.021	47.310
	아파트수	-.012	.008	-.791	-1.585	.115	.004	260.203
	연립주택수	-.006	.020	-.014	-.284	.777	.404	2.473
	다세대주택수	-.015	.022	-.034	-.652	.515	.349	2.864
	비거주용건물내주택수	.054	.043	.084	1.267	.207	.220	4.553
	주택이외거처수	-.018	.035	-.029	-.520	.604	.299	3.344
	숙박업소수수준	6.261	1.143	.427	5.480	.000	.158	6.326
	주점수수준	-.720	.751	-.036	-.959	.339	.665	1.504
연면적6평미만주택수	.162	.092	.115	1.768	.079	.226	4.419	
노후주택수	.016	.006	.155	2.802	.006	.311	3.211	

* p < .05, ** p < .01, *** p < .001

다음으로 2군집의 경우에는 사회특성의 영향력이 다른 요인들에 비해서 높은 지역적 특성을 보였다. 이러한 지역에는 광역시 중 울산광역시 외에도 충청남도과 제주도가 혼재되어 있어 일반적인 시도의 구분에 의해서 범죄발생의 특성을 구분하기는 어렵다는 것을 보여주고 있다. 이들 지역은 특별시나 거대 광역시, 그리고 비도시 지역을 많이 포함하고 있는 시도지역과는 차이가 있는 것으로 보인다.

<그림 28> 군집 2의 요인별 설명량 비율



그 사례로 충청남도에 대한 회귀분석 결과에 따르면 3단계 모델을 기준으로 인구구성요인 중 60대 이상의 비율이 부적 효과, 고졸이하의 비율이 정적인 효과를 보이는 것으로 나타났다. 사회특성 요인 중에서는 인구밀도, 노령화지수, 1인가구수 등 도시화와 관련된 변인들이 유의미한 효과를 나타내고 있었다. 마지막으로 주거환경 변인 중에서는 숙박업소수, 주점수 등이 유의미한 변인으로 확인되었다. 즉, 충청남도 지역의 행정동들 중에서 유흥업소가 많은 지역의 경우에 다른 지역에 비해서 성범죄 발생량이 많은 것으로 나타났다.

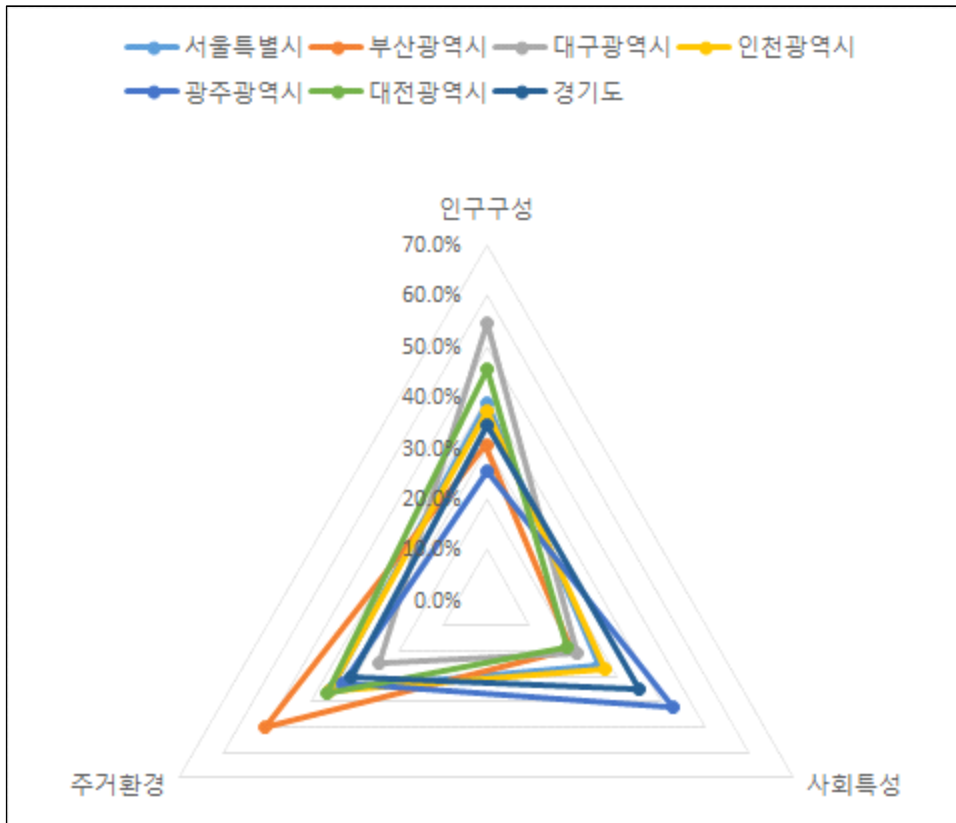
<표 21>강간·강제추행발생률에 대한 위계적 회귀분석 결과(충청남도)

모형	비표준화 계수		표준화 계수	t	p	공선성 통계량	
	B	표준오차	베타			공차	VIF
(상수)	47.527	52.008		.914	.362		
성비	-.225	.177	-.042	-1.273	.205	.669	1.494
20-30대비율	-131.563	83.507	-.186	-1.575	.117	.052	19.068
40-50대비율	15.721	81.657	.009	.193	.848	.325	3.074
60대이상비율	-240.542	69.762	-.703	-3.448	.001	.018	56.915
고졸이하비율	84.773	40.712	.196	2.082	.039	.083	12.110
인구수	.002	.002	.351	1.001	.318	.006	168.343
인구밀도	.002	.001	.122	3.305	.001	.534	1.872
노령화지수	.046	.015	.243	3.175	.002	.125	8.001
1인가구수	.033	.004	.818	8.884	.000	.086	11.599
수급가구수	.012	.016	.039	.739	.461	.262	3.821
범죄경력지수	-.010	.007	-.120	-1.288	.199	.084	11.961
단독주택수	-.007	.004	-.154	-1.531	.127	.072	13.834
아파트수	-.009	.005	-.507	-1.737	.084	.009	116.485
연립주택수	-.030	.022	-.075	-1.380	.169	.245	4.085
다세대주택수	.007	.011	.030	.613	.541	.300	3.335
비거주용건물내주택수	-.067	.046	-.111	-1.444	.150	.123	8.146
주택이외거처수	.013	.012	.050	1.075	.284	.337	2.964
숙박업소수수준	4.380	.795	.308	5.510	.000	.233	4.288
주점수수준	1.845	.450	.120	4.103	.000	.860	1.163
연면적6평미만주택수	-.007	.022	-.017	-.319	.750	.262	3.811
노후주택수	.007	.006	.055	1.104	.271	.298	3.354

* p < .05, ** p < .01, *** p < .001

마지막으로 1군집에 해당하는 지역에는 서울특별시와 대부분의 광역시 그리고, 수도권으로 분류되는 경기도가 포함되었다. 이들 지역은 세 가지 요인의 효과가 골고루 나타나는 지역이기는 하지만 다른 군집들에 비해서 주거환경의 효과가 상대적으로 큰 지역으로 나타났다.

