

GWZINBR의 적용: 서울시 불법주정차를 사례로

이용채 (경희대학교 지리학과 석사과정)



경희대학교 일반대학원 총학생회
KYUNG HEE UNIVERSITY GRADUATE STUDENTS' ASSOCIATION

I. 연구 개요

- 일반적으로 공간적 의존성을 갖는 데이터로 회귀분석을 할 경우, 일반선형회귀보다 공간회귀분석 모형을 적용하는 것이 더 좋은 설명력을 나타내는 것으로 알려져 있음.
- 불법주정차 또한 공간현상의 하나로, 공간적 의존성을 나타내기 때문에 이를 제어할 수 있는 분석 모형을 쓰는 것이 더 적합할 수 있음.

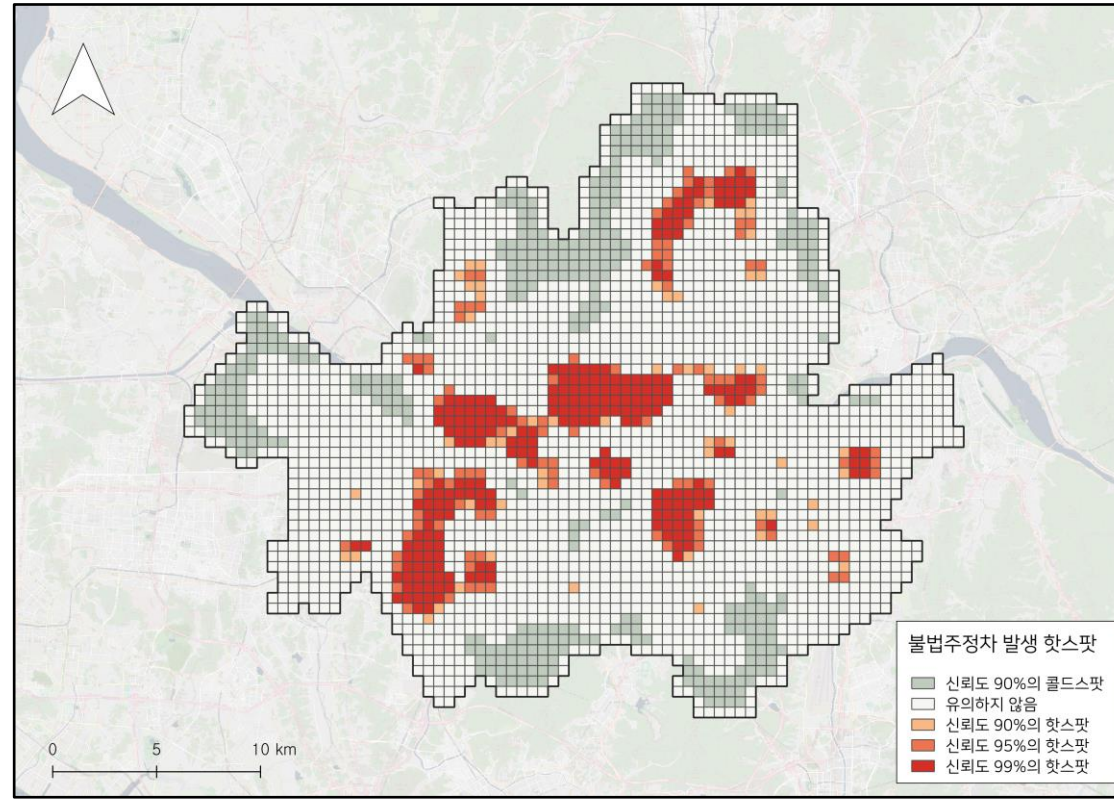


그림 1. 불법주정차 발생 핫스팟

- 하지만 불법주정차는 도로 위에서만 발생할 수 있다는 특이점을 가지고 있기 때문에 행정동보다 작은 격자, 집계구 등 소규모 공간단위로 집계할 경우, 영과잉(zero-inflated)이 발생할 우려가 크고, 이는 일반적인 공간 회귀분석 적용의 어려움과 부정확한 결과로 이어질 수 있음.
- 따라서 본 연구에서는 영과잉 공간현상을 위한 회귀 분석 모형인 GWZINBR을 적용하여 불법주정차 발생 영향요인을 분석하고자 함.

