

점사상의 지역단위 집계가 K-지표에 미치는 영향

An Effect of Aggregation of Point Features to Areal Units on K-Index

이병길¹⁾

Lee, Byoung-Kil

Abstract

Recently, data gathering and algorithm developing are in progress for the GIS application using point feature. Several researches prove that verification of the spatial clustering and evaluation of inter-dependencies between event and control are possible. On the other hand, most of the point features as GIS data are gathered by indirect method, such as address geo-coding, rather than by direct method, such as field surveying. Furthermore, lots of statistics by administrative district based on the point features have no coordinates information of the points. In this study, calculating the K-index in GIS environment, to evaluate the effect of aggregation of raw data on K-index, K-indices estimated from raw data (parcel unit), topographically aggregated data (block unit), administratively aggregated data (administrative district unit) are compared and evaluated. As a result, point feature, highly clustered in local area, is largely distorted when aggregated administratively. But, the K-indices of topographically aggregated data is very similar to the K-indices of raw data.

Keywords : Point pattern, Areal Unit, K-index, D-index

초 록

최근 점사상을 활용하는 GIS 분야에서 많은 양의 점사상 축적과 함께 점분포 패턴을 정량적으로 평가하기 위한 알고리즘의 개발이 이루어지고 있다. 여러 연구에서 K-지표를 활용하여 점사상의 공간적 밀집 여부의 검증이 가능하며, 사건과 배경의 상호 관련성 평가가 가능함을 증명하고 있다. 한편 GIS 데이터로서의 점사상은 측량에 의해 실좌표가 관측된 사상보다는 주소와 같은 위치참조에 의해 간접적으로 좌표가 주어지는 경우가 많으며, 경우에 따라서는 통계자료와 같이 행정구역과 같은 지역단위의 집계자료로 대표되어 점사상 각각이 좌표를 가지지 못하는 경우도 많다. 본 연구에서는 GIS를 이용한 공간 분석 기법으로서 K-지표를 계산할 때, 집계자료의 사용이 K-지표의 산출에 미치는 영향을 평가하기 위하여, 원데이터(지번단위), 지형적인 집계(블록 단위), 행정적인 집계(행정구역 단위) 등 세 가지 형태의 데이터로부터 산출된 K-지표를 비교, 분석하였다. 연구결과 가까운 거리에서 밀집이 심하게 일어나는 점사상의 경우에는 행정구역과 같은 큰 지역단위를 이용하면 결과의 왜곡이 심하게 발생하여 활용이 곤란하나, 블록단위의 K-지표는 원데이터의 K-지표와 거의 유사함을 알 수 있었다.

핵심어 : 점 패턴, 지역단위, K-지표, D-지표

1. 서 론

조경, 생태와 같이 전통적인 GIS의 응용분야 외에도 보건, 환경, 비즈니스 분야와 같이 GIS가 최근에 적용되고 있는 분야에서는 기본도 정보에 부가된 다양한 내용의 점사상 정보가 폭발적으로 증가하면서 이를 이용한 다양한 분석 알고

리즘이 연구되고, 새로운 응용분야가 생겨나고 있다. 최근 비즈니스 GIS에서 활용되는 정보에는 고객정보, 매출정보, 업종별 POI 등의 실시간으로 변동되는 동적정보가 포함되며, 이미 상당수의 업체에서는 이러한 정보를 축적하고 있다. 보건환경분야에서는 전염병 발병, 환경에 영향을 받는 암 발병과 같은 현상의 공간적 분포와 환경과의 상관관계를

1) 정회원 · 환경대학교 토목공학과 겸임교수(E-mail:basillee68@daum.net)

구명하기 위한 공간통계의 적용이 활발히 연구되고 있다.

점사상의 분포를 통계적으로 검증하는 목적은 크게 세가지로, 첫째 점사상이 밀집하여 분포하고 있는지에 대한 평가, 둘째 밀집이 환경적 요인에 의해 발생하는 것인지에 대한 평가와 마지막으로 밀집이 시계열적인 연관성을 가지고 있는가에 대한 평가이다. 이 세가지 목적 중 첫 번째의 목적은 어느 한 시점의 데이터를 활용하면 용이하게 달성될 수 있으나, 두 번째와 세 번째의 목적은 비교해야 할 대상이 되는 정보도 점사상으로 존재해야 한다는 전제가 충족되어야 한다. 그러나 실제 모든 응용분야에서 직접 관리하고 활용하는 정보는 점사상으로 확보 되지만, 그렇지 않은 정보로서 참조하고 비교하는 대부분의 정보는 통계정보이거나 별도의 조사를 통해 구축하여야 한다. 예를 들어 소아암환자의 분포와 14세 이하 소아인구의 분포를 비교해야 할 경우, 소아암환자의 발병 정보는 점사상으로 획득 가능하지만 소아인구의 정보는 점사상으로 획득하는 것이 현실적으로 불가능하다. 또한, 편의점과 같은 소매체인 업체의 분포와 슈퍼마켓을 포함한 전체 소매점의 분포를 비교해야 할 경우, 소매체인 본사는 가맹 프랜차이즈에 대한 정보는 점사상으로 획득 가능하지만, 전체 소매점의 정보를 점사상으로 획득하기 위해서는 상당한 시간과 비용을 투입해야 하기 때문에 지역단위의 통계자료를 사용하게 된다. 이처럼 사건(event)인 점사상에 비해 획득이 어려운 환경(environment or control)이 되는 점사상은 일정 지역단위로 묶여진 집계 또는 통계자료를 사용하게 되나 이 경우에는 공간적 집계의 편의(spatial aggregation bias)가 발생하게 된다.