<그림 29> 군집 1의 요인별 설명량 비율



서울특별시를 기준으로 유의미한 변인들의 효과를 확인해본 결과, 앞서 유동 인구를 포함한 분석에서의 결과와 거의 동일한 결과가 나타났다. 즉, 인구구성 요인 중 고졸이하비율이 부적의 영향을 미치고 있었고, 사회특성 요인 중 노령화지수가 정적인 효과가 있는 것으로 나타났다. 주거환경 변인에 있어서도 숙박업소, 주점이 많은 지역은 성범죄 발생량이 많고, 비거주용건물내주택수, 노후주택이 많은 지역은 성범죄 발생량이 적었다.

<표 22> 공간·강제추행발생률에 대한 위계적 회귀분석 결과(서울특별시)

모형	비표준화 계수		표준화 계수	t	p	공선성 통계량	
	B	표준오차	베타			공차	VIF
(상수)	39.931	223.487		.179	.858		
성비	-1.179	.754	-.063	-1.563	.119	.568	1.759
20-30대비율	357.692	281.393	.143	1.271	.204	.072	13.802
40-50대비율	409.005	366.717	.089	1.115	.265	.145	6.916
60대이상비율	-294.040	361.235	-.071	-.814	.416	.119	8.391
고졸이하비율	-402.450	93.957	-.291	-4.283	.000	.198	5.053
인구수	.010	.008	.632	1.273	.204	.004	270.802
인구밀도	-1.702E-5	.001	-.001	-.034	.973	.626	1.598
노령화지수	.594	.246	.209	2.411	.016	.122	8.216
1인가구수	.035	.025	.382	1.372	.171	.012	85.204
수급가구수	-.001	.024	-.002	-.033	.973	.327	3.059
범죄경력자수	.033	.017	.198	1.878	.061	.082	12.152
단독주택수	-.036	.029	-.509	-1.229	.220	.005	188.081
아파트수	-.038	.028	-.766	-1.339	.181	.003	359.689
연립주택수	-.010	.034	-.021	-.297	.767	.179	5.594
다세대주택수	-.039	.027	-.396	-1.444	.150	.012	82.680
비거주용건물내주택수	-.127	.050	-.182	-2.524	.012	.175	5.711
주택이외거처수	-.030	.030	-.125	-.982	.327	.056	17.846
숙박업소수수준	14.138	1.151	.572	12.278	.000	.420	2.379
주점수수준	12.326	5.086	.116	2.423	.016	.395	2.533
연면적6평미만주택수	-.045	.031	-.068	-1.463	.144	.416	2.405
노후주택수	-.012	.005	-.082	-2.392	.017	.776	1.288

* p < .05, ** p < .01, *** p < .001

4. 시간적 요인의 고려

본 연구에서 고려하지 못한 부분은 시간적 요인에 대한 부분이다. 시공간적 연구를 통해서 정밀한 범죄예측을 할 수 있는데 자료의 부족으로 인해서 시간적 요인에 대한 효과를 확인하지 못했다. 실제 범죄발생은 계절, 월, 요일, 시간대에 따라서 큰 차이가 있기 때문에 시간적인 요인의 효과를 동시에 살펴봐야 정확한 범죄예측이 가능해질 것이다.

따라서 범죄예측이 치안활동에 유용하게 활용되기 위해서는 상황 변화에 따라서 다른 예측값을 제시할 필요가 있는데, 시간은 중요한 변인이 될 수 있다. 현재의 범죄위험도 지수 역시 시간대별로 다른 예측치를 보여주고 있다. 범죄 발생과 시간의 상호작용, 그리고 기온, 습도 등 날씨와 관련된 요인들 역시 시시각각으로 변화가 크기 때문에 시간적 요인의 관점에서 연구가 필요한 영역이다.

시간주기의 변화는 사람들의 생리적인 상태, 기분과 같은 심리적 상태에 영향을 미치기 때문에 결과적으로 행동변화를 유발한다. 다만, 시간적인 변화가 절도나 폭력 등 특정 범죄행위를 직접 유발하는 것은 아니다. 시간적인 변화에 수반된 날씨 요소가 범죄 충동이나 공격성과 같은 생리적 변화와 심리적 변화를 유발하고, 주변적인 기회요인이 뒷받침될 때 범죄가 발생하게 된다. 시간적인 요인은 특정 범죄와 밀접하게 관련되어 있는 경향이 있다. 소매치기는 사람들이 붐비는 시간대에 주로 발생한다. 사람들 간의 밀도가 높아서 다른 사람의 접촉에 대한 경계심이 약화되고 다른 곳에 주의를 팔기 쉬운 상황적·공간적 특성에 기인한 것이지만, 시간대라고 하는 객관적인 시간적 특성으로 표현할 수 있다.

<표 23> 시간적 요인 유형과 관련 현상

연번	분류방식	관련 현상
1	오전, 오후, 저녁, 밤, 새벽	기온, 일조량 등에 의한 생리적인 변화
3	월~일요일	생활주기와 기분의 변화
4	평일, 주말	근무여부에 따른 활동방식의 변화
5	1~12월	날씨 변화에 따른 생리적 변화
6	봄, 여름, 가을, 겨울	기후 변화에 따른 생리적 변화

출처: 윤상연 (2017). 발생일자 기준 계절별·월별 범죄발생특성. 치안정책연구소. p.25.