II. 데이터 및 변수

- 본 연구의 연구 지역은 서울시이며 행정동과 500M 격자로 불법주정차 발생 건수를 집계했을 때, 두 단위 모두에서 유의한 양의 공간상관관계가 나타남.
- 행정동 단위에서는 발생 건수가 0인 지역이 존재하지 않았지만, 격자단위에서는 약 28% 지역의 값이 0인 영과잉이 발생했고, 전체적으로는 좌측 편포의 분포를 나타냄.

공간단위	Global Moran's I	0인 지역의 개수
행정동 (424개)	0.257956***	없음
500M 격자 (2634개)	0.499845***	738개 (전체의 약 28%)

표 1. 공간단위 별 Moran's I 통계량과 영과잉 발생 여부

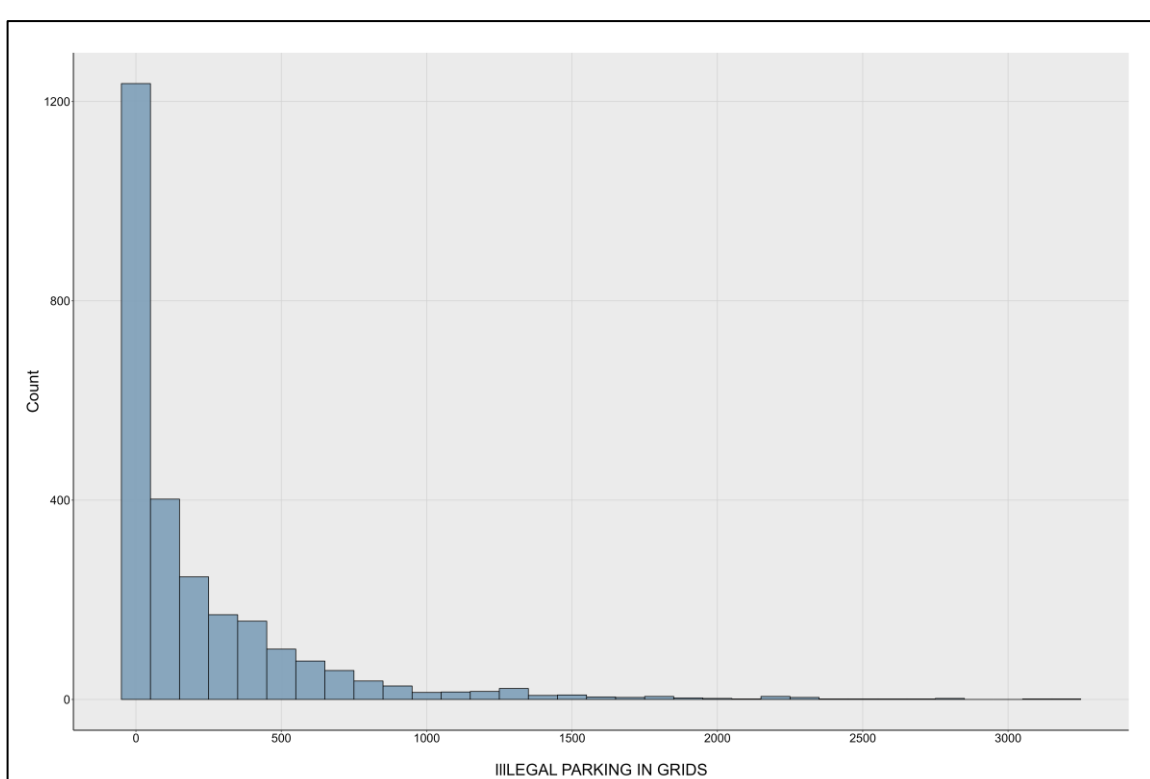


그림 2. 격자단위 불법주정차 발생 히스토그램

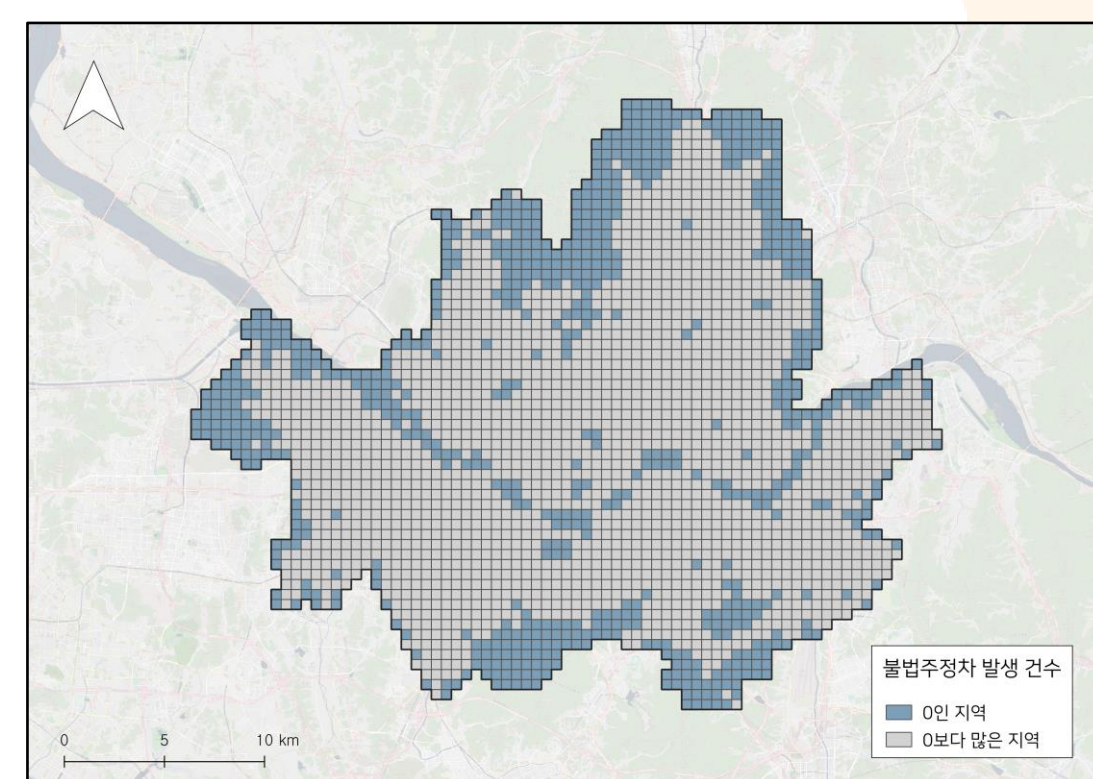


그림 3. 격자단위 발생 건수 0인 지역의 분포

- 탐색 결과 공간적 의존성과 영과잉의 제어가 필요하다고 판단하여 격자 단위를 대상으로 지리 가중 영과잉 음이향 회귀모형(GWZINBR)을 적용하여 분석함.

연구 변수	구분	자료형	데이터 출처
생활인구	독립 변수	가산형	서울열린데이터광장
CCTV 개수		가산형	국가공간정보포털
음식점 개수 (휴게음식점, 일반음식점)		가산형	공공데이터포털
도로 연장		연속형	Open Street Map
2023년 불법주정차 단속 데이터	종속변수	가산형	서울열린데이터광장

표 2. 연구 변수 및 데이터 출처

* 여기서 생활인구는 행정동과 집계구 단위로만 제공되기 때문에 다음과 같은 방법으로 재집계하였음.
- 건물이 존재하는 격자만 추출 → 해당 격자와 중첩되는 집계구의 면적 / 해당 집계구의 전체 면적으로 비율 계산 → 해당 집계구 생활인구 * 비율로 산출)