한편, 대부분의 경우 점사상 정보는 측량에 의해 결정된 절대좌표를 가지고 있기 보다는 주소 정보를 통해 부

여된 좌표를 가지게 된다. 즉, 주소와 지번도의 매칭을 통하여 좌표를 부여하게 되며, 부여된 좌표를 점사상으로 변환하여 GIS에서 활용하게 된다. 주소를 이용하게 될 경우 주소정보와 지번도의 완결성이 부족하기 때문에 모든 정보가 점사상으로 변환되지 못하고 정보의 누락이 발생하게 된다. 주소정보의 완결성 부족을 보완하기 위해 주소를 행정구역이나 우편번호에 매칭시킬 수 있으나 이 경우에도 역시 공간적 집계의 편이가 발생하게 된다 (Feser and Sweeney, 2000).

본 연구에서는 실제 점사상으로 부터 계산한 K-지표(D-지표)와 점사상을 도로에 의해 형성된 블록 단위로 집계하여 계산한 K-지표(D-지표), 그리고 점사상을 행정구역 단위로 집계하여 계산한 K-지표(D-지표) 등 세가지 값을 서로 비교, 평가하여 점사상을 지역단위로 집계하는 것이 K-지표에 어떠한 영향을 주는지를 알아보고자 하였다.

2. 연구 동향

점사상은 하나의 XY 좌표로 위치가 정의되는 지리적 사상(유근배, 1998)으로 실제 공간상의 면적을 점유하고 있는 건물, 사업체 등을 주소를 이용하여 위치를 지정하는 경우, 또는 교통사고, 전염병 발생과 같은 사건을 지도 공간 상에 표시하는 경우에 점사상을 사용하게 된다. 점사상은 동질적인 여러 점사상과 함께 특정한 분포패턴을 나타내게 된다. 분포패턴에는 임의적(random) 분포, 규칙적(regular) 분포, 군집(clustered) 분포 등의 세가지가 있는 것으로 알려져 있으며, 이들 각각은 포아송 분포, 이항 정규분포, 부적이항분포와 같은 특정한 확률분포에 의해

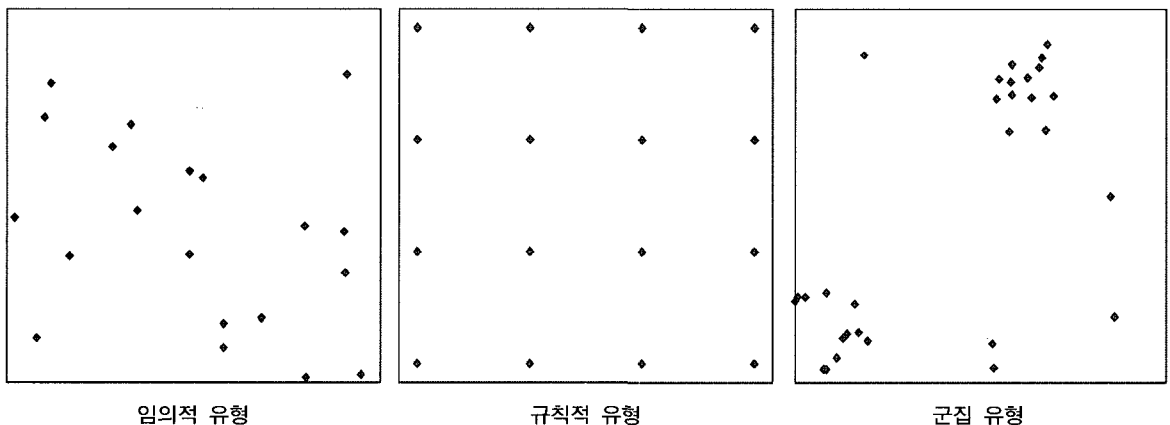


그림 1. 점 패턴의 유형

형성된 분포유형이다(이희연, 1989)(그림 1 참조). 일반적으로 근집 분포와 규칙적 분포는 시각적으로도 판단이 용이하나, 임의적 분포와 근집 분포는 시각적으로 판단하기 어려운 경우가 많다.

점 패턴 분석은 1950년대에서 60년대에 이르는 기간동안 생태학자들의 분석 기법을 지리학자들이 체계화하면서 지리학의 한 분야로 자리 잡게 되었다(Gatrell 등, 1996). 점 패턴을 분석하는 방법은 크게 2가지 방식으로 분류할 수 있다(Haggett 등, 1977). 그 중 하나는 거리기반(distance-based) 기법으로 최근린 분석법과 같이 점간의 거리정보를 사용하는 방법이고, 다른 하나는 영역기반(area-based) 기법으로 방격법과 같이 규칙적으로 배열된 격자내에서 관측되는 점의 빈도를 이용하는 방법이다. 최근에는 점 분포의 형성과정까지를 설명할 수 있는 여러 가지 통계적 분석기법이 연구되고 있다.