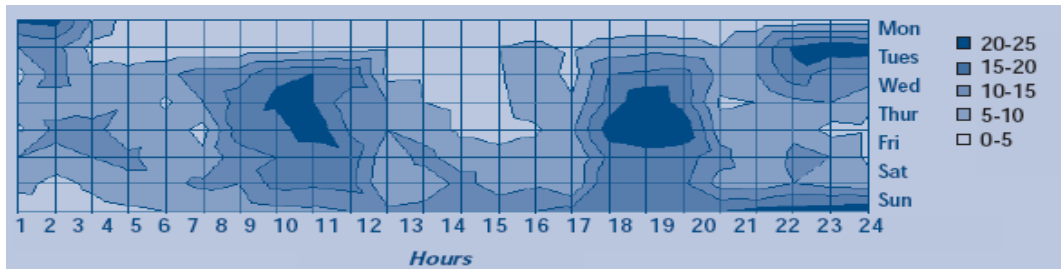
또한 가중치에 대한 효과는 요인들 간에 상호작용이 있을 수 있기 때문에 단순히 가중치를 곱하는 형태의 적용은 부정확한 결과를 낳을 수 있다. 예를 들어서 요일에 따라서 범죄의 주요 발생 시간이 달라질 수 있다. 또한 아래의 실

제 사례에서도 확인할 수 있는 것처럼 계절에 따라서 기후가 달라지면, 기온이나 습도의 의미는 달라진다.⁵¹⁾

<그림 30> 요일 및 시간별 사건발생 수에 관한 예시

Days	Hours																								Mean
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	
Mon	7	3	6	9	9	11	10	17	16	17	5	6	12	7	9	5	20	18	16	8	7	10	8	7	10.1
Tues	10	9	10	11	7	6	13	15	15	18	6	12	7	7	8	5	16	18	17	12	5	11	10	4	10.5
Wed	4	11	12	6	8	3	18	17	24	14	10	7	4	2	4	9	23	24	24	7	5	12	3	4	10.6
Thur	8	8	7	9	4	10	17	21	20	11	10	5	4	3	9	6	17	24	18	4	5	4	10	6	10.0
Fri	3	12	6	7	9	12	13	19	20	19	3	4	3	4	9	3	15	16	17	8	10	18	12	13	10.6
Sat	10	12	3	1	1	1	1	1	2	3	4	3	4	5	6	4	7	8	7	5	10	21	24	23	6.9
Sun	21	22	10	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3.7
Mean	9.0	11.0	7.7	6.4	5.6	6.3	10.4	13.0	14.0	11.9	5.6	5.4	5.0	4.1	6.6	4.7	14.1	15.6	14.3	6.4	6.1	11.0	9.7	8.3	

<그림 31> 요일 및 시간별 리듬 예시



따라서 기존의 범죄예측에서 월별, 요일별, 기온이나 습도에 대해서 가중치를 곱하는 형태로는 정확한 범죄예측이 힘들다. 왜냐하면 기온을 예로 들었을 때 여름철과 겨울철에 동일한 기온이 행동상의 변화에 주는 의미는 다르기 때문이다. 따라서 각 시간적 요인의 효과에 대해서는 보다 구체적인 연구를 통해서 그 효과를 적용하는 방법을 찾을 필요가 있다.

51) Clarke, R. V., & Eck, J. E. (2005). Crime Analysis for Problem Solvers in 60 Small Steps, Washington, (역) 윤민우 (2010), 문제해결을 위한 범죄분석 60단계, 형사정책연구원, 144쪽.

제5장 결론

제1절 연구 의의

범죄예방이 치안활동의 중요한 패러다임이 된 현 시점에서 범죄예측의 중요성은 부정할 수 없다. 그러나 그동안 국내외에 구축된 많은 예측 시스템들이 충분한 사전 연구 없는 개발로 인해 그 효과성에 대해 많은 의문이 제기되었다. 본 연구는 경찰의 대표적인 범죄예측 기능을 탑재하고 있지만 다양한 한계를 가지고 있는 GeoPros의 범죄위험도 지수를 분석하여 그 한계를 확인하고, 이를 개선할 수 있는 방법을 발견하고자 하였다. 이러한 과정을 거쳐서 얻은 연구결과가 국내 범죄예측 시스템에 반영된다면, 범죄예측의 신뢰성과 타당성을 제고할 수 있을 것으로 기대하였다.

본 연구는 성범죄를 주요 범죄유형으로 하여 진행된 만큼 연구 과정에서 성범죄와 관련된 여러 가지 발생 특성들을 확인할 수 있었다.

우선 성범죄는 인지, 감정상의 정신적 문제와 관련이 있는 것으로 알려져 있지만 정신질환자의 지역적 분포를 알 수 없는 이상 성범죄의 발생에 어느 정도 영향을 미치는지 알 수 없다. 따라서 지역별 자료를 토대로 성범죄를 예측하는데 있어서 심리적인 원인은 큰 의미를 못 가지는 것으로 보인다.

보다 직접적으로 성범죄의 발생에 영향을 미치는 요인에 대한 단서를 제공하는 것은 범죄발생 특성에 대한 통계적 분석이었다. 기존의 성범죄 발생 특성을 살펴보면 성범죄는 유형에 따라서 차이가 있지만 강간에 비해서 많은 비중을 차지하는 강제추행의 경우에는 노상 등 개방성이 높은 장소에서의 발생비율이 높은 것으로 확인되었다. 이는 유동인구가 밀집한 지역에서의 성범죄발생 가능성을 시사하며, 지하철, 버스정류장 등 각종 대중교통수단이나 역 등의 장소에

서 성범죄 발생 가능성이 높다는 것을 의미한다. 또한 시간적인 관점에서 볼 때 성범죄는 봄철에 증가하기 시작해서 여름철까지 높은 수준으로 발생하다가 가을과 겨울까지 감소하는 형태를 보이고 있다. 또한 시간대별로 성범죄의 발생수준 차이가 상당히 크다. 따라서 계절, 월별, 시간대별 다른 범죄대응 전략을 펼칠 필요가 있으며, 공간과 시간에 맞는 적절한 범죄위험도를 제공하는 것이 필요하다.

일부 샘플 지역에 대한 현장조사를 통해서 확인된 것은 구체적인 물리적 환경의 효과보다는 거시적인 환경 요인의 효과였다. 즉, 노상 성범죄에 한정되기는 했지만 성범죄가 발생한 장소는 주로 유동인구와 숙박업소, 주점 등이 밀집해 있는 특성을 보였다. 이는 일반적으로 성범죄가 사람들이 접근하기 어렵고 어두운 장소에서 발생할 것이라는 인식과 상반된 결과다. 주로 심야시간에 발생하는 경우가 많기 때문에 타인의 시선이 감시의 역할을 제대로 못하는 경우가 있을 수 있고, 음주 등의 상태로 가해자는 합리적인 판단능력이 떨어지고, 피해자의 경우에는 제대로 대응하기 어려운 상태에 있다는 점도 범죄가 발생하는 원인이 될 수 있다.