IV. 연구 방법

- 지리 가중 영과잉 음이향 회귀모형(GWZINBR)은 영과잉 음이향 분포를 나타내는 데이터에 지리 가중 회귀 모형을 적용할 수 있도록 고안된 분석 방법으로, 본 연구에서는 영과잉을 고려하지 않은 일반 지리 가중 음이향 회귀 모형(GWNBR)과 결과를 비교하였음.
- 결합된 모형에 대한 설명은 다음과 같음.
 - ✓ 지리 가중 회귀(GWR): 일반 회귀 모형의 가정을 만족하기 어려운 공간현상의 의존성을 제어하기 위한 회귀 모형. 대역폭을 사용하여 인접성에 따른 가중치를 부여하고, 지역마다 다른 회귀계수를 산출함 (조동기, 2009).
 - ✓ 영과잉 음이향 회귀(ZINBR): 특정 현상을 경험할 가능성이 극단적으로 낮거나 전혀 없는 표본을 분석할 수 있도록 고안된 모형. 가산형 데이터에서 0이 차지하는 비율이 지나치게 크고, 과분산이 발생했을 때 사용 (이용일, 2018).

V. 분석 결과

Goodness-of-fit measures	GWZINBR	GWNBR
Bandwidth	Fixed bandwidth (10km)	
Log likelihood	-270.2257	-270.2257
R^2	0.5639	0.4521
R^2 Adj	0.3457	0.2276
AIC	571.0741	566.1887
AICc	572.2772	566.3249

표 3. 모형 적합도

- 분석 결과, 각 모형의 적합도를 나타내는 지표는 [표 3]과 같음.
 - ✓ 10km의 고정 대역폭을 사용하여 분석하였음. (AICc 값을 최소화하는 대역폭의 근사값)
 - ✓ 두 모형의 Log likelihood는 같은 값을 나타내지만, GWZINBR 모형이 더 큰 수정 R^2 값을 나타내기 때문에, 모형 적합도가 더 높다고 할 수 있음.

VI. 결론

- 본 연구에서는 공간단위에 따라 불법주정차 발생 데이터에 영과잉이 발생한다는 것을 확인하고, 이를 제어할 수 있는 지리 가중 영과잉 음이향 회귀 모형을 적용하여 결과를 비교하였음.
- 분석 결과 일반적인 음이향 지리 가중 회귀 모형에 비해 영과잉을 고려했을 때 모형의 적합도가 더 높다는 사실을 확인함.
- 본 연구의 한계점과 추후 보완할 사항은 다음과 같음.
 - ✓ 모형의 적합도는 확인하였지만, 지역 별 회귀계수의 유의성은 확인하지 못함.
 - ✓ 다양한 크기의 대역폭으로 추가 분석을 수행하여 모형의 과적합 여부를 파악하고 방지할 필요가 있음.

참고문헌

- 조동기. (2009). 지역 단위 조사연구와 공간정보의 활용: 지리정보시스템과 지리적 가중 회귀분석을 중심으로. 조사연구, 10(3), 1-19.
- 이용일. (2018). 항공사 이용객의 수요함수 추정을 위한 포아송모형, 음이향 회귀모형, 0 과잉 포아송모형, 0 과잉 음이향모형의 비교 분석 연구: 국내선 FSC 중심. 관광연구저널, 32(1), 113-123.
- da Silva, A. R., & de Sousa, M. D. R. (2023). Geographically weighted zero-inflated negative binomial regression: A general case for count data. Spatial Statistics, 58, 100790.

제4회 일반대학원 총학생회 공동학술 세미나 발표집



경희대학교 일반대학원 총학생회
KYUNG HEE UNIVERSITY GRADUATE STUDENTS' ASSOCIATION



경희대학교 대학원 혁신기획팀
KYUNG HEE UNIVERSITY
GRADUATE SCHOOL INNOVATION PLANNING TEAM



경희대학교 대학원
GRADUATE SCHOOL
KYUNG HEE UNIVERSITY



BrainKorea21
FOUR