GIS의 도입 이전에 점 패턴 분석이 활발히 사용되지 못한 이유는 대략 두가지 정도로 해석할 수 있다. 먼저 들 수 있는 (아마도 더욱 중요할) 원인은 귀무가설이 현실세계에 적용하기 곤란했었다는 것이다. 특히, 초기의 분석 기법들은 공간적 임의성을 검증하지 못했다. 다른 원인 하나는 자료를 처리하고 분석하기 위한 좋은 소프트웨어가 없었다는 사실이다(Gatrell 등, 1996). 최근린 분석과 방격법을 사용할 수 있는 몇몇 소프트웨어에서 조차 지도와 같은 형식의 출력은 기대할 수 없었으며, 단순히 요약된 통계수치 만이 제시될 뿐이었다. 최근 GIS의 도입과 함께 다양한 공간상의 점 패턴으로 취급될 수 있는 많은 공간정보와 GIS의 강력한 분석기능이 결합되어 많은 연구가 진행되고 현장에 적용되고 있다(Gatrell 등, 1996).

점 패턴 분석이 응용되고 있는 분야는 촌락의 분포와 같이 소축척에서 이루어지는 분석에서부터, 질병(전염병) 분석, 범죄분석 및 상점 분포에 대한 분석과 같이 대축척에서 이루어지는 분석에 이르기까지 다양한 영역이 포함되어 있다(이희연, 1989 ; Gatrell 등, 1996 ; 이미혜, 2004 ; 전용완, 2002 ; Feser and Sweeney, 2000).

한편, Feser and Sweeney(2000)은 zip code에 의한 집계정보는 전체 조사자료를 거의 누락없이 사용할 수 있기 때문에 장점이 있으나, 짧은 거리에서는 K-지표(D-지표)에 대해 상당한 영향을 줄 수 있으며, 일반적으로 가까운 거리(0-5km)에서 D-지표가 과대하게 나오는 경향이 있음을 보고하고 있다. 이에 더해 Feser and Sweeney(2000)은 공간적 밀집 경향을 분석하고자 하는 관심이 주로 짙

은 거리에 맞춰져 있기 때문에 D-지표의 사용 시, 대도시 권역 이하의 지역에서는 지번단위의 데이터를 사용할 때 근거리에서의 밀집 패턴이 반영될 수 있고, 원거리 대상에서는 zip code 중심점의 사용이 합리적이라고 하였다.

3. 점 패턴 분석기법

점 패턴이란 특정 대상공간 R 에서 사건이 발생한 일련의 위치 (s_1, s_2, \dots) 로 정의되며, s_i 는 벡터 형식의 표현으로 공간상에서는 i 번째 관측된 사건의 좌표를 의미한다. '사건'이란 R 내에서 관측된 위치를 임의의 위치로부터 구별하기 위해서 사용되는 점 형성과정(point process)에서 사용되는 표준적인 용어이다(Diggle, 1983). R 은 사각형이거나 다각형일 수 있으며, 어느 경우이건 경계선 효과가 발생하므로 이를 보완하기 위해 완충지역을 사용하거나 분석 시에 경계의 형상을 고려해야 한다.

공간상의 점 패턴에 대한 가장 단순한 통계적 모델은 CSR(Complete Spatial Randomness)이다. CSR은 사건이 대상지역 R 내에서 동일한 확률분포에 따라 독립적으로 분포되었다는 것을 의미한다. 점 패턴이 근집 분포 또는 규칙적 분포의 특성을 명확하게 나타내는 경우에는 시각적으로도 판별이 가능하다. 따라서, 임의적 분포인지 근집 분포 또는 규칙적 분포인지 시각적으로 판단되지 않을 때, 점 패턴이 CSR이라는 귀무가설을 통계적으로 검증한다. 그런데, 독립적으로 놓고 보면 밀집이 발생하여 있는 사상이 그 배경이나 환경을 고려할 때, 전혀 밀집이 발생하지 않는 경우나, 그 반대의 경우가 발생한다. 예를 들어, 주점이 특정 지역에 밀집해 있는 것으로 보이지만, 사무실, 주거지, 유동인구와 같은 배경의 밀집과 비교할 때 특별한 밀집이 발생하지 않는 경우가 있다.

점 패턴의 밀집 여부와 그 특성을 파악하기 위해서는 1차특성과 2차특성을 분석해야 한다. 1차특성이란 공간상에서 발생한 사상의 기댓값(평균)에 대한 기술(description)이고, 2차특성이란 공간상의 서로 다른 영역에서 발생한 값 간의 공분산(또는 상관관계)에 대한 기술이다. 공간상에서 관측된 데이터의 패턴을 이해하기 위해서는, 전역적인 추세(1차특성의 변동) 또는 상관관계 구조(2차특성의 변동) 또는 두가지의 혼합에 의해 패턴이 만들어 진다는 것을 알아야 한다.

1차특성은 사상 s 를 중심으로 한 단위면적에서 발생한

사건의 평균인 사상의 밀도($\lambda(s)$)로 표시되며, 식 (1)과 같다.

$$\lambda(s) = \lim_{ds \rightarrow 0} \frac{E(Y(ds))}{ds} \quad (1)$$

여기서, ds 는 사상 s 주위의 미소면적, $Y(ds)$ 는 ds 에서 관측된 사건의 수, $E()$ 는 기댓값이다.