그러나 이러한 결과는 성범죄의 신고율이 매우 낮다는 점을 고려하면 성급하게 내리기 힘든 결론이다. 실제로 CCTV나 가로등이 없는 장소에서는 가해자 얼굴 인식, 증거확보 등의 어려움으로 인해 성범죄에 대한 보고가 안 될 가능성이 높다. 이러한 결과는 성범죄와 관련된 모든 범죄분석, 범죄예측, 그리고 범죄대응에 있어서 중요하게 고려해야 할 부분이다. 왜냐하면 인지된 범죄발생 수준에 따라서 CCTV와 가로등 설치의 우선순위가 정해지는데, 범죄발생 인지율이 높은 지역은 계속해서 설치를 많이 하고, 따라서 범죄에 대한 인지율이 높아질 가능성이 있다. 결과적으로 실제 범죄가 발생하고 있지만 인지율이 낮은 지역에서는 성범죄에 대한 인지율이 낮아서 환경 개선에 대한 투자도 적을 수 있다.

범죄예측의 다양한 기법들을 확인하고 검토하여 기본적으로 회귀분석을 통해서 범죄예측 위험도를 산출하는 것이 적합한 것으로 판단하였다. 또한 단순한 회귀식만으로는 정확한 범죄예측을 할 수 없다는 판단 하에 회귀식을 보정할

수 있는 다양한 방법도 제안하였다.

이를 위해 성범죄에 대한 범죄예측 산식을 도출하기 위한 여러 과정을 거쳤다. 기본적으로 범죄예측에 있어서 가장 중요한 요인으로 보이는 유동인구의 의미를 심층적으로 분석하였다. 상관관계 분석을 통해서 유동인구는 성범죄뿐만 아니라 다른 모든 범죄의 발생량을 설명할 수 있는 요인이라는 점이 확인되었고, 범죄의 발생량을 예측하는데 가장 핵심적인 요인으로 생각된다. 기존의 범죄위험도 예측에서 한 범죄 유형의 발생량을 예측하기 위해서 다른 범죄의 발생량을 예측 요인으로 사용하였는데, 사실 살인, 강도, 강간·강제추행, 절도, 폭력 등 5대 범죄만 하더라도 상호 간에 상관계수가 높게 나타난다고 하더라도 이들 범죄들 간의 관계는 기본적으로는 독립적이다. 어느 한 범죄가 다른 범죄의 원인이 되지는 않는다. 유동인구가 이 모든 범죄 유형들에 대한 공통의 원인이 된다고 보는 것이 더 정확할 것이다.

GeoPros에서는 치안블럭을 단위로 하는 범죄위험도지수의 경우에는 유동인구가 예측변인으로 사용되고 있지만 기초경계구역을 단위로 하는 범죄위험도의 경우에는 유동인구가 예측변인으로 사용되고 있지 않다는 점에서 범죄예측에서 유동인구의 의미에 대해서는 연구가 필요한 부분이다. 서울 지역 이외의 지역을 포함해서 전국적으로 유동인구의 효과에 대한 분석을 통해서 유동인구의 의미를 확인할 필요는 있지만, 다소 예측 방식이 복잡해지더라도 예측변인에 유동인구를 기본적으로 포함할 필요는 있을 것이다.

그러나 추가적인 분석들을 통해서 확인된 것처럼 유동인구를 비롯하여 특정 변인의 효과가 범죄발생에 선형적인 영향을 미치는 것이 아니라는 점을 고려하여 다른 변인들의 효과를 예측 산식에 반영할 수 있어야 할 것이다. 가령 본 연구에서 수행한 것처럼 범죄경력자의 수는 유동인구 수준에 따라서 범죄발생량에 미치는 효과가 달라진 것처럼 변인들의 효과는 다른 변인들과의 상호작용이 발생할 수 있다. 또한 지역별로 인구구성, 사회특성, 주거환경이 미치는 영향의 크기가 다르기 때문에 유사한 특성을 가진 지역들을 군집으로 분류하여 범죄예측을 할 경우에 보다 정교한 예측 산식을 얻을 수 있다.

또한 유형별 지역 분류는 범죄발생 원인에 대한 분석을 전제로 하기 때문에

유형별로 다른 범죄 예방 전략의 수립에 기여할 수 있다. 범죄예측 방식의 현실화를 위해서 범죄예측을 지역 등의 특성으로 구분하여 실제로 연구의 결과를 현실화해서 적용하게 되면 세부 지역 단위 검토, 시군구 이하의 치안 활동 단위를 구체화, 개별 지역 블록별 범죄 위험도 개발을 통해서 경찰관의 현장 활동에 효과적으로 적용될 수 있는 시스템을 개선하는데 기여할 것으로 기대된다.

또한 본 연구는 범죄위험도지수의 개발(개선) 및 고도화를 위한 기초연구로서 의미를 가진다. 이러한 연구가 가지는 의미는 실증적인 연구를 통해서 근거에 기반한 개선방안을 도출하는 것이다. 그동안 국내의 범죄예측은 빅데이터 전문 업체가 범죄발생 자료와 각종 인구·사회 지표들을 사용하여 주로 통계적인 기법에 의존하여 시스템을 구축하였다. 대부분 가용한 통계자료를 기초로 회귀분석 기법을 사용하고 있다.

그러나 범죄예측 산식의 도출 과정에서 범죄의 발생 원리에 관한 전문가가 거의 개입하지 않기 때문에 도출된 회귀식의 정확성에 대해서는 의문의 여지가 있다. 통계에 대한 전문가도 중요하지만 범죄학에 대한 전문가가 있어야 범죄예측 산식의 결정 과정에서 발생하는 다양한 문제들을 해결할 수가 있기 때문이다. 종속변인에 해당하는 범죄통계 자체의 한계, 즉 정확성에 대한 문제들을 정확하게 인식하고 있어야 데이터의 전처리 과정에서 부정확한 데이터들을 정확하게 걸러낼 수 있다.