1차특성 분석을 하는 방법으로 방격법이 많이 사용되어 왔으나, GIS 도입 이후, 이동 kernel을 사용한 분석기법(예를 들어, DENCLUE(이병길, 2003))이 널리 사용되고 있다. 이동 kernel을 이용한 분석에서 결과적으로 얻어지는 산출물은 라스터 형태의 데이터로 방격법과 유사하지만, 방격법과 비교할 때 격자의 크기는 최소한 작게 하고, 여러 크기의 윈도우(검색 영역)를 사용한 분석결과를 비교할 수 있는 장점이 있다.

2차특성을 평가하는 여러 가지 방법 중 대표적인 것은 K-지표이다. K-지표는 특정지점으로부터 일정거리 내에 실제로 존재하는 점의 수와, 이론적으로 기대되는 점의 수를 비교하여, 점의 분포가 임의적인지 아닌지를 판정한다.

$$\hat{K}(h) = \frac{R}{n^2} \sum \sum \frac{I_h(d_{ij})}{w_{ij}} = \frac{1}{\lambda^2 R} \sum \sum \frac{I_h(d_{ij})}{w_{ij}} \quad (2)$$

여기서, R 은 대상지역의 면적, n 은 사건의 수, $\lambda (= \frac{n}{R})$ 는 사건의 밀도, d_{ij} 는 사건 s_i 와 s_j 간의 거리, $I_h()$ 는 indicator 함수($d_{ij} < h$ 이면 1, 그렇지 않으면 0)이고, w_{ij} 는 경계선 효과를 제거하기 위한 가중치이다.

임의적 패턴일 경우 모든 위치에 점이 있을 확률은 동일하며 서로 독립적이다. 따라서 어떤 지점으로부터 일정한 거리 h 이내에 발견될 것으로 기대되는 평균적인 점의 수는 $\lambda\pi h^2$ 이다. 즉, 공간적 상호작용이 존재하지 않는 등질적 상황(isotropic)에서 $K(h) = \pi h^2$ 이며, 군집 분포일 경우 $K(h) > \pi h^2$ 이고, 규칙적 분포일 경우 $K(h) < \pi h^2$ 로 나타난다.

이 때, K함수는 상수인 π 와 반경 h 의 영향만을 받기

때문에 두개의 서로 다른 사상의 패턴이 각각 임의적이라면 식 (3)이 성립한다.

$$K_{11}(h) = K_{22}(h) = K_{12}(h) \quad (3)$$

여기서, $K_{11}()$ 은 분석 대상 사상의 K함수, $K_{22}()$ 은 배경 또는 환경이 되는 사상의 K함수이고, $K_{12}()$ 은 대상과 배경 간의 K함수이다.

식 (3)으로부터 다음의 식을 유도할 수 있다.

$$\hat{D}(h) = \hat{K}_{11}(h) - \hat{K}_{22}(h) \quad (4)$$

여기서, $\hat{D}()$ 는 대상과 배경과의 차이로 양의 값은 배경에 비해 더욱 밀집되어 있는 패턴을 음의 값은 배경에 비해 규칙적인 패턴을 나타내는 것이다.

$\hat{K}()$ 와 $\hat{D}()$ 가 해당 거리 축척에서 업종간의 밀집정도를 비교할 수 있는 지표로 사용할 수 있는데 비해, $\hat{D}()$ 의 값을 거리(h)로 나누어 거리에 대해 표준화하면, 하나의 업종 내에서 거리에 따른 밀집의 정도를 상대비교하는 지표로 사용될 수 있다(Feser and Sweeney, 2000).

4. 실험 적용 및 평가

4.1 대상 데이터 및 사용 소프트웨어

본 연구에서 분석을 위해 사용된 데이터는 2003년 2월 기준 전화번호부의 서초, 강남 업종편 정보로부터 추출한 3개 업종 2,927개 상호(표 1 참조)이며, 데이터의 처리에는 MS Access와 범용 GIS 소프트웨어인 ArcGIS8.3과 ArcInfo8.1.2를 이용하였고, 소프트웨어에서 지원되지 않는 주소정제 및 주소매칭은 Visual Basic을 이용하여 프로그래밍하였다.

업종은 약간의 밀집을 보이는 일용품 업종으로 한식당을, 밀집이 심한 선매품 업종으로 변호사 사무실을, 밀집하지 않고 공간적으로 적절히 분포하는 동사무소를 선정하였다. 선정된 업종은 주소정제 및 주소 매칭을 통해 약 95.8%의 상호가 좌표를 부여받았다. 또한, CSR 여부의

표 1. 사용 데이터

구분	일용품	선매품	기타	계	CSR점
업종	한식당	변호사 사무실	동사무소		
상호수	1,761	1,251	44	3,056	
좌표부여된 상호수	1,692	1,191	44	2,927	1,890