그리고 더 중요한 것은 범죄예측 과정에서 예측변인을 발굴하고 최종적으로 적용하는 방식에 대해서 판단을 내리는데 중요한 역할을 할 수 있다. 우선 다양한 통계지표들 중에서 어떤 변인들이 예측변인으로 활용될 수 있는지에 대해서 범죄학의 이론에 근거하여 판단할 수 있다. 그리고 조절효과와 가중치, 변인의 조합 등 변인의 다양한 변용에 대해서 제안 및 판단을 내릴 수 있다.

마지막으로 가장 중요한 것은 범죄발생을 예측할 것으로 예상한 변인이 실제로는 효과가 없거나 예상과 다른 방향으로 효과를 보이는 경우에 이에 대한 해결방식을 찾는 것이다. 이런 경우에 통계 전문가는 데이터의 분석 결과에 의존하기 때문에 통계분석에서 관계가 확인되지 않는다면 그 원인이나 해결책을 찾

을 수 없다. 그러나 범죄학에 대한 전문가라면 그러한 상황에서 왜 일반적인 이론과 다른지, 그리고 범죄예측 산식 개발 과정에서 예상된 가설과 다른지에 대해서 그 이유를 탐색하고 수정할 수 있는 역량을 가질 수 있다.

그러나 앞서 언급된 연구의 결과와 의의에도 불구하고 본 연구는 여러 측면에서 한계를 가지고 있다.

첫째, 실증적인 연구의 주요 방법으로 현장조사 결과를 중요한 연구방법으로 적용하였으나, 의미 있는 결과를 충분히 얻지 못하였다. 범죄예측 연구가 어려운 가장 큰 이유 중 하나로 데이터 확보의 어려움을 꼽을 수 있는데, 범죄환경에 대한 평가를 위한 표준화된 방법의 확립이 향후에도 매우 중요하다는 점을 다시 확인하였다. 향후에는 범죄예측에 가장 직접적인 변인들을 중심으로 CPO가 직접 데이터를 수집하는 방법을 통해 데이터를 축적함으로써 범죄예측을 위한 고급 자료로 활용해 나갈 필요가 있다.

둘째, 데이터 분석의 결과를 시스템에 그대로 적용하기는 어렵다. 범죄예측 산식에서 무엇보다 중요한 것은 산식 도출을 위해 사용한 데이터이다. 범죄예측 산식은 사용한 데이터에 최적화되어 있으며, 데이터가 다를 경우에 산식도 부정확할 수 있다. 당연히 과거의 데이터를 가지고 현재나 미래를 예측하는데 있어서도 오차가 발생할 수 있다. 그러나 본 연구에서 사용한 데이터는 무엇보다도 예측의 단위가 된 지역의 범위가 행정동이기 때문에 치안블럭이나 기초경계구역을 대상으로 하는 범죄위험도 또는 범위위험도지수와는 차이가 있다. 또한 데이터의 날짜가 현행 시스템과는 달리 최신의 것이 아니라 최소 몇 년 이전의 것이기 때문에 한계가 있다. 그리고 본 연구에서 사용한 변인들이 실제 예측 시스템에서 활용 중인 변인들과는 차이가 있어 회귀계수 값들을 그대로 사용할 수 없다. 그러나 본 연구는 처음부터 이러한 한계를 고려한 상태에서 현행 시스템에 대해서 새로운 제안을 하기 위한 시범적인 노력이었던 만큼 최종적으로 도출된 분석결과나 회귀식보다도 그 취지를 잘 이해할 필요가 있다.

셋째, 데이터의 정확성, 데이터의 가용성 등 범죄예측 시스템 개발에 있어서 본질적인 문제들은 여전히 해결되지 않고 남아 있다. 특히 범죄통계 분석에 있어서의 한계의 문제로 제기된 성범죄의 신고율이 매우 낮은 점은 범죄예측에

있어서도 결정적인 한계가 될 수 있다. 통계를 통한 범죄의 분석과 대응에 있어서 중요한 문제로 인식할 필요가 있다. 경찰이 범죄예측을 정확히 하기 위해서는 발생 사건의 위치와 시간 등 기초정보를 포함하여 범죄와 관련된 다양한 정보를 정확하게 수집하고 관리해 나가야 할 것이다.

제2절 범죄위험도 개선을 위한 제언

본 연구에서 논의된 사항 이외에도 범죄예측 또는 범죄위험도지수의 개선을 위해서는 몇 가지 개선되어야 할 사항들이 있다.

첫째, 범죄 환경 요인에 따른 예측 산식의 도출 시 지역의 특수성을 충분히 고려하지 않는다는 점이다. 통계적인 방법에 의해 도출된 범죄위험도에서 각 예측 요인의 영향력은 개별 사례에서는 부적합한 경우가 발생한다. 이 때문에 개별 지역에서의 범죄위험도 산출이 정확하지 못한 경우가 많다.⁵²⁾ 예를 들어 범죄가 실제로 거의 발생하지 않는 공원 지역에 유동인구가 많고, 가로등이 적다는 이유로 범죄위험도가 높게 나타나는 경우가 있다. 이런 경우 거주인구 대비 유동인구의 비율이 비정상적으로 매우 높은 지역 특성을 고려하여 유동인구, 가로등과 같은 요인들의 가중치를 별도로 적용하는 등의 보정이 필요하다.

둘째, 자료 업데이트 및 단일 알고리즘 적용 문제이다. 가령 인구수, 소득과 같은 고정 자료와 범죄 발생, 카드 사용 빈도 등의 실시간 데이터가 동일시점에 업데이트 되는 경우, 그리고 재개발 혹은 주거단지 개발로 대규모 택지가 들어서는 경우, 일시에 다수의 환경 정보 변화가 발생하지만 인구 통계와 같은 주기적 업데이트 자료는 변경 없이 고정적으로 투입되기 때문에 실시간 범죄 환경을 반영하지 못한다는 단점이 있다. 더욱 심각한 점은 이런 경우 새로운 데이터에 부합하도록 범죄발생 예측 알고리즘을 변경하지 않고 동일한 알고리즘을 적용하여 범죄 발생을 예측하고 있다는 것이다.

52) 이는 빅데이터 연구가 가지고 있는 근본적인 한계이기도 하다. 따라서 범죄학 이론과 실증적인 검증을 통해서 예측요인과 범죄발생 간의 관계에 대한 확인이 지속적으로 요구된다.

이러한 문제는 향후 인공지능을 활용하여 투입되는 데이터의 성격이 바뀌는 경우에는 예측 산식을 자동으로 업데이트 하는 방식으로 개선해 나갈 필요가 있다. 기존 시스템 개발 과정에서 적용된 단순하고 일원화된 방식은 다수의 부정확한 결과를 산출할 가능성이 있다. 따라서 시스템의 고도화를 위해서는 실제 데이터를 기반으로 다양한 가능성을 검토할 필요가 있다.

셋째, 본 연구에서 논의가 되지 않는 않지만 범죄예측을 위한 기술적인 문제 이외에 추가로 고려해야 할 것은 데이터 수집에 걸림돌이 되는 법적인 제약이다. 특정 데이터 활용을 위한 기술적인 해결 방안을 고려하더라도 번번이 실패하는 이유가 범죄예측의 자료가 되는 관련 데이터 활용이 어렵기 때문이다. 예를 들어 범죄데이터의 경우 ‘형사사법절차전자화촉진법’의 제한을 받기 때문에 자료 활용에 한계가 있다. 그 이외에 활용 가능성이 높은 지역 자료들이 개인정보보호법 등의 제한으로 인해서 사용되지 못하는 경우가 많다. 이러한 문제는 범죄예측 연구의 발전을 위해서 기술적 접근방식 뿐만 아니라 관련 법률과 제도 연구도 병행되어야 함을 보여주는 부분이다.

제3절 범죄예측 기술에 대한 제언

예측을 포함, 범죄분석에 관한 오늘날의 추세는 보다 쉽고 직관적인 관측을 요구하고 있다. 그러나 범죄라는 현상은 매우 다양한 원인과 복잡한 과정을 통해서 발생한다. 이 때문에 빅데이터의 활용은 많은 왜곡이 발생할 수 있는 영역이기도 하다.

그동안 예측적 경찰활동은 언론과 연구 문헌에서 상당한 관심을 받았다. 그러나 이러한 기술에 관한 일종의 신화들이 확산되기도 했다.⁵³⁾ 이것은 부분적으로 비현실적인 기대에서 비롯된다. 예측적 경찰활동은 너무 과장되어 현실적이지 않다. 예를 들어 발전된 수학 및 계산 능력이 범죄를 줄이기 위해 필요하

53) Perry, W. L., (2013), Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations., RAND Corporation.

고 충분하다는 근본적인 잘못된 가정이 있다. 여기서는 예측적 경찰활동에 대한 가장 일반적인 신화 4가지를 알아보았다.⁵⁴⁾ 일반인뿐만 아니라 법집행기관이 가지기 쉬운 범죄예측에 관한 맹목적인 믿음은 전략의 운용에 지장을 줄 수 있으므로 범죄예측에 관한 정확한 인식을 가질 필요가 있다.

- 신화 1: 컴퓨터는 실제로 미래를 안다
- 신화 2: 컴퓨터가 모든 것을 처리한다
- 신화 3: 고사양(비싼) 모델이 필요하다
- 신화 4: 정확한 예측은 자동으로 주요범죄 감소로 이어진다

다음으로 범죄예측 결과가 법 집행에 유용하게 쓰이기 위해서는 종합적인 방법 전략의 일환으로 적용해야한다. 예측 방법이 범죄 감소에 기여하기 위해서는 범죄예측 기법이 가지고 있는 몇 가지 함정을 피해야 한다.⁵⁵⁾

- 함정 1 : 전술적 유용성보다 예측 정확도에 초점을 맞춘다
- 함정 2: 품질이 떨어지는 데이터에 의존한다
- 함정 3: 예측의 배경에 있는 요인을 오해한다

54) 예측적 치안에 대한 신화 유형과 구체적인 설명은 부록1에서 설명하였다.

55) 예측적 치안 활동을 위해 피해야할 함정의 유형과 구체적인 설명은 부록2에서 설명하였다.

- 함정 4: 측정과 평가를 과소평가한다

- 함정 5: 민사 및 개인 정보 보호의 권리를 간과한다

위와 같은 함정은 사실 범죄예측 기술이 제대로 작동하기 위해서 갖추어야 할 조건들을 소개한 것이기도 하다. 범죄예측 기술을 효율적으로 활용하기 위해서는 기술적으로 가능한 수준에서 분석 결과를 활용하는 유연한 접근 방식이 필요하다. 그러나 결과의 정확성이 보장되기 위해서 통계적인 관점에서의 엄격성을 무시해서는 안 된다. 결과의 기초가 되는 데이터의 품질을 확보하기 위한 노력, 상관관계와 인과관계를 구분하려는 노력, 현상의 측정과 평가를 정확하게 하려는 노력 등이 전제되어야만 정확한 결과를 얻을 수 있다. 그리고 마지막으로 범죄예측 기술의 활용 시 개인의 프라이버시를 보장하지 않는다면 기술 적용 과정에서 한계가 발생하게 된다. 프라이버시를 보장하면서도 범죄예측이 가능하도록 법제도에 관한 연구 역시 소홀히 해서는 안 될 것이다.

참 고 문 헌

I. 단행본

- Clarke, R. V., & Eck, J. E., (2005), Crime Analysis for Problem Solvers in 60 Small Steps, (역) 윤민우(2010), 문제해결을 위한 범죄 분석 60단계, 한국형사정책연구원.
- Hanson, R. K., & Thornton, D. (1999). Static 99: Improving actuarial risk assessments for sex offenders (Vol.2). Ottawa, Ontario: Solicitor General Canada.
- Howitt, D. (2002). Forensic and criminal psychology. Harlow: Pearson Education.
- Russell, D. E. H. (1984). Sexual exploitation-rape, child sexual abuse, and workplace harassment.

II. 논문

- 강선호, 양진경, 한별, 이동희 (2017). 국가기초구역을 이용한 범죄예측모델: 안양시 사례를 중심으로, 한국경영과학회지, 42(4), 135-146.
- 노성훈 (2015). 시공간 분석과 위험영역모델링을 활용한 범죄예측모형의 예측력 검증. 형사정책연구, 26(3), 239-266.
- 윤상연, 신상화 (2018). 범죄예측의 현재와 미래: 한국형 범죄예측 시스템 개선·도입에 대한 제언. 한국공안행정학회, 27(3), 11-38.