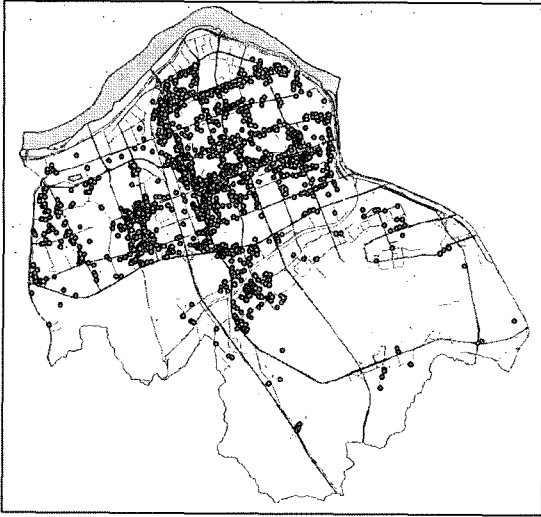


그림 2. 한식당의 분포도

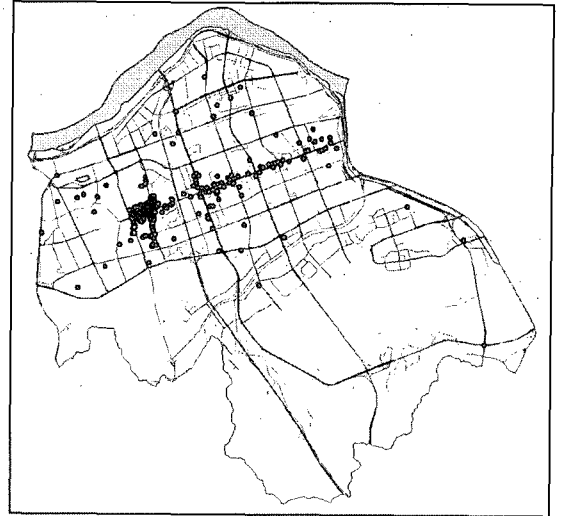


그림 3. 변호사 사무실의 분포도

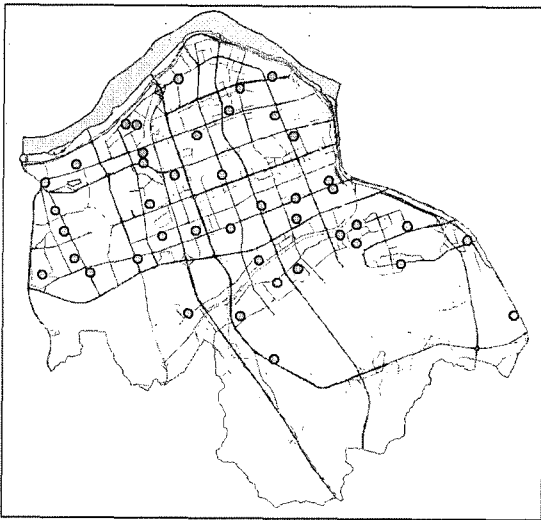


그림 4. 동사무소의 분포

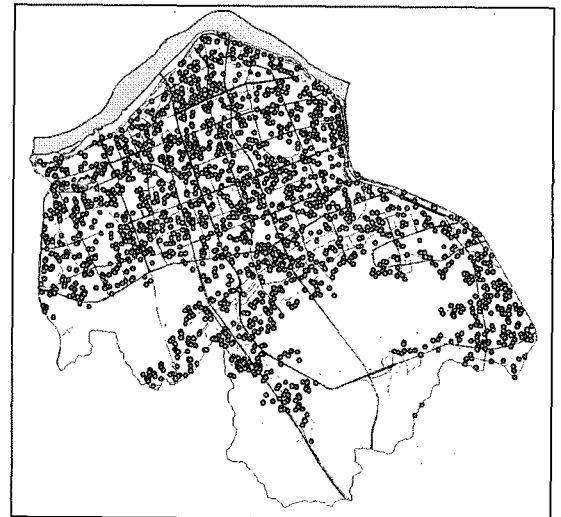


그림 5. CSR점(임의 생성 좌표)의 분포

평가를 위해서 대상지역 내에서 녹지, 도로, 하천을 제외한 지역에서 점사상 1,890개를 임의의 위치에 생성하였다(표 1 참조).

그림 2~5는 각각 본 연구에 사용된 한식당, 변호사 사무실, 동사무소와 CSR점의 분포도이다.

4.2 점사상의 집계

점사상이 분포하는 특성을 알아보기 위하여 우편번호를 이용하는 방법을 고려해 보았으나, 우편번호를 기준으로 한 집계구역은 대상지역을 배타적으로 분할하지 못하

고 중복되기 때문에 점사상 분포 특성을 알아보는 데는 부적합하였다. 따라서, 향후 새주소 사업에 따라 활용이 가능할 것으로 보이는 블록과 현행 행정구역을 이용하였다. 블록은 국토지리정보원의 도로중심선 레이어와 연구대상지역인 서초, 강남구의 경계를 이용하여 생성하였다. 연구대상 지역 내에 블록은 총 5,039개, 행정구역은 총 44개였다(그림 6, 7 참조).