- 이수정, 김경옥. (2005). 성범죄 재범율에 관한 바른 이해와 재범 방지 방안 모색. *한국심리학회지: 사회 및 성격*, 19(3), 83-99.
- 주일엽 (2018). 경제지표의 범죄발생에 대한 선행관계. *경찰학논총*, 13(3), 9-37.
- 허선영, 김주영, 문태헌 (2017). 베이지안 확률 기반 범죄위험영역 예측 모델 개발, *한국지리정보학회지*, 20(4), 89-101.
- Bagley, C., & Shewchuk-Dann, D. (1991). Characteristics of 60 children and adolescents who have a history of sexual assault against others: evidence from controlled study. *Journal of Child and Youth Care, Special Issue*, 43-52.
- Becker, J. V., & Hunter, Jr. J. A. (1997). Understanding and treating child and adolescent sexual offenders. *Advances in Clinical Child Psychology*, 19, 177-197.
- Bushman, B. J., Wang, M. C., & Anderson, C. A. (2005). Is the Curve Relating Temperature to Aggression Linear or Curvilinear? Assaults and Temperature in Minneapolis Reexamined. *Journal of Personality and Social Psychology*, 89(1), 62-66.
- Carich, M. S., Newbauer, J. F., & Stone, M. H. (2001). Sexual offenders and contemporary treatment. *Journal of Individual Psychology*, 57(1), 3-17.
- Dietz, P. E., Hazelwood, M. S., Warren, D. S. W. (1990). The sexually sadistic criminal and his offences. *Bulletin of American Academy of Psychiatry and the Law*. 16, 163-178.
- Ford, M. E., & Linney, J. A. (1995). Comparative analysis of juvenile sexual offenders, violent nonsexual offenders, and status offenders. *Journal of Interpersonal Violence*, 10(1), 56-70.

- Grubin, D. (1999). Actuarial and clinical assessment of risk in sex offenders. *Journal of Interpersonal Violence*, 14(3), 331-344.
- Hall, G. K., & Barongan, C. (1997). Prevention of sexual aggression: sociocultural risk and protective factors. *American Psychologist*, 52(1), 5-14.
- Kavoussi, R. J., Kpalan, M., & Becker, J. V. (1988). Psychiatry diagnoses in adolescent sex offenders. *Journal of the American Academy of Child and Adolescent Psychiatry*, 27(2), 241-243.
- Koper, C. S. (1995). Just Enough Police Presence: Reducing Crime and Disorderly Behavior by Optimizing Patrol Time in Crime Hot Spots. *Justice Quarterly*, 12(4), 649-672.
- Malleson, N., & Andresen, M. A. (2015). Spatio-temporal crime hotspots and the ambient population. *Crime science*, 4(1), 1-10.
- Mohler, G. O., Short, M. B., Brantingham, P. J, Schoenberg, F. P., & Tita, G. E. (2011). Self-Exciting Point Process Modeling of Crime. *Journal of the American Statistical Association*, 106(493), 100-108.
- Prentky, R. A., Harris, B., Frizzell, K., & Righthand, S. (2000). An actuarial procedure for assessing risk with juvenile sex offenders. *Sexual Abuse: a Journal of Research and Treatment*, 12(2), 71-93.
- Ryan, G., Miyoshi, T. J., Metzner, J. L., Krugman, R. D., & Fryer, G. E. (1996). Trends in a national sample of sexually abusive youths. *Journal of the American Academy of Child & Adolescent Psychiatry*, 35(1), 17-25.
- Zou H., & Hastie, T. (2005). Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistics Society: Series B (Statistical Methodology)*, 67(2), 301-320.

III. 보고서

박준휘, 김한균, 유진, 한민경, 이성규, 최지선, 홍문기, 이예담, 강용길, 박재풍, 윤상연, 이춘삼, 이연지, 김도형, 조윤오, 박형아, 박준승, (2017), “국민 안전 보장을 위한 형사정책의 실효성 제고 방안 연구(II): 1인가구 밀집 지역의 안전실태와 개선방안 연구.” 한국형사정책연구원.

윤상연 (2017). 범죄환경 위험도지수 개발 연구 I. 치안정책연구소.

윤정숙, Marshall, W. L., Marshall, L. E., Knight, R. A., Sims-Knight, J. E., 이수정. (2012). 성범죄자를 위한 치료프로그램 개발 및 제도화방안 (I). 한국형사정책연구원 연구총서.

Perry, W. L., (2013), “Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations.”, RAND Corporation.

IV. 기타

중앙일보, “52억짜리 AI 수사관 ‘클루’가 ‘살인의 추억’ 재발 막는다”, 2017. 12. 08.

한국일보, “빅데이터로 범죄예측.. 한국관 ‘마이너리티 리포트’ 구축한다”, 2018. 09. 07.

Gorr, W., & Olligschlaeger, A. (2002). Crime Hot Spot Forecasting: Modeling and Comparative Evaluation, prepared for the U.S. Department of Justice, <https://www.ncjrs.gov/App/abstractDBDetails.aspx?id=195167>

부록 1. 예측적 치안의 4가지 신화⁵⁶⁾

■ 신화 1: 컴퓨터는 실제로 미래를 안다

예측적 경찰활동에 대한 설명은 마치 컴퓨터가 미래를 예측할 수 있는 것처럼 착각하게 만든다. 많은 뉴스가 예측적 경찰활동이 마치 미래를 들여다볼 수 있는 수정 구슬처럼 표현하지만, 이러한 알고리즘은 사건 발생 자체가 아니라 미래 사건의 위험을 예측하는 것에 불과하다. 컴퓨터는 도구로서 패턴 검색을 극적으로 단순화 수 있지만, 이러한 모든 기법은 과거와 다른 방식으로 미래를 예측한다. 또한 예측은 데이터를 생성하는 데 사용된 기본 데이터 수준만큼만 효과적이다.

■ 신화 2: 컴퓨터가 모든 것을 처리한다

예측적 경찰활동에 대한 만능 해결책으로 소프트웨어 패키지를 홍보하는 것이 일반적이지만 예측적 경찰활동 프로세스에서 인간은 여전히 가장 중요한 요소이다. 가장 완벽한 소프트웨어 제품군을 사용하더라도 인간이 직접 관련 데이터를 찾아 수집하고 데이터를 진처리하여 끊임없이 변화하는 범죄 상황에 대응하여 분석, 설계 및 분석을 수행하고 이러한 분석 결과를 검토 및 해석하고 오류를 제외해야 한다. 결과를 분석하고, 통합된 연구 결과를 분석하고, 연구 결과를 활용하고 그러한 영향을 평가하기 위한 조치를 취해야 한다.

■ 신화 3: 고사양(비싼) 모델이 필요하다

대부분의 경찰 부서는 예측적 경찰활동 프로그램을 시작하기 위해 가장 비

56) Perry, W. L., (2013), Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations., RAND Corporation. pp.xix-xx.

싼 소프트웨어 패키지나 컴퓨터가 필요하지 않다. 표준 작업 환경 소프트웨어(ex. Microsoft Office) 및 지리 정보 시스템(ex. ArcGIS)에 내장된 기능은 다양한 예측 방법을 지원할 수 있다. 일반적으로 모델의 복잡성과 예측력 사이에는 상관관계가 있지만 예측력은 가격이 증가하는 만큼 비례해서 증가하지 않는다. 단순한 경험적 연구(heuristics)는 분석 소프트웨어만큼이나 성능이 좋은 것으로 나타났다. 이러한 특성은 크고 정교한 모델을 지원하기에 불충분한 데이터를 보유한 소규모 부서에서 특히 중요하다.

■ **신화 4: 정확한 예측은 자동으로 주요범죄 감소로 이어진다**

분석과 소프트웨어에 중점을 두면 예측은 단지 예측일 뿐이라는 사실을 망각하게 만들 수 있다. 범죄의 감소는 그러한 예측을 바탕으로 실제 조치를 취해야 가능하다.

부록 2. 예측적 치안의 5가지 함정⁵⁷⁾

■ 함정 1 : 전술적 유용성보다 예측 정확도에 초점을 맞춘다

분석가가 강도에 대한 예측을 가능한 ‘정확하게’ 제공하도록 요청 받았다고 가정해 보자. 즉, 가능한 범죄 발생 수준을 분석하는 설계를 통해서 예측된 범죄 위험도가 높은 지역은 범죄 위험도가 높다고 진단한다. 이를 달성하기 위한 한 가지 방법은 전체 도시를 거대한 ‘위험 지역’으로 지정하는 것이다. 그러나 그러한 지정에는 거의 전술적인 효용성이 없다. 예측 가능한 핫스팟이 전술적으로 실행 가능할 정도로 작은 규모가 되기 위해서는 범죄 발생률 형태로 예측되는 결과의 정확도에 몇 가지 한계가 있다는 것을 받아들여야 한다.

■ 함정 2: 품질이 떨어지는 데이터에 의존한다

데이터 품질에 영향을 미칠 수 있는 세 가지 일반적인 결함으로는 데이터 검열(censoring), 체계적인 오류 및 높은 상호 관련성 등이 있다. 데이터 검열은 관심 있는 사건에 대해 특정 장소(그리고 특정 시간)의 데이터를 생략하는 것을 포함한다. 데이터가 검열되면 특정 지역에 범죄가 없는 것으로 나타나기도 한다. 체계적인 오류는 데이터가 수집되는 방식에서 비롯된다. 예를 들어 오전 7-8시 사이에 특별히 강도가 신고가 많은 경우 해당 시간 동안 실제로 많은 수의 강도 사건이 발생했는지, 아니면 집주인과 관리자가 야간에 발생한 강도 사건을 뒤늦게 발견하고 신고했는지 여부가 명확하지 않을 수 있다. 마지막으로 관련성은 데이터의 유용성과 관련이 있다. 일부 범죄 집단의 경우 몇 달 또는 몇 년 전의 사건 데이터를 재입력하는 것이 현재의 범죄예측에 매우 유용할 수 있다. 반대로, 같은 범죄자에

57) Perry, W. L., (2013), Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations., RAND Corporation. pp.xx-xxi.

의해 저질러질 만한 유사한 강도 사건이 발생하면 수 개월간의 데이터가 별로 쓸모가 없을 것이다.

■ **합정 3: 예측의 배경에 있는 요인을 오해한다**

범죄발생에 대한 관찰자, 특히 핫스팟을 만드는 실무자는 핫스팟의 어떤 요인이 위험을 유발하고 있는지 궁금해 할 수 있다. 그러나 일반적으로 예측 도구는 특정 영역에 존재하는 위험 요소를 강조하는 것이 불가능하지는 않더라도 어렵도록 설계되었다. 다만 회귀 또는 데이터 마이닝 변형과 같은 기법을 적용 할 때 상식을 사용하여 모델에 통합 된 요소를 확인하면 가짜 관계를 피할 수 있다.

■ **합정 4: 측정과 평가를 과소평가한다**

실무자들 중 극소수의 사람들은 자신이 내놓은 예측의 효과성이나 예상치에 따라 개발 된 중재를 평가했다고 말했다. 분석 및 중재의 효율성은 데이터를 최신 상태로 유지하기 위한 전체 노력의 일환으로 평가되어야 한다. 측정은 개량을 위한 영역을 식별하고 개입을 수정하며 자원을 분산시키는 데 중요하다.

■ **합정 5: 민사 및 개인 정보 보호의 권리를 간과한다**

더 많은 법 집행에 관심을 기울일만한 가치가 있는 지역 및 사람들에게 본질적으로 시민의 자유와 개인 정보 보호 권리에 대한 우려가 제기되고 있다. '위험에 처한' 것으로 영역을 표시하는 것은 개인이 직접 대상으로 지정되지 않기 때문에 문제가 더 적게 나타난다. 미 연방 대법원은 합법적인 의심을 구성하는 기준이 우범지역에서 완화되고 있다고 판결했다. 그러나 공식적으로 '범죄율이 높은' 지역을 구성하는 것은 무엇이며, 완화된 합리적 의심 원칙에 따라 그러한 영역에서 취할 수 있는 조치가 무엇인지는 여전히 숙제로 남아있다.

책임과제연구보고서 2018-10

범죄환경 위험도 지수 개발 연구 II -성범죄 발생의 환경요인 분석-

발 행 : 2019년 5월

인 쇄 : 2019년 5월



발 행 인 : 치안정책연구소장

발 행 처 : 치안정책연구소

충남 아산시 신창면 황산길 100-50

홈페이지 : www.psi.go.kr

이 책의 무단 복제를 금합니다.

이 책자에 게재된 내용은 연구자 개인의 의견이며
치안정책연구소 공식 견해가 아님을 밝혀둡니다.