블록과 행정동 레이어에 점사상을 중첩하여 공간 조인(spatial join)하여 각 폴리곤에 포함된 점사상의 수를 집계하였다(표 2 참조). 그림 8, 9는 블록레이어에 한식당을

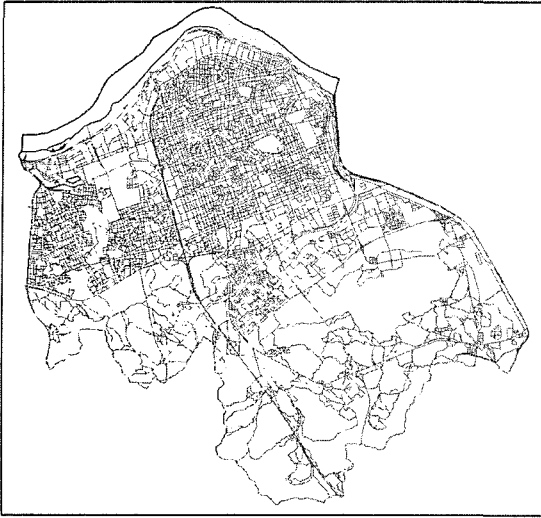


그림 6. 지형에 의한 분할(블록)

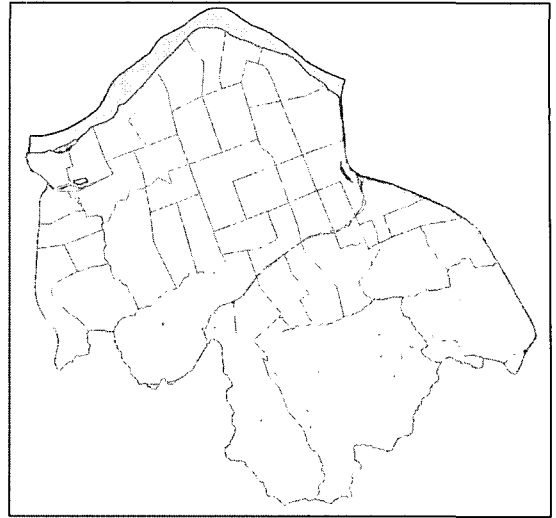


그림 7. 행정구역에 의한 분할(행정동)

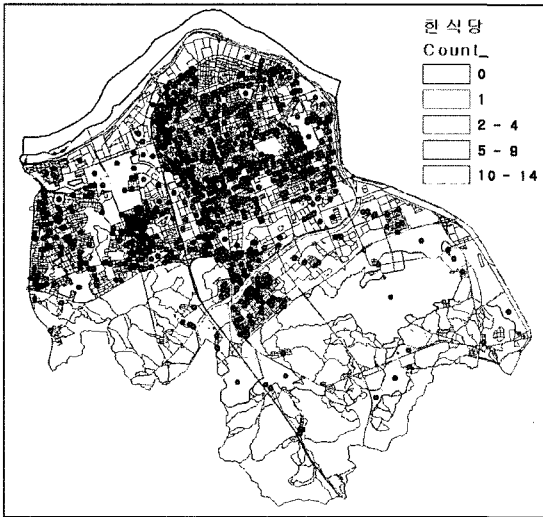


그림 8. 블록과 한식당 공간 조인 결과

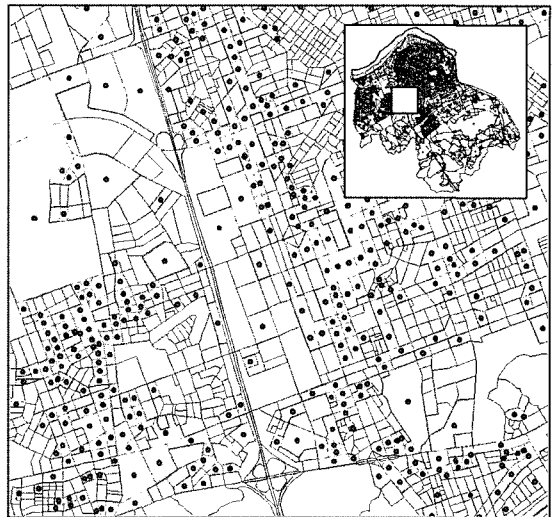


그림 9. <그림 8> 부분 확대

표 2. 폴리곤의 수

업종	한식당	변호사 사무실	동사무소	CSR점
공간 조인 대상 점사상 수	1,692	1,191	44	1,890
점사상을 포함한 블록 수	1,012	153	44	1,073

공간조인한 결과로서 가장 진한 폴리곤에 10-14개의 한식당이 포함되어 있다. 이 그림에 표시되어 있는 포인트 심볼은 폴리곤의 중심점이다.

4.3 K-지표 계산

K-지표 계산을 위해 블록과 행정구역의 무게중심점을 이용하여 점간 거리를 계산하고, 각 거리 구간에 포함된 점사상의 쌍(시점, 종점)의 수는 폴리곤 당 집계된 점사상 개수를 곱하여 계산하였다. 무게중심점 간의 거리는 ArcInfo 8.1.2의 거리계산 함수를 이용하였으며, 대상지역의 크기를 고려하고, 분석의 편의를 위하여 최대 3,000m 범위 내에 위치한 점까지의 거리를 산출하였다. 계산된 거리를 MS-Access에서 거리구간 별로 집계하여 K-지표를 산출하였다(그림 10 참조).

산출된 K-지표로부터 CSR의 K-지표를 기준으로 D-지표를 계산한 후 거리에 대해 표준화하여 표준화된 D-지표를 유도해 내었다(그림 11, 12 참조).

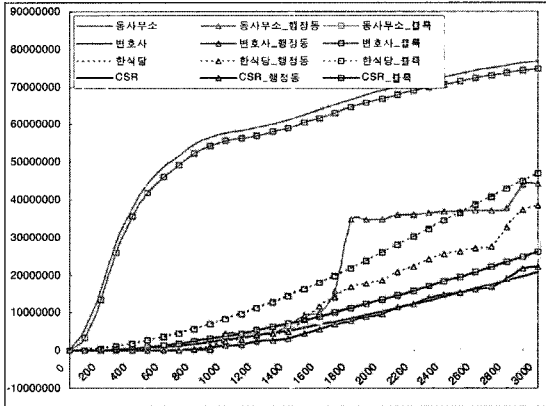


그림 10. K-지표 산출 결과

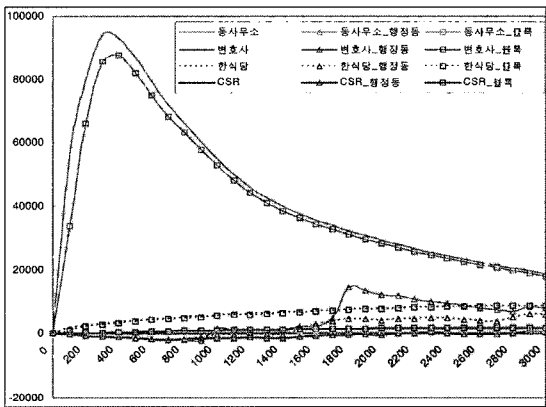


그림 11. 표준화된 D-지표 산출 결과

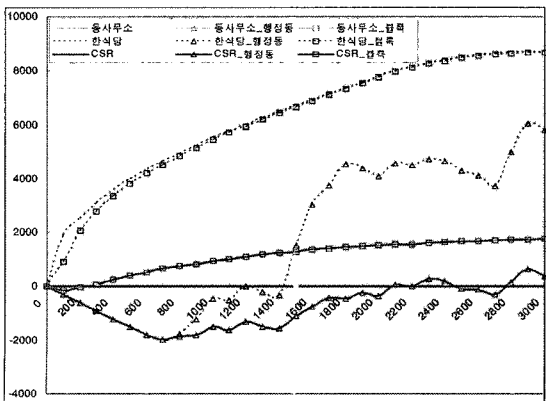


그림 12. 표준화된 D-지표 일부확대(변호사 사무실 제외)

그림 10, 11, 12에서 보는 바와 같이 블록 단위로 점사상을 집계하여 계산된 K-지표는 원데이터와 매우 유사한 평가 결과를 보이고 있으나, 행정구역의 경우 집계시 발생하는 편이에 의해 매우 심한 평가의 오류를 보이고 있음을 알 수 있다.

행정구역으로 집계된 데이터의 경우 공간 집계의 편이에 의한 영향으로 가까운 거리에서 K-지표가 낮게 나타나고 있다. 특히 행정구역의 무게 중심점 간의 최소 거리인 약 647m(방배본동과 방배4동 간)에 이르기까지는 모든 데이터에 대해 정상적인 K-지표의 산출이 불가능하여 점분포의 밀집도에 대한 어떠한 정보도 제공해 줄 수 없음을 알 수 있으며, 행정구역 무게 중심점 간의 평균 거리(약 1851m) 전후의 구간에서 대부분의 K-지표가 원데이터에 근접하게 증가하는 모습을 볼 수 있다.

반면, 블록으로 집계된 데이터의 경우에는 집계 편이에 의한 영향이 상대적으로 작게 나타나고 있으며, 표본이 많은 변호사 사무실, 한식당, CSR의 경우 원데이터의 K-지표를 대신해서 사용할 수 있을 정도의 결과를 나타내고 있으나, 소수의 점사상을 포함하고 있는 동사무소의 경우 원데이터에서 구해진 K-지표에 비해 상대적인 값의 변동 폭이 크게 나타나고 있음을 알 수 있다. 블록단위로 산출된 D-지표가 원자료의 것과 일치하는 지 여부를 평가하기 위해 피어슨의 상관계수를 구하였다(표 3 참조). 표 3에서 보이는 것처럼 원자료와 블록은 거의 1에 가까운 상관관계를 보이고 있어 이의 유용성을 평가하기 위해 양측 T-검정을 시행하였다. 양측 T-검정을 시행한 결과 세가지 업종 모두에서 블록으로부터 산출된 표준화된 D-지표는 원자료의 것과 동일한 것으로 나타난 반면, 행정구역의

표 3. 피어슨 상관계수

업종	한식당	변호사 사무실	동사무소	CSR점
원자료 대 블록	1.00	0.98	0.98	/
원자료 대 행정구역	0.87	-0.79	0.95	

표 4. 양측 T-검정 결과

업종	한식당	변호사 사무실	동사무소	CSR점
원자료 대 블록	0.22	0.53	0.21	/
원자료 대 행정구역	7.24	9.16	3.25	
유의수준 0.05의 기각역	2.00	2.00	2.00	/

것은 다른 것으로 나타났다(표 4 참조).

한편, 본 연구의 결과는 비록 작은 지역에서 이루어지기는 했으나 Feser and Sweeney(2000)의 연구 결과와는 상반되는 결과를 보여주고 있다. Feser and Sweeney(2000)의 연구에서는 평가대상 제조업종의 대부분이 3km 내에서 주소매칭된 데이터보다 zip code로 집계된 데이터로부터 산출된 D-지표가 높게(업종에 따라서는 매우 높게) 나타났으나, 본 연구에서는 행정구역은 물론 블록에서도 주소매칭된 원자료의 D-지표에 비해 집계된 자료의 D-지표가 낮게 나타나고 있다. 이는 집계단위의 특성과 집계대상인 점사상의 분포특성에 따라 D-지표가 나타나는 경향이 달라지기 때문인 것으로 판단되며 도시공간 내에서의 점사상의 분포를 평가하기 위해서는 집계자료의 사용은 가급적 피해야 함을 보여주는 것이다.

5. 결 론

본 연구에서는 우선 최근 관심의 대상이 되고 있는 점사상 분포 패턴의 평가에 측량을 통해 절대좌표를 가지지 못하고 주소매칭에 의해 점사상을 생성하여 사용하는 다양한 GIS 분야에서 발생할 수 있는 문제점에 대해 고찰하고 이에 대한 대안으로서 일정 지역단위로 집계된 데이터가 분포패턴 평가에 유용한지 여부를 K-지표(D-지표)를 이용하여 평가하였다.

주소매칭을 이용하여 생성된 점사상을 블록단위와 행정구역 단위로 집계하여 K-지표를 산출하여 비교한 연구 결과, 블록단위의 집계는 주소단위와 거의 차이가 없는 것으로 나타났다. 따라서 블록단위의 GIS 데이터(블록 레이어와 블록별 주소목록 테이블로 구성됨)가 구축되고, 이를 기준으로 한 통계자료의 입수가 가능하다면 개별 점사상에 좌표를 부여하기 위한 부가적인 노력을 감소시키면서 점사상의 분포패턴을 분석해 낼 수 있을 것으로 기대된다. 반면 행정구역은 데이터의 밀집 상태나 지역적인 범위와 무관하게 점사상의 분포특성을 왜곡시키고 있음을 알 수 있었으며, 특히 근거리에서의 D-지표 평가결과는 도시지역 내에서의 국지적인 밀집에 대한 평가가 전혀 불가능함을 보여주고 있다.

블록단위는 현 시점에서 블록별 주소목록 테이블과 같은 정보가 없기 때문에 실제 활용할 수 없는 단위라는 문제점이 있다. 그러나 현재 입법이 추진되고 있는 도로명주소(새주소)사업은 2009년에 대부분의 지자체에 대해 시설을 완료하고 2011년까지 공법관계 주소변경을 완료하여 2015년 새주소로 완전 전환 활용하는 것을 목표로 하고 있으며, 그 시행결과를 이용하여 정보통신부에서는 GIS 집배관리시스템을 구축하고 통계청에서는 인구주택센서스시 기초조사 도면을 구축할 계획에 있어, 향후 새주소로 현재의 법정동과 지번에 의한 주소체계가 대체되면, 새주소로부터 용이하게 블록 정보를 구축할 수 있고, 이를 이용한 점사상의 집계와 통계자료의 발표가 가능해질 것으로 기대된다.

참고문헌

- 유근배 (1998), 점패턴분석을 이용한 수치지형도의 점사상 일반화, 한국GIS학회지, 한국GIS학회, 제6권, 제1호, pp. 11-23.
- 이미혜 (1994), 우리나라 소아암의 공간적 분포에 관한 연구 - 군집성을 중심으로-, 서울대학교 지리학석사학위논문, 서울대학교.
- 이병길 (2003), 비즈니스 GIS에서 공간 데이터마이닝(Spatial Data Mining)기법을 이용한 상권추출, 한국GIS학회지, 한국GIS학회, 제11권, 제2호, pp. 171-184.
- 이희연 (1989), 지리통계학, 법문사, pp. 644-646.
- 전용완 (2002), GIS와 공간통계를 이용한 범죄 분석에 관한 연구 -서울시를 사례로-, 서울대학교 교육학석사학위논문, 서울대학교.
- Diggle P.J. (1983), *Statistical analysis of spatial point patterns*, Academic Press, London.
- Feser E.J. and Sweeney S.H. (2000), A test for the coincident economic and spatial clustering of business enterprises, *Journal of Geographical Systems*, Vol. 2, No. 4, pp. 349-373.
- Gatrell A.C., Bailey T.C., Diggle P.J. and Rowlingson B.S. (1996), Spatial point pattern analysis and its application in geographical epidemiology, *Royal Geographical Society (with the Institute of British Geographer)*, NS 21, pp. 252-746.
- Haggett P., Cliff A.D. and Frey A.E. (1977), *Locational methods in human geography*, Edward Arnold, London, in Anthony C. Gatrell, et al.(1996), Spatial point pattern analysis and its application in geographical epidemiology, *Royal Geographical Society (with the Institute of British Geographer)*, NS 21, pp. 252-746.

(접수일 2006. 3. 2, 심사일 2006. 3. 14, 심사완료일 2006. 3. 